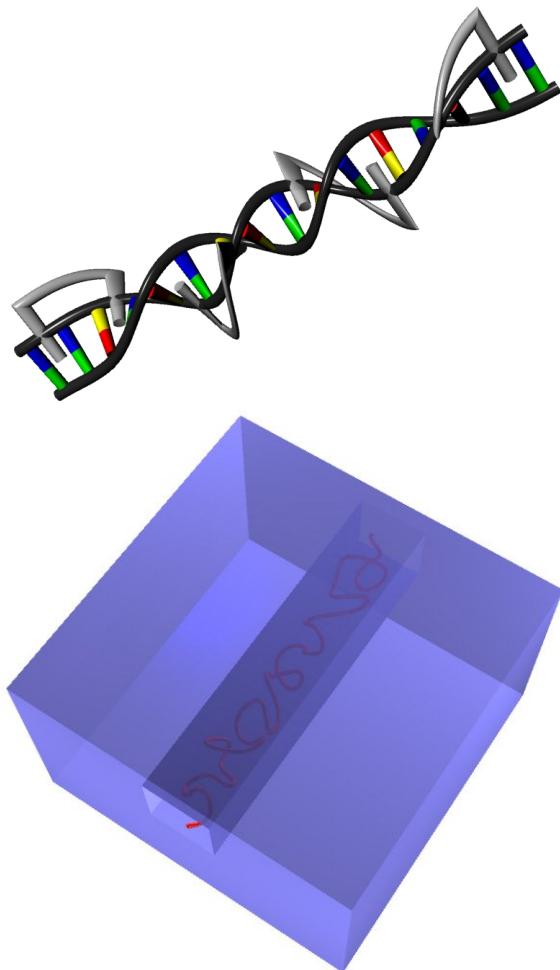


---

# Biopolymer Physics for Engineers

—With an emphasis on DNA

---



---

Lasse Højlund Thamdrup  
Department of Micro and Nanotechnology (MIC)  
Technical University of Denmark

First edition, 2008

# Preface

---

In 1920 Hermann Staudinger proposed the macromolecular hypothesis [1] in which polymers<sup>1</sup> were described as large entities consisting of covalently bonded monomer units. Before the introduction of this hypothesis, the conventional assumption was that macromolecules consisted of physically associated clusters of small-scale molecules. Staudingers pioneer work effectively started a scientific avalanche within the field of polymer physics. The change from a colloidal to a long-chain perception of macromolecule structure marked a paradigm shift which catalyzed the development of polymer physics as an independent discipline.

In the following 40 years the foundation of polymer physics was developed: Werner Kuhn, Otto Kratky and Günther Porod contributed to a better understanding of macromolecular size and conformation which, among other things, lead to the introduction of the ideal chain and worm-like chain models. Especially the latter has proven efficient in describing the conformation of DNA that, as proven by Francis Crick and James D. Watson in 1953, has a semiflexible double helical structure. Furthermore the work of Paul J. Flory, Prince E. Rouse and Bruno H. Zimm, which lead to a better understanding of the conformation and dynamics of single polymer molecules in solution, also represents a major contribution.

As an example of more recent scientific achievements, Pierre-Gilles de Gennes' work on the conformation of confined semiflexible self-avoiding polymers (e.g. DNA) should be mentioned. Based on scaling arguments, de Gennes derived expressions for the free energy and extension length of macromolecules subject to confinement inside structures with dimensions larger than the polymer persistence length (i.e.  $\sim 50$  nm for DNA). The work by de Gennes was subsequently extended by Theo Odijk to cover the case of strong confinement in which the dimensions of the confining structure are smaller or comparable to the persistence length. Mediated by the progress and refinement of nanofabrication techniques, it has been possible to conduct experimental measurements on nano-scale confined DNA and the results are in good agreement with the theories proposed by de Gennes and Odijk.

---

<sup>1</sup>The term "polymer" is derived from Greek, poly=many and meros=parts.

The primary objective of this report is to provide engineers endeavoring the field of nano-scale investigation of DNA with the necessary theoretical background knowledge. Chapter 1 covers the basics of polymer physics and crucial concepts regarding the properties of flexible and semiflexible polymers are introduced. Chapter 2 proceeds by describing the chemical composition and physical structure of linear DNA. Furthermore DNA bending, mediated by specific nucleotide sequences, is considered. Chapter 3 introduces the confinement-induced conformational response of DNA based on the theoretical framework developed by de Gennes and Odijk. Furthermore a measurement technique known as epifluorescence microscopy, which is commonly used for conducting measurements on confined DNA, will be described and some experimental results will be presented.

*Tempora mutantur, nos et mutamur in illis*

# Contents

<b>1 Biopolymer Physics</b>	<b>6</b>
1.1 The Structure of Macromolecules . . . . .	7
1.1.1 Synthetic Polymers . . . . .	7
1.1.2 Biopolymers . . . . .	9
1.1.3 Primary, Secondary and Tertiary Structure . . . . .	12
1.2 Ideal Flexible Chains . . . . .	12
1.2.1 End-to-End Distance . . . . .	13
1.2.2 Elasticity . . . . .	15
1.3 Ideal Worm-Like Chains . . . . .	16
1.3.1 End-to-End Distance . . . . .	17
1.3.2 Elasticity . . . . .	19
1.4 Excluded Volume Effects . . . . .	21
1.5 Chapter Summary . . . . .	24
<b>2 Unravelling DNA</b>	<b>26</b>
2.1 The Double Helical Structure of DNA . . . . .	27
2.1.1 The Basic DNA Building Blocks . . . . .	28
2.1.2 Hybridization and Different Conformations . . . . .	30
2.2 Localized DNA Bending . . . . .	32
2.2.1 Experimental Results . . . . .	32
2.2.2 DNA Bending Models . . . . .	34
2.3 Chapter Summary . . . . .	36
<b>3 DNA in Nanoconfinements</b>	<b>38</b>
3.1 Confinement Induced Effects . . . . .	39
3.1.1 The de Gennes Regime . . . . .	39

---

3.1.2	The Odijk Regime . . . . .	41
3.2	Investigating DNA in Nanofluidic Structures . . . . .	43
3.2.1	Epifluorescence Microscopy . . . . .	43
3.2.2	Experimental Results . . . . .	45
3.3	Chapter Summary and Concluding Remarks . . . . .	49
	<b>Bibliography</b>	<b>51</b>

# Chapter 1

## Biopolymer Physics

---

One might intuitively comprehend the extreme degree of complexity associated with establishing fundamental analytical models describing the physics of macromolecules. Polymers, synthetic or biological, exhibit an overwhelming variety in both structure and chemical composition. From relatively simple monomer building blocks, complex fractal-like molecules<sup>1</sup> are formed. The exact polymer structure may e.g. be linear, circular, branched or dendrimic and the degree of flexibility and self-avoidance for different polymers is subject to large variations.

This chapter will initially provide a brief over-all description of polymer structures with an emphasis on biological macromolecules. Furthermore the analytical concepts of ideal flexible and worm-like chains will be introduced. Certain linear polymers may be modelled as ideal flexible chains whereas semiflexible biopolymers, such as DNA, are more adequately described using the Kratky-Porod worm-like chain model. Finally a description of excluded volume effects will be given. Unless otherwise stated, the information in this chapter has been adopted from Refs. [2] and [3].

---

<sup>1</sup>Fractals are characterized by self-similarity and a dimensionality which does not unambiguously assume integer values. The most famous fractals such as the Cantor set, the Sierpinski triangle and the Koch curve are mathematical constructs. However, numerous objects such as snowflakes, electrical discharges, certain plants and polymers also exhibit a fractal-like structure.

### 1.1. THE STRUCTURE OF MACROMOLECULES

## 1.1 The Structure of Macromolecules

Before commencing a description of biopolymers it is beneficial to briefly consider the structure of synthetic homo- and heteropolymers which are produced in carefully controlled and highly optimized polymerization reactions.

### 1.1.1 Synthetic Polymers

Commonly known man-made homopolymers, consisting of only one type of monomer, include polystyrene (PS), polydimethylsiloxane (PDMS), polymethyl-methacrylate (PMMA) and polyethylene (PE) which have all been shown in Fig. 1.1 (a). Heteropolymers are composed of two or more different monomer entities as seen schematically in Fig. 1.1 (b).

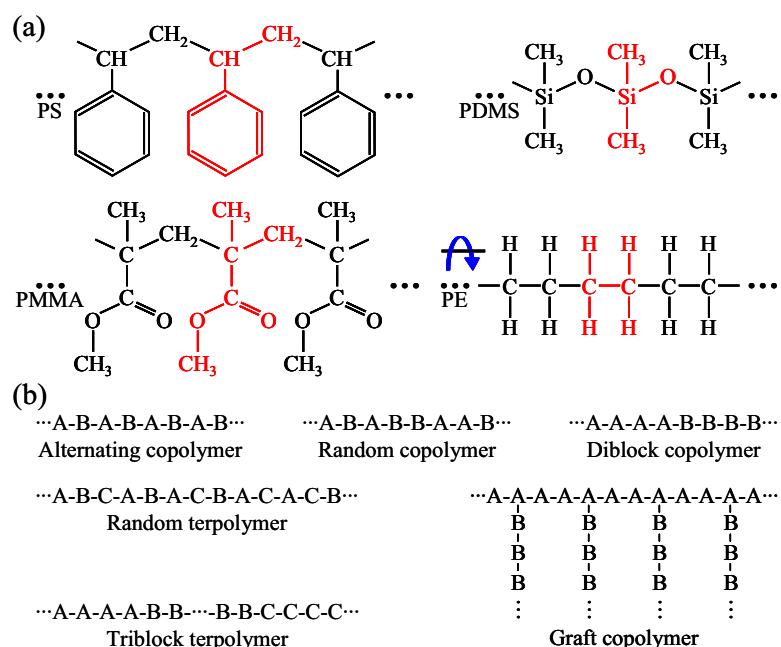


Figure 1.1: (a) Chemical composition of the synthetic homopolymers PS, PDMS, PMMA and PE. The repeating molecular entities have been highlighted in red. Notice that the hydrogen atoms on the phenyl groups in PS have been left out. The direction of the torsion angle around  $\sigma$ -bonds has been indicated on PE. (b) Schematics showing different types of heteropolymers. A, B and C denote different monomers. Copolymers consists of two different monomers whereas terpolymers are composed of three. The different monomers may be placed randomly, in alternating periodic sequences or in discrete homogenous blocks.

## 1.1. THE STRUCTURE OF MACROMOLECULES

---

PS	Pure PS is a transparent and rigid polymer commonly used for packaging.
PDMS	PDMS is a silicon-based organic polymer used for e.g. contact lenses, medical devices and heat resistant tiles
PMMA	PMMA, commonly referred to as Plexiglas, Lucite and Perspex, is a highly versatile polymer which, on the more exotic side, is used as a resist in nanoimprint and electron-beam lithography.
PE	PE has found widespread application in e.g. gas and water pipes, containers, detergent bottles, cable covering, toys and plastic carrier bags.

Table 1.1: Descriptions of the homopolymers PS, PDMS, PMMA and PE.

Synthetic polymers, in which the monomers are covalently joined by single  $\sigma$ -bonds along the macromolecule backbone, are generally rather flexible. This inherent flexibility, which is primarily mediated by a high degree of variation in the torsion around  $\sigma$ -bonds connecting neighboring monomers, enables polymers to assume a multitude of conformations. Based on the exact polymerization conditions and chemical composition of the monomers, polymers having extremely different overall structures or architectures may be synthesized as seen in Fig. 1.2. The linear and circular polymer structures are the simplest and presumably most well-known. Grafted copolymers or polymers in which e.g. hydrogen atoms have been replaced by covalently bonded chain segments may adopt comb, star or branched structures. Dendrimers are a special type of non-randomly branched polymers exhibiting beautiful fractal-like architectures as evident from the image of a polyphenylene dendrimer presented in Fig. 1.3 (a).

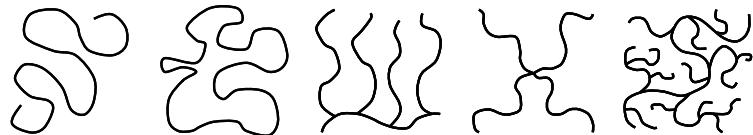


Figure 1.2: Commonly observed polymer structures. From left to right: Linear, circular, comb, star and randomly branched.

Taking the wealth of different polymer structures and compositions into account it should not be surprising that the abundance of synthetically made macromolecules is extraordinary. Some applications of the aforementioned homopolymers (i.e. PS, PDMS, PMMA and PE) have been included in Tab. 1.1. We will return to a

## 1.1. THE STRUCTURE OF MACROMOLECULES

---

physical model useful for explaining the conformation of certain single linear polymer molecules in Sec. 1.2.

### 1.1.2 Biopolymers

In contrast to artificial macromolecules, biopolymers are structurally complex entities which have been continuously developed and refined as a result of evolution in biological systems. Prominent examples of biological macromolecules include polysaccharides, nucleic acids and polypeptides which have all been described in Tab. 1.2.

Polysaccharides	Carbohydrate molecules consisting of long and potentially branched chains of one or more monosaccharides coupled together by covalent glycosidic bonds. Common examples include cellulose, starch and chitin. Cellulose is the primary constituent of plant matter and is usually extracted from wood pulp and cotton. Starch is the most important carbohydrate in terms of human nutrition. Chitin is the main building block in the exoskeletons of e.g. crabs, lobsters, beetles and ants and has also found medical and industrial applications.
Nucleic acids	The basic building blocks of the nucleic acids DNA and RNA are nucleotides composed of an aromatic nitrogenous base coupled to a phosphorylated sugar molecule. Within cells, RNA is usually found as single stranded macromolecules. RNA plays a crucial role in translating the genetic information contained in DNA into proteins and enzymes.
Polypeptides	Proteins and enzymes are examples of polypeptides which are composed of amino acids. Structural proteins include collagen which is found in connective tissues and keratin found in human hair and nails. Other proteins are involved in cell signaling (e.g. insulin) or ligand transport (e.g. hemoglobin and myoglobin). Enzymes are proteins that catalyze biochemical reactions involved in metabolism, DNA replication and repair and RNA synthesis.

Table 1.2: description of various groups of biological macromolecules.

Polysaccharides, which are carbohydrate molecules, are the group of biopolymers bearing most resemblance to synthetic polymers. This is evident from the image of cellulose presented in Fig. 1.3 (b). They may be composed of several different monosaccharide units and often possess a branched macromolecular structure.

Nucleic acids include ribonucleic acids (RNA) and deoxyribonucleic acids (DNA) which are examples of biological heteropolymers consisting of specific sequences of approximately four different monomer units known as nucleotides. The main difference between the nucleotides of RNA and DNA is the phosphorylated sugar moi-

## 1.1. THE STRUCTURE OF MACROMOLECULES

---

eties constituting the molecular backbone. Furthermore RNA is usually found as single strands however, certain RNA molecules, such as the ribozyme depicted in Fig. 1.3 (c), has a duplex structure.

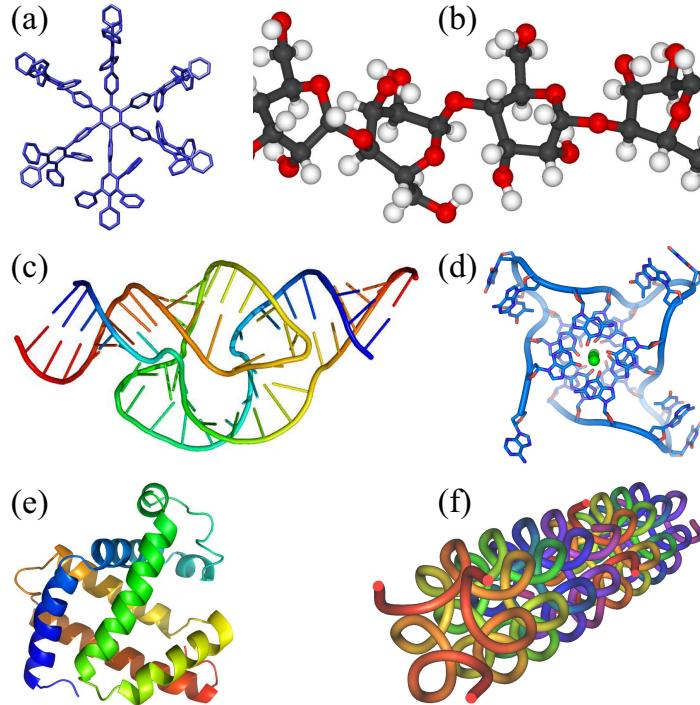


Figure 1.3: Images showing the structure of a selected ensemble of polymers [4]. Most images are based on X-ray crystallography data. (a) Polyphenylene dendrimer composed of aromatic pentagonal monomers [5]. (b) Cellulose which is a polysaccharide commonly found in wood pulp and cotton. (c) Hammerhead ribozyme consisting of complementary RNA strands held together by hydrogen bonds. Ribozymes catalyze biochemical reactions within cells. (d) Telomere DNA quadruplex structure formed by four single stranded DNA molecules chelated to the **metal ions** in the center. (e) Oxymyoglobin protein responsible for distributing oxygen in muscle tissue. The protein is composed of approximately 150 amino acids arranged into distinct  $\alpha$ -helices. (f) Collagen protein composed of three left-handed and intertwined polypeptide strands. Collagen constitutes approximately 25 % of the overall protein content in mammals.

The structure of DNA will be properly introduced in Chap. 2 and here it will suffice to state that DNA commonly has a double helical structure in which two complementary strands, composed of nucleotides, are intimately entwined. However, different structures such as cruciform and triple-stranded DNA has been observed and furthermore quadruplex structures called telomeres are known to exist at the termini

## 1.1. THE STRUCTURE OF MACROMOLECULES

---

of linear chromosomal DNA (refer to Fig. 1.3 (d)). In terms of molecular chain flexibility, duplex RNA and DNA deviate significantly from synthetic polymers and they behave as semiflexible macromolecules.

The last group of biopolymers to be considered in this report are polypeptides which consists of  $\sim 20$  different monomeric units, known as amino acids, joined by covalent peptide bonds. A peptide bond is formed in a condensation reaction between an amino group ( $-NH_2$ ) on one monomer and a carboxyl group ( $-COOH$ ) attached to another monomer. Amino acids contain both amine (organic derivatives of ammonia) and carboxyl functional groups which mediate biopolymeric interaction. The primary members of the polypeptide family of biopolymers are proteins and enzymes. Proteins solve different tasks such as cell signalling and ligand transport and furthermore constitute the structural molecules of connective tissue, hair and nails. In Fig. 1.3 (e) and (f) images of myoglobin (responsible for transporting oxygen) and collagen (a structural protein) are presented. The characteristic right-handed helices seen in the image of myoglobin are called  $\alpha$ -helices and they represent a commonly observed structural motif in diverse polypeptides. The conformation of the polypeptide chain in  $\alpha$ -helices is stabilized by hydrogen bonds, directed along the helical axis, between non-adjacent amino acids. Another commonly observed structural motif is the so called  $\beta$ -sheet which consists of two pleated amino acid chains joined by hydrogen bonds as seen in Fig. 1.4.

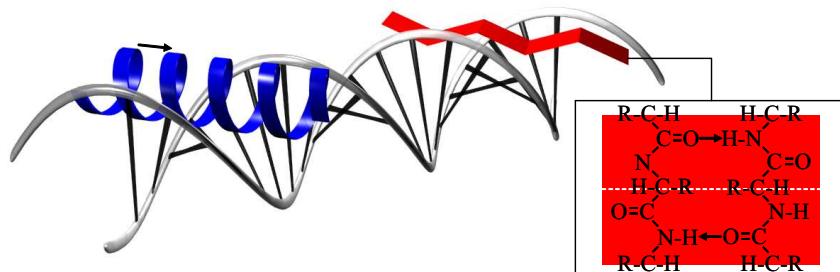


Figure 1.4: Image showing an  $\alpha$ -helix and a  $\beta$ -sheet interacting with the base pairs of DNA by aligning in the grooves formed between two complementary strands. The inset shows how amino acids are aligned linearly in a side-by-side fashion in the slightly pleated  $\beta$ -sheet. The dashed white line has been included to emphasize that not all carbon atoms are confined within the same plane and  $R$  designate side-chains. Arrows indicate the direction of hydrogen bonds.

### 1.1.3 Primary, Secondary and Tertiary Structure

Based on the material presented in this section one might conclude that synthetic and biological macromolecules are remarkably similar as both types of polymers consist of repeating monomer units arranged into overall architectures exhibiting varying degrees of complexity. Such a conclusion would be highly erroneous. Biopolymers such as DNA, RNA, proteins and enzymes are biologically synthesized in template-based processes which effectively means that biopolymers are monodisperse, or in other words, consist of the exact same sequence of specific monomeric units. The chemical composition and sequential arrangement of the monomers are commonly referred to as the primary structure of biopolymers. In the case of linear DNA and RNA, the primary structure is merely defined by the sequence of nucleotides. Especially in proteins and enzymes, local biopolymer segments may form complex structures such as the aforementioned  $\alpha$ -helix and  $\beta$ -sheet thus giving rise to an inherent secondary structure which is of immense biological importance. The exact monomeric composition and structure of such secondary motifs largely determine the biological functionality of e.g. proteins. The spatial arrangement and composition of secondary structural motifs constitute the tertiary structure which represents a complete three-dimensional description of a biopolymers architecture. Within the field of structural biology, people are struggling to predict the complex tertiary structure of individual biopolymers from the vastly more simple primary structure—a task that is far from being trivial. Unlike artificially man-made polymers, which exhibit a high degree of polydispersity and stochastic variation in the overall molecular architecture, biopolymers are monodisperse and refined macromolecules exhibiting an overwhelming structural integrity.

## 1.2 Ideal Flexible Chains

Within the framework of polymer physics, the concept of an ideal chain represents a convenient way of modelling the conformation of flexible linear macromolecules in which interactions between monomers, separated by a large distance along the backbone of the chain, may be neglected. Ideal chain models do not consider subtle and intricate steric effects associated with the spatial extent of the individual monomers and furthermore local chain segments are allowed to intersect (i.e. the diameter/width of the chain is 0). In this section, the freely jointed chain model

## 1.2. IDEAL FLEXIBLE CHAINS

---

will be described in order to introduce important aspects of the conformation and elasticity of polymers.

### 1.2.1 End-to-End Distance

A freely jointed polymer may be modelled as a series of  $N$  rigid links which constitute the chemical bonds between  $N + 1$  monomer entities as shown in Fig. 1.5. Throughout this section it will be assumed that  $N \gg 1$ . The bond vector  $\vec{r}_i$  simply denotes the linkage between neighboring monomers  $M_{i-1}$  and  $M_i$  for  $i = 1-N$  and it is assumed that  $|\vec{r}_i| = l \forall i$ . This is a reasonable assumption in a simple model, since the distance between neighboring monomers is rather constant and only subject to small fluctuations. In the freely jointed chain model it is furthermore assumed that the individual bond vectors may have any spatial direction which effectively means that the polymer is perfectly flexible.

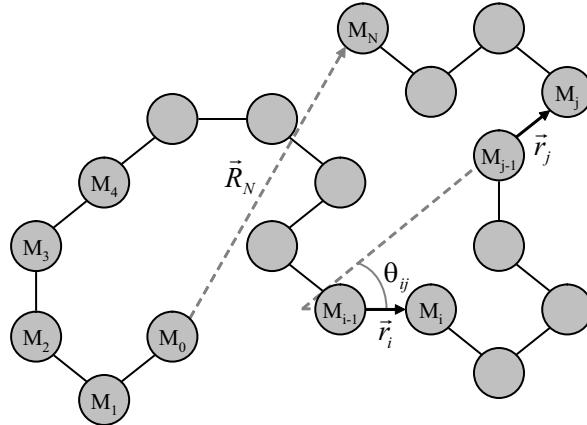


Figure 1.5: 2D Schematic showing one specific conformation of a freely jointed ideal polymer chain. Notice the direction of the bond vectors between neighboring monomers and the end-to-end vector  $\vec{R}_N$  which may be used as a measurand of the polymer chain size. The distance between neighboring monomers is  $l$  and a typical value is  $l = 1.54 \text{ \AA}$  which is the distance between carbon atoms in PE [3].

The overall end-to-end vector  $\vec{R}_N$ , connecting the termini of a polymer chain, is simply given by the vectorial sum of the  $N$  bond vectors

$$\vec{R}_N = \sum_{i=1}^N \vec{r}_i. \quad (1.1)$$

## 1.2. IDEAL FLEXIBLE CHAINS

---

Clearly  $\vec{R}_N$  is not a constant vector unless the polymer ends are pinned and therefore the ensemble average  $\langle \vec{R}_N \rangle$  will be considered. As the individual bond vectors  $\vec{r}_i$  have no preferred spatial direction, the freely jointed chain may be modelled in terms of a 3D random walk of step length  $l$ . This effectively ensures that  $\langle \vec{R}_N \rangle = 0$ . Therefore the mean-square magnitude of  $\vec{R}_N$  should be considered

$$\langle R^2 \rangle \equiv \langle \vec{R}_N^2 \rangle = \sum_{i=1}^N \sum_{j=1}^N \langle \vec{r}_i \cdot \vec{r}_j \rangle = l^2 \sum_{i=1}^N \sum_{j=1}^N \langle \cos \theta_{ij} \rangle. \quad (1.2)$$

The last equality in Eqn. 1.2 is ensured by the constant bond length  $l$ . As seen in Fig. 1.5,  $\theta_{ij}$  merely specifies the angle between  $\vec{r}_i$  and  $\vec{r}_j$ . A freely jointed chain is characterized by the lack of correlation between the spatial directions of different bond vectors (i.e.  $\langle \cos \theta_{ij} \rangle = 0$  for  $i \neq j$ ) which allows for a substantial simplification of Eqn. 1.2

$$\langle R^2 \rangle = l^2 \sum_{i=1}^N \sum_{j=1}^N \delta_{ij} = Nl^2. \quad (1.3)$$

$\delta_{ij}$  is simply Kronecker's delta defined by

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad (1.4)$$

In real polymer chains, steric hindrance and restricted bond angles induce a certain degree of correlation between different bond vectors which effectively serve as to increase the chain stiffness. More detailed models incorporate local bond correlations however, it is a common feature of all ideal chain models that  $\langle \cos \theta_{ij} \rangle = 0$  for  $|i - j| \gg 1$ . It may be shown, that the end-to-end distance of polymers behaving as ideal flexible chains is well approximated by

$$\langle R^2 \rangle \approx C_\infty Nl^2 \quad (1.5)$$

where Flory's characteristic ratio  $C_\infty$  is a constant describing the local stiffness of the polymer. Values of  $C_\infty$  have been included for PS, PDMS, PMMA and PE in Tab. 1.3. From both Eqn. 1.3 and 1.5 it is evident that  $\langle R^2 \rangle \propto Nl^2$  which is a generic property of ideal chains. Before proceeding with a description of the elasticity of a freely jointed chain, the probability distribution  $P(\vec{R}_N)$  will be considered.  $P(\vec{R}_N)$  describes the probability of a chain adopting a specific end-to-end vector

## 1.2. IDEAL FLEXIBLE CHAINS

---

$\vec{R}_N$ . For a freely jointed chain consisting of  $N \gg 1$  links,  $P(\vec{R}_N)$  is merely a normal distribution [6]

$$P(\vec{R}_N) = \left( \frac{3}{2\pi\langle R^2 \rangle} \right)^{\frac{3}{2}} \exp \left( -\frac{3\vec{R}_N^2}{2\langle R^2 \rangle} \right). \quad (1.6)$$

From Eqn. 1.6 it is seen that the width of the distribution is determined by  $\langle R^2 \rangle$ . Given a specific fixed  $\vec{R}_N$ , the contour between the end points may redistribute spatially in a number of ways denoted by  $\Omega$ . Therefore it should be clear that

$$\Omega = P(\vec{R}_N) \int \Omega d\vec{R}' \quad (1.7)$$

where the integral represents the total number of conformations that a chain with  $N$  links may assume. It should be emphasized that the integral solely depends on  $N$  and not the specific choice of  $\vec{R}_N$ . Furthermore it should be noted that  $\Omega$  assumes a maximum value for  $\vec{R}_N = \vec{R}_0 = \vec{0}$  (i.e. when the chain ends are located at the same position).

	PS	PDMS	PMMA	PE
$C_\infty$	9.5	6.8	9.0	7.4

Table 1.3: Characteristic ratios  $C_\infty$  of linear homopolymers [3].

### 1.2.2 Elasticity

In order to investigate the elasticity of a freely jointed chain we may consider the Helmholtz free energy:

$$F = U - TS. \quad (1.8)$$

$U$  is the internal energy,  $S$  is the entropy and  $T$  is the absolute temperature. As the chain is perfectly flexible and monomer-monomer interactions have been neglected, the internal energy assumes a constant value  $U_0$  which is independent of the polymer conformation. The entropy is directly related to  $\Omega$

$$S = k_B \ln \Omega \Rightarrow \\ S = -\frac{3k_B}{2\langle R^2 \rangle} \vec{R}_N^2 + k_B \ln \left[ \left( \frac{3}{2\pi\langle R^2 \rangle} \right)^{\frac{3}{2}} \int \Omega d\vec{R}' \right] \quad (1.9)$$

### 1.3. IDEAL WORM-LIKE CHAINS

---

where  $k_B$  is Boltzmann's constant. For a specific  $N$ , the second term is merely a constant which will be denoted  $S_0$ . Combining Eqn. 1.8 and 1.9 allows for stating the following simple expression for the free energy of a freely jointed chain

$$F = \frac{3k_B T}{2\langle R^2 \rangle} \vec{R}_N^2 + U_0 - TS_0. \quad (1.10)$$

From Eqn. 1.10 it is seen that  $F$  assumes a minimum value for  $\vec{R}_N = \vec{R}_0$ . In order to maintain a fixed end-to-end vector  $\vec{R}_N \neq \vec{R}_0$ , an external force  $\vec{f}$  must be applied

$$\vec{f} = \frac{\partial F}{\partial \vec{R}_N} = \frac{3k_B T}{\langle R^2 \rangle} \vec{R}_N = \overbrace{\frac{3k_B T}{Nl^2}}^k \vec{R}_N. \quad (1.11)$$

The constant of proportionality  $k$  is referred to as the entropic spring constant and it is evident that  $k \propto N^{-1}$  which means that longer chains are more easily stretched. The elasticity is purely entropic and is a direct result of the decrease in available conformations  $\Omega$  as  $|\vec{R}_N|$  is increased. It is also noticed that  $k \propto T$  which means that a freely jointed chain, counter-intuitively, gets softer at lower temperatures. It must be mentioned that Eqn. 1.11 is only valid for small extensions  $|\vec{R}_N| \ll Nl = l_c$  since highly nonlinear effects arise as the chain is stretched towards its full contour length<sup>2</sup>. In the remainder of this report the focus will primarily be on unilateral elongations of macromolecules. Therefore Eqn. 1.11 has been stated in a scalar version in Eqn. 1.12 where  $r$  is the extension and  $Nl$  has been replaced by  $l_c$

$$f = \frac{3k_B T}{l_c l} r. \quad (1.12)$$

## 1.3 Ideal Worm-Like Chains

It should be no surprise that all macromolecules will appear rigid if probed on a sufficiently small length scale. Even the freely jointed chain, which surely must be conceived as highly flexible, will be rigid on a scale comparable to the inter-monomer bond distance  $l$ .

---

<sup>2</sup>The contour length  $l_c$  will be used to denote the maximum end-to-end distance of a polymer and is not unambiguously given exactly by  $Nl$ .

### 1.3.1 End-to-End Distance

All polymers have a distinct scale, abbreviated the Kuhn length  $l_K \geq l$ , below which the end-to-end distance of a chain segment is well approximated by the contour length. These Kuhn segments of length  $l_K$  will be relatively uncorrelated and therefore an equivalent freely jointed chain may be build from a number  $N_K$  of these discrete segments. Clearly the mean-square end-to-end distance  $\langle R^2 \rangle$  and the contour length  $l_c$  of the equivalent and real chain should be equal. Therefore it must be demanded that  $N_K l_K = l_c$  and  $\langle R^2 \rangle = N_K l_K^2$  which allows for stating an expression for  $l_K$

$$l_K = \frac{\langle R^2 \rangle}{l_c}. \quad (1.13)$$

In the case of a freely jointed chain, Eqn. 1.13 merely reduces to  $l_K = l$ . However, for less flexible polymers  $l_K$  will be larger than  $l$ . In the case of DNA  $l_K \approx 100$  nm which is an astonishingly 300 times larger than the distance between two neighboring base pairs! The Kuhn length is actually a good measurand of the flexibility of any given polymer chain and for long semiflexible worm-like chains, such as DNA,  $l_K = 2l_p$  as will be shown later in this section.  $l_p$  is the persistence length which will be considered now.

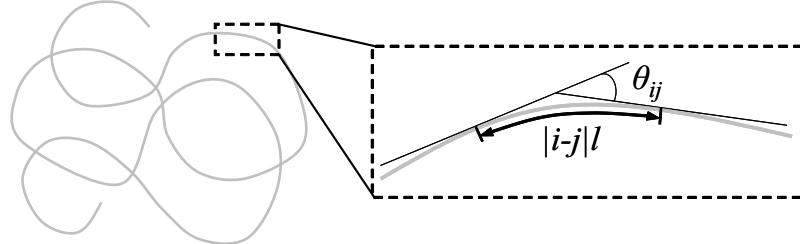


Figure 1.6: Schematic of a coiled-up semiflexible polymer and a zoom in on a small segment of the chain. Notice the angle between the tangent vectors  $\theta_{ij}$  and the length of the curved segment. As in the last section,  $i$  and  $j$  denote two different bonds and  $l$  is the bond length.

If we imagine zooming in on a segment of a worm-like polymer as depicted in Fig. 1.6, a smooth chain, consisting of what appears to be highly correlated monomer units, would be discovered. Initially it is not really feasible to consider the intricate relationship between individual bond vectors when trying to describe a chain that is continuously flexible. For this task a mesoscopic approach seems more appealing. As seen in Fig. 1.6, the tangent vectors of a curved segment of length  $|i - j|l$  will

### 1.3. IDEAL WORM-LIKE CHAINS

---

form an angle  $\theta_{ij}$  and we may consider  $\cos \theta_{ij}$ . The exact value of  $\cos \theta_{ij}$  will be subject to fluctuations and therefore it is more consistent to address the average  $\langle \cos \theta_{ij} \rangle$ . It may be shown that if  $|i - j|l$  is sufficiently large then  $\langle \cos \theta_{ij} \rangle$  will decay exponentially [7]

$$\langle \cos \theta_{ij} \rangle = \exp\left(-\frac{|i - j|l}{l_p}\right). \quad (1.14)$$

Eqn. 1.14 represents an exact definition of the persistence length. For  $|i - j|l \ll l_p$  it is seen that  $\langle \cos \theta_{ij} \rangle \approx 1$  which means that the segment is approximately linear thus suggesting a high degree of correlation between the monomers. Conversely, for  $|i - j|l \gg l_p$  it is clear that  $\langle \cos \theta_{ij} \rangle \approx 0$  which directly indicates that the angle between the tangent vectors may assume random values from  $0$ – $2\pi$  and therefore the monomers are highly uncorrelated. Quantitatively speaking, the persistence length may therefore be described as the length scale over which the correlation between monomers is retained. Some common values of  $l_p$  for flexible and semiflexible polymers have been included in Tab. 1.4.

	PMMA	ssDNA	DNA	RNA
$l_p$ [nm]	0.85	0.75	50	63

Table 1.4: Approximate values for the persistence length of PMMA, ssDNA, DNA and double stranded RNA. Values for PMMA and DNA are taken from Ref. [3] and values for ssDNA and RNA are from Refs. [8] and [9] respectively.

In order to determine  $\langle R^2 \rangle$  for a worm-like chain, Eqn. 1.14 may be substituted directly into Eqn. 1.2

$$\langle R^2 \rangle = l^2 \sum_{i=1}^N \sum_{j=1}^N \exp\left(-\frac{|i - j|l}{l_p}\right). \quad (1.15)$$

Instead of considering a summation over discrete bonds, Eqn. 1.15 may conveniently be changed into an integration along the contour length

$$l \sum_{i=1}^N \rightarrow \int_0^{l_c} dx_i \quad \text{and} \quad l \sum_{j=1}^N \rightarrow \int_0^{l_c} dx_j \Rightarrow$$

$$\langle R^2 \rangle = \int_0^{l_c} dx_i \int_0^{l_c} dx_j \exp\left(-\frac{|x_i - x_j|}{l_p}\right) \Rightarrow$$

### 1.3. IDEAL WORM-LIKE CHAINS

---

$$\langle R^2 \rangle = 2l_p l_c - 2l_p^2 \left[ 1 - \exp \left( -\frac{l_c}{l_p} \right) \right] \quad (1.16)$$

From Eqn. 1.16 it is immediately evident that the mean-square end-to-end distance of semiflexible polymers with  $l_c \gg l_p$  can be reduced to  $\langle R^2 \rangle \approx 2l_p l_c$  which resembles the result for the freely jointed chain. This means that a long worm-like chain may be modelled as a 3D random walk with a step size  $2l_p$ . It should be noted that this step size coincides with the Kuhn length as stated earlier in this section. In the so called rod-like limit for  $l_c \ll l_p$  the exponential may be expanded [10]

$$\exp \left( -\frac{l_c}{l_p} \right) = \sum_{i=0}^N \frac{1}{i!} \left( -\frac{l_c}{l_p} \right)^i \approx 1 - \frac{l_c}{l_p} + \frac{1}{2} \left( -\frac{l_c}{l_p} \right)^2 \quad (1.17)$$

which allows for reducing Eqn. 1.16 to  $\langle R^2 \rangle \approx l_c^2$ . This result was not highly enlightening as chains, shorter than the inherent persistence length, were expected to behave as rather rigid entities.

#### 1.3.2 Elasticity

When describing the elasticity of worm-like polymers it is instructive to consider both low and high extensions since semiflexible DNA molecules confined inside nanofluidic structures may be stretched to a degree where  $r \approx l_c$ . Owing to the fact that long worm-like chains may be conceived as 3D random walks, the force-extension relationship for  $r \ll l_c$  may be deduced directly from Eqn. 1.12 by simply inserting the proper step size for a semiflexible chain (i.e.  $2l_p$ )

$$f = \frac{3k_B T}{2l_p l_c} r. \quad (1.18)$$

It is important to notice that the inherently large persistence lengths of semiflexible macromolecules drastically reduces the entropic spring constant. By simple insertion,  $k$  of  $\lambda$ -DNA (with 48500 base pairs) which has  $l_p \approx 50$  nm and  $l_c \approx 16.5$   $\mu$ m may be estimated. At  $T = 298$  K,  $\lambda$ -DNA has an entropic spring constant  $k \approx 7.5$  nN/m hence this macromolecule is easily stretched under the influence of modest forces. As mentioned in the previous section, the chain elasticity at low extensions is solely entropy related and strictly linear. However, when a polymer chain is stretched sufficiently, the intrinsic rigidity will come into play thereby inducing nonlinearities<sup>3</sup>. In

---

<sup>3</sup>A somewhat farfetched but relatively good analogy would be pulling at the ends of a coiled-up piece of thread. Initially the thread just unravels until at some stage, when extended to its full

### 1.3. IDEAL WORM-LIKE CHAINS

---

order to quantitatively estimate when the transition between predominantly linear and nonlinear elasticity occurs during polymer stretching, the characteristic force  $f_{\text{ch}} = k_B T / l_p$  can be considered. For flexible polymers such as PMMA  $f_{\text{ch}} \approx 4$  pN and for semiflexible chains  $f_{\text{ch}} \approx 0.08$  pN. Under an applied force  $f < f_{\text{ch}}$  the elasticity is predominantly governed by entropy since  $r \ll l_c$ . In this regime Eqn. 1.18 adequately describes the elastic response of semiflexible macromolecules. For  $f > f_{\text{ch}}$  nonlinear elastic behavior must be anticipated. No analytical expression, which is equally valid in both regimes, exist. However, John F. Marko and Eric D. Siggia have proposed an interpolation formula which is asymptotically correct in the limits  $f \ll f_{\text{ch}}$  and  $f \gg f_{\text{ch}}$  [11]

$$f = \frac{k_B T}{l_p} \left[ \frac{1}{4 \left( 1 - \frac{r}{l_c} \right)^2} - \frac{1}{4} + \frac{r}{l_c} \right]. \quad (1.19)$$

At low extensions Eqn. 1.19 predicts a linear relationship between  $f$  and  $r$  and at high extensions  $f$  diverges strongly  $f \propto (1 - r/l_c)^{-2}$  which may be seen in Fig. 1.7. The formula proposed by Marko and Siggia agrees well with empirical data on DNA stretching [11, 12]. Other analytical models dealing with the elasticity of worm-like chains have been presented and interested readers should refer to Ref. [13] in which a selected variety is conveniently presented.

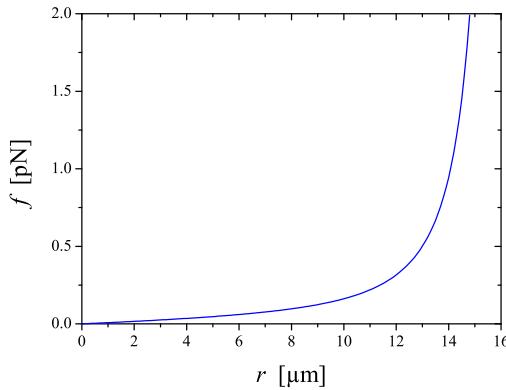


Figure 1.7: Graph showing  $f$  as a function of  $r$  according to Eqn. 1.19:  $l_p = 50$  nm,  $l_c = 16.5$   $\mu\text{m}$  and  $T = 298$  K. At low extensions  $f \propto r$  and at high extensions  $f$  diverges strongly.

---

contour length, it abruptly becomes rigid.

## **1.4. EXCLUDED VOLUME EFFECTS**

---

Before ending this section, a noticeable difference in the elasticity of freely jointed and worm-like chains should be commented. In the high force regime  $f$  diverges less strongly for freely jointed chains  $f \propto (1 - r/l_c)^{-1}$  and this might seem counterintuitive. The explanation is buried in a subtle difference in the rigidity of the persistent Kuhn segments composing the two types of chains. In a freely jointed chain, the Kuhn segments (i.e. the inter-monomer bonds) are per definition completely rigid. Conversely, the persistent segments in a worm-like chain may exhibit some degree of flexibility since fluctuations and discrete bending modes are not suppressed. This lack of complete rigidity of the Kuhn segments effectively enhances the divergence of  $f$  at high extensions.

## **1.4 Excluded Volume Effects**

The description of polymer conformations in terms of random walks merely represents a mathematically convenient simplification which allows for deducing rough analytical relations between e.g. the end-to-end distance and the number of monomers. In the freely jointed and worm-like chain models, the monomer units are mere mathematical points with no spatial extent or physical attributes. Therefore ideal chains are basically one dimensional constructs wriggling around in space in a highly stochastic manner. This non-physical depiction of a polymer is most strongly manifested in the ideal freely jointed chain in which correlation between bond vectors is entirely neglected. This in turn infers that even neighboring monomers may co-exist at the exact same spatial coordinate. Real macromolecules are composed of monomers with a spatial extension. In addition, the monomers may not unambiguously be charge-neutral in aqueous solutions. Polymers which are not charge neutral in solution are called polyelectrolytes and prominent members of this family of macromolecules are proteins and DNA. More specifically nucleotides will be negatively charged in solution and therefore tend to repel each other. Due to steric hindrance and potential electrical repulsion, a certain excluded volume  $\chi$  will be associated with each monomer which basically means that chain segments cannot intersect. A real polymer is therefore more adequately modelled by a selfavoiding walk. In the following we will investigate how excluded volume may alter the conformation of a single polymer molecule in a good solvent. In a good solvent, the monomers have no energetic incitement to gather in clusters in order to expel solvent

## 1.4. EXCLUDED VOLUME EFFECTS

---

molecules and  $\chi > 0^4$ .

The freely jointed and worm-like chain models both predict an end-to-end distance  $R \propto l_c^\nu$  with  $\nu \approx 1/2$  when  $N \gg 1$ . Taking the selfavoidance of real polymers, suspended in a good solvent, into consideration it is clear that  $\nu = 1/2$  must be a minimum value. In order to incorporate the influence of selfavoidance in a model for describing real polymers, Flory argued that the size of a flexible coil in solution is the result of a subtle balance between selfavoidance and entropy. Excluded volume interactions surely serve as to increase the coil size whereas entropy favors small end-to-end distances as evident from Eqn. 1.9. Flory assumed that the polymer chain could be approximated by an ensemble of  $N$  uncorrelated monomers evenly distributed within a volume  $r^3$  (i.e. the polymer is modelled as a "gas" of monomers)<sup>5</sup>. The number density of monomers within this volume is  $c = N/r^3$  and the probability of excluded volume overlap between monomers is given by  $\chi N/r^3$ . For a single monomer, the free energy cost associated with excluded volume interactions is of the order  $k_B T$  and therefore the overall energy cost  $F_{\text{ex}}$  is given by

$$F_{\text{ex}} \approx \sum_{i=1}^N k_B T \chi \frac{N}{r^3} = k_B T \chi \frac{N^2}{r^3}. \quad (1.20)$$

The entropic contribution to the free energy  $F_S$  is simply the energy required to induce an extension  $r$  and may be estimated from Eqn. 1.10. Therefore the total free energy of the polymer is

$$F = F_{\text{ex}} + F_S \approx k_B T \chi \frac{N^2}{r^3} + \frac{3k_B T r^2}{2\langle R^2 \rangle} \approx k_B T \left( \chi \frac{N^2}{r^3} + \frac{r^2}{Nl^2} \right). \quad (1.21)$$

In the final expression, the constant pre-factor (i.e.  $3/2$ ) of the entropic term has been omitted and  $\langle R^2 \rangle$  has been replaced by  $Nl^2$  which holds for an ideal freely jointed chain. Flory's prediction of the coil size  $R_F$  is found by minimizing the free energy

$$\begin{aligned} \frac{\partial F}{\partial r} &= k_B T \left( -3\chi \frac{N^2}{r^4} + 2 \frac{r}{Nl^2} \right) = 0 \Rightarrow \\ R_F &\approx \chi^{1/5} N^{3/5} l^{2/5}. \end{aligned} \quad (1.22)$$

---

<sup>4</sup>In poor or non-solvents  $\chi$  may actually assume negative values and the polymer chain is compacted.

<sup>5</sup>Surely this represents a tremendous simplification of the intrinsic structure of polymers and in the analysis within this section, constant pre-factors of order unity will be neglected.

## 1.4. EXCLUDED VOLUME EFFECTS

---

In the case of a freely jointed polymer consisting of charge neutral and isotropic monomers it may be assumed that the excluded volume  $\chi \approx l^3$  and therefore  $R_F \approx N^{3/5}l$  which effectively shows that selfavoidance tends to swell the coil. Considering a charged semiflexible polymer such as DNA, another estimate of  $\chi$  is clearly needed. Such macromolecules may roughly be perceived as a sequence of selfavoiding and persistent segments as schematically shown in Fig. 1.8 (a).

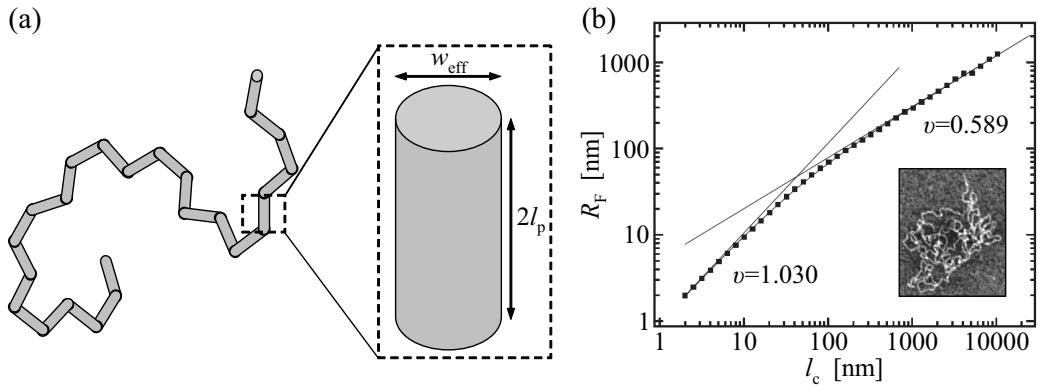


Figure 1.8: (a) Schematic showing how a semiflexible charged polymer may be perceived as a sequence of discrete rods each with a length  $2l_p$  and width  $w_{\text{eff}}$ . (b) Log-log plot showing the end-to-end distance of  $\lambda$ -DNA (with different number of base pairs) as a function of  $l_c$ . Notice the smooth interpolation between the rod-like and selfavoiding regime. The value of  $\nu$  has been indicated for each regime. The inset shows an AFM image of a specific DNA molecule [15].

Each segment may be represented by a rod having a length  $\sim 2l_p$  and an effective width  $w_{\text{eff}}$ . The effective width is the intrinsic width  $w$  plus a contribution which depends in a complicated manner on the linear charge density of the polymer and the ionic content of the aqueous solution<sup>6</sup>. Based on Ref. [14], the excluded volume of a charged and persistent rod may be estimated

$$\chi \approx \pi w_{\text{eff}} l_p^2 \approx w_{\text{eff}} l_p^2. \quad (1.23)$$

In order to directly apply Eqn. 1.22, the  $N$  uncorrelated monomers should be substituted with  $l_c/(2l_p)$  independent rods and furthermore the bond length  $l$  must be

---

<sup>6</sup>The intrinsic and effective width of DNA will be properly addressed in Chap. 2 but for the analysis in this section a more thorough description is immaterial.

## 1.5. CHAPTER SUMMARY

---

replaced with  $2l_p$

$$R_F \approx (w_{\text{eff}} l_p^2)^{1/5} \left( \frac{l_c}{2l_p} \right)^{3/5} (2l_p)^{2/5} \approx (w_{\text{eff}} l_p)^{1/5} l_c^{3/5}. \quad (1.24)$$

As compared to  $R = l_p^{1/2} l_c^{1/2}$ , which holds for a non-selfavoiding long semiflexible polymer, it is seen from Eqn. 1.24 that excluded volume effects adds a weak dependence on  $w$  and swells the coil by increasing the exponent of  $l_c$  from  $\nu = 1/2$  to  $\nu = 3/5$ . In the rod-like limit where  $l_c \approx l_p$  selfavoidance should have negligible influence on the conformation due to the persistent nature of semiflexible polymers. Therefore it may be assumed that  $R_F = R \approx l_c$  for short chains (i.e.  $\nu = 1$ ). Francesco Valle *et al.* conducted atomic force microscopy measurements on DNA (deposited on mica glass) with  $l_c$  ranging from 2–10000 nm in order to determine the end-to-end distance [15]. As seen from Fig. 1.8 (b), two distinct regimes were identified for which  $\nu = 1.030$  and  $\nu = 0.589$  respectively. The transition between the rod-like and selfavoiding regime occurs at  $l_c \approx 40$  nm which approximately coincides with the persistence length of DNA. Eqn. 1.24 agrees rather well with the experimental results presented by Valle *et al.*: (i) The predicted scaling exponent  $\nu = 3/5 \approx 0.589$  and (ii) the pre-factor  $(w_{\text{eff}} l_p)^{1/5}$  deviates only by a multiplication factor of order unity.

Leaning back and revising the crudeness of the approach towards deducing an analytical solution for  $R_F$  of semiflexible and selfavoiding polymers, it is rather intriguing that Eqn. 1.24 provides such a competent estimate. We will return to the conformation and elasticity of confined semiflexible and selfavoiding macromolecules in Chap. 3.

## 1.5 Chapter Summary

Actually counting the vast number of macromolecules surrounding us would be a futile Sisyphus work. Polymers are characterized by their monomer units and the explicit internal architecture. Synthetic polymers, such as dendrimers or other highly branched polymers, may exhibit a high degree of structural complexity, but they are still polydisperse and characterized by stochastic variations. Conversely biopolymers are commonly monodisperse and their exact primary, secondary and tertiary structures have been highly optimized to serve specific purposes.

## **1.5. CHAPTER SUMMARY**

---

Different layers of complexity is associated with the analytical description of polymers. Initially it may be assumed that polymers behave as ideal flexible entities which allows for extracting quantitative information on e.g. the end-to-end distance and elasticity by utilizing simple random walk statistics. However, macromolecules such as DNA and double stranded RNA are semiflexible and may therefore more adequately be described in terms of persistent random walks composed of larger step sizes. Roughly speaking the inclusion of semiflexibility increases the end-to-end distance  $R$  and reduces the effective spring constant  $k$ . Taking the excluded volume of the individual monomers into consideration adds yet another layer of complexity. In good solvents, long polymer chains (flexible or semiflexible) will swell as a direct result of the increase in the scaling exponent  $\nu$  associated with the contour length. In the case of semiflexible polymers a slight dependence on the width is also added.

In the next chapter, the chemical composition, double helical structure and sub-persistence length bending of DNA will be introduced.

# Chapter 2

## Unravelling DNA

---

In 1953 Francis Crick and James D. Watson presented the double helical model of deoxyribonucleic acid (DNA) [16] which was partially based on X-ray diffraction images acquired by Rosalind Franklin. DNA is a highly complex biopolymer composed of monomer entities known as nucleotides<sup>1</sup> which effectively constitute the genome of prokaryotic and eukaryotic cells<sup>2</sup>. As an example, the human genome is encoded by the specific nucleotide sequence contained within 23 chromosomes and the fully extended macromolecule has an astonishing contour length of approximately 1 m.

In this chapter the chemical composition and physical structure of double stranded DNA will be presented. Furthermore localized DNA bending, induced by specific nucleotide sequences, will be introduced in a cursory manner. The main focus throughout the chapter will be on linear B-form DNA and unless otherwise stated information has been adopted from Ref. [17].

---

<sup>1</sup>It is a matter of preference whether the term nucleotides or base pairs (i.e. pairs of complementary nucleotides) is used to designate the monomer entities of DNA. In this report the former term will be used.

<sup>2</sup>In prokaryotic (pro=before and karyon=kernel) cells, the DNA is usually circular and not confined within a nucleus. In eukaryotic (eu=true and karyon=kernel) cells such as human cells, the DNA is usually linear and confined within a nucleus.

## 2.1 The Double Helical Structure of DNA

Before commencing a thorough description of the chemical composition and physical structure of DNA it is instructive to perform a systematic macromolecule dissection in order to identify the basic constituents of this complex biopolymer. One might call this "Democritus' approach towards understanding DNA". DNA consists of two complementary single strands which are held together by hydrogen bonds, hydrophobic interactions and Van der Waals forces. Each strand is composed of nucleotides coupled together by covalent phosphodiester bonds. The individual nucleotides, which constitute the primary structure of DNA, are made up of (i) an aromatic base, (ii) a deoxyribose sugar and (iii) a phosphate group, which effectively means that DNA is composed of carbon, hydrogen, oxygen, nitrogen and phosphate atoms. In the following the chemical composition of the aromatic bases and the deoxyribose sugar will be described. Subsequently the formation of nucleotides and the arrangement of these in the double helical DNA structure will be addressed.

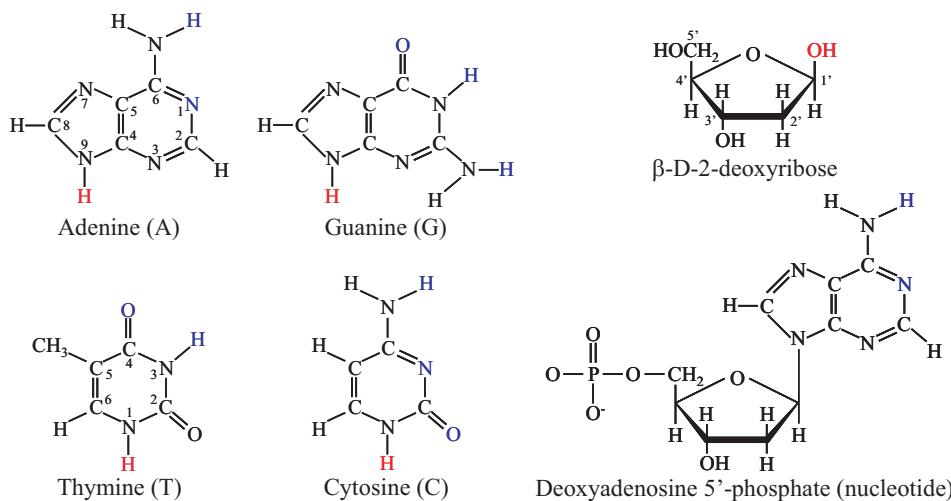


Figure 2.1: Left: Schematic representation of the four aromatic bases. The ring-molecule numbering system for the purine and pyrimidine bases are shown on adenine and thymine respectively. When the **hydrogen atoms** on the bases react with the **hydroxyl group** on the deoxyribose sugar, a glycosidic bond is formed. The **blue** atoms are involved in the hydrogen bonds between complementary bases. Right: Schematic of  $\beta$ -D-2-deoxyribose. The numbering system is indicated on the deoxyribose molecule. The four carbon atoms C1'-C4' are not shown. Furthermore a specific nucleotide has been included. Notice the attachment of the phosphate group at C5' and the glycosidic bond between C1' and the base.

### 2.1.1 The Basic DNA Building Blocks

In Fig. 2.1, the four heterocyclic aromatic bases most commonly observed in DNA may be seen<sup>3</sup>. Adenine (A) and guanine (G) are purine bases exhibiting a high degree of similarity. The major difference is the chemical groups associated with C6 and C2. Thymine<sup>4</sup> (T) and cytosine (C) are pyrimidine bases that differ mainly in the groups connected to C4 and C5. Due to the aromatic ring structure, both purine and pyrimidine bases are rigid and planar molecules that may easily be stacked within the double helix. In 1952 Erwin Chargaff presented empirical results, based on analysis of DNA, showing that the ratio between the contents of A and T and furthermore G and C assumed near-unity values [18]. This experimental finding was a good indication of the complementarity between A and T as well as G and C. Due to differences in electronegativity (refer to Tab. 2.1), hydrogen atoms connected to nitrogen have a partial positive charge. Conversely, the carbonyl oxygen atoms and nitrogen atoms have a net negative charge. This allows for the formation of non-covalent directional hydrogen bonds between the complementary bases as seen in Fig. 2.2 (b). The bases A and T form two Watson-Crick hydrogen bonds whereas G and C form three which effectively makes G·C<sup>5</sup> base pairs more stable. As hydrogen bonds are weak, they may be deformed and twisted with relative ease thereby increasing the flexibility of the DNA molecule.

	H	C	N	P	O
$\chi_{el}$	2.1	2.5	3.0	2.1	3.5

Table 2.1: Values of the electronegativity  $\chi_{el}$  for the five different elements in the nucleotides. Values are adopted from [19]

The sugar moiety present in nucleotides is a pentose abbreviated  $\beta$ -D-2-deoxyribose. A schematic of the molecule is provided in Fig. 2.1. Deoxyribose is a derivative of  $\beta$ -D-ribose which is found in e.g. RNA and adenosine 5'-triphosphate (ATP). ATP functions as a vital energy carrier within the cell. In ribose, a hydroxyl group is connected to the C2' atom instead of the hydrogen atom found in deoxyribose. A

<sup>3</sup>Due to tautomerization other chemical isomeric forms of the bases exist. In this report the most common forms, keto and amine, are considered.

<sup>4</sup>In ribonucleic acid (RNA), thymine has been exchanged with uracil which is another base.

<sup>5</sup>In this report, base pairs will be denoted by e.g. A·T whereas nucleotide sequences are written e.g. AATGC in the 5'→3' direction (refer to Fig. 2.2).

## 2.1. THE DOUBLE HELICAL STRUCTURE OF DNA

covalent glycosidic bond between an aromatic base and deoxyribose can be formed in a condensation reaction between the hydrogen atom situated at N1 or N9 (depending on whether the base is a pyrimidine or a purine) and the hydroxyl group on C1'. This reaction results in the formation of a nucleoside. By further introducing a negatively charged phosphate group as a substitute for the hydroxyl group at C5', a nucleotide is formed. It is worth noticing that the nucleotides are not charge neutral and therefore DNA molecules are negatively charged polyelectrolytes. The lack of charge neutrality allows for manipulating DNA using electrical fields—this has especially proven beneficial in gel electrophoresis where multidisperse DNA samples may be sorted according to size.

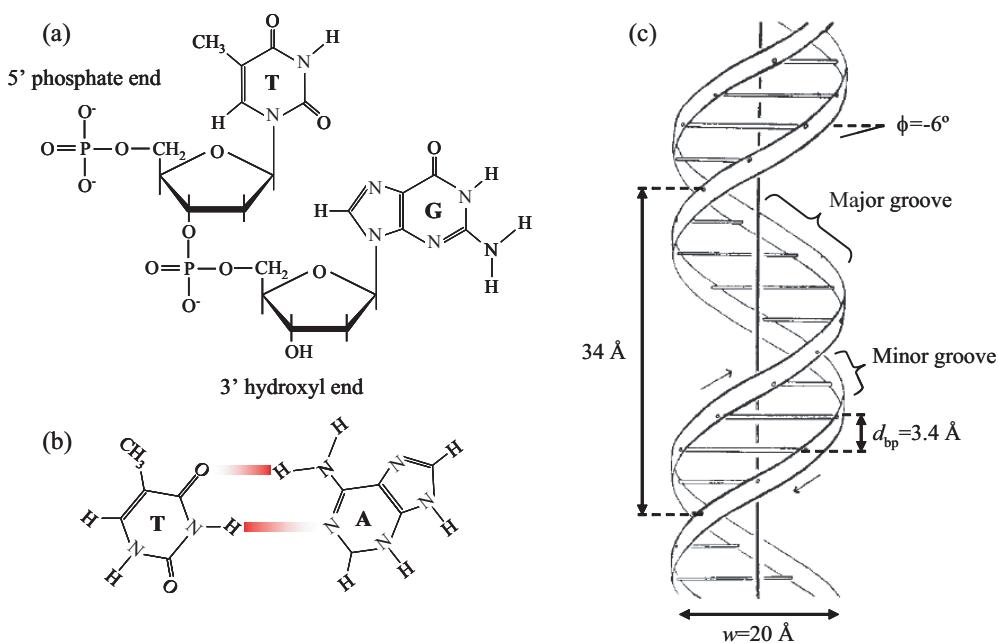


Figure 2.2: (a) Schematic of a TG dinucleotide showing the phosphodiester bond and the chemically and biologically different 5' and 3' ends. The bases are connected to deoxyribose via a glycosidic bond. (b) Schematic of the hydrogen bonds between A and T. The O-N/N-N distance in the upper/lower hydrogen bond is approximately 2.85 Å. (c) The original schematic of the B-DNA structure presented by Watson and Crick [16]. Helical parameters and descriptions have been added. As evident, the distance between base pairs  $d_{bp} = 3.4 \text{ \AA}$ , the helix pitch is 34 Å, the base pair tilt  $\phi = -6^\circ$  and the diameter  $w = 20 \text{ \AA}$ .

The individual nucleotides, that constitute a single stranded DNA molecule, are connected via phosphodiester bonds formed between C5' and C3' on neighboring

## 2.1. THE DOUBLE HELICAL STRUCTURE OF DNA

---

nucleotides as shown in Fig. 2.2 (a). Single DNA strands have two distinct ends, abbreviated 5' and 3', which exhibit highly different chemical and biological properties<sup>6</sup>. The 5' and 3' ends are terminated by a phosphate and a hydroxyl group respectively. During the *in vivo* biochemical polymerization, additional nucleotides are added to the 3' end with the aid of an enzyme known as DNA polymerase. Further chain elongation may effectively be prevented if at some stage a dideoxynucleotide is added to the 3' end. Dideoxunucleotides resembles nucleotides but the hydroxyl group at C3' has been substituted with a hydrogen atom which effectively suppresses formation of the phosphodiester bond.

### 2.1.2 Hybridization and Different Conformations

Two complementary strands may combine in a hybridization process<sup>7</sup> which results in the formation of a double helical DNA molecule as seen in Fig. 2.2 (c). Due to the aforementioned polarity, the two hybridizing strands align in an antiparallel manner. The double helix is stabilized by the hydrogen bonds between complementary bases and Van der Waals forces and hydrophobic interactions associated with base stacking. The energy related to the formation of hydrogen bonds is predominantly dependent on base composition (i.e. A·T or G·C) and is fairly constant. The energy associated with base stacking is conversely sensitive to the exact base pair sequence which ultimately infers that the intrinsic shape of the double helix must be heterogenous. The most common double helical structure is known as B-DNA in which the helices are right-handed. The spacing between neighboring base pairs  $d_{bp} = 3.4 \text{ \AA}$  and the diameter is  $w = 20 \text{ \AA}$ . Since the sugar-phosphate backbone is negatively charged, the effective width  $w_{eff}$ , taking electrical screening of the negative phosphate groups into account, of DNA in solution is highly dependent on the ionic content. In Ref. [20], it was observed that  $w_{eff}$  assumed values of  $29.5 \text{ \AA}$  and  $157 \text{ \AA}$  at monovalent salt concentrations of  $1 \text{ M}$  and  $0.01 \text{ M}$  respectively. The helices of B-DNA goes through a complete  $360^\circ$  turn within a sequence of 10 successive nucleotides<sup>8</sup> thus giving rise to a helix pitch of  $34 \text{ \AA}$  as seen in Fig. 2.2 (c). The base pairs do not align perpendicular to the helical axis but are slightly inclined as seen

---

<sup>6</sup>The existence of two distinct ends gives DNA strands a direction/polarity.

<sup>7</sup>The opposite process is termed denaturation and is favored by e.g. alkaline or acidic environments and high temperatures.

<sup>8</sup>For B-DNA in solution 10.5 base pairs constitute a complete helical turn.

## 2.1. THE DOUBLE HELICAL STRUCTURE OF DNA

---

by  $\phi = -6^\circ$ . As a result of the interwinding of the strands, two distinct grooves are formed: The major groove which is 22 Å wide and the minor groove having a width of 12 Å. These grooves allow e.g. proteins to access the genetic information.

It should be mentioned that linear B-DNA is just one out of many different DNA structures. Circular DNA, in which the double helix forms a loop, is for example found in the prokaryotic cells of bacteria. Besides the linear and circular conformations, cruciform and triplet DNA structures have also been observed. In the latter, three single strands are intertwined. As examples of DNA structures bearing resemblance to B-DNA, the characteristics of A-DNA and Z-DNA, which are also believed to be biologically active, should be briefly mentioned. A-DNA occurs for example in dehydrated B-DNA samples and the helices are right-handed. A-DNA has a slightly larger diameter and the spacing between neighboring base pairs is reduced to 2.55 Å. As seen from Tab. 2.2, the base pair tilt angle is  $\phi = 20^\circ$  which differs substantially from the tilt angle in B-DNA. As a result of this drastic tilt angle, A-DNA is characterized by a shallow and wide minor groove and a more narrow and deep major groove. Z-DNA is believed to be facilitated by specific alternating nucleotide sequences, high salt content and DNA supercoiling<sup>9</sup>. The most prominent difference between B-DNA and Z-DNA is the orientation of the double helix: The helices in Z-DNA are left-handed. Furthermore the diameter of Z-DNA is slightly smaller and  $d_{bp}$  is increased to 3.7 Å as seen in Tab. 2.2.

	A-DNA	B-DNA	Z-DNA
Helix orientation	Right	Right	Left
$d_{bp}$ [Å]	2.55	3.4	3.7
$w$ [Å]	23	20	18
$\phi$	20	-6	7
$N_{hel}$	11	10	12

Table 2.2: Helix parameters for A, B, and Z-DNA based on X-ray diffraction data.  $d_{bp}$  is the base pair spacing,  $w$  is the double helix diameter,  $\phi$  is the tilt angle of the constituent base pairs and  $N_{hel}$  is the number of nucleotides per helical turn.

---

<sup>9</sup>Supercoiling e.g. allows for compacting DNA in spatially separated topological domains (the human genome is divided into 23 different domains called chromosomes) thereby decreasing the effective occupied volume. For further information on supercoiling refer to Refs. [17] and [21]

## 2.2. LOCALIZED DNA BENDING

---

The double helical structure of DNA constitutes one of natures sophisticated, intriguing and marvellous accomplishments. The genetic information encoded in DNA is sufficiently protected from chemical degradation while still allowing enzymes and proteins to address the bases of the nucleotides thereby enabling e.g. transcription and replication. Furthermore the double helical structure ensures, that if a single strand is damaged, by for example genotoxic chemicals or ultraviolet radiation, the undamaged complementary strand provides a fresh template carrying the exact same information (i.e. a biological backup copy).

## 2.2 Localized DNA Bending

The double helical structure makes DNA a rather rigid biopolymer with a persistence length  $l_p$  of approximately 50 nm corresponding to 150 base pairs (bp). Genomic-length DNA are long macromolecules composed of several thousands or millions of base pairs. As an example, a certain segment of bacteriophage  $\lambda$ -DNA, found in *Escherichia coli* (i.e. *E. coli*) which is a bacteria harbored in the lower intestine, consists of 48.5 kbp and has a fully stretched contour length  $l_c = N_{bp}d_{bp} \approx 16.5 \mu\text{m}$  ( $N_{bp}$  is the number of base pairs). Short DNA molecules with  $\sim 150$  bp or local segments of genomic-length DNA will behave as rigid rods. However, specific nucleotide sequences may lead to unexpected bending on a length scale  $\sim l_p$ .

### 2.2.1 Experimental Results

Some of the first experimental results showing how specific nucleotide sequences may induce DNA bending was provided by Marini *et al.* in 1982-83 [22, 23]. During investigation of a 490 bp fragment of kinetoplast DNA<sup>10</sup> from the parasite known as *Leishmania tarentolae*, an anomalous electrophoretic behavior was observed: The 490 bp fragment migrated through a polyacrylamide gel, which has a pore size of  $\sim 2$  nm, at rate which suggested that the fragment consisted of 1380 bp. Clearly the migration of bent DNA molecules through a gel-matrix with narrow pores would be cumbersome, and Marini *et al.* correctly suggested that the anomalous behavior could be the result of a curved DNA conformation induced by specific base pair sequences. Numerous experiments conducted using different techniques have ver-

---

<sup>10</sup>Kinetoplast DNA is found inside mitochondria (the cells power plants) in parasites.

## 2.2. LOCALIZED DNA BENDING

---

ified that the existence of periodically occurring A-tracts consisting of continuous sequences of 4–9 A·T base pairs do in fact cause double helix bending. In the following, the focus will be on the results presented in Refs. [24, 25, 26].

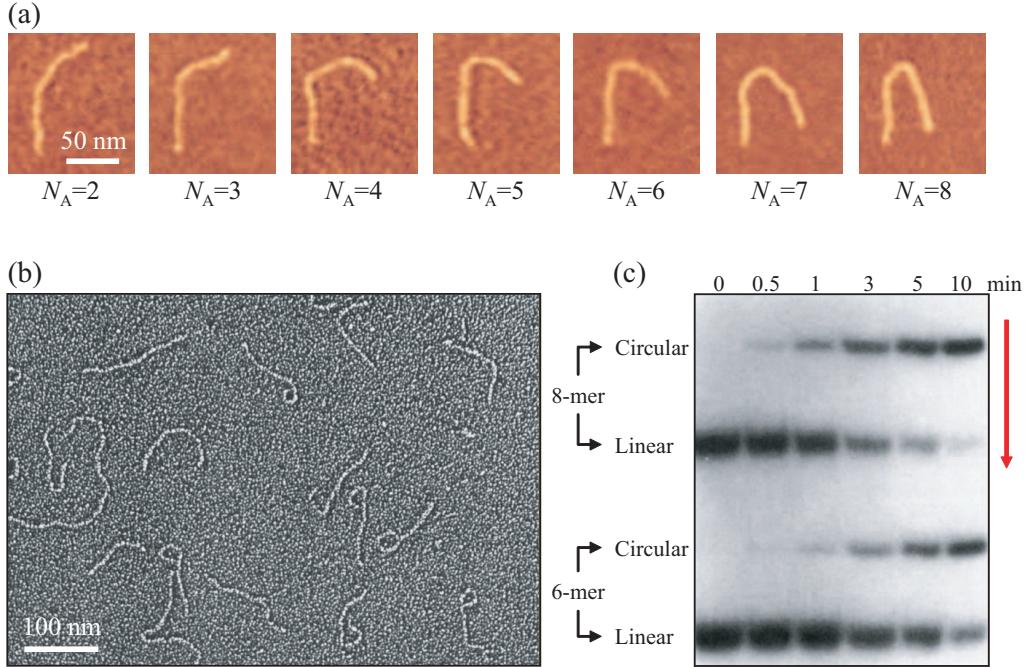


Figure 2.3: (a) Scanning force microscopy images of DNA with  $\sim 410$  bp deposited onto mica substrates [24]. The image sequence shows an increase in DNA bending as the number of phased A-tracts  $N_A$  in the middle of the molecule is raised from 2–8. The average bending angle per A-tract is  $13.5^\circ$ . Each A-tract consists of six A·T base pairs followed by four G·C spacer base pairs. Notice how the end-to-end distance is drastically decreased from  $N_A = 2$  to  $N_A = 8$ . (b) Transmission electron microscope image of DNA molecules with 492 bp deposited onto an extremely thin carbon film and coated with tungsten [25]. The bent end-regions consist of 219 bp with approximately 18 phased A-tracts each composed of 4–6 A·T base pairs. It is evident that the majority of the bent segments have a curvature of  $360^\circ$  thus yielding an average bending angle of  $\sim 20^\circ$  per A-tract. (c) Image taken from an experimental cyclization kinetics study in which linear multimers of synthetic 21 bp oligonucleotides have been allowed to cyclize under the influence of T4 DNA ligase for 0–10 min [26]. Each oligonucleotide contains two phased A-tracts with 6 A·T base pairs. The 6-mer and 8-mer contains 126 bp and 168 bp respectively. Linear and circular molecules have been separated by means of polyacrylamide gel electrophoresis. The **arrow** indicates the direction of migration and the shading of discrete regions within the gel is a measurand of DNA concentration. It is evident that the 8-mers more readily form cyclic entities as compared to the 6-mers. The induced bending angle per A-tract was estimated to be  $17\text{--}21^\circ$ .

## 2.2. LOCALIZED DNA BENDING

---

In Fig. 2.3 (a), scanning force microscopy images of bent DNA with approximately 410 bp may be seen. As the number of phased A-tracts near the center of the DNA molecule is incrementally increased from  $N_A = 2$  to 8, the conformation changes from linear to semicircular. Fig. 2.3 (b) shows a transmission electron micrograph of DNA fragments with 492 bp. The fragments have been obtained by cloning a highly bent 219 bp segment into plasmid DNA<sup>11</sup>. The 219 bp segment contains approximately 18 phased A-tracts that mediate a complete 360° bent which is rather extreme considering the fact that  $l_c \approx 74$  nm. The image in Fig. 2.3 (c) has been taken from a paper on the cyclization kinetics of DNA molecules with 105–210 bp. Multimers consisting of 5–10 synthetic oligonucleotides (i.e. short sequences of synthesized nucleotides) with 21 bp were investigated. Each oligonucleotide has two A-tracts composed of six A-T base pairs. The multimers were suspended in a solution containing the enzyme T4 DNA ligase, which mediates the cyclization, for varying amounts of time. The degree of cyclization depends primarily on (*i*) the end-to-end distance and (*ii*) the angular orientation of the two ends. Therefore the cyclization kinetics must be intimately coupled to the number of A-tracts. The degree of cyclization for each multimer was analyzed by means of gel electrophoresis which allowed for separating linear and circular segments. As evident from Fig. 2.3 (c), the 8-mers more easily adopted cyclic conformations under the influence of T4 DNA ligase.

### 2.2.2 DNA Bending Models

Different models have been proposed in order to explain the unexpected bending of DNA with sizes comparable to the persistence length. Two of these models will be described in a cursory manner after having listed some important features of primary structures that induce systematic DNA bending:

- Continuous sequences of A-tracts with a periodicity of  $\sim 10.5$  bp<sup>12</sup> will effectively result in cumulative rather than random bending (refer to Fig. 2.4 (c) for a specific sequence which induce bending). If the A-tracts are phased at  $\sim 16$  bp intervals, the DNA will exhibit a zig-zag structure thereby increasing

---

<sup>11</sup>Plasmid DNA is extrachromosomal and is most commonly found in prokaryotic organisms such as bacteria.

<sup>12</sup>This specific value only applies when considering B-DNA in solution where 10.5 bp constitutes a complete helical turn.

## 2.2. LOCALIZED DNA BENDING

---

the end-to-end distance as compared to DNA in which A-tracts are non-present or have a random periodicity.

- Based on experiments and numerical calculations it has been shown that the specific composition of sequences flanking the A-tracts may, to some extent, influence the DNA curvature.
- To ensure substantial DNA bending, the number of A·T base pairs per A-tract should be 3–9. Maximal bending is observed for the median value of 6.
- DNA molecules without A-tracts in the primary structure may also exhibit bending, albeit to a far lesser extent. The exact mechanisms governing the bending of such DNA molecules are less well determined.

The two most prominent models, used for explaining the sequence dependant DNA bending, are the junction and wedge models which have been schematically presented in Fig. 2.4 (a) and (b).

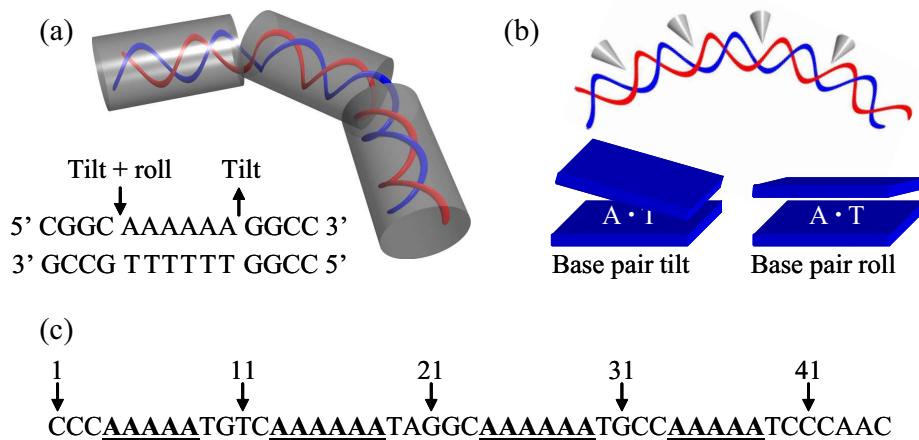


Figure 2.4: (a) Schematic showing a bent DNA segment as envisioned in the junction model. The three cylinders indicate DNA segments with different secondary structures. The inset shows a specific DNA sequence containing an A-tract. The arrows point in the directions of the tilt deflections at the 5' and 3' ends respectively. (b) Schematic showing how "biological wedges" may induce DNA bending according to the wedge model. Furthermore the concepts of base pair tilt and roll between neighboring A·T base pairs have been illustrated. (c) A specific nucleotide sequence constituting a bend locus [17]. Notice the periodic spacing of the A-tracts containing 5–6 A·T base pairs. One A-tract is found for every 10 bp. The complementary strand has not been included.

### **2.3. CHAPTER SUMMARY**

---

In the junction model it is assumed that bending occurs at the interface between DNA segments having different helical structures. In 1983, Arnott *et al.* presented results showing that the secondary structure of polynucleotide duplexes, solely containing A-T base pairs, was heteronomous and quite different from the structure of B-DNA [27]. This experimental finding lead to the suggestion that bending occurs at junctions between B-form DNA and non-B-form DNA containing phased A-tracts. Within the framework of the junction model, DNA bending is believed to be the combined result of tilt and roll deflections of the double helix at the 5' end of an A-tract and tilt deflection at the 3' end. The directions of the tilt deflections may be seen in Fig. 2.4 (b) and the roll causes an opening of the double helix towards the minor groove.

The wedge model constitutes a different approach towards understanding and explaining DNA bending. In this model it is assumed that each AA-dinucleotide in the complementary strands represents a "biological wedge" which serve as to deflect the double helical axis of DNA locally. According to the wedge model, individual AA-dinucleotides are associated with base pair tilt and roll angles of approximately  $2.4^\circ$  and  $8.4^\circ$ . The base pair tilt stretches the sugar-phosphate backbone between the two neighboring adenine bases and the roll causes a widening of the major groove (refer to Fig. 2.4 (b)). Localized DNA bents may occur as a result of the cumulative vectorial contributions from a vast number of phased AA-dinucleotide sequences.

Conceptually the junction and wedge models are very similar when it comes to explaining DNA curvature induced by phased A-tracts and both models have provided adequate predictions of the conformations of many DNA molecules with well determined nucleotide sequences. Before ending this chapter it should be mentioned that the interaction between DNA and several proteins may also lead to rather extreme changes in the local conformation of DNA. The experimental investigation of the implications and mechanisms related to DNA-protein interactions is a scientific field in fast progress.

## **2.3 Chapter Summary**

Despite the high degree of structural complexity, DNA molecules are biopolymers consisting of rather simple building blocks. Purine and pyrimidine bases attached to

### **2.3. CHAPTER SUMMARY**

---

a deoxyribose sugar by a covalent glycosidic bond forms nucleosides, which may be phosphorylated thereby producing nucleotides. Single stranded DNA is composed of sequences of nucleotides coupled together by phosphodiester bonds in a biological polymerization process. By hybridization of two complimentary single strands, double helical DNA is formed. Despite the anticipated rigidness of short DNA molecules or local segments with approximately 150 bp, specific base pair sequences, such as phased A-tracts, may induce local bending on a length scale comparable to the inherent macromolecule persistence length. In order to explain such abnormal or unanticipated conformations, the junction and wedge models have been developed.

In the next chapter the conformation and behavior of DNA molecules confined inside nanofluidic structures will be considered.

# Chapter 3

## DNA in Nanoconfinements

---

The conventional empirical way of studying DNA is by performing macro-scale *in vitro*<sup>1</sup> experiments such as gel electrophoresis and X-ray diffraction measurements. In the former technique, an electrical potential is applied across a porous gel, consisting of purified agarose or polyacrylamide, and based on differences in mobility a multidisperse DNA sample may be sorted according to size and/or conformation. The tremendous development within the field of nanotechnology has allowed for fabricating fluidic structures with dimensions comparable to the persistence length of DNA, thereby providing the possibility of addressing and investigating single DNA molecules on an entirely different and perhaps more appropriate length scale.

When DNA is confined inside fluidic structures which prohibits retainment of the coiled-up conformation, the statics and dynamics of the confined molecules will differ significantly from what is observed in bulk solution. This chapter will initially describe how nano-scale confinement influences DNA conformation and elasticity. Subsequently a commonly used measurement technique, known as epifluorescence microscopy, will be described and a brief survey of recent scientific research within this fast-growing field will be provided.

---

<sup>1</sup>in vitro (within the glass) means that experiments are performed under controlled conditions in e.g. a test tube. On the other hand *in vivo* (within the living) experiments involves tissue from living organisms.

## 3.1 Confinement Induced Effects

Imagine confining a long DNA molecule inside a rigid and repelling cylindrical fluidic structure with a diameter  $D$ . Clearly the DNA will simply adopt a curled-up conformation if  $D \geq R_F$ . However if  $D$  becomes comparable to the persistence length, things change dramatically.

### 3.1.1 The de Gennes Regime

If the cylinder is scaled down such that  $R_F \gg D \gg l_p$ , the DNA will effectively be stretched along the axis of the cylinder and the extension  $r > R_F$  will depend on  $D$ . It is important to realize that the confinement induced stretching purely is a result of the inherent selfavoiding nature of DNA. The confinement increases the monomer number density thereby enhancing self-exclusion effects. As  $D \gg l_p$ , an equivalent non-selfavoiding ideal worm-like chain would be able to back-fold and  $r$  would be totally independent of  $D$  since the random walk components parallel to the cylinder axis would be unaffected [28]. Based on scaling arguments, de Gennes derived a simple expression for the extension of a confined selfavoiding polymer [28, 29]. Imagine the confined polymer as being stretched out in a series of blobs each storing contour length  $l_b$  as seen in Fig. 3.1. Due to selfavoidance, the individual blobs will (i) repel like hard spheres and (ii) behave as discrete Flory coils (i.e. Eqn. 1.24 applies) with an end-to-end distance  $R_b$ . It is reasonable to assume that  $R_b \approx D$  and therefore

$$R_b \approx D \approx (w_{\text{eff}} l_p)^{1/5} l_b^{3/5} \Rightarrow \\ l_b \approx \frac{D^{5/3}}{(w_{\text{eff}} l_p)^{1/3}}. \quad (3.1)$$

As each blob stores the same amount of contour  $l_b$ , the extension is merely given by  $l_c/l_b$  multiplied by the blob size  $R_b$

$$r \approx \frac{(w_{\text{eff}} l_p)^{1/3}}{D^{2/3}} l_c. \quad (3.2)$$

As seen from Eqn. 3.2, confinement increases the scaling exponents of  $l_c$ ,  $w_{\text{eff}}$  and  $l_p$ . A good measurand of the degree of extension is the ratio  $\varepsilon \equiv r/l_c$ . Considering DNA with  $l_p \approx 50$  nm confined inside a cylinder of diameter  $D = 200$  nm, it is seen

### 3.1. CONFINEMENT INDUCED EFFECTS

---

that characteristic values of  $\varepsilon$  are 0.16 and 0.27 for  $w_{\text{eff}} = 3 \text{ nm}$  and  $16 \text{ nm}$  respectively<sup>2</sup>. Extensions amounting to 16–27% of the full contour length are substantial and merely by decreasing  $D$  the confinement induced stretching may be enhanced. It has been shown in Ref. [30], that Eqn. 3.2 is equally valid when considering extension inside confinements with rectangular cross sections provided that  $D$  is replaced by the geometrical average  $\sqrt{hw_c}$  where  $h$  and  $w_c$  are the height and width respectively<sup>3</sup>. As will be described more thoroughly in the next section, experimental measurements on DNA extension inside fused silica and PMMA nanochannels with a rectangular cross section are in reasonable agreement with Eqn. 3.2 [31, 32, 33].

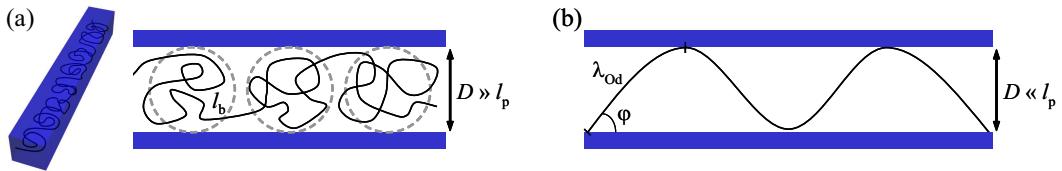


Figure 3.1: (a) Schematic of a confined polymer in the de Gennes regime where  $D \gg l_p$ . The extended polymer is represented by a series of interconnected blobs each storing a contour length  $l_b$ . (b) Schematic of a confined polymer in the Odijk regime where  $D \ll l_p$ . The polymer extension is no longer a result of excluded volume effects and as the polymer is stretched along the axis of the confinement it undergoes a series of deflections with the boundaries. The deflection angle  $\varphi$  and the characteristic Odijk scale  $\lambda$  are indicated.

It is interesting to derive an analytical expression for the effective spring constant  $k_{\text{dG}}$  of confined DNA, since this will enable a description of how thermally induced fluctuations influences the extension length. Initially the increase in free energy as a result of confinement  $F_{\text{con}}$  must be estimated. In a rather simple and elegant scaling analysis, de Gennes argued that each blob is associated with an energy of the order  $k_B T$  and therefore  $F_{\text{con}}$  is merely the sum of  $l_c/l_b$  discrete energetic contributions [28]

$$F_{\text{con}} \approx k_B T \frac{(w_{\text{eff}} l_p)^{1/3}}{D^{5/3}} l_c. \quad (3.3)$$

By inserting values for  $\lambda$ -DNA with  $l_c \approx 16.5 \mu\text{m}$ ,  $l_p \approx 50 \text{ nm}$  and  $w_{\text{eff}} \approx 10 \text{ nm}$ , we may estimate a confinement free energy  $F_{\text{con}} \approx 20 k_B T$  for  $\lambda$ -DNA confined inside

---

<sup>2</sup>These values for  $w_{\text{eff}}$  are measured values of the effective width in solutions with high and low monovalent salt concentrations (refer to Sec. 2.1.2).

<sup>3</sup>When stating values for  $D$  in the next section they will refer to a geometrical average.

### 3.1. CONFINEMENT INDUCED EFFECTS

---

a structure with  $D = 200$  nm. Clearly this is a significant increase in free energy. Utilizing Eqn. 3.2 allows for restating Eqn. 3.3 in terms of the extension  $r$

$$F_{\text{con}} \approx \frac{k_{\text{B}}T}{(w_{\text{eff}}l_{\text{p}})^{1/2} l_{\text{c}}^{3/2}} r^{5/2}. \quad (3.4)$$

The effective spring constant of the confined macromolecule as a function of  $D$  is then determined by differentiating Eqn. 3.4 twice with respect to  $r$  and substituting  $D$  back into the equation

$$k_{\text{dG}} \approx \frac{k_{\text{B}}T}{l_{\text{c}}} \frac{1}{(w_{\text{eff}}l_{\text{p}}D)^{1/3}}. \quad (3.5)$$

The mean-square average extension fluctuations  $\langle \delta r^2 \rangle$  may then be determined [31]

$$\langle \delta r^2 \rangle = \frac{k_{\text{B}}T}{k_{\text{dG}}} \approx (w_{\text{eff}}l_{\text{p}}D)^{1/3} l_{\text{c}} \quad (3.6)$$

From Eqn. 3.6 it is evident that stronger confinement effectively suppresses length fluctuations. This is a rather important result since the standard deviation  $\sigma = \sqrt{\langle \delta r^2 \rangle}$  largely determines the degree of accuracy associated with a single experimental measurement of the extension length  $r$ .

#### 3.1.2 The Odijk Regime

In the de Gennes regime where  $R_{\text{F}} \gg D \gg l_{\text{p}}$ , the extension is a result of the increased impact of excluded volume effects. If the confining fluidic structure is scaled further down such that  $D \ll l_{\text{p}}$ , excluded volume effects must be negligible since persistent macromolecules would not be allowed to coil. Therefore the extension of DNA will be governed by boundary interactions and the intrinsic molecular rigidity. As seen in Fig. 3.1 (b), Theo Odijk envisioned that the contour length was stored in a series of polymer deflections mediated by hard-wall repulsion at the confining surfaces. Furthermore he argued that these deflections would occur over a characteristic length scale  $\lambda_{\text{Od}}$  abbreviated the Odijk length. Without going into the subtle mathematical details, Odijk derived an expression for  $\lambda_{\text{Od}}$  using two different approaches [34, 35]

$$\lambda_{\text{Od}} \approx l_{\text{p}}^{1/3} D^{2/3}. \quad (3.7)$$

Using simple trigonometry, the projected length of an Odijk segment along the axis of the confinement may be approximated by  $l_{\text{Od}} \approx \lambda_{\text{Od}} \cos \varphi$  where  $\varphi$  is the deflection

### 3.1. CONFINEMENT INDUCED EFFECTS

---

angle (refer to Fig. 3.1 (b)). It is reasonable to assume that the deflection angle is small and therefore  $\varphi \approx D/\lambda_{\text{Od}} \approx D^{1/3}/l_p^{1/3} \ll 1$  and the cosine may be expanded in order to determine  $l_{\text{Od}}$

$$\cos \varphi \approx 1 - \frac{1}{2}\varphi^2 \approx 1 - \frac{1}{2}\left(\frac{D}{l_p}\right)^{2/3} \Rightarrow$$

$$l_{\text{Od}} \approx \lambda_{\text{Od}} \left[ 1 - \frac{1}{2}\left(\frac{D}{l_p}\right)^{2/3} \right]. \quad (3.8)$$

From Eqn. 3.8, the confinement induced extension can simply be estimated by multiplying the total number of Odijk segments  $l_c/\lambda_{\text{Od}}$  with the projected length of a single segment

$$r \approx \frac{l_c}{\lambda_{\text{Od}}} l_{\text{Od}} \approx l_c \left[ 1 - A \left( \frac{D}{l_p} \right)^{2/3} \right]. \quad (3.9)$$

In Eqn. 3.9,  $A$  is a numerical pre-factor of order unity which depends e.g. on the cross sectional geometry of the confinement. Typical values of  $A$  ranges from 0.2–0.36 [32, 36]. In the Odijk regime where  $D \ll l_p$ , the extension length gradually approaches  $l_c$ . However, the degree of extension  $\varepsilon$  depends less strongly on  $D$  in comparison to what was found in the de Gennes regime.

Lets now consider the analytical derivation of expressions for the confinement free energy  $F_{\text{con}}$  and the spring constant  $k_{\text{Od}}$  of DNA confined in sub-persistence length structures. After having recognized that in the Odijk regime we are dealing with contour segments instead of contour blobs which effectively means that  $l_b \mapsto \lambda_{\text{Od}}$ , the derivations are rather straightforward. Assuming that each segment contributes with an energy of order  $k_B T$ , the total confinement free energy  $F_{\text{con}}$  is simply given by

$$F_{\text{con}} \approx k_B T \frac{l_c}{\lambda_{\text{Od}}} = B k_B T \frac{1}{l_p^{1/3} D^{2/3}} l_c. \quad (3.10)$$

The constant numerical pre-factor  $B$  in Eqn. 3.10 have a slight dependence on the cross sectional geometry but in general  $B = 1\text{--}2.8$  [37]. It is immediately seen that  $F_{\text{con}}$  is drastically increased in the Odijk regime. Considering  $\lambda$ -DNA (48.5 kbp) confined in a structure with  $D = 20$  nm, values of  $F_{\text{con}}$  are on the order of  $1000 k_B T$ ! The approach towards estimating  $k_{\text{Od}}$  is identical to the aforementioned

route towards determining  $k_{dG}$  and the result will be stated without further ado

$$k_{Od} \approx \frac{k_B T}{l_c} \frac{l_p}{D^2}. \quad (3.11)$$

By comparing Eqn. 3.5 and 3.11 it is evident that  $k_{Od} \gg k_{dG}$  and this nicely reflects the fact, that the elasticity of strongly confined semiflexible polymers is predominantly governed by the intrinsic rigidity. It is possible to derive force-extension relations for confined polymers. However, it is outside the scope of this report to deal with this subject and interested readers should consult Ref. [36].

## 3.2 Investigating DNA in Nanofuidic Structures

Currently, lab-on-a-chip (LOC) applications, which may be perceived as miniaturized laboratories, with passive nanofuidic structures are being utilized for investigating DNA confined inside structures that have dimensions comparable to the biomolecule persistence length. The work entails real-time contour length measurements of DNA stretched inside nanochannels [31, 32, 33, 38], filtering of DNA samples [39], investigation of DNA diffusion in nanoslits [40], entropic recoil effects observed at the interface between micro- and nanofuidic structures [41] and restriction mapping inside nanochannels using certain enzymes which may fragment DNA [42]. The major advantage of incorporating nanofuidic structures is the possibility of probing and investigating single DNA molecules on the appropriate length scale. This allows for extracting information which is not accessible when the DNA assumes its bulk solution coiled-up conformation.

### 3.2.1 Epifluorescence Microscopy

Commonly epifluorescence microscopy is used for visualizing confined DNA and the basic principles of this specific measurement method will be outlined. The incident light from a bright illumination source (e.g. a metal-halide lamp) is initially passed through an excitation filter in order to narrow down the wavelength distribution (refer to Fig. 3.2 (a)). The transmitted light is subsequently reflected off a dichroic mirror and passed into an objective with a high magnification and numerical aperture. In general water or oil immersion objectives are preferred as they reduce refraction. The excitation light is focused onto the confined DNA which is stained

### 3.2. INVESTIGATING DNA IN NANOFUIDIC STRUCTURES

with fluorescent molecules. Commonly dimeric cyanine dyes such as YOYO-1 (excitation/emission wavelength=490 nm/510 nm [43]) are used to stain the DNA at a ratio of one dye molecule per 5–10 base pairs. Dimeric cyanine dyes usually binds by bisintercalation, which simply means that the dye molecules slide in between neighboring base pairs as seen in Fig. 3.2 (a)<sup>4</sup>.

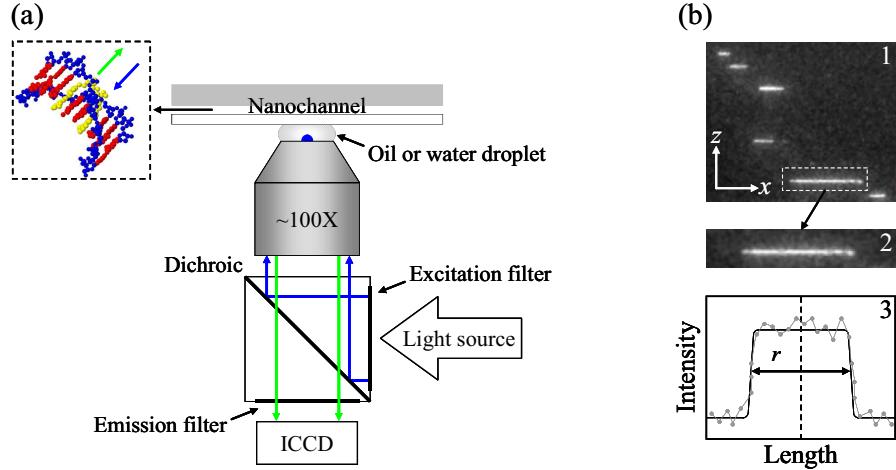


Figure 3.2: (a) Schematic of an epifluorescence measurement setup. Notice the position of the excitation and emission filters. The inset shows a YOYO-1 dimer bound to DNA via bisintercalation. (b) The three discrete steps in the analysis procedure. 1: A selected DNA molecule is isolated and boxed. 2: The intensity is transversely summed ( $z$ -direction) along the DNA center line ( $x$ -direction). 3: The average intensity trace is fit to Eqn. 3.12 which enables the extraction of  $r$ . By analyzing multiple frames the dynamic properties of the confined DNA may be investigated.

Upon DNA binding, the fluorescence of YOYO-1 is enhanced by a factor of  $\sim 3000$  [44] making it ideal for epifluorescence measurements. As may intuitively be understood, the attachment of rather bulky dye molecules to a substantial amount of the nucleotides must have an effect on both  $l_p$  and  $l_c$ . It has been shown, that both the persistence and contour length may be increased by as much as 30% at a saturating staining ratio of one dye molecule per 4 base pairs [45, 46, 47]. In order to avoid bleaching of the dye molecules and photo-nicking of the DNA,  $\beta$ -mercaptoethanol,  $\beta$ -D-glucose, glucose oxidase and catalase are added to the experimental solution in which the DNA is suspended [32]. The emitted fluorescence light is transmitted through the objective and the dichroic and emission filters and collected by an

<sup>4</sup>An artistic impression of bisintercalation may actually be seen on the front page of this report.

intensified charged-coupled device (ICCD). The statics and dynamics of the confined molecules may subsequently be investigated by analyzing the intensity data as outlined in Fig. 3.2 (b). Initially a DNA molecule in a single frame is selected and isolated in a boxing procedure. By transversely summing the intensity along the DNA centerline, a one dimensional average intensity trace is obtained. In Ref. [31] it was argued that the average intensity  $I(x)$  of the extended DNA could be modelled as a convolution of a step function  $I_0$  of length  $r$  (i.e. the true extension length) with a Gaussian point spread function  $R(x) = 1/2\pi\sigma_0^2$ .  $\sigma_0$  is directly associated with the resolution of the optical setup. Therefore the fitting function in Eqn. 3.12, would allow for an extraction of  $r$  in a single frame

$$I(x) = \frac{I_0}{2} \left[ \text{Erf} \left( \frac{x}{\sqrt{2}\sigma_0} \right) - \text{Erf} \left( \frac{x-r}{\sqrt{2}\sigma_0} \right) \right]. \quad (3.12)$$

By analyzing multiple frames, an ensemble of extensions lengths may be obtained which allows for investigating the dynamics of the confined DNA. We will proceed by investigating some experimental results.

#### 3.2.2 Experimental Results

In Ref. [32], the extension of T2 DNA (166 kbp) and  $\lambda$ -DNA (48.5 kbp), inside fused silica nanochannels with  $D$  ranging from 35 nm to 440 nm, was experimentally measured in order to investigate the transition between the de Gennes and Odijk regimes. As immediately evident from the fluorescent intensity traces of confined  $\lambda$ -DNA in Fig. 3.3 (a), the extension increases steadily in the smallest channels until reaching a plateau when  $D$  approaches the dye-adjusted persistence length  $l_p = 57.2$  nm. Recalling the differences in dependency on  $D$  of Eqn. 3.2 and 3.9, this plateau could indicate that a transition between the de Gennes regime and the Odijk regime occurs at  $D \approx 80$  nm. By analyzing the relaxation time of confined DNA as a function  $D$ , a critical crossover length scale  $D_c = 110$  nm was identified.  $D_c$  determines the degree of confinement at which the intrinsic rigidity of DNA influences the extension and elasticity significantly and it should be noted that  $D_c$  coincides with the Kuhn length (i.e.  $2l_p$ ). From the log-log plot of the extension factor  $\varepsilon = r/l_c$  (the dye-adjusted contour length is 18.6  $\mu\text{m}$ ) for confined  $\lambda$ -DNA versus  $D$ , it is evident that  $\varepsilon$  for  $D = 35$  nm clearly deviates significantly suggesting that the extension is predominantly governed by the intrinsic rigidity. The fact that

### 3.2. INVESTIGATING DNA IN NANOFUIDIC STRUCTURES

the data points for  $D = 69\text{--}135\text{ nm}$  are well fit by both de Gennes' and Odijk's models and 3.9) suggests a smooth transition between the two regimes rather than an abrupt jump. A power-law fit to the data for  $D = 69\text{--}440\text{ nm}$ , resulted in an exponent of  $-0.85 \pm 0.05$  which deviates from the expected value of  $-2/3$  and effectively suggests that the DNA extension is more sensitive to the degree of confinement than what would be expected based on Eqn. 3.2.

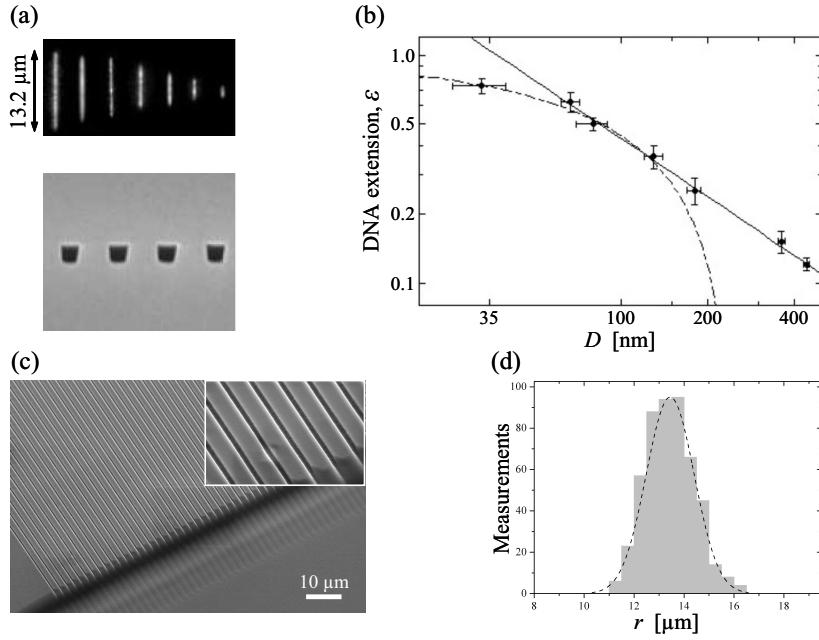


Figure 3.3: (a) Averaged intensity of  $\lambda$ -DNA confined inside fused silica nanochannels with  $D=35, 69, 80, 135, 186, 363$  and  $440\text{ nm}$  (from left to right) and cross-sectional SEM image of  $60\times 80\text{ nm}$  (i.e.  $D = 69\text{ nm}$ ) fused silica nanochannels. (b) Log-log plot of the extension degree  $\varepsilon$  for  $\lambda$ -DNA as a function of  $D$ . The bold line is a power-law fit to the data from the six largest nanochannels (the exponent is  $-0.85 \pm 0.05$ ). The dashed line, which is seen to fit the data from the 3–4 smallest nanochannels, is the Odijk prediction according to Eqn. 3.9 with  $l_p = 52 \pm 5\text{ nm}$  and  $A \approx 0.34$ . Notice how the data point for  $D = 35\text{ nm}$  deviates strongly from the bold line. (c) SEM image of PMMA nanochannels width  $D = 250\text{ nm}$  and a period of  $1\text{ }\mu\text{m}$ . The scale bar is  $10\text{ }\mu\text{m}$ . The inset is merely a magnification. (d) Histogram of the measured extension lengths of a single T4 DNA molecule ( $166\text{ kbp}$ ) confined inside a PMMA nanochannel with  $D = 250\text{ nm}$ . The average extension length, based on analysis of 500 consecutive frames,  $r = 13.4\text{ }\mu\text{m}$  ( $\varepsilon = 0.19$ ) and the standard deviation  $\sigma = 1.0\text{ }\mu\text{m}$ . The dashed line shows the Gaussian curve fit.

Conventionally, LOC applications for investigating confined DNA are made by planar processing of fused silica which is both expensive and rather time consum-

### 3.2. INVESTIGATING DNA IN NANOFUIDIC STRUCTURES

---

ing. In Ref. [33], polymer devices with both micro- and nanofuidic structures were made by means of nanoimprint lithography (NIL, a hot embossing technique [48]) in PMMA. In order to investigate the functionality of the fabricated devices, the extension and thermal fluctuations of T4 DNA (166 kbp), confined inside PMMA nanochannels with  $D = 250$  nm (refer to Fig. 3.3 (c)), was investigated. The dye-adjusted persistence and contour lengths were  $l_p = 62$  nm and  $l_c = 70 \mu\text{m}$  respectively and the estimated effective width was  $w_{\text{eff}} = 7.4$  nm. At these values, an extension length of  $r \approx 13.6 \mu\text{m}$  was anticipated based on de Gennes' model. By sampling and analyzing the intensity traces in 500 frames for 10 different DNA molecules, an overall average extension  $r = 13.5 \mu\text{m}$  was determined. Surely this experimentally determined value agrees extremely well with the estimate based on Eqn. 3.2 and suggests that for PMMA nanochannels with  $D \gg D_c$ , the scaling behavior predicted by de Gennes is valid. As seen from Fig. 3.3 (d), thermal fluctuations in the extension length  $\delta r$  results in a Gaussian distribution of  $r$ . Recalling that the standard deviation  $\sigma = \sqrt{\langle \delta r^2 \rangle}$ , it could be anticipated that  $\sigma \approx 1.8 \mu\text{m}$  which compares favorably with the actually measured values  $\sigma \approx 1.0 \mu\text{m}$ . Since the uncertainty in the average extension is given by  $u = \sigma / \sqrt{N_f}$ , where  $N_f$  denotes the number of frames analyzed, it should be possible to determine the contour length of a confined DNA molecule with an arbitrary precision [31, 36].

The ramifications of being able to stretch DNA inside nanofuidic structures naturally goes beyond experimental investigations of contour length and elasticity. Since the extension is known to be directly proportional to the contour length, it is furthermore possible to accurately probe the genomic interaction sites of e.g. proteins. This section will be concluded by presenting two exotic examples which elegantly illustrates what may be accomplished when merging nanofabrication and biology.

Certain enzymes may fragment DNA by cutting it at specific locations known as restriction sites and these sites are intimately related to the specific base pair sequence<sup>5</sup> [17]. By size-sorting the individual restriction fragments it is possible to gain useful insight on the genomic content of a given DNA sample. This way of analyzing DNA is known as restriction mapping. In Ref. [42], restriction mapping

---

<sup>5</sup>Usually restriction enzymes recognize specific sequences composed of 4-10 base pairs. Whenever such a characteristic sequence is encountered, the enzymes effectively cut the DNA double helix.

### 3.2. INVESTIGATING DNA IN NANOFUIDIC STRUCTURES

of DNA was conducted inside nanofuidic channels as seen in Fig. 3.4 (a) where  $\lambda$ -DNA (48.5 kbp) is being fragmented by SacI. SacI cuts  $\lambda$ -DNA into three discrete fragments composed of 22.6, 0.9 and 24.8 kbp respectively. As seen from the intensity time trace, two large fragments (corresponding to 22.6 and 24.8 kbp) are clearly evident, while the smallest fragment could not be resolved. The distinct fragments cannot interpenetrate inside the nanochannel and merely separate as a result of diffusion. An important thing to notice in Fig. 3.4 (a) is the time scale: The sizes of the restriction fragments may actually be determined within approximately one minute. Furthermore the ordering of the fragments is maintained inside nanofuidic channels. Therefore it is also possible to determine the position of the cut sites along the contour. Conventional ways of performing restriction mapping usually involves size-sorting by means of gel-electrophoresis which may take several hours. Furthermore the cut sites along the contour may not be directly identified.

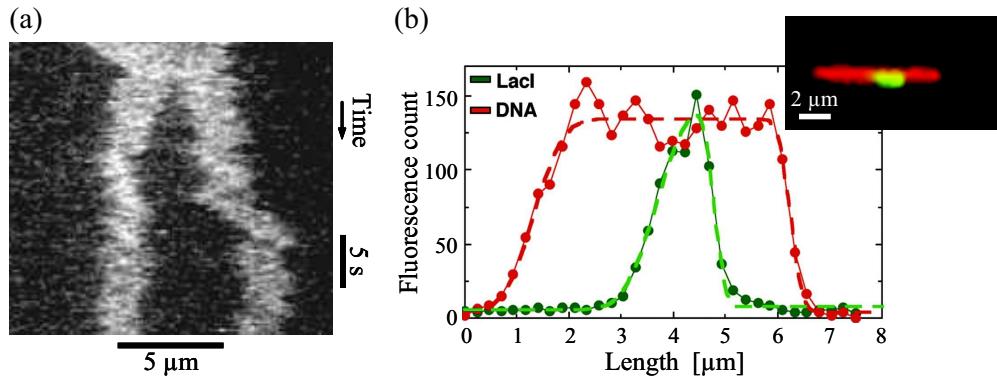


Figure 3.4: (a) Restriction mapping of  $\lambda$ -DNA (48.5 kbp) inside nanochannels with  $D = 159$  nm. The restriction enzyme used was SacI [42]. Two distinct fragments are clearly evident. (b) Fluorescence intensity profiles of GFP-LacI and LacO<sub>256</sub>-DNA. Fits to a function bearing resemblance to Eqn. 3.12 are shown as dashed lines. The inset shows the corresponding time-averaged fluorescence trace. Based on image analysis it was estimated that  $\sim 20$  GFP-LacI proteins were bound to the DNA [49].

In a process called gene expression, well-defined sequences of the genetic information encoded in DNA is translated into functional biopolymers such as proteins. Gene expression is basically composed of two major steps [17]: (i) An enzyme, RNA polymerase, initially transcribes the DNA sequence into an equivalent RNA blueprint known as messenger RNA (mRNA). (ii) By means of enzymatic synthe-

### 3.3. CHAPTER SUMMARY AND CONCLUDING REMARKS

---

sis, mRNA is translated into e.g. proteins by RNA-protein complexes known as ribosomes. Gene expression may be suppressed and regulated by repressor proteins that bind to DNA thereby effectively reducing or preventing the transcription step. Like restriction enzymes, repressor proteins recognize specific DNA segments abbreviated operators. In Ref. [49], the interaction between a repressor protein (LacI) and LacO<sub>256</sub>-DNA was investigated in nanochannels with  $D \approx 173$  nm. LacO<sub>256</sub>-DNA is derived from  $\lambda$ -DNA by insertion of a specific sequence containing 256 LacI operators thereby enabling repressor attachment. The individual operators within this sequence are separated by 15 non-specific base pairs. Green fluorescent proteins (GFP) were attached to LacI whereas the DNA was stained using the dimeric cyanine dye BOBO-3 which emits red light. As seen from Fig. 3.4 (b), GFP-LacI repressor proteins have attached to the operator sequence of LacO<sub>256</sub>-DNA. However, only a small percentage of the operators have been occupied by GFP-LacI complexes. This experimental finding clearly suggests a mutually excluding effect in the binding mechanism governing repressor-operator interactions. As argued in Ref. [49], a single binding event could induce a significant bending of the DNA which effectively prohibits the attachment of other repressor proteins within a certain proximity.

## 3.3 Chapter Summary and Concluding Remarks

When considering the extension of semiflexible and selfavoiding polymers inside nanofluidic structures, two conceptually different regimes may be identified: (i) The de Gennes regime in which  $D \gg l_p$  and the macromolecule stretches in a series of blobs due to excluded volume effects and (ii) the Odijk regime where  $D \ll l_p$  and the extension is governed by the intrinsic elasticity and the extension may be perceived as an oscillation composed of a continuous series of boundary deflections. In both regimes, the extension  $r \propto l_c$  which allows for extracting the contour length and therefore also the number of base pairs since  $l_c \approx N d_{bp}$ . Furthermore the interaction sites of polypeptides such as proteins and enzymes may be resolved. Based on experimental results, the degree of extension  $\varepsilon(D)$  appears to interpolate smoothly in the transition region where  $D = D_c \approx 2l_p = l_K$ .

Being able to stretch DNA inside nanochannels not only allows for probing the intricate structural and mechanical properties of DNA, it also enables real-time analysis of the interactions between DNA and functional biopolymers such as pro-

### **3.3. CHAPTER SUMMARY AND CONCLUDING REMARKS**

---

teins and enzymes. One might easily be tantalized by the experimental results presented within the previous section and unambiguously applaud the benefits of implementing nanofabrication for performing biopolymer research. However, using e.g. optical tweezers the elasticity, contour length and interaction sites of proteins may also be investigated [50, 51]. Furthermore conventional gel-electrophoresis or planar stretching of DNA on a smooth surface allows for performing restriction mapping with a high accuracy [52, 53]. Therefore it is only reasonable to question why the merging of nanofabrication and biology should offer a substantial improvement. From the authors point of view the answer is straightforward: Being able to stretch DNA and accurately probe both the intrinsic properties and sequence-specific interaction sites on a single macromolecule within a couple of minutes represents a drastic amelioration. The confinement induced stretching is primarily related to persistence and excluded-volume effects in the de Gennes regime and therefore e.g. double stranded RNA may also be investigated experimentally using nanochannels with a relatively large cross-sectional geometric average  $D \gg D_c$ . Clearly the nanoconfinement-approach also represents a vastly more flexible method, as passive nanofluidic structures may easily be implemented in more sophisticated devices. One could imagine analyzing a polydisperse DNA sample by size-sorting it using bumper arrays or entropic filters [54, 55, 39] and subsequently measure the contour length and protein/enzyme interaction sites. Using LOC applications, this could be done on a highly limited sample volume within a time-span on the order of minutes rather than hours or days.

# Bibliography

- [1] H. Staudinger 1920 *Ber. Deut. Chem. Ges.* **53** 1073.
- [2] A. Y. Grosberg and A. R. Khokhlov: *Giant Molecules- Here, There and Everywhere...*, Academic Press, 1997.
- [3] M. Rubinstein and R. H. Colby: *Polymer Physics*, Oxford University Press, 2003.
- [4] The images have been acquired from Wikipedia, <http://en.wikipedia.org>.
- [5] R. E. Bauer, V. Enkelmann, U. M. Wiesler, A. J. Berresheim and K. Müllen 2002 *Chem. Eur. J.* **8** 3858.
- [6] D. W. Jordan and P. Smith: *Mathematical Techniques*, Oxford University Press, 2002.
- [7] M. Doi and S. F. Edwards: *The Theory of Polymer Dynamics*, Oxford University Press, 1986.
- [8] S. B. Smith, Y. Cui and C. Bustamante 1996 *Science* **271** 795.
- [9] J. A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker and N. H. Dekker 2005 *Biophys. J.* **88** 2737.
- [10] M. R. Spiegel and J. Liu: *Mathematical Handbook of Formulas and Tables*, McGraw Hill, 1999.
- [11] J. F. Marko and E. D. Siggia 1995 *Macromolecules* **28** 8759.
- [12] C. Bustamante, J. F. Marko, E. D. Siggia and S. Smith 1994 *Science* **265** 1599.

## BIBLIOGRAPHY

---

- [13] M. D. Wang, H. Yin, R. Landick, J. Gelles and S. M. Block 1997 *Biophys. J.* **72** 1335.
- [14] D. Stigter 1977 *Biopolymers* **16** 1435.
- [15] F. Valle, M. Favre, P. Rios, A. Rosa and G. Dietler 2005 *Phys. Rev. Lett.* **95** 158105.
- [16] J. D. Watson and F. Crick 1953 *Nature* **171** 737.
- [17] R. R. Sinden: *DNA Structure and Function*, Academic Press, 1994.
- [18] S. Zamenhof, G. Brawerman and E. Chargaff 1952 *Biochim. et Biophys. Acta* **9** 402.
- [19] P. Atkins: *The Elements of Physical Chemistry*, Oxford University Press, 2001.
- [20] A. Vologodskii and N. Cozzarelli 1995 *Biopolymers* **35** 289.
- [21] J. W. Pavlicek, E. A. Oussatcheva, R. R. Sinden, V. N. Potaman, O. F. Sankey and Y. L. Lyubchenko 2004 *Biochemistry* **43** 10664.
- [22] J. C. Marini, S. D. Levene, D. M. Crothers and P. T. Englund 1982 *Proc. Natl. Acad. Sci. USA* **79** 7664.
- [23] J. C. Marini, S. D. Levene, D. M. Crothers and P. T. Englund 1983 *Proc. Natl. Acad. Sci. USA* **80** 7678.
- [24] C. Rivetti, C. Walker and C. Bustamante 1998 *J. Mol. Biol.* **280** 41.
- [25] J. D. Griffith, M. Bleyman, C. A. Rauch, P. A. Kitchin and P. T. Englund 1986 *Cell* **46** 717.
- [26] H.-S. Koo, J. Drak, J. A. Rice and Donald M. Crothers 1990 *Biochemistry* **29** 4227.
- [27] S. Arnott, R. Chandrasekaran, I. H. Hall and L. C. Puigjaner 1983 *Nucleic Acids Res.* **11** 4141.
- [28] P. G. de Gennes: *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979.

## BIBLIOGRAPHY

---

- [29] M. Daoud and P. G. de Gennes 1977 *J. Physique* **38** 85.
- [30] L. Turban 1984 *J. Physique* **45** 347.
- [31] J. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan and R. H. Austin 2004 *Proc. Natl. Acad. Sci. USA* **101** 10979.
- [32] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Chou, E. Frey and R. H. Austin 2005 *Phys. Rev. Lett.* **94** 196101.
- [33] L. H. Thamdrup, A. Klukowska and A. Kristensen 2008 *Nanotechnology* In press.
- [34] T. Odijk 1983 *Macromolecules* **16** 1340.
- [35] T. Odijk 1984 *Macromolecules* **17** 502.
- [36] W. Reisner: *Statics and Dynamics of DNA in Nanofabricated Devices*, Ph. D. thesis, Princeton University, 2006.
- [37] T. W. Burkhardt 1997 *J. Phys. A: Math. Gen.* **30** L167.
- [38] W. Reisner, J. Beech, N. Larsen, H. Flyvbjerg, A. Kristensen and J. Tegenfeldt 2007 *Phys. Rev. Lett.* **99** 058302.
- [39] J. Fu, J. Yoo and J. Han 2006 *Phys. Rev. Lett.* **97** 018103.
- [40] A. Balducci, P. Mao, J. Han and P. Doyle 2006 *Macromolecules* **39** 6273.
- [41] J. T. Mannion, C. H. Reccius, J. D. Cross and H. G. Craighead 2006 *Biophys. J.* **90** 4538.
- [42] R. Riehn, M. Lu, Y. M. Wang, S. F. Lim , E. C. Cox and R. H. Austin 2005 *Proc. Natl. Acad. Sci. USA* **102** 10012.
- [43] The stated excitation and emission wavelengths are quoted from <http://en.wikipedia.org>.
- [44] A. N. Glazer and H. S Rye 1992 *Nature* **359** 859.

## BIBLIOGRAPHY

---

- [45] O. Bakajin, T. A. J. Duke, C. F. Chou, S. S. Chan, R. H. Austin and E. C. Cox 1998 *Phys. Rev. Lett.* **80** 2737.
- [46] S. B. Smith, L. Finzi and C. Bustamante 1992 *Science* **258** 1122.
- [47] T. Perkins, D. E. Smith, R. G. Larson and S. Chu 1995 *Science* **268** 83.
- [48] S. Y. Chou, P. R. Krauss and P. J. Renstrom 1996 *J. Vac. Sci. Tehcnol. B* **14** 4129.
- [49] Y. M. Wang, J. Tegenfeldt, W. Reisner, R. Riehn, X.-J. Guan, L. Guo, I. Golding, E. C. Cox, J. Sturm and R. H. Austin 2005 *Proc. Natl. Acad. Sci. USA* **102** 9797.
- [50] M. D. Wang, H. Yin, R. Landick, J. Gelles and S. M. Block 1997 *Biophys. J.* **72** 1335.
- [51] Y. Kimura and P. R. Bianco 2006 *Analyst* **131** 868.
- [52] D. C. Schwartz, X. Li, L. I. Hernandez, S. P. Ramnarain, E. J. Huff and Y.-K. Wang 1993 *Science* **262** 110.
- [53] J. Lin, R. Qi, C. Aston, J. Jing, T. S. Anantharaman, B. Mishra, O. White, M. J. Daly, K. W. Minton, J. C. Venter and D. C. Schwartz 1999 *Science* **285** 1558.
- [54] J. Herrick and A. Bensimon 1999 *Chromosome Research* **7** 409.
- [55] J. Tegenfeldt, C. Prinz, H. Cao, R. L. Huang, R. H. Austin, S. Y. Chou, E. C. Cox and J. C. Sturm 2004 *anal. Bioanal. Chem.* **378** 1678.

# Nano-2 lecture notes part2

Rodolphe Marie and Jonas Nyvold Pedersen\*

*Department of Health Technology, Technical University of Denmark, Building 345C,  
DK-2800 Kongens Lyngby, Denmark*

E-mail: [R.M:rcwm@dtu.dk](mailto:R.M:rcwm@dtu.dk),[JNPE:jnpe@dtu.dk](mailto:JNPE:jnpe@dtu.dk)

August 25, 2025

## Contents

<b>1 How to use those lecture notes</b>	<b>2</b>
<b>2 Transport in a temperature gradient: Microscale thermophoresis.</b>	<b>3</b>
<b>3 Transport in salt gradient: Diffusio-osmosis and -phoresis.</b>	<b>5</b>
3.1 Introduction . . . . .	5
3.2 Some quantities, assumptions, and numbers. . . . .	8
3.3 Reminder of notations . . . . .	9
3.4 The case of a neutral solute. . . . .	9
3.5 The case of a charged solute . . . . .	13
3.6 Diffusio-osmosis in a salt gradient . . . . .	15
3.7 Diffusio-phoresis of particles. . . . .	27
<b>References</b>	<b>29</b>

# 1 How to use those lecture notes

Read those notes before the lectures. Text in blue means those are keywords you might look up in other literature.

Reading those notes you can focus on for example:

- understanding the physical origin of thermophoresis and diffusioosmosis.
- explain the role of the wall-solute potential in diffusioosmosis.
- explain the main difference between the diffusioosmosis due to a neutral and charged solute.
- explain the structure of the electrical double layer.
- explain diffusiophoresis by comparing it to diffusioosmosis.
- point at the two contributions to the diffusiophoresis mobility.
- understand what tools are used for the derivations.

Some notation we will try to stick to:

Table 1: Notations

Symbol	Name	Units	Value (typical)
$k_B$	Bolzmann's constant	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{K}^{-1}$	$1.38 \cdot 10^{-23}$ at 298 K.
$a$	Radius	m	n.a.
$D$	Diffusion coefficient	$\text{m}^2 \cdot \text{s}^{-1}$	-
$\gamma$	Drag coefficient	-	-
$\eta$	Dynamic viscosity	Pa.s	$8.90 \cdot 10^{-4}$ , water at 298 K.

## 2 Transport in a temperature gradient: Microscale thermophoresis.

**Introduction.** The second law of thermodynamics tells us that wherever there is a gradient of concentration, particles will tend to rearrange to even out this gradient. So if one can 1) maintain a gradient of concentration by applying a force on the particles and 2) measure this gradient, we may get to measure either the diffusion coefficient or some property of the particle that the force is depending on, or both. As an example, in a sedimentation gradient experiment, the gravitation balances out diffusion and one can get the size and density of the particles.

In this section we look at one way of maintaining a concentration gradient of particles (molecules) by imposing a phoretic force using a temperature gradient. In a temperature gradient, a force called thermophoresis acts on a particles so that this particle drifts with a velocity  $\vec{v} = -D_T \nabla T$ . The thermophoresis strength is described by the Soret coefficient  $S_T$  that rather describes the balance between diffusion and thermophoresis as  $S_T = D_T/D$ . At steady state the concentration in a temperature gradient is defined by  $c/c_0 = \exp[-S_T(T - T_0)]$  with  $c_0$  the concentration at temperature  $T_0$ .

**Microscopic theory of thermophoresis.** The microscopic theory of thermophoresis is only recent and has been established by Duhr and Braun in their paper from 2006<sup>1</sup>. Here we should mention that a particle moves in the temperature gradient if it's free energy gets lower. In particular, the entropy of the particle-solvent system changes. The microscopic theory establishes that the Soret coefficient has two contributions, one from hydration entropy and one from the ionic shielding. The hydration entropy is the entropy cost to arranging water molecules on the surface of a particle compare to if the water molecules could be 'wherever they like to be'. You may know ionic shielding as the spontaneous formation of the Electric Double Layer (EDL) on a charged surface and is described by the Debbye length. Since the screening by ions results from the balance between the potential at the surface and the

diffusion of the ions, this depends on temperature. More about the EDL formation in the next section on diffusio-phoresis. In conclusion, the temperature dependency of the hydration entropy and the EDL formation cause the particle to move in a temperature gradient in order to minimize it's free energy.

**Instrumentation.** This work lead to a commercial technique called microscale thermophoresis and an instrument to measure binding affinity of antibodies and generally binding reactions<sup>5</sup>. The instrument is basically an optical microscope with two functionalities: creating a thermal gradient in a  $500\text{-}\mu\text{m}$  diameter capillary using the adsorption of a focused infrared laser and imaging the fluorescence of the sample for reading the concentration gradient (Figure 1). While the application of microscale thermophoresis highlighted in the 2006 paper is the measurement of the hydration entropy, the commercial applications seem to be the titration of antibodies and affinity from serum<sup>2</sup> which is of immediate relevance to the field of drug discovery.

**Data analysis.** The measurement typically yield a time curve of the fluorescence intensity at the position of the hot-spot in three phase before heating, during heating and during cooling. The hot-spot creates a temperature increase from  $T_0$  before heating and  $T_{max}$  at steady state during heating. While the exact temperature profile in the capillary does not need to be known, the temperature increase  $\Delta T$  needs to be precisely controlled and calibrated. In particular it is kept low (5-10 K) in order to keep the depletion low (15 percent) and ensures that depletion can be considered linear with temperature changes. Next, the relative fluorescence increase caused by a temperature increase  $\Delta T$  is measured. In a typical application where a labeled molecule is in the presence of a (non labeled) binding partner, the relative concentration of bound and unbound molecules is calculated from the relative fluorescence intensity change at the hot-spot using that 1) bound and unbound molecules have different thermophoretic mobility 2) the fluorescence signal of bound and unbound molecules are independent and add up, and 3) the total concentration of bound and unbound molecules is

---

<sup>5</sup>Commercial instrument from NanoTemper Technology: <https://nanotempertech.com/monolith/>

constant. From there the binding properties of the molecule of interest is probed using a standard [binding isotherm](#) where the fraction of bound molecules is measured at increasing concentration to extract the dissociation constant  $K_D = \frac{k_{OFF}}{k_{ON}}$  using a fit of the isotherm where  $k_{OFF}$  is the dissociation rate in seconds and  $k_{ON}$  is the association rate constant in  $M^{-1}.s^{-1}$ .

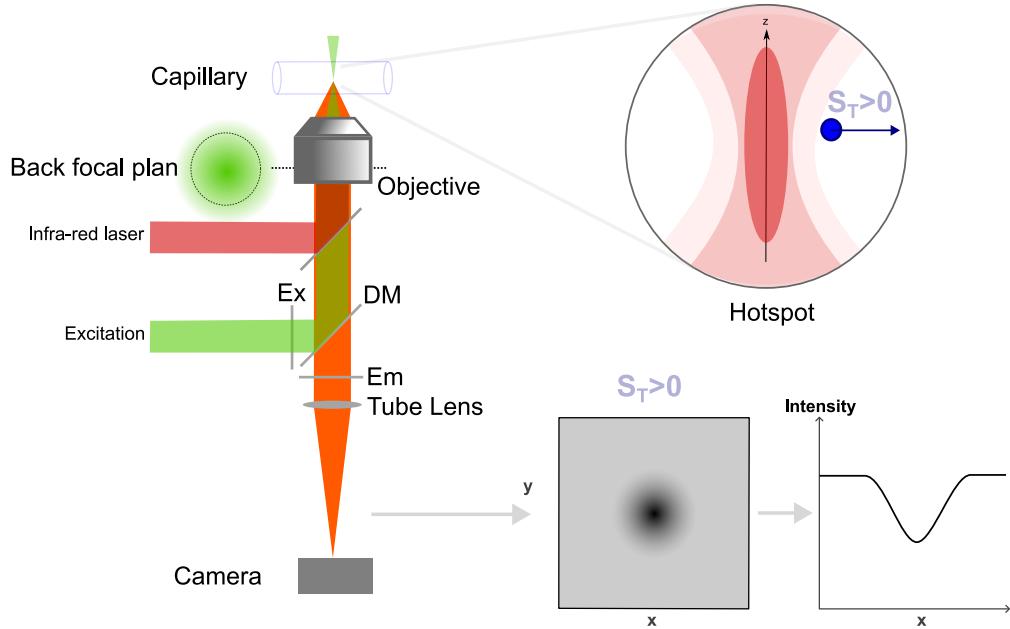


Figure 1: Set-up for Microscale thermophoresis based on a epifluorescence microscope with an infra-red laser spot that creates a temperature gradient via absorption of the water. Depending on the Soret coefficient of the molecules, the infra-red laser spot appears in the fluorescence image as a brighter or darker spot as the fluorescent molecules accumulate or escape in/from the hot-spot (adapted from <sup>2</sup>).

### 3 Transport in salt gradient: Diffusio-osmosis and -phoresis.

#### 3.1 Introduction.

In this section we look at the phenomena of diffusio-osmosis and -phoresis as an introduction to the research paper presented in the lecture. Diffusio-osmosis (-phoresis) is the transport of fluid (particle) in a chemical gradient. Already now we should note that the suffixes

-osmosis and -phoresis refer to the object moving but that the physical origin of the motion is the same in both cases, it is the motion of the fluid along the surface. Diffusio-osmosis and Diffusio-phoresis are part of a larger collection of interfacial phenomena that lead to transport and that are increasingly used in miniaturization.

In the following, We will describe diffusio-osmosis in the case of a gradient of non-charged solute by Andersson (1989)<sup>3</sup>, then we will first describe the case of a charged solute, starting with the electric double layer, making a detour by the electro-osmosis because it helps understanding diffusio-osmosis. For complete derivation of the equations, we refer to chapter 8 of 'theoretical microfluidics' by H. Bruus for the electro-osmosis<sup>4</sup>. Diffusio-osmosis for charged solute by Prieve (1984)<sup>5</sup> and a general more recent description of osmosis in general by Marbach and Bocquet (2019).<sup>6</sup>

**Physical origin of diffusioosmosis.** The diffusio-osmotic flow is the flow of liquid emerging above a surface when imposing a chemical gradient in the bulk of the solution. Note that a chemical gradient is in broad term a gradient of 'something' in 'something else'. Generally, the chemical gradient is created by a solute, dispersed in a solvent (water) that is in a much larger concentration. And so diffusio-osmosis can occur in many different systems such as gradients of electrolytes, polymer, CO<sub>2</sub>, or ethanol in a solvent like water. From this list we can distinguish two cases, one general where the solute is non-charged, and a second case where the solute is charged. In the examples presented in class, the solute is table salt *NaCl*. The physical origin of diffusio-osmosis is found at the surface ie. the solid–liquid interface although the chemical gradient is in the bulk (away from the surface). The condition for diffusio-osmosis is that the solute experiences a potential at the surface.

And so the first physical phenomenon to describe is how a surface can create a potential  $\Phi$ . This potential varies along the direction perpendicular to the surface  $y$ . In some cases the exact shape of the potential  $\Phi(y)$  is not known, in some others it can be calculated. For example in the case of an electrolyte and a charge surface, the shape of the electrical

potential can be calculated from the Poisson equation.

This potential creates a 'diffuse layer' where the solute concentration varies. We thus extract the concentration profile of the solute in the  $y$ -direction. It is always a Boltzmann distribution, but where the exact shape of the potential is either known or not known. For example, we can either define a known potential (example of the 'hard wall potential' or we can calculate the potential in the case of the electrical potential of a charged surface,. In other cases, we might model the exact shape of the potential using some other physics (neutral solutes such as entropy exclusion of polymers). In doing this we will assume that the potential is constant in the other directions ie. along the  $(z, x)$ -plane. This can be surprising at first since the gradient of solute in the  $x$ -direction is the very reason for the diffusioosmosis in the first place, but in order to decompose the mechanisms, we assume that the surface potential acts on a length scale that is small (the thickness of the diffuse layer at the surface) compare to the length scale of the gradient.

Next we look at the implications of the concentration gradient in the diffuse layer, along the  $y$ -direction, for the motion of the fluid. To this end we use the Navier-Stokes equation to extract the velocity of the fluid along the surface  $v(x)$ , because the potential varies with  $y$ , and that this gradient of potential in  $y$  means there is a force acting on the fluid perpendicularly to the surface. This force is a body force, equivalent to an osmotic pressure, and depends on the concentration of solute. But in the case of an electrolyte, there is an additional body force that is similar to an electrostatic force and that is also depending on the concentration of solute in the diffusive layer. In both cases a gradient of solute along the surface (it depends on  $x$ ) makes the solution move along the surface.

Calculating the solution of the NS equation with the correct boundary conditions, we will obtain an expression of the velocity of the fluid. Since the NS equation is notoriously difficult, an analytical solution might not always be attainable. And so in the following we will consider simple geometries.

The Navier-Stokes equation is difficult to solve but also intuitively understand because of

the Laplacian of the viscous term. The term means that the velocity of the fluid in a given direction ( $x$ ) depends on the how fast this component of the velocity field varies along the two other directions. The negative sign of the viscous term means that the fluid viscosity tends to even out gradients of velocities.

### 3.2 Some quantities, assumptions, and numbers.

- Diffusion coefficients of electrolytes:<sup>7</sup>  $D_{\text{Na}^+} = 1330 \frac{(\mu\text{m})^2}{\text{s}}$ ,  $D_{\text{K}^+} = 1960 \frac{(\mu\text{m})^2}{\text{s}}$ ,  $D_{\text{Cl}^-} = 2030 \frac{(\mu\text{m})^2}{\text{s}}$ .
- Diffusion coefficient of fluorescein:<sup>8</sup>  $D_{\text{fluorescein}} = 350 \frac{(\mu\text{m})^2}{\text{s}}$ .
- Debye layer:  $\kappa^{-1} = \lambda_{DB} = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2e^2 N_A I}} = \sqrt{\frac{\epsilon k_B T}{2(Ze)^2 c_0}}$  ( $\epsilon$  dielectric constant,  $N_A$  Avogadro's number,  $I$  ionic strength,  $e$  elementary charge,  $c_0$  is the concentration).
- Bjerrum length:  $\ell_B = \frac{e^2}{4\pi\epsilon_0\epsilon_r k_B T}$ .
- Debye-Bjerrum length:  $\kappa^{-1} = \frac{1}{\sqrt{8\pi\ell_B N_A I}}$ .

Example: in water at 298 K, at physiological salt concentration,  $I = 0.156 M$ ,  $\kappa^{-1} = 0.304/\sqrt{I}$ , where  $[\kappa^{-1}]$  is in nm, and  $\ell_B = 0.7 \text{ nm}$ .

Specific to the Diffusioosmotsis/phoresis literature:

$$\beta = \frac{D_+ - D_-}{D_+ + D_-} \quad (\simeq -0.20 \text{ for NaCl}), \quad (1)$$

$$D = \frac{2D_+ D_-}{D_+ + D_-} \quad (\simeq 1607 (\mu\text{m})^2/\text{s} \text{ for NaCl}), \quad (2)$$

$$\text{Pe} = \frac{\epsilon}{8\pi\eta D} \left( \frac{k_B T}{Ze} \right)^2 \quad (\simeq 0.1, \text{ Peclet number. NB: Definition of permittivity!}) \quad (3)$$

$$\zeta, \text{ zeta potential.} \quad (4)$$

$$\bar{\zeta} = \frac{Ze\zeta}{k_B T} \quad (\simeq -2.7 \text{ for } \zeta = -70 \text{ mV}, Z = 1, T = 300 \text{ K}), \quad (5)$$

$$\gamma = \tanh \frac{Ze\zeta}{4k_B T} = \tanh \frac{\bar{\zeta}}{4} \quad (\simeq \tanh[-0.677] = -0.59) \text{ is a drag coefficient.} \quad (6)$$

### 3.3 Reminder of notations

In the following we will consider scalars like the pressure  $p$ , and vectors like the velocity  $\mathbf{v}$ .

A vector  $\mathbf{v}$  has components  $(v_x, v_y, v_z)$ .

The  $\nabla$  operator of a scalar is a vector with component  $\frac{\partial p}{\partial x}$  along  $x$ . Same for the gradient of a vector field ie. has component  $\frac{\partial \mathbf{v}_x}{\partial x}$  along  $x$ .

The operator  $\nabla^2$  of a vector has the component  $\frac{\partial^2 v_x}{\partial^2 x} + \frac{\partial^2 v_x}{\partial^2 y} + \frac{\partial^2 v_x}{\partial^2 z}$  along  $x$

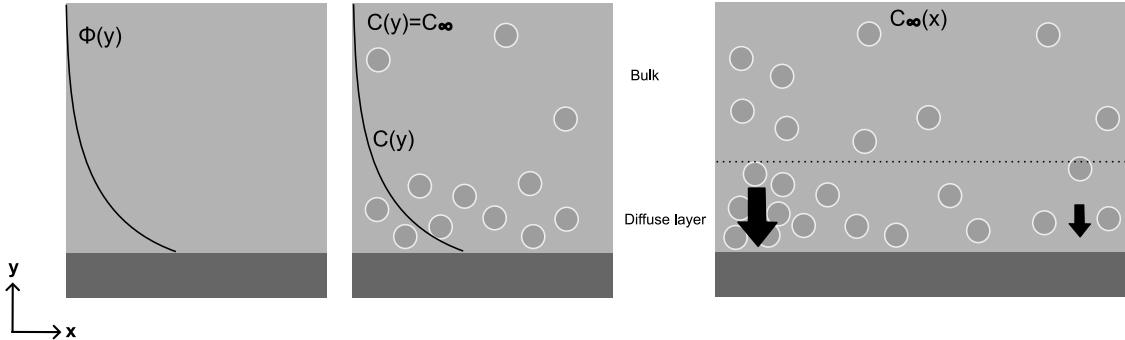


Figure 2: The solute gradient in the bulk  $\nabla C_\infty$  is along  $x$ , The interface is in the  $(x, z)$ -plane. The surface potential  $\Phi$  is constant along  $x$  and  $z$ . The diffusioosmotic flow velocity is positive (from left to right) because the potential makes the solute concentration higher in the diffuse layer than in the bulk.

### 3.4 The case of a neutral solute.

Using this general 'recipe' we first look at the case of neutral solute and we follow Andersson (1989)<sup>9</sup>. See also Ref. AndersonSPM1984 for a derivation.

We consider an infinite  $xz$ -plane and let  $y$  define the height over the plane. The concentration at  $y \rightarrow \infty$  is denoted  $C_\infty(x)$  (independent of  $z$ ).

**Key elements:** A wall-solute potential  $\Phi$  (due to exclusion and, e.g., dipole or van der Waal's forces) and the Stokes flow equation ('creeping flow') Eq. (8).

Let  $\Phi(y)$  denote the 'potential of mean force' so  $-\nabla\Phi$  is the force at a distance  $y$ . Through solute-fluid interactions, a body force is transmitted to the fluid  $-C\nabla\Phi$ , where  $C(x, y)$  is

the local concentration. An important point is, see Ref. AndersonSPM1984, that  $\Phi$  is the sum of solid-fluid-solute interactions relative(!) to the fluid-solute interactions, so  $\Phi(y) \rightarrow 0$  for  $y \rightarrow \infty$ .

[Since  $\Phi$  has a length scale  $\delta$ , which is the thickness of the interface] The variation in  $y$  is given by Boltzmann statistics,

$$C(x, y) = C_\infty(x) e^{-\Phi(y)/k_B T}. \quad (7)$$

The Stokes limit of the Navier-Stokes equations for an incompressible Newtonian fluid is (no inertia) ie. 'creeping flow'.

$$\eta \nabla^2 \mathbf{v} - \nabla p + \mathbf{f} = 0, \quad (8)$$

where  $\mathbf{v}$  is the velocity,  $p$  is the pressure, and  $\mathbf{f}$  is the body force.

For our system there is no velocity in the  $y$ -direction and no force in the  $x$ -direction, and we get

$$\frac{\partial p}{\partial y} + C(x) \frac{\partial \Phi}{\partial y} = 0 \quad (9)$$

$$\eta \frac{\partial^2 v_x}{\partial y^2} - \frac{\partial p}{\partial x} = 0. \quad (10)$$

We integrate along  $y$  starting from infinity to a coordinate  $y$  we get

$$\int_{\infty}^y \frac{\partial p}{\partial y} dy = - \int_{\infty}^y C(x) \frac{\partial \Phi}{\partial y} dy \quad (11)$$

Inserting the expression for  $C(x, y)$  we get:

$$\int_{\infty}^y \frac{\partial p}{\partial y} dy = - \int_{\infty}^y C_\infty(x) e^{-\Phi(y)/k_B T} \frac{\partial \Phi}{\partial y} dy \quad (12)$$

On the left hand side, the integral is simply  $[p]_{\infty}^y$ . On the right hand side, we recognize the

form  $\int ae^{bf(x)}f'dx$  with  $a = C_\infty(x)$  and  $b = -1/k_B T$ , and so we write:

$$[p]_\infty^y = k_B T C_\infty(x) [e^{-\Phi(y)/k_B T}]_\infty^y \quad (13)$$

Then remembering that  $\Phi(y) \rightarrow 0$  for  $y \rightarrow \infty$  we get

$$p(x, y) - p_\infty = k_B T C_\infty(x) [e^{-\Phi(y)/k_B T} - 1] = k_B T [C(x, y) - C_\infty(x)]. \quad (14)$$

Writing this in Eq. (10) for  $v_x(y)$  we get:

$$\eta \frac{\partial^2 v_x}{\partial y^2} = k_B T \frac{dC(x)}{dx} [e^{-\Phi(y)/k_B T} - 1], \quad (15)$$

which is an inhomogeneous second-order differential equation with the general solution

$$v_x(y) = c_1 + c_2 y + \frac{k_B T}{\eta} \frac{dC(x)}{dx} y \int_0^y dy' [e^{-\Phi(y')/k_B T} - 1] - \frac{k_B T}{\eta} \frac{dC(x)}{dx} \int_0^y dy' y' [e^{-\Phi(y')/k_B T} - 1] \quad (16)$$

The boundary conditions are

$$v_x(0) = 0 \text{ for } y = 0 \text{ (no slip)}, \quad (17)$$

$$\frac{\partial v_x}{\partial y} \rightarrow 0 \text{ for } y \rightarrow \infty \text{ (no pressure gradient in the bulk fluid)}. \quad (18)$$

The first condition gives  $c_1 = 0$ . Calculating  $\partial v_x / \partial y$  and using the second boundary condition gives that

$$v_x(y) = -\frac{k_B T}{\eta} \frac{dC(x)}{dx} y \int_y^\infty dy' [e^{-\Phi(y')/k_B T} - 1] - \frac{k_B T}{\eta} \frac{dC(x)}{dx} \int_0^y dy' y' [e^{-\Phi(y')/k_B T} - 1] \quad (19)$$

as

$$c_2 = -\frac{k_B T}{\eta} \frac{dC(x)}{dx} \int_0^\infty dy' [e^{-\Phi(y')/k_B T} - 1]. \quad (20)$$

This gives for  $y \rightarrow \infty$

$$v_x(\infty) = -\frac{k_B T}{\eta} \frac{dC(x)}{dx} \int_0^\infty dy' y' \left[ e^{-\Phi(y')/k_B T} - 1 \right]. \quad (21)$$

So the slip velocity  $v^s = v_x(\infty)$  is

$$\begin{aligned} v^s &= -\frac{k_B T}{\eta} \frac{dC}{dx} \int_0^\infty y \left[ e^{-\Phi(y)/k_B T} - 1 \right] dy \\ &= -\frac{k_B T}{\eta} \frac{dC}{dx} K L^* \end{aligned} \quad (22)$$

where  $K = \int_0^\infty [e^{-\Phi(y)/k_B T} - 1] dy$  and  $L^* = \int_0^\infty y [e^{-\Phi(y)/k_B T} - 1] dy / K$ . The integrand is  $e^{-\Phi(y)/k_B T} - 1 = c/c_\infty - 1$ , i.e., the excess-concentration profile.  $K$  is the amount of solution adsorbed on the surface per surface area, divided by the bulk concentration at equilibrium. It is positive if solute adsorbs to the surface.

**Exercise: The 'hard wall potential' example.** See separate exercise note.

### 3.5 The case of a charged solute

**The Electric Double Layer.** In this course, we consider several devices that can be seen as a capillary that we fill with a solution containing one or more electrolytes. In a classic example, the capillary is made of glass and so the surface is charged. When the solution is in contact with the charged surface of the capillary, charges on the surface attract/repel the ions in the solution. Counter-ions accumulate close to the surface and co-ions are repelled; this is the formation of the electric double layer (EDL). Close to the surface, counterions are tightly bound to the surface. They constitute the Stern layer. Further into the solution, the ions are free to move and they are distributed according to the Boltzmann's distribution, this is the Debye layer and constitute the 'diffuse layer' in the diffusioosmosis terminology mentioned above. The 'depth' of the EDL is characterized by the Debye length  $\lambda_D$  that tells how the Boltzmann's distribution decays in the solution away from the surface. In fact, diffusion tends to even-out the order created by the surface electrical potential ie. the separation of the positive and negative ions near the charged surface. Close to the surface, the electrical potential wins over diffusion and ions are ordered. Far from the surface the diffusion wins over the electrical potential and ions are randomly distributed (thus maximizing the entropy). The Debye length  $\lambda_D$  is, as described in Section 3.2:

$$\lambda_{DB} = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2e^2 N_A I}} = \sqrt{\frac{\epsilon k_B T}{2(Ze)^2 c_0}}$$

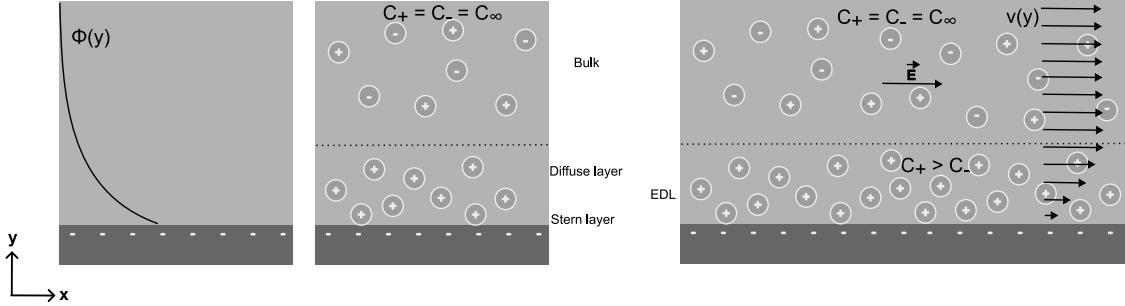
with  $c_0$  is the bulk salt concentration  $C_\infty$  and  $l_B$  the Bjerrum length expressed as  $l_B = \frac{e^2}{4\pi\epsilon k_B T}$ ,  $\epsilon$  is the dielectric permitivity of water. Here we note that  $c_0$  is in ions/m<sup>3</sup>, while the ionic strength is usually given in Molar concentration.<sup>1</sup>

A consequence of the EDL formation is that although the solution is charge-neutral far from the surface ie. at a distance larger than the Debye length, it is not within the EDL so the solution has a volumetric net charge. In our example of a glass capillary, the solution

---

<sup>1</sup>Equation 8.7 in <sup>4</sup> is then quite handy:  $\lambda_D \approx \sqrt{\frac{1mM}{Z^2 c_0}} \times 9.6\text{ nm}$  where the concentration of ions is inserted in mM and the Debye length is in nm.

within the EDL is positively charged. This has in particular implications when an electric field is applied.



**Figure 3: The EDL and electroosmosis (EO) in absence of solute gradient.** The salt concentration in the bulk  $C_\infty$  is constant along  $x$ . The surface potential  $\Phi$  is created by the surface charges. Surfaces charges attract a layer of immobile counterions (Stern layer) and a layer of weakly bound counterions (Debye layer or diffuse layer). Together they form the EDL. The liquid in the EDL has a net charge and so it can be moved by applying an electric field  $E$  leading to an electro-osmotic flow.

**The Nernst-Planck equation.** The Nernst-Planck equation describes the flux of ions in the solution under the action of diffusion, electric potential and convection.

It has the form:

$$J_i = -D_i \nabla C_i - D_i \frac{Z_i e}{k_B T} C_i \nabla \Phi + C_i \mathbf{v}$$

with  $J_i$  Flux of ion  $i$  [ $\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ],  $D_i$  Diffusion coefficient of ion  $i$  [ $\text{m}^2\cdot\text{s}^{-1}$ ],  $C_i$  Concentration of ion  $i$  [ $\text{mol}\cdot\text{m}^{-3}$ ],  $\Phi$  Electrostatic potential [V],  $Z_i$  Charge of ion  $i$ ,  $\mathbf{v}$  Fluid velocity [ $\text{m}\cdot\text{s}^{-1}$ ].

In the following, the first use of the NP equation is to calculate the distribution of ions in the electrical potential of the charged surface that is along  $y$ . The Boltzmann distribution is the solution of the NP equation in the direction perpendicular to the surface  $z$  at steady state ( $J_i = 0$ , no net flux), and in the absence of convection ( $\mathbf{v} = 0$ ) that is only considering the electric potential from the charged surface and the diffusion acting on the ions.

The second use of this equation is to predict the velocity of the liquid in the directions along the surface and we can consider two situations leading to fluid transport of interest here. The first is when an electric field applied along the surface (say along  $x$ ) leading to the

electro-osmotic flow, as we will see in the next section. The second is when a solute gradient is present in the bulk leading to a diffusioosmotic flow.

**Electro-osmosis.** The electro-osmotic flow is the flow obtained when an electric field is applied in a capillary, in the absence of a solute gradient in the bulk. The EDL formation implies that the fluid carries a net charge that can experience a body force if an electric field is applied.<sup>2</sup> Hence, the electroosmotic flow is a flow that originates from a force acting only on the liquid within the EDL, similarly to the example of the diffusioosmotic flow in the 'neutral solute' case. Due to the viscous force, the moving solution in the EDL pulls the liquid in the whole capillary but it is counterbalanced by the viscous drag of the fluid along the surface since the liquid has a zero velocity at the surface (in the no-slip boundary condition). Such flow has an even flow velocity towards the bulk of the liquid and is thus called a 'plug-flow' as opposed to pressure driven flows that are characterized by parabolic flow velocity profiles.

### 3.6 Diffusio-osmosis in a salt gradient

**Main differences with the neutral solute case.** In the previous section we looked at the consequence of the formation of the EDL on a surface when an electric field is applied along the surface in the absence of solute gradient in the bulk. We now look at the situation with a solute gradient in the bulk  $C_\infty(x)$ .

The first difference with the 'neutral solute' case is that the potential at the surface: it is well defined as the electrical potential and can be calculated from the Poisson equation. The second difference is that the solute has actually two entities, the positive and the negative ions and so instead of calculating one solute distribution along  $y$ , we consider two, one for the positive ions and one of the negative ions as they will behave differently in the potential so we get  $C^-(y)$  and  $C^+(y)$ . The calculation can be done using the Nernst-Planck equation that links diffusion, electric potential and convection. Here we note that in doing this we take

---

<sup>2</sup>The flow velocity profile can be calculated using Eq. (8) using a body force  $\mathbf{f} = q\mathbf{E}$ , also see for example here<sup>4</sup>

into consideration that positive and negative ions may have different diffusion coefficients. From this will the difference in diffusion coefficients appear in the equations in the form of the factor  $\beta$  as defined in Section 3.2. The third difference is also that in order to get the velocity of the fluid, we need to consider the net charge of the diffuse layer that is resulting from the charge separation. This leads to an additional body force in the NS equation. The 'neutral' osmotic pressure contribution to the total body force is still present in the expression of the slip velocity. In the end we will see that the slip velocity is proportional to  $\frac{d \ln C_\infty(x)}{dx}$  while for the neutral-solute gradient case it is proportional to the gradient  $\frac{dC_\infty(x)}{dx}$ .

In applying the NP and the NS equations, we will work under the approximation that the diffusive layer is much thinner ( Debye length is smaller) than the distance over which the gradient is applied in the bulk.

In the end, the gradient of osmotic pressure in the EDL makes the liquid move from higher salt concentration to lower salt concentration – that is along the ionic strength gradient – within the EDL. In a way the osmotic pressure force only acts on the fluid within the EDL while the pressure in the bulk is constant similarly to the case of the electro-osmosis where the electric field applies a force to the liquid of the diffusive layer of the EDL.<sup>3</sup>

In this case, it can be shown that the diffusio-osmotic flow rate in a rectangular capillary  $Q$  is:

$$Q = \frac{wh}{L} D_{DO} \ln \frac{C_L}{C_R} \quad (23)$$

with  $w$  and  $h$  the width and height of the capillary, with  $w > h$ .  $L$  is the length of the capillary where a gradient of solute is imposed with  $C_L$  and  $C_R$  the concentration in the left and right ends of the capillary. Note that there is no pressure driven flow through the capillary so the pressure in the bulk at both ends is equal. The diffusio-osmotic mobility

---

<sup>3</sup>Here we note that the assumption that the pressure in the bulk is constant, implies that there is a gradient of hydrostatic pressure that balances out the gradient of osmotic pressure expressed at  $\Pi = k_B T C_\infty(x)$ . In Marbach 2019 Eq. (39) this is  $p_\infty = cte$  along x.

$D_{DO}$  is then expressed as the sum of two contributions:

$$D_{DO} = \frac{\epsilon k_B T}{Z e} \left[ \beta \zeta - 2 \frac{k_B T}{Z e} \ln(1 - \gamma^2) \right] \quad (24)$$

with parameters defined in Section 3.2. In particular,  $\zeta$  is the zeta potential of the surface, and  $\beta$  expresses the difference in diffusion coefficient between the two ions of the electrolyte. In our example of  $NaCl$  as solute,  $Z = 1$ ,  $\beta = -0.20$ . In summary, to predict the diffusio-osmotic flow in a capillary we need to know the geometry of the capillary, the charge of the surface, the concentration of the gradient of electrolyte and the diffusion coefficients of the ions and that includes knowing the viscosity and temperature.

Here we note that the chemio-osmotic contribution to the velocity in the case of an electrolyte (see Eq. (3.6)) does not have the same dependency on the solute gradient as in the case of the neutral solute (See Eq. (3.4)) although the origin of this contribution is the same ie. accumulation/depletion of solute in the diffuse layer. This difference is a result of the interaction between the ions and the potential that makes the concentration of solute in the diffuse layer vary with the  $\ln$  of the concentration rather than the concentration.

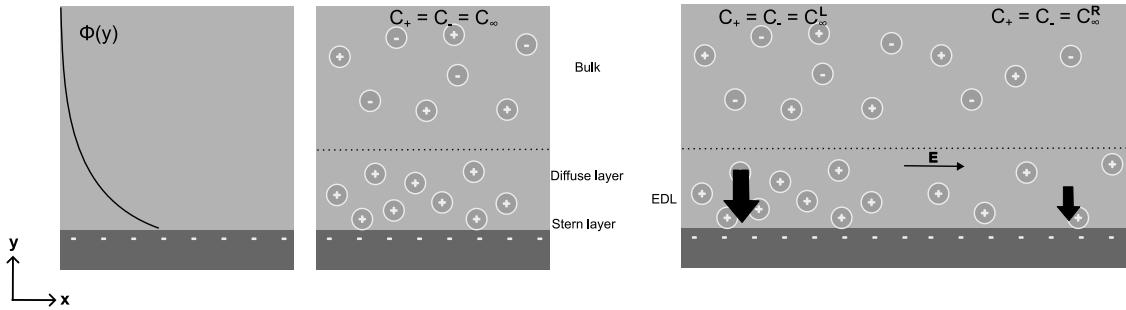


Figure 4: The salt gradient in the bulk  $\nabla C_\infty$  is along  $x$ , The surface potential  $\Phi$  is created by the surface charges. Surfaces charges attract a layer of immobile counterions (Stern layer) and a layer of weakly bound counterions (Debye layer or diffuse layer). Together they form the EDL. The gradient of salt in the bulk leads to a chemio-osmotic flow (same as for neutral solute) and a flow contribution emerging from the separation of ions in the EDL similar to an electro-osmotic flow.

### Derivation according to Prieve (1984)

Here we follow/copy mainly Ref. PrieveJFM1984. The main ingredients are the Nernst-

Planck equation (= flux of ions under a concentration gradient and an electric field), the Boltzmann distribution, the Poisson equation for the charge distribution, and Stokes' equation.

Again we consider an infinite  $xz$ -plane. Let the subscript '+' cations and '-' anions, respectively.  $C_i(x)$  ( $i = +, -$ ) denotes the local number density concentration, and  $C_\infty(x)$  the concentrations for  $y \rightarrow \infty$ .

**Density distribution in absence of a solute gradient.** Even when  $C_\infty$  is independent of  $x$  and  $y$ , the charge surface gives a nonuniform distribution of ions, as counterions experience attraction (force towards the surface), and Brownian motion which tends to make their distribution uniform.

For an ideal solution (i.e. dilute) the (particle!, not current flux) flux  $J_i$  is in steady-state (Nernst-Planck)<sup>4</sup>

$$\begin{aligned}
\mathbf{J}_i &= \mathbf{J}_{\text{diff}} + \mathbf{J}_{\text{el}} \\
&= -D_i \nabla C_i + \frac{\sigma_{\text{el}}}{Z_i e} \mathbf{E} \quad \text{NB: Divides by } Z_i e \text{ as particle current} \\
&= -D_i \nabla C_i - \frac{Z_i e C_i}{\gamma_i} \nabla \psi \\
&= -D_i \left[ \nabla C_i + \frac{Z_i e C_i}{k_B T} \nabla \psi \right].
\end{aligned} \tag{25}$$

Here  $Z_i e$  is the charge of a single ion,  $\mathbf{E} = -\nabla \psi$  where  $\psi$  is the electrostatic potential,  $\sigma_{\text{el}} = Z_i e C_i \mu_{\text{ion}}$  is the electric conductivity of the ions,  $\mu_{\text{ion}} = Z_i e / \gamma_i$ , and  $\gamma_i$  is the drag coefficient.

---

<sup>4</sup>Flux is noted  $N$  in <sup>5</sup>.

In equilibrium there is no flux,  $\mathbf{J}_i = \mathbf{0}$ , and the ionic distribution is

$$C_{\pm} = C_{\infty} e^{\mp \Phi} \quad (26)$$

assuming  $Z_+ = -Z_- = Z$  and introducing the dimensionless electrostatic potential  $\Phi = Ze(\psi - \psi_{\infty})/k_B T$ .

With Poisson's equation, we get an expression between the charge distribution of the ions  $ZeC_i$  and the electric potential

$$\nabla^2 \psi = \frac{4\pi Ze}{\epsilon} (C_+ - C_-). \quad (27)$$

Combining Eqs. (26,27) with the boundary conditions  $\psi = \psi_s$  at  $y = 0$  and  $\psi \rightarrow \psi_{\infty}$  as  $y \rightarrow \infty$  we get :

$$\tanh(\Phi/4) = \gamma e^{-\kappa y} \quad (28)$$

where  $\kappa$  (inverse Debye length), and  $\gamma$  are defined in Section 3.2.

This is the so-called Chapman-Gouy solution <sup>5</sup>. The  $y$ -dependence of the electrostatic potential  $\psi(y)$  is shown in Figure 5. Note that the electrical potential in the bulk is explicitly  $\psi_{\infty} \rightarrow 0$  and so the dimensionless potential is  $\phi(x, y) = \frac{Ze}{k_B T} \psi(x, y)$ . The electrical potential at the surface  $\psi(x, 0)$  is the zeta potential  $\zeta$ .

Now assume that  $C_{\infty}(x)$  is not constant, but depends on  $x$ . Far from the surface the solution is electrically neutral, and we expect no current arising from flow of ions, so we must require  $\mathbf{J}_+ = \mathbf{J}_-$  (i.e., the ions flow in the same direction, otherwise the fluid would be

---

<sup>5</sup>See Eq. (8.25) in H. Bruus 'theoretical microfluidics'<sup>4</sup> for full derivation and assumptions.

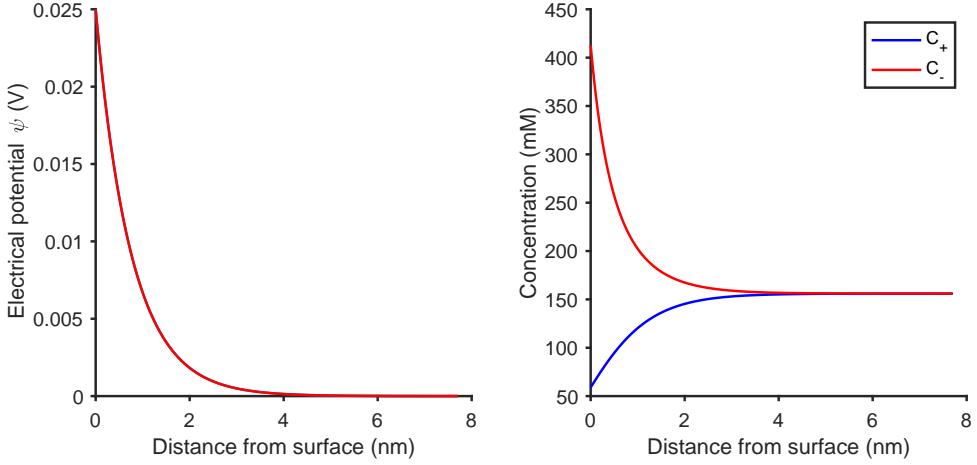


Figure 5: Gouy Chapman solution for the electrical potential  $\psi(y)$  and the ions concentration ( $C_+(y)$  and  $C_-(y)$ ) in water. The parameters used are  $\psi(0) = 25 \text{ mV}$ ,  $T = 298 \text{ K}$ ,  $c_\infty = 156 \text{ mM}$  and  $Z = 1$  ie. the ionic strength of a 1 x phosphate buffered saline solution (PBS). The Debye length is calculated to be 0.76 nm. The ion density used in the expression of the Debye length is  $9.4 \cdot 10^{25} \text{ ions/m}^3$ .

electrically polarized?!), which is only possible if there is a spontaneous electrical field  $\mathbf{E}_\infty$ ,

$$\mathbf{J}_+^\infty = \mathbf{J}_-^\infty \quad (29)$$

$$-D_+ \left[ \nabla C_+^\infty + \frac{ZeC_+^\infty}{k_B T} \nabla \psi_\infty \right] = -D_- \left[ \nabla C_-^\infty - \frac{ZeC_-^\infty}{k_B T} \nabla \psi_\infty \right] \quad (30)$$

$$(D_+ - D_-) \nabla C_\infty + \frac{ZeC_\infty}{k_B T} (D_+ - D_-) \nabla \psi_\infty = 0 \quad (31)$$

such that

$$\mathbf{E}_\infty(x) = -\nabla \psi_\infty(x) = \frac{k_B T}{Ze} \beta \nabla \ln C_\infty(x), \quad (32)$$

where  $\beta$  is defined in Section 3.2, and  $\nabla \ln C_\infty = \nabla C_\infty / C_\infty$ . Notice that the electric field is parallel to the surface (i.e. along  $x$ ).

This gives rise to a net flux of electrolytes

$$\mathbf{J}_+^\infty = -D_+ \left[ \nabla C_\infty - \frac{ZeC_\infty}{k_B T} \left( \frac{k_B T}{Ze} \beta \nabla \ln C_\infty \right) \right] \quad (33)$$

$$= -D_+ \nabla C_\infty + D_+ \frac{D_+ - D_-}{D_+ + D_-} \nabla C_\infty \quad (34)$$

$$= -\frac{2D_+ D_-}{D_+ + D_-} \nabla C_\infty \quad (35)$$

$$= -D \nabla C_\infty, \quad (36)$$

$$= \mathbf{J}^\infty \quad (37)$$

where  $D$  is defined in Section 3.2. Similarly,

$$\mathbf{J}_-^\infty = -D_- \left[ \nabla C_\infty + \frac{ZeC_\infty}{k_B T} \left( \frac{k_B T}{Ze} \beta \nabla \ln C_\infty \right) \right] \quad (38)$$

$$= -D_- \nabla C_\infty - D_- \frac{D_+ - D_-}{D_+ + D_-} \nabla C_\infty \quad (39)$$

$$= -\frac{2D_+ D_-}{D_+ + D_-} \nabla C_\infty \quad (40)$$

$$= -D \nabla C_\infty \quad (41)$$

$$= \mathbf{J}^\infty. \quad (42)$$

One can argue that even though the fluid is not at rest (non-equilibrium), the concentration at nearby points may still be approximated by the Boltzmann distribution (more specifically  $|\nabla \ln C_\infty| \ll \kappa$ ),

$$C_\pm(x, y) = C_\infty(x) e^{\mp \Phi(x, y)}, \quad (43)$$

$$\Phi(x, y) = \frac{Ze}{k_B T} [\psi(x, y) - \psi_\infty(x)]. \quad (44)$$

Furthermore, if the tangential component of the field is much smaller than the normal

component due to the surface charge, i.e. if

$$|\mathbf{E}_\infty(x)| \ll \kappa |\zeta(x)| \quad (45)$$

with the zeta potential defined as the potential energy at the surface compared to the bulk (far away from the surface)  $\zeta(x) = \psi(x, 0) - \psi_\infty(x)$ , that is then same as

$$|\nabla \psi_\infty(x)| \ll \frac{|\psi(x, 0) - \psi_\infty(x)|}{\lambda_{DB}} \quad (46)$$

then  $\partial^2/\partial x^2 \ll \partial^2/\partial y^2$  in Poisson's equation, and we are left with the same Chapman-Gouy solution for  $\Phi$  as before. Notice that  $\kappa |\zeta(x)| = |\zeta(x)|/\lambda_{DB}$  is the total variation in the potential from the wall to infinity over a Debye length.

Fluid elements inside the double layer are charged, which results in an electrostatic body force due to the electrical potential  $\psi$ . Inserting this as a body force  $\mathbf{f}$  in the Stokes equation (Equation 8) gives

$$-\eta \nabla^2 \mathbf{v} + \nabla p + (C_+ - C_-) Z e \nabla \psi = \mathbf{0}. \quad (47)$$

Scaling argument applied to the continuity equation (??!!) gives that there is only an  $x$ -component of the velocity, and thus we get for the  $y$ -component of Stokes equation:

$$\frac{\partial p}{\partial y} + C_\infty(x) [e^{-\Phi(x,y)} - e^{\Phi(x,y)}] Z e \frac{\partial \psi(x, y)}{\partial y} = 0 \quad (48)$$

$$\frac{\partial p}{\partial y} = 2k_B T C_\infty(x) \sinh[\Phi(x, y)] \frac{\partial \Phi(x, y)}{\partial y} \quad (49)$$

that is

$$p(x, y) - p_\infty = 2k_B T C_\infty(x) \{\cosh[\Phi(x, y)] - 1\} \quad (50)$$

where  $p(x, y) \rightarrow p_\infty$  for  $y \rightarrow \infty$  as the pressure equilibrates far from the surface, and  $\Phi(x, y) \rightarrow 0$  for  $y \rightarrow \infty$  and  $\cosh(0) = 1$ .

Here, see also Eq. (9) and note that the integration is different because of the related but different dimensionless potential  $\Phi$  and the electrostatic potential  $\psi$ .

Now we look at the  $x$ -component in Eq. (47) and get

$$-\eta \frac{d^2 v_x(y)}{dy^2} + \frac{dp(x, y)}{dx} + [C_+(x, y) - C_-(x, y)] Ze \frac{d\psi(x, y)}{dx} = 0 \quad (51)$$

The derivative of the pressure with respect to  $x$  is

$$\frac{dp(x, y)}{dx} = 2k_B T \left( \frac{dC_\infty(x)}{dx} \{ \cosh[\Phi(x, y)] - 1 \} + C_\infty(x) \sinh[\Phi(x, y)] \frac{d\Phi(x, y)}{dx} \right) \quad (52)$$

Inserting the expression for  $dp/dx$  gives

$$\begin{aligned} 0 &= -\eta \frac{d^2 v_x(y)}{dy^2} + \frac{dp(x, y)}{dx} - [C_+(x, y) - C_-(x, y)] Ze \frac{d\psi(x, y)}{dx} \\ &= -\eta \frac{d^2 v_x(y)}{dy^2} + 2k_B T \left( \frac{dC_\infty(x)}{dx} \{ \cosh[\Phi(x, y)] - 1 \} + C_\infty(x) \sinh[\Phi(x, y)] \frac{d\Phi(x, y)}{dx} \right) - 2C_\infty(x) \sinh[\Phi(x, y)] \left( k_B T \frac{d\Phi(x, y)}{dx} - Ze \frac{d\psi(x, y)}{dx} \right) \\ &= -\eta \frac{d^2 v_x(y)}{dy^2} + 2k_B T \frac{dC_\infty(x)}{dx} \{ \cosh[\Phi(x, y)] - 1 \} + 2C_\infty(x) \sinh[\Phi(x, y)] \left( k_B T \frac{d\Phi(x, y)}{dx} - Ze \frac{d\psi(x, y)}{dx} \right) \end{aligned} \quad (53)$$

As

$$\frac{d\Phi(x, y)}{dx} = \frac{Ze}{k_B T} \frac{d}{dx} [\psi(x, y)] - \psi_\infty(x) \quad (54)$$

we get

$$k_B T \frac{d\Phi(x, y)}{dx} - Ze \frac{d\psi(x, y)}{dx} = Ze \frac{d}{dx} [\psi(x, y)] - \psi_\infty(x) - Ze \frac{d\psi(x, y)}{dx} = -Ze \frac{d\psi_\infty(x)}{dx}. \quad (55)$$

Thus the differential equation for  $v_x(y)$  becomes

$$\eta \frac{d^2 v_x(y)}{dy^2} = 2k_B T \frac{dC_\infty(x)}{dx} \{ \cosh[\Phi(x, y)] - 1 \} - 2C_\infty(x) \sinh[\Phi(x, y)] Ze \frac{d\psi_\infty(x)}{dx}. \quad (56)$$

Quite beautiful, and similar to the expression 2.14 in Ref. PrieveJFM1984, where the difference is due to the expression for the concentrations.

Inserting the expression for  $\frac{d\psi_\infty(x)}{dx}$  using Eq. (32) gives

$$\eta \frac{d^2 v_x(y)}{dy^2} = 2k_B T \frac{dC_\infty(x)}{dx} [\{\cosh[\Phi(x, y)] - 1\} + \beta \sinh[\Phi(x, y)]]. \quad (57)$$

This is the form they have in Ref. AndersonSPM1984, Eq. (26).

Now recall Eq. (28) and notice that  $\gamma$  and  $\kappa$  only depend on  $x$ , not on  $y$ , i.e.  $\tanh[\Phi(x, y)/4] = \gamma(x)e^{-\kappa(x)y} = q$ . We can use that

$$\sinh[\Phi/4] = \sinh[\tanh^{-1}(q)] = \frac{q}{\sqrt{1-q^2}} \quad (58)$$

$$\cosh[\Phi/4] = \cosh[\tanh^{-1}(q)] = \frac{1}{\sqrt{1-q^2}} \quad (59)$$

and

$$\cosh(\Phi) = 8 \cosh^4(\Phi/4) - 8 \cosh^2(\Phi/4) + 1 \quad (60)$$

$$\sinh(\Phi) = 8 \sinh^3(\Phi/4) \cosh(\Phi/4) + 4 \sinh(\Phi/4) \cosh(\Phi/4). \quad (61)$$

This gives

$$\cosh(\Phi) - 1 = 8 \cosh^4(\Phi/4) - 8 \cosh^2(\Phi/4) \quad (62)$$

$$= 8 \left( \frac{1}{(\sqrt{1-q^2})^4} - \frac{1}{(\sqrt{1-q^2})^2} \right) \quad (63)$$

$$= 8 \frac{q^2}{(1-q^2)^2} \quad (64)$$

and

$$\begin{aligned}
\sinh(\Phi) &= 8 \sinh^3(\Phi/4) \cosh(\Phi/4) + 4 \sinh(\Phi/4) \cosh(\Phi/4) \\
&= 8 \frac{q^3}{(\sqrt{1-q^2})^4} + 4 \frac{q}{(\sqrt{1-q^2})^2} \\
&= \frac{4q(1+q^2)}{(1-q^2)^2}.
\end{aligned} \tag{65}$$

Now we integrate the expression for  $\sinh$ , which turned out to be tricky. but Notice that  $q = e^{-\kappa y} \gamma = \tanh[\Phi/4]$  belongs to the interval  $[-1, 1]$ . Naively inserting the expression for  $q$  and integrating the expression with  $\sinh$  once with Mathematica gives

$$\int dy \sinh(\Phi) = -4 \frac{\gamma e^{-\kappa y}}{\kappa(1-\gamma^2 e^{-2\kappa y})} + c_1 \tag{66}$$

and integrating once more,

$$\int dy \left[ -4 \frac{\gamma e^{-\kappa y}}{\kappa(1-\gamma^2 e^{-2\kappa y})} + c_1 \right] = 4 \frac{\tanh^{-1}(e^{\kappa y}/\gamma)}{\kappa^2} + c_1 y + c_2. \tag{67}$$

But  $e^{\kappa y} \rightarrow \infty$  for  $\kappa y \rightarrow \infty$  as  $\kappa > 0$ . However, one can show that  $\frac{\tanh^{-1}(e^{-\kappa y} \gamma)}{\kappa^2}$  is also a solution to Eq. (65), and this solution is well-behaved for  $\kappa y \rightarrow \infty$ . This is due to the fact that  $\coth^{-1}(x) = \tanh^{-1}(1/x)$  and the derivatives of  $\tanh^{-1}$  and  $\coth^{-1}$  have the same form, but are valid in different intervals.  $\coth^{-1}$  on  $|x| > 1$  and  $\tanh^{-1}$  on  $[-1, 1]$ . I am not sure how to get the other solution from integration with Mathematica...

Similarly, the expression for  $\cosh - 1$  gives

$$\int dy [\cosh(\Phi) - 1] = 4 \frac{\gamma^2}{\kappa(e^{2\kappa y} - \gamma^2)} + d_1 \tag{68}$$

and integrating once more

$$\int dy \left[ 4 \frac{\gamma^2}{\kappa(e^{2\kappa y} - \gamma^2)} + d_1 \right] = \frac{2}{\kappa^2} [\ln(e^{2\kappa y} - \gamma^2) - 2\kappa y] + d_1 y + d_2. \tag{69}$$

However, taking the derivative twice of this expression gives minus the expression in Eq. (64), so the function should be

$$-\frac{2}{\kappa^2} [\ln(e^{2\kappa y} - \gamma^2) - 2\kappa y] + d_1 y + d_2. \quad (70)$$

So the final solution to Eq. (57) is

$$v_x(y) = \frac{2k_B T}{\eta} \frac{dC_\infty(x)}{dx} \left\{ -\frac{2}{\kappa^2} [\ln(e^{2\kappa y} - \gamma^2) - 2\kappa y] + 4\beta \frac{\tanh^{-1}(e^{-\kappa y}\gamma)}{\kappa^2} \right\} + h_1 y + h_2, \quad (71)$$

with the boundary conditions

$$v_x(0) = 0, \quad (\text{no slip at the wall}) \quad (72)$$

$$v_x(\infty) = \text{cst.} \quad (73)$$

The second condition gives  $h_1 = 0$ . Inserting  $y \rightarrow \infty$  gives zero for the first two terms, so  $h_2 = v_x(\infty)$ ;

$$v_x(y) = \frac{2k_B T}{\eta} \frac{dC_\infty(x)}{dx} \left\{ -\frac{2}{\kappa^2} [\ln(e^{2\kappa y} - \gamma^2) - 2\kappa y] + 4\beta \frac{\tanh^{-1}(e^{-\kappa y}\gamma)}{\kappa^2} \right\} + v_x(\infty), \quad (74)$$

For  $y = 0$ , the second term becomes

$$\frac{4}{\kappa^2} \tanh^{-1}(\gamma) = \frac{4}{\kappa^2} \tanh^{-1}\{\tanh[\Phi(x, 0)/4]\} = \frac{1}{\kappa^2} \frac{Ze\zeta}{k_B T} \quad (75)$$

and the other term returns  $-2 \ln(1 - \gamma^2)/\kappa^2$ . So

$$\begin{aligned}
v_x(0) &= \frac{2k_B T}{\eta} \frac{dC_\infty(x)}{dx} \left\{ -\frac{2}{\kappa^2} [\ln(1 - \gamma^2)] + \frac{\beta}{\kappa^2 k_B T} Z e \zeta \right\} + v_x(\infty) \\
&= -\frac{2\epsilon}{\eta} \left( \frac{k_B T}{Z e} \right)^2 \ln(1 - \gamma^2) \frac{d \ln C_\infty(x)}{x} + \frac{2\beta}{\eta \kappa^2} Z e \zeta \frac{dC_\infty(x)}{dx} + v_x(\infty) \\
&= -\frac{2\epsilon}{\eta} \left( \frac{k_B T}{Z e} \right)^2 \ln(1 - \gamma^2) \frac{d \ln C_\infty(x)}{x} + \frac{\epsilon k_B T}{\eta Z e} \beta \zeta \frac{d \ln C_\infty(x)}{dx} + v_x(\infty) \\
&= \frac{\epsilon k_B T}{\eta Z e} \left[ \beta \zeta - 2 \frac{k_B T}{Z e} \ln(1 - \gamma^2) \right] \frac{d \ln C_\infty(x)}{dx} + v_x(\infty).
\end{aligned} \tag{76}$$

Due to the boundary condition  $v_x(0) = 0$ , we get

$$v_x(\infty) - v_x(0) = v_x(\infty) = -\frac{\epsilon k_B T}{\eta Z e} \left[ \beta \zeta - 2 \frac{k_B T}{Z e} \ln(1 - \gamma^2) \right] \frac{d \ln C_\infty(x)}{dx} \tag{77}$$

which is exactly the expression Eq. (2.16) in Ref. PrieveJFM1984 and can be reduced to:

$$v_x(\infty) = -D_{DO} \frac{d \ln C_\infty(x)}{dx} \tag{78}$$

with  $D_{DO}$  diffusio-osmotic mobility is expressed as the sum of two contributions:

$$D_{DO} = \frac{\epsilon k_B T}{\eta Z e} \left[ \beta \zeta - 2 \frac{k_B T}{Z e} \ln(1 - \gamma^2) \right] \tag{79}$$

with parameters defined in the 'assumptions and definitions' Section 3.2.

This describes the slip velocity on a surface in the presence of a gradient of electrolytes.

In the lecture we will look at the flow rate induced in a channel open at both ends using Lee et. al<sup>8</sup>.

### 3.7 Diffusio-phoresis of particles.

Diffusio-phoresis, the motion of particles in a chemical gradient, should rather be seen as a motion of the liquid along the surface of the particle since the forces involved are acting on

the liquid, not the particle. Because of the pressure within the interfacial layer of solution at the 'front' and at the 'back' of the particle, the liquid moves along the surface and propels the particle. In the case of a charged particle in a gradient of one or more electrolyte, the interaction of the solute with the charged surface of the particle is electrostatic in nature and the diffusio-phoretic mobility has two contributions, one referred to as the chemio-osmosis and one extra contribution from electro-osmosis characterized by the difference of diffusion coefficient of the two ions of the electrolyte  $\beta$ .

The particle velocity can be expressed:

$$v_x(\infty) = -D_{DP} \frac{d \ln C_\infty(x)}{dx} \quad (80)$$

with  $D_{DP}$  diffusio-phoretic mobility that has some similarities with  $D_{DO}$  and is expressed as the sum of two contributions:

$$D_{DP} = \frac{\epsilon}{2\eta} \left( \frac{k_B T}{Ze} \right)^2 [u_o + \lambda u_1] \quad (81)$$

with  $\lambda = \lambda_{DB}/a$ ,  $a$  being the particle radius and all other parameters defined in the 'assumptions and definitions' Section 3.2.

The functions  $u_0$  and  $u_1$  are:

$$u_0 = 2\beta \frac{Ze\zeta}{k_B T} - 4 \ln(1 - \gamma^2) \quad (82)$$

$$u_1 = F_0 + \beta F_1 + Pe[F_2 + \beta(F_3 + F_5) + \beta^2 F_4] \quad (83)$$

with  $Pe$  the Peclet number and  $F_i$  a tabulated function.

Here we note that  $\lambda$  is a dimensionless number and we assume  $\lambda \ll 1$  to satisfy the assumption that the diffuse layer is thin compare to the 'length' of surface over which the salt gradient is acting. This is the case for particles that are in the order of 100 nm and the Debye layer is 0.7 nm. This will be put into use when looking at the paper by Rasmussen et

al.<sup>10</sup>.

## References

- (1) Duhr, S.; Braun, D. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 19678–19682.
- (2) Lippok, S.; Seidel, S. A.; Duhr, S.; Uhland, K.; Holthoff, H. P.; Jenne, D.; Braun, D. Direct detection of antibody concentration and affinity in human serum using microscale thermophoresis. *Analytical Chemistry* **2012**, *84*, 3523–3530.
- (3) Anderson, J. L. Colloid Transport by Interfacial Forces. *Annual Review of Fluid Mechanics* **1989**, *21*, 61–99.
- (4) Bruus, H. *Theoretical microfluidics*; Oxford University Press, 2007; Vol. 18; pp XVI, 346 S. (unknown).
- (5) Prieve, D. C.; Anderson, J. L.; Ebel, J. P.; Lowell, M. E. Motion of a particle generated by chemical gradients. Part 2. Electrolytes. *Journal of Fluid Mechanics* **1984**, *148*, 247–269.
- (6) Marbach, S.; Bocquet, L. Osmosis, from molecular insights to large-scale applications. *Chemical Society Reviews* **2019**, *48*, 3102–3144.
- (7) Shin, S.; Um, E.; Sabass, B.; Ault, J. T.; Rahimi, M.; Warren, P. B.; Stone, H. A. Size-dependent control of colloid transport via solute gradients in dead-end channels. *Proceedings of the National Academy of Sciences* **2016**, *113*, 257–261.
- (8) Lee, C.; Cottin-Bizonne, C.; Biance, A.-L.; Joseph, P.; Bocquet, L.; Ybert, C. Osmotic Flow through Fully Permeable Nanochannels. *Phys. Rev. Lett.* **2014**, *112*, 244501.

- (9) Anderson, J. L. Colloid transport by interfacial forces. *Annual Review of Fluid Mechanics* **1989**, *21*, 61–99.
- (10) Rasmussen, M. K.; Pedersen, J. N.; Marie, R. Size and surface charge characterization of nanoparticles with a salt gradient. *Nature communications* **2020**, *11*, 1–8.

# Analyzing nano- and biosystems using liquid phase transmission electron microscopy

By Murat Nulati Yesibolati, Hongyu Sun and Kristian Mølhav, DTU Nanolab

08.2023

## Contents

1	Introduction to Liquid Phase Electron Microscopy .....	2
2	Introduction to Electron Microscopy .....	3
2.1	Why use electrons .....	3
2.2	What happens when electrons hit solid matter? .....	6
2.3	The scanning and transmission electron microscope.....	6
2.4	The Components and types of electron microscopes .....	7
2.4.1	How can you make an electron gun? .....	8
2.4.2	How does an electro-magnetic lens work and how good is it?.....	8
2.5	Observing processes by In situ electron microscopy.....	11
2.6	Liquid phase electron microscopy.....	12
2.7	Radiolytic processes .....	14
3	Introduction to Synthesis Mechanisms - with a Direct View to the Processes.....	17
3.1	Nucleation: forming a critical nucleus .....	18
3.2	Nucleation rate .....	19
3.3	Heterogeneous Nucleation on a substrate.....	20
3.4	Growth mechanisms: growth of nanoparticles in solution .....	23
3.5	Growth mechanisms: growth on a substrate.....	27

# 1 Introduction to Liquid Phase Electron Microscopy

Liquid phase reactions are present all around us in nature, everyday life, industrial processes and science. Yet the ability to directly observe such reactions with molecular or atomic scale image resolution in the liquid environment has not been possible until around 2010.

Liquid phase electron microscopy uses ultrathin liquid samples that are encapsulated in-between two ultrathin electron transparent membranes as, in total, only some 10s-1000 nanometers thick. The thin sample allows the electron beam to pass through and provide a direct view to the on-going processes inside at the nano/atomic scale. Such ultrathin liquid samples are often created using microfluidic chip systems that provide a nanoscale laboratory to “*in situ*” observe the liquid phase processes inside the vacuum chamber of an electron microscope. Recent books and reviews introduce the emerging field<sup>1,2,3</sup>.

In the first part of the lecture notes introduce liquid phase electron microscopy (LPEM). As a foundation to understand the method, we give

- a short introduction to the electron microscope;
- the requirements to create a liquid cell for LPEM;
- how this in turn imposes important constraints on the method that one should be cautious about when using it.

LPEM is an emerging method still under rapid development, and there are more established methods of analysis using optical microscopy, x-rays or neutrons that we briefly touch upon to guide the selection of methods to use or to get complementary data to verify LPEM results. One competing method to LPEM is cryogenic transmission electron microscopy (cryo TEM) on frozen samples. The development of this important technique for studying biological structures on the nanoscale was rewarded the Nobel prize in Chemistry 2017<sup>4</sup> and the cryo TEM is rapidly growing in importance by providing insights into e.g. the atomic structure of proteins. We briefly review this method for comparison. We also describe the influence the high energy electron radiation on the studied processes. This will guide when LPEM provides truly unique data over other methods and a starting point for selecting the best possible methods, that each has advantages and unavoidable possible artifacts and assumptions required in the analysis.

The primary type of analysis done so far with LPEM is on liquid phase synthesis of nanoparticles. There are several theories describing liquid phase nucleation and growth. In the second part of the lecture, we review the theories together with experimental examples of LPEM studies that illustrate them or provide new insights. The field of nucleation and growth is a fundamental aspect of material science.

---

<sup>1</sup> <https://www.cambridge.org.proxy.findit.dtu.dk/core/books/liquid-cell-electron-microscopy/A668214DAFA539E0682ADF8672FE8C6C>

<sup>2</sup> <https://www.nature.com/articles/nmat944>

<sup>3</sup> <http://science.sciencemag.org/content/350/6267/aaa9886>

<sup>4</sup> [https://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2017/](https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2017/)

## 2 Introduction to Electron Microscopy

### 2.1 Why use electrons

Optical microscopy has limited resolution due to the interference of light waves as they form the image of the object through the optical system of the microscope. The achievable resolution for a wavelength  $\lambda$  is often given by the Abbe diffraction limit  $d$  as

$$d = \lambda / (n \sin \alpha)$$

with angle of incidence  $\alpha$ , and what is called the “numerical aperture” of the objective  $NA = n \sin \alpha$ . In air, NA can be approximated by the angle of incidence  $\alpha = R_{ob}/R_{wd}$  where  $R_{ob}$  is the radius of the objective lens aperture and  $R_{wd}$  the working distance, and hence NA is determined by the objective lens and is often written on the lenses, indicating their performance. Optical microscopes can often reach a resolution of about  $d = 200$  nm with  $\lambda = 400$  nm for green light.

The Abbe diffraction limits is given by the minimal distance between two points that can create constructive interference to form an image as in Figure 1.

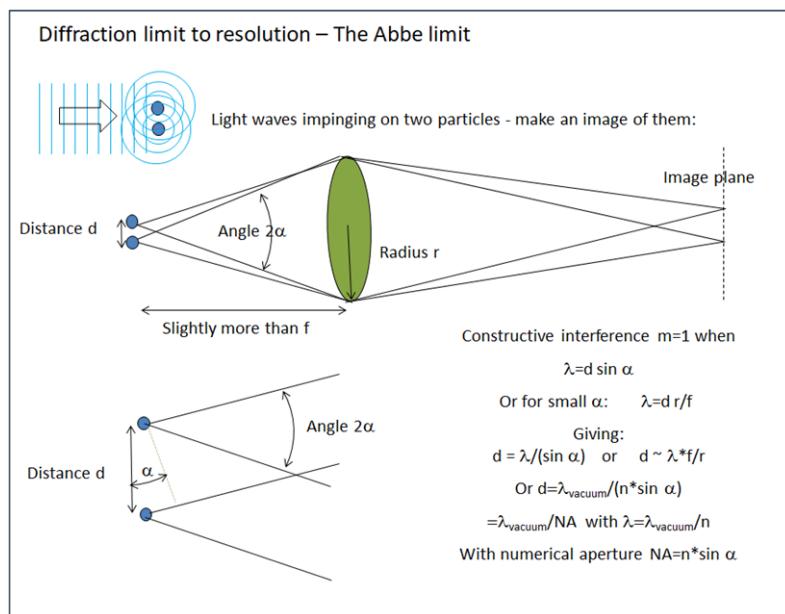


Figure 1: Interference of the light waves scattered from the object results in a diffraction limitation to the resolution in a microscope

Using radiation with smaller wavelength one can achieve a corresponding improvement in resolution. To go towards atomic scale Ångstrøm resolution we would need x-ray wavelengths from kilo electron volt radiation rather than optical wavelengths of about 1-2 eV.

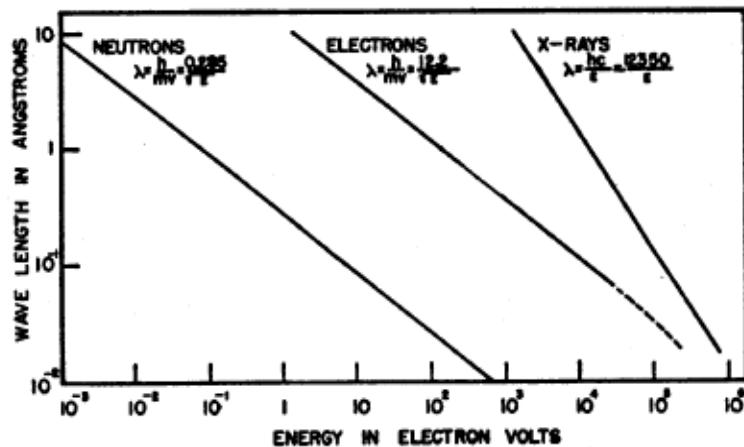


Figure 2: Comparison of the relationship between particle wave length and energy for X-rays, electrons, and neutrons.<sup>1</sup>

X-ray microscopes do exist but are very expensive as x-ray optics with high NA are difficult to create. Considering that X-rays can easily pass through your body, it is obvious that x-rays scatter very weakly on especially light atoms and small microscopic samples will not scatter very much and hence form very weak signals, and hence require large doses for detection. Similarly neutrons could be used as they can have very short wavelengths, but neutron sources are even more expensive (nuclear reactors and e.g. the new ESS being built in Sweden<sup>2</sup>), and neutron optics almost impossible to make.

Both x-rays<sup>3</sup> and neutrons<sup>4</sup> can be used to obtain a lot of information when doing diffraction and through the diffraction pattern one can study the atomic order and crystal structure of materials. X-ray diffraction is a widely used method that provides a wealth of information about the atomic scale crystal structure, even from a reasonable low cost table-top instrument and will remain an essential method in material science.

Compared to x-rays and neutrons, electrons are scattered orders of magnitude more strongly on atoms, and only can pass through few nano-micrometers of material depending on their energy.

The scattering of incoming radiation along the z-direction makes the incoming beam intensity reduce exponentially with distance. For each atom encountered there is a certain probability of being scattered and the number of scattering events of an incoming radiation follows a Poisson distribution decaying with decay constant called the mean free path  $\lambda_{mfp}$ :

$$P(z, n) = \frac{1}{n!} \left( \frac{x}{\lambda_{mfp}} \right)^n e^{-\frac{x}{\lambda_{mfp}}}$$

<sup>1</sup> <http://science.sciencemag.org/content/108/2795/69>

<sup>2</sup> <https://europeanspallationsource.se/>

<sup>3</sup> Cullity, B. D., Elements Of X Ray Diffraction. Addison-Wesley Publishing Company, Inc. 1956.

<sup>4</sup> Kisi, Erich H.; Howard, Christopher J., Applications of Neutron Powder Diffraction. Oxford Science Publications, 2008.

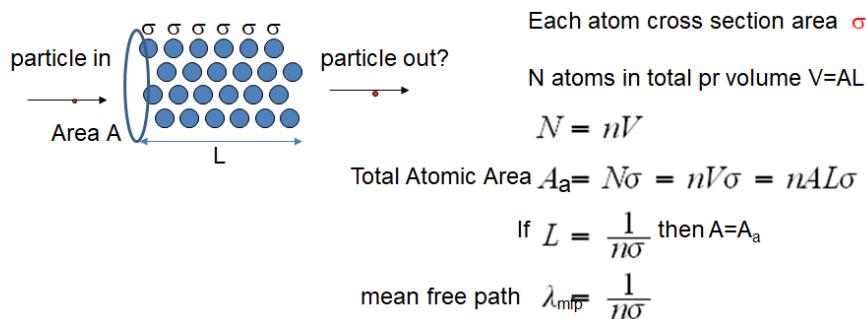
And to find the transmitted unscattered  $n=0$  beam, we find the exponentially decaying Beer-lambert law

$$P(z, n=0) = e^{-\frac{x}{\lambda_{mfp}}}$$

The mean free path in a material describes how strongly the material scatters incoming radiation. Considering the scattering is due to individual atoms, we can assign an equivalent cross section which is the average area an individual atom appears to have that appears to block the incoming radiation.

If one considers an incoming electron meeting a sample with atoms with a density  $n$ , then as illustrated below, the scattering cross section gives the total effective area the electron will experience a chance of absorption from.

Will the particle pass through?



*Figure 3 The scattering cross section and the relation to the mean free path of radiation passing through scattering particles.*

For electrons the cross section is about a billion times larger than for neutrons and x-rays<sup>1</sup>. Hence if one would like to image with molecular scale resolution, one also needs molecular scale strong signals, and electrons can hence provide a far better atomic scale image quality than x-rays and neutrons. But of course linked with a very short mean free path as shown in the figure below

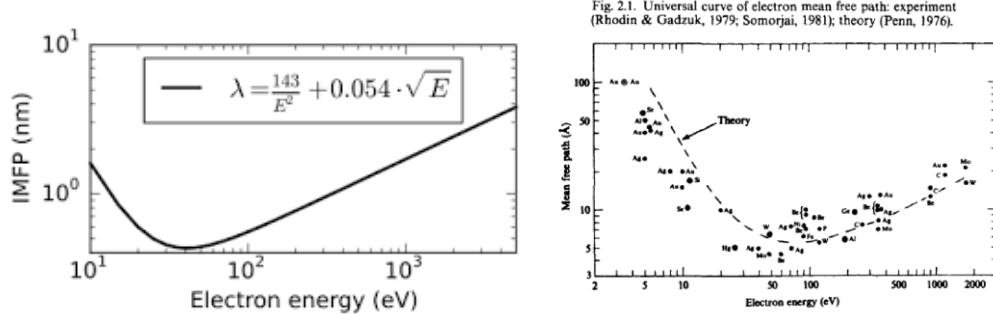
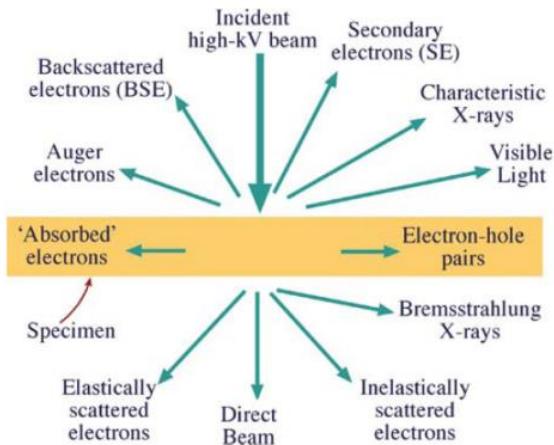


Figure 4 The mean free path as function of energy for an average material. The formula is often called 'the universal formula, as it makes a fairly good fit to the experimentally observed MFPs. Figure references: left figure, Seah, M. P.; Dench, W. A. (1979), Surface and Interface Analysis, 1: 2–11, doi:10.1002/sia.740010103. right figure, A. Zangwill, Physics at surfaces, Cambridge University Press, Cambridge, 1988, page 21.

<sup>1</sup> <http://digitalcollections.library.cmu.edu/awweb/awarchive?type=file&item=383731>

In addition electron imaging also gives the lowest damage to the sample for a given information level or image quality<sup>1</sup>. But of course imaging with electrons requires an ultra high vacuum chamber as the electrons would scatter strongly in any gas within the microscope given the large cross section for scattering.

## 2.2 What happens when electrons hit solid matter?



*Figure 5 The interactions between electrons and solid matter. Different kinds of signals are generated and can be detected. Figure source, adapted from text book "Transmission Electron Microscopy" by David B. Williams, C. Barry Carter.*

Since electrons are one type of ionizing radiation, during their interaction with solid matter, even the tightly bound, inner-shell electrons of the nucleus can be removed accompanying by transferring some of electron energy to individual atoms in the specimen. In this complex process, a wide range of signals are generated as shown in Figure 5. Many of the signals can be detected by a range of different detectors all depending on the type of electron microscope (such as high-angle annular dark-field, annular bright-field, energy-dispersive X-ray, electron energy loss detectors), giving us detail structural and chemical composition information about the solid matter.

## 2.3 The scanning and transmission electron microscope<sup>2,3</sup>

Electron microscopes are made in two main types: The scanning electron microscope (SEM) and the transmission electron microscope (TEM).

<sup>1</sup> <https://www.ncbi.nlm.nih.gov/pubmed/7568675>

<sup>2</sup> Goldstein, G. I.; Newbury, D. E.; Echlin, P.; Joy, D. C.; Fiori, C.; Lifshin, E. Scanning electron microscopy and x-ray microanalysis. New York: Plenum Press. 1981.

<sup>3</sup> Williams, David B., Carter, C. Barry. Transmission Electron Microscopy A Textbook for Materials Science. Springer, 2009.



Figure 6 SEM (left) and TEMs (center/right) in DTU Nanolab. Figure source from <https://www.nanolab.dtu.dk/english/microscopes/microscopes/transmission-electron-microscopes>.

In an SEM, a focused electron beam is scanned over the surface of a sample, and detectors continuously measures how many electrons are scattered back from the sample. A set of coils can create magnetic fields to move the beam in a raster pattern over the sample; and as it moves, the electrons scattered off the surface are detected to form an image with the scattered signal as function of position. The acceleration voltage in an SEM is in the range 1-30 kV, and such ‘low’ energy electrons do not penetrate very deep (max ~1 µm from the IMFP Figure 4). The samples can be as big as the vacuum chamber allows. The scattered electrons used for imaging however comes from a spread out region of the sample, so the resolution is limited to at best some 3 nm. Often an additional detector on the SEM allows simultaneous detection of the x-rays emitted from the sample that can be analyzed by a spectrometer to give information on what types of elements are present in the sample and map them.

In a TEM, the energy is typically much higher at 60-300 keV and the samples are so thin, 10-200 nm, that electrons can pass through the sample. Then the beam passes through a projector system of multiple electromagnetic lenses to achieve a magnification of up to several million times before the beam impinges on a phosphor screen for direct view by the operator through a window in the vacuum system, or on to a CCD camera or some other detector system. The beam can also be scanned in a raster pattern as in the SEM, and the scattered electrons detected to form a scanning transmission electron microscope (STEM) image.

In the TEM, the resolution is far better than SEM as the image can be made from unscattered electrons, and today reach 0.4 Å or just 40 picometers in current top models<sup>1</sup>. The TEM can be operated in many different ways, both as a bright field TEM (imaging with direct beam, where the electron beam functions analogous to an optical transmission microscopy), dark field TEM (imaging with the diffracted part of the beam), or it can be scanned and images built up from detecting the scattered electrons in different positions of the sample (STEM), or images formed from other signals such as the spectroscopic information from x-ray emission or energy loss.

Using electron beams unavoidably requires working in a vacuum environment, and all electron microscopes work under at least low pressures and usually in ultra-high vacuum chambers to avoid scattering of the electrons in the gas.

## 2.4 The Components and types of electron microscopes

---

<sup>1</sup> <https://www.cambridge.org/core/journals/microscopy-and-microanalysis/article/div-classtitledetection-of-single-atoms-and-buried-defects-in-three-dimensions-by-aberration-corrected-electron-microscope-with-05-aa-information-limitdiv/E83FE6E6AC9EEBD2901B4E4267C08404>

### 2.4.1 How can you make an electron gun?

SEM as well as TEMs have an electron gun held at a high and negative acceleration potential and emitting a well defined beam of electrons, that are accelerated across a gap towards a grounded electrode with a hole where the electrons then enter into the magnetic lens system which is called the column of the microscope all at a ground potential.

Extracting electrons from materials requires overcoming the work function. Traditionally this was done with a ‘thermionic gun’ based on a heated filament where the high ~2000 K temperature made a sufficient fraction of electrons reach energies to be emitted from the material and form an electron beam in an electric extraction field. The energy spread in kinetic energy from such an electron gun is however 1-2 eV given the thermal energy spread, and this can limit the TEM resolution because an effect called chromatic aberration which is when electrons with different energies will take different paths through the electron optical system of the microscope due to imperfections in the optics.

For the best possible control of energy and hence minimal influence of imperfections in the optics, a ‘cold field emitter’ can be used. A field emitter is an extremely sharp – ideally atomically sharp - needle in a very high extractor field, and in this configuration electrons can make direct tunneling through the work function barrier of the needle material and into the accelerating extraction field. The FEG cannot provide as high beam currents as a thermionic gun, but provides a far more concentrated electron source with a far more narrow energy spread, making the electron beam very bright and coherent, which is the ideal source for optimal image quality in a TEM and high resolution imaging.

Emitter type	Thermionic Tungsten filament	Cold field emitter
Working Temperature (K)	2700	300
Current density (A/m <sup>2</sup> )	10 <sup>4</sup>	10 <sup>10</sup>
Brightness (A/sr m <sup>2</sup> )	10 <sup>9</sup>	10 <sup>13</sup>
Energy spread (eV)	3	0.3
Vacuum (Pa)	10 <sup>-2</sup>	10 <sup>-8</sup>
Lifetime (h)	100	>1000
Cost	20 USD	10000 USD

Figure 7 Characteristics of the three typical electron sources operating at 100 kV.

### 2.4.2 How does an electro-magnetic lens work and how good is it?

Magnetic lenses are used to focus the electron beam and form the illuminating and imaging beams. These are electromagnets with a magnetic field designed to work as a lens on the moving electrons through the Lorentz force  $F=q \mathbf{v} \times \mathbf{B}$ . The field in the lenses can be up to almost two tesla which is about the maximal achievable in a nonsuperconducting electromagnet.

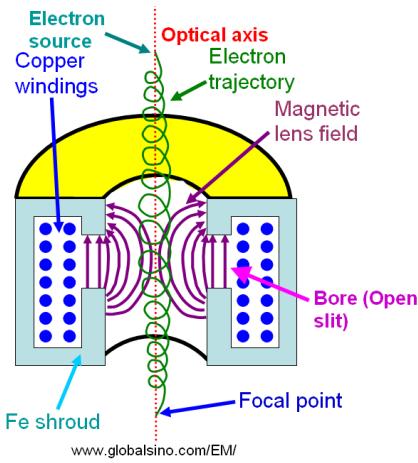
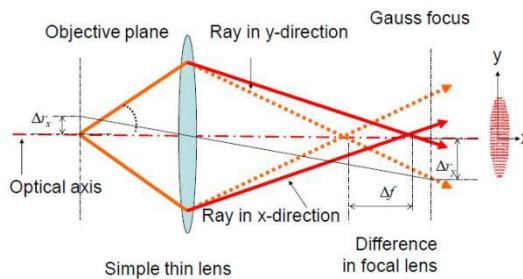


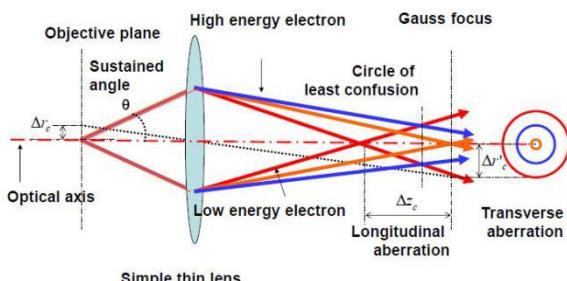
Figure 8 Working principle of an electro-magnetic lens. Image source, [www.globalsino.com](http://www.globalsino.com)

An electromagnetic lens is however not a very good lens, often said to be marginally better than the bottom of a glass bottle as a magnifying glass. The electromagnetic lens introduces numerous imperfections in the optical system of the microscope:

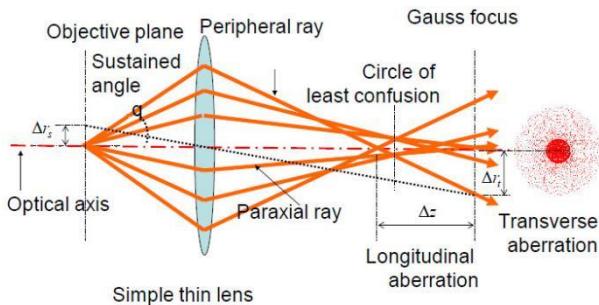
- It gives astigmatism, which is when the two directions perpendicular to the beam has different focus length. Special coils in the column have also been included in SEM and TEMs to provide adjusting fields to counter this imperfection.



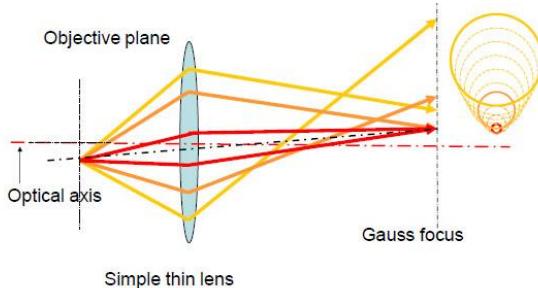
- Chromatic aberration is when different electron energies give different focus length. This is typically reduced by minimizing the energy spread from the electron gun by choosing a FEG source.



- Spherical aberration is when the focus length depends on the distance from the beam central axis, so central parts of the beam have different focus lengths compared to the outermost parts. Recently 'spherical aberration correctors' that require a very complicated set of correction coils and elaborate control systems have been developed and the reduction of spherical aberration has significantly improved the TEM resolution to become subatomic at 0.4 Å. Aberration corrected microscopes are however also quite expensive as the correction system is highly complicated.



- Coma, or comatic aberration is when off-axis point sources such are smeared out.



Today the resolution of TEM has reached the subatomic scale and the best TEMs provide about 0.4 Å resolution. Resolution can be improved further, but the added information below 0.4 Å will come at extremely high costs, but will provide for instance better images of light atoms.

### Question: Why would better TEM imaging of light atoms be so interesting?

**The resolution in TEM:** Once a FEG, low energy spread beam in an aberration corrected microscope is used, the main resolution-limiting factor is chromatic aberration caused by inelastic scattering of electrons in the sample. The full-width at half-maximum of the energy distribution  $\Delta E$  of transmitted electrons follows from calculations of the inelastic electron scattering cross-section and is given by the following equation:

$$\Delta E = N_A e^4 Z \rho T / 2\pi \epsilon_0^2 W m_0^2 v^2 \quad (1)$$

Here,  $N_A$  is Avogadro's number,  $e$  is the elementary charge,  $Z$  the atomic number,  $\rho$  the density and  $T$  the thickness of the liquid layer,  $\epsilon_0$  the permittivity of space,  $W$  the atomic weight,  $m_0$  the mass and  $v$  the velocity of the electrons.

This energy broadening affects the resolution of the image because of the chromatic aberration of the objective lens. The image resolution  $d_c$  can be approximated by<sup>1</sup>:

$$d_c = \alpha C_c \Delta E / 2E \quad (2)$$

where  $\alpha$  is the objective semi-angle,  $C_c$  is the chromatic aberration coefficient and  $E$  is the beam energy.

---

<sup>1</sup> Reimer, L. & Kohl, H. Transmission Electron Microscopy: Physics of Image Formation. Springer, 2008.

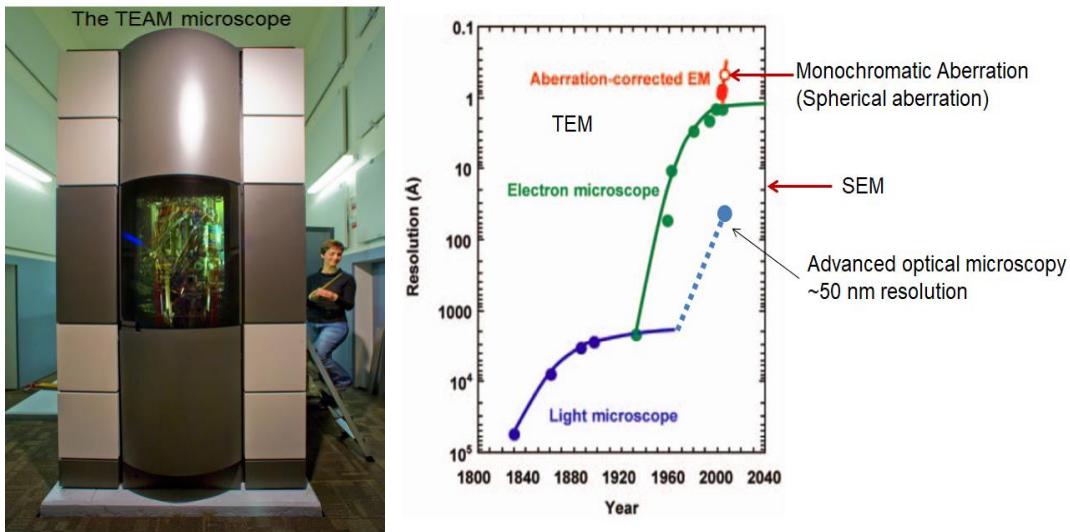


Figure 9 The evolution of resolution in microscopy. Adapted from S. Pennycook et al.<sup>1</sup>

**Question: How do they achieve super-resolution optical microscopy at some 50 nm resolution and how does this compare to TEM?**

## 2.5 Observing processes by In situ electron microscopy

One of the major ongoing TEM developments is currently for better cameras that can detect individual electrons with high quantum yield and at high frame rate to both allow imaging fast process and to avoid electron beam induced sample damage by reducing the required amount of radiation to the absolute minimum. Another aspect is to enable imaging processes live rather than just a sample taken from a process – this is in situ TEM.

Today, given the ability to image with atomic resolution, there is increased focus on developing the microscope into a micro/nano-laboratory<sup>2</sup> where external stimuli can be applied to the sample and the response observed directly.

In so-called ‘environmental electron microscopes’ ETEM/ESEM, differential pumping systems are used to make it possible to have a very small few millimeter large confined region with moderate pressures up to ~0.01 Atm of gasses present around the sample together with the electron beam, which allow imaging many gas phase chemical reactions that are important in e.g. catalysis, but not readily to have water present and study liquid samples<sup>3,4</sup>.

In these, a set of small apertures around the beam with very strong vacuum pump systems in between work to limit the gas flow from the sample region into the electron gun and

<sup>1</sup> <https://www.cambridge.org/core/journals/mrs-bulletin/article/div-classtitlematerials-advances-through-aberration-corrected-electron-microscopydiv/3C5F085F80F48063062ADC493CC68003>

<sup>2</sup> E.P. Butler, In situ experiments in the transmission electron microscope. Rep Prog Phys **42**(5), 833–895 (1979). doi:[10.1088/0034-4885/42/5/002](https://doi.org/10.1088/0034-4885/42/5/002)

<sup>3</sup> <http://www.sciencedirect.com/science/article/pii/S030439919600099X>

<sup>4</sup> <http://www.sciencedirect.com/science/article/pii/S0968432812000510>

projection system. In this way the required ultra-high vacuum can be maintained throughout the system apart from a few millimeters in the sample region where the electron beam still will be able to pass through gasses at pressures below 10 mBar, to provide high resolution imaging conditions.

Heating samples has been standard for many years using special sample holders with small heating systems, and often used in ETEMs to image e.g. catalytic nanoparticle reactions. More complex sample holders are now available with even scanning tunneling microscopy(STM) and atomic force microscopy(AFM) commercially available that fit into TEMs and similarly small robotic manipulation systems that can mechanically or electrically contact nanostructrues to probe their response in SEMs.

### Question: What would it require to have liquid water in the ETEM?

#### 2.6 Liquid phase electron microscopy

Given the ubiquitous presence of liquid phenomena, creating devices to contain liquids and attempting to image liquid phase processes was attempted even back in the 40's shortly after the invention of the TEM.<sup>1</sup> The actual resolution of the liquid cell systems was never better than optical microscopy, until microfabrication methods were used to make electron transparent silicon nitride membranes on microchips around 2003 at IBM.<sup>2</sup> It took another 10 years before real atomic scale resolution was achieved in 2012 with three papers in the journal Science achieving this and showing new details of nanoparticle synthesis processes in the same year,<sup>3</sup> 70 years after the first attempts of liquid cell imaging. This kick started the field of liquid phase electron microscopy (or liquid cell electron microscopy) which has rapidly developed in the past years.

The pressure from the liquid sample causes an outward pressure on the containing system in the TEM vacuum chamber. The thicker liquid layer in the bulging regions will deteriorate the microscope resolution and contrast given that at sufficiently large thickness no beam will be transmitted, and what does pass through has been scattered multiple times. To obtain a good TEM resolution, the sample must remain thin enough and the encapsulating material must be as strong as possible to reduce outwards bulging, but at the same time with the thinnest membrane possible, that also must consist of light elements, in order to reduce beam scattering in the membrane itself to make it maximally electron transparent. Finally it must be chemically inert and able to withstand extended electron radiation. Silicon based materials such as silicon dioxide, silicon nitride, silicon rich silicon nitride, or hexagonal boron nitride, and graphene are good candidates. Graphene could in many ways seem to be ideal, but is not yet readily available as a standard microfabrication method although development is very fast towards this.

---

<sup>1</sup> I.M. Abrams, J.W. McBain, A closed cell for electron microscopy. *Science* 100, 273–274 (1944)

<sup>2</sup> M.J. Williamson, R.M. Tromp, P.M. Vereecken, R. Hull, F.M. Ross, Dynamic microscopy of nanoscale cluster growth at the solid–liquid interface. *Nat Mater* 2, 532–536 (2003)

<sup>3</sup> J.M. Yuk, J. Park, P. Ercius, K. Kim, D.J. Hellebusch, M.F. Crommie, J.Y. Lee, A. Zettl, A.P. Alivisatos, High-Resolution EM of Colloidal Nanocrystal Growth Using Graphene Liquid Cells. *Science* 336, 61 (2012)

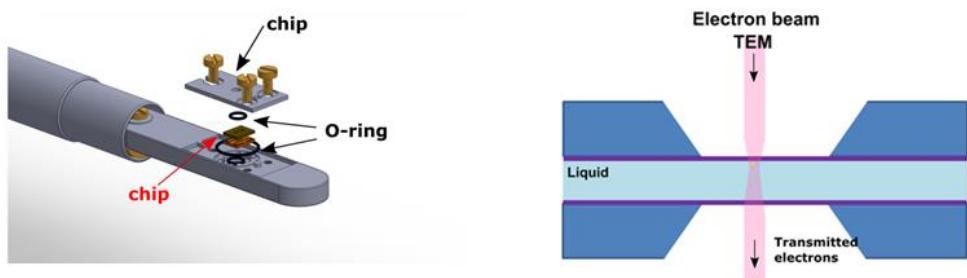


Figure 11 Liquid cell TEM holder, and the cross section of a sandwich setup under the irradiation of electron beam.

The most common system that is also commercially available from several vendors is made up of two chips with silicon nitride membranes that are clamped together mechanically. An optional ‘spacer layer’ can be included between the chips to maintain a minimal liquid thickness layer if for instance flow through the liquid cell is wanted. The mechanical alignment of the two chips and the microfabrication and dicing of the silicon chips gives an uncertainty in the position of the two membranes and typically membranes of at least 50  $\mu\text{m}$  are used.

The shortest membrane length determines the maximal bulge and in these systems the bulge often is about 2  $\mu\text{m}$  liquid layer in the middle due to the 50  $\mu\text{m}$  membrane length. This is a considerable thickness with little chance of doing TEM in the thick region, although STEM can reach a few nm resolution in 1  $\mu\text{m}$  thick samples. A thinner liquid layer thickness has been demonstrated in custom built devices, such as microfabricated nanochannel chips or in liquid droplets caught between graphene layers.

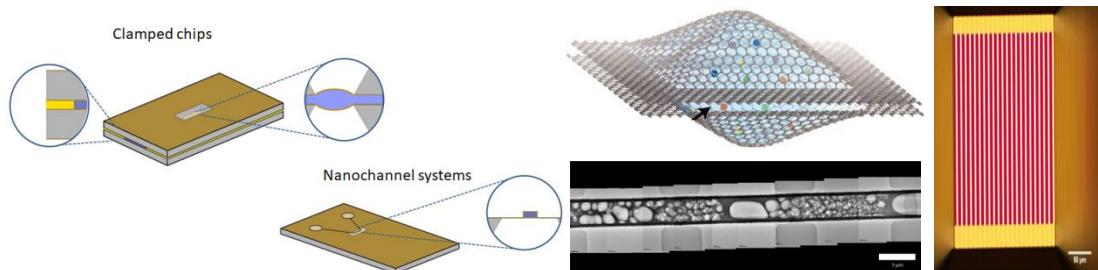


Figure 12 (Left) The two basic types of microfabricated in situ liquid TEM setups. The drawings are not to scale. (a) A cross-sectional view of a sandwich setup. Two silicon chips are placed together separated by a spacer material (yellow). Each chip has a hole etched in the center over which an electron transparent membrane (brown) is suspended. Liquid (blue) fills the volume between the windows. (b) A monolithic chip with a fabricated channel that encloses the liquid. The membrane material also forms the channel walls. In both cases, the chips are typically each 200 – 300  $\mu\text{m}$  thick, with the windows and liquid on the order of 50 nm thick. (Right) Graphene liquid cell and direct bonded nanochannel cell. Figures from Eric Jensen thesis, <https://doi.org/10.1016/j.mee.2017.01.038> and 10.1126/science.1217654

**Resolution in LPTEM:** The achievable resolution of electron microscopy in liquids is limited by the mass thickness of combined membranes and liquid layer. The limiting factor is the chromatic aberration as the beam energy spread becomes wide due to inelastic scattering in the sample<sup>1</sup>.

In TEM imaging mode, for water, the effective Z = 4.7, combining the above two equations, we obtain the resolution expression for TEM:

$$d_{\text{TEM}} = 6 \times 10^{12} \alpha C_e T / E^2 \quad (3)$$

<sup>1</sup> <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2818.2010.03484.x/full>

with  $E$  in eV.

For typical values,  $\alpha = 10$  mrad,  $C_c = 2$  mm,  $E = 200$  keV and  $T = 1 \mu\text{m}$ , we obtain  $d_{\text{TEM}} = 4$  nm. At a high tension of 300 keV, a higher  $d_{\text{TEM}}$  of  $\sim 1$  nm has been obtained in organic solvent<sup>1</sup>.

In STEM imaging mode, the best resolution can be obtained when the sample is on the top membrane of the liquid cell where the electron beam is optimally focused. Therefore, nanometer scale resolution can easily be obtained even for micrometers-thick liquid. If the sample is in a deeper area of the liquid, imaging takes place with a scattered and blurred probe, which results in a worse resolution. The STEM resolution in a liquid sample can be thought of as the diameter of the smallest nanoparticle visible above the background noise<sup>2</sup>:

$$d_{\text{STEM}} = 5l_{\text{np}}(T/N_0 l_{\text{liquid}})^{1/2} \quad (4)$$

$l_{\text{np}}$  and  $l_{\text{liquid}}$  are the mean free path lengths for elastic scattering of nanoparticle and liquid, respectively;  $N_0$  and  $T$  are the total electron numbers and temperature.

**The resolution in SEM:** The resolution follows from the beam broadening equation:

$$d_{\text{SEM}} = 6.1 \times 10^3 Z T^{3/2} / E \cdot (\rho/W)^{1/2} \quad (5)$$

where  $T$ ,  $Z$ ,  $\rho$  and  $W$  are as in Eqn 1 parameters of the membrane (S)TEM.

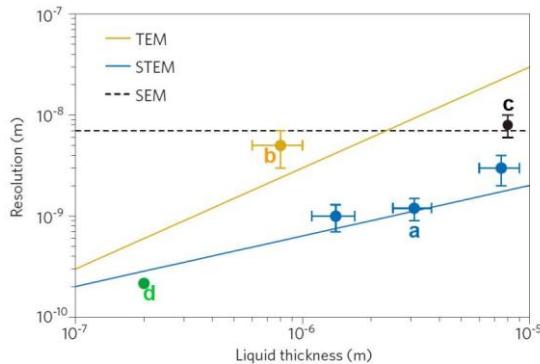


Figure 13 Resolution of different forms of electron microscopy in liquid. Theoretical maximal resolution versus water thickness for TEM, STEM and SEM. The resolution was calculated for typical TEM and STEM instrument parameters at 200 keV beam energy<sup>3</sup>, and for the imaging of Au nanoparticles at the bottom of a layer of water for TEM, and at the top of the layer for STEM. The resolution obtained in SEM just below the liquid-enclosing membrane does not depend on the liquid thickness. Experimental data points are shown for Au nanoparticles in TEM (b: J. Microsc. 2011, 242, 117-123), STEM (a: Ultramicroscopy 2010, 110, 1114-1119) and SEM with a 30-nm-thick SiN window (c: J. Struct. Biol. 2010, 169, 438-449), and for PbS nanoparticles in water imaged with STEM (d: Nano Lett. 2011, 11, 2809-2813). The error bars represent experimental errors. Figure from Liquid phase electron microscopy by F. Ross et. al. chapter 8, <https://doi.org/10.1017/9781316337455>

## 2.7 Radiolytic processes

During imaging of a sample in liquid, the electron beam interacts with silicon nitride membrane, the liquid, and the sample. As we mentioned above, this process produces many different signals, and also results in both generation of heat and changes in chemistry. Since

<sup>1</sup> <http://science.sciencemag.org/content/324/5932/1309>

<sup>2</sup> <http://www.pnas.org/content/106/7/2159.abstract>

<sup>3</sup> <https://www.nature.com/articles/nmat944>

the sample-liquid interaction is essential for the processes being investigated, it is therefore important to study the effect of electron beam irradiation on both the liquid and sample properties.

In the past radiolysis has been extensively studied to ensure safety of nuclear reactors etc. The radiation conditions in a focused electron microscope beam are however extremely more intense than what is found typical nuclear facilities. The radiation dose measured as deposited energy pr mass has the unit Gray in joules radiation energy absorbed pr kilo matter. A deadly dose is around 10 Gy, as e.g. given to tumors in radiation therapy and in nuclear reactors dose rates are of the order Gy per second. But in a focused TEM beam, the dose rate can become gigaGy pr second. Although much is known about the complex radiation chemistry in water at nuclear dose rates, there is so far little quantitative and verified knowledge about the radiation chemistry in the range relevant for TEM/SEM.

Based on the measured reactions occurring at nuclear dose rates for water, with some 80 different reaction constants measured, water radiolysis in TEM was systematically modelled by Schneider et al<sup>1</sup> by using a finite element model (FEM). This approach has the implicit limitation that it is extrapolated to rates a million times higher than that reference conditions.

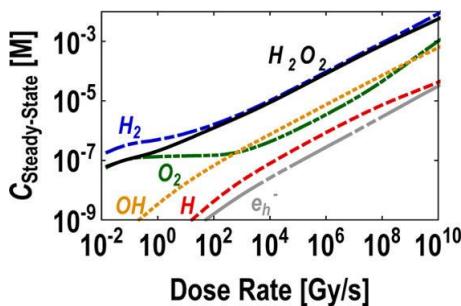


Figure 14 The steady state concentrations of  $e_h^-$ ,  $H^\cdot$ ,  $H_2$ ,  $H_2O_2$ ,  $OH^\cdot$ , and  $O_2$  as functions of dose rate <sup>2</sup>

Schneider et al. concluded that:

1. The radiolysis products reach a steady state concentrations at the high dose rates associated with electron microscope imaging. As the dose rate increases, these steady state concentrations increase as shown in Fig 14, while the time to achieve the steady state decreases; The main products are hydrogen gas and hydrogenperoxide.
2. The initial properties of solution, such as pH and oxygen concentration, have a large effect on the steady state concentrations of the radiolysis products. Significantly, irradiation has a strong effect on the local pH of a solution;
3. The radiolysis effect must be carefully considered in LPEM experiments to achieve correct understanding the results.

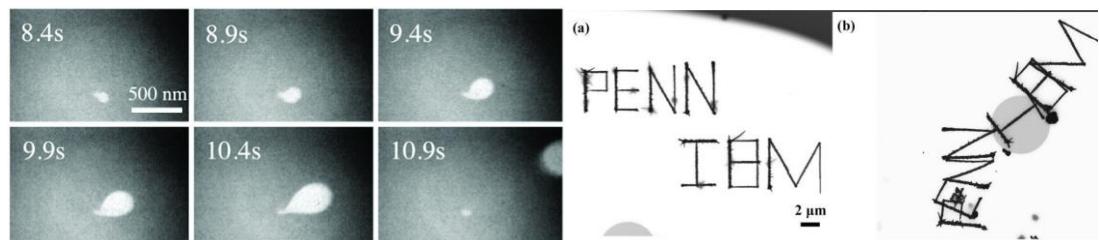
Experimentally the radiolysis can have strong effects on LPEM results:

1. Primarily the formation of bubbles upon intense radiation is often observed, and this is taken as a sign of radiolytic hydrogen formation and bubble create when the solution becomes supersaturated as shown in the image below;
2. Local heating are expected but difficult to measure and quantify in TEM;
3. Changes in pH or charging effects are also apparent ;

<sup>1</sup> <http://pubs.acs.org/doi/full/10.1021/jp507400n>

<sup>2</sup> <http://pubs.acs.org/doi/full/10.1021/jp507400n>

4. The secondary electrons can create hydrated electrons that is a very strong reducing agent that can cause numerous reactions. For experiments the electron beam is often used to initiate or drive the chemical process that is being studied, such as gold nanoparticle nucleation. By moving the beam, gold particles can be nucleated in patterns to write letters as shown below.



*Nucleation, growth, and migration of radiolytic bubbles at a serendipitous nucleation site<sup>1</sup> and Beam-induced precipitation of gold from 20 mM HAuCl<sub>4</sub> solution<sup>2</sup>*

---

<sup>1</sup> <http://pubs.acs.org/doi/full/10.1021/nl404169a>

### 3 Introduction to Synthesis Mechanisms - with a Direct View to the Processes

In the drive toward greener and more efficient processes, nanoparticle engineering is an increasingly important requirement in functional materials such as catalysts for chemical synthesis, batteries, fuel cells, hydrogen production and energy storage etc. In order to achieve and optimize the desired functionality of the many essential nanostructured components in such processes, it is crucial to have the best possible control the fundamental properties, such as nanoparticles' crystal perfection, crystal size, the characteristic shape (habit), chirality if there are left/right symmetry, and of course the overall synthesis cost and increasingly also environmental impact of the process.

So far, many synthesis strategies have been established to obtain a wide range of crystal structures based on physical (physical and chemical vapor deposition, laser ablation, ball milling, lithographic, etc), chemical (gas phase reaction, various solution phase synthesis), or biological methods. Comparing to other methods, the solution phase route shows unique advantages, such as low cost, being both low in energy consumption and equipment costs, easier scalability to larger volumes, and overall ease of handling. Most of the solution phase reactions occur under mild condition with a relatively low temperature ( $< 200\text{ }^{\circ}\text{C}$ ), which also often will be a less energy consuming synthesis route than the often very high temperatures and low volume production from physical methods. Therefore, solution phase synthesis has and will be an essential process that is likely to be the best alternative of the many methods if in any way possible for a given nanostructure. Typical solution phase synthesis includes precipitation, microemulsions, hydrothermal/solvothermal, sol-gel, electrochemical deposition, chemical bath deposition, and so on.

The formation of any solid material from a liquid phase includes nucleation and growth processes. Nucleation can be either homogeneous, in the absence of foreign particles or surfaces in the solution, or heterogeneous, in the presence of foreign particles or surfaces in the solution. Both types of nucleation are collectively known as primary nucleation for the initial formation of a new phase of a material.

This chapter sets out by analyzing the initial nucleation and presenting a model for how this proceeds, and from this explains why heterogeneous nucleation is faster. Based on this foundation we describe the fundamental models of nucleation and growth that are currently considered, as well as the nucleation and growth that can take place on a substrate. Together this theoretical foundation of models provide a 'tool box' of modes a material can grow in liquids and often also in other physical processes, and has a very wide applicability in physics, chemistry and materials science. Throughout this model overview, we provide a series of experimental verifications and observations of the phenomena from liquid cell TEM studies that are uniquely suited to provide new information in this fundamentally important scientific area.

**Question: Can you list some important homogenous and heterogeneous nucleation processes?**

### 3.1 Nucleation: forming a critical nucleus

The formation of a new solid entity from a solution starts through the nucleation process, which represents the very first stage of any crystallization process. Nucleation is defined as the series of atomic or molecular processes by which the atoms or molecules of a reactant phase rearrange into a cluster of the product phase that is large enough to have the ability to grow irreversibly to a larger size particle. Initially the nuclei of solids are very unstable, because of their small size as we show below and heterogeneous occurs much easier than homogeneous, since a stable nucleating surface is already present in heterogeneous nucleation.

For simplicity, we first consider the thermodynamics of homogeneous nucleation. The total free energy of nucleation can be viewed as the sum of the energy released by the free single atoms in forming a particle by condensation of atoms. The gained energy will be different depending on whether the atom is in the bulk (giving a volume term) and on the particle surface (giving a surface term) as their surroundings will be different. One can think of this for instance in terms of the bonds each atom can make to neighboring atoms.

**Question: This mentions ‘solid entity’, but would the same apply to say ice forming in water, carbon dioxide bubbles when you open a drink, or vapor bubbles when boiling water?**

The total free energy change  $\Delta G$ , which must be negative for the nucleation to proceed<sup>1</sup>:

$$\Delta G = \frac{4}{3}\pi r^3 * \Delta G_v + 4\pi r^2 * \gamma \quad (1)$$

For a spherical particle of radius  $r$ , surface energy  $\gamma$  and the free energy of the bulk crystal<sup>2</sup>

$$\Delta G_v = -k_B T \ln S / V_0 \quad (2)$$

with  $k_B$ ,  $T$ ,  $S$ , and  $V_0$  are Boltzmann’s constant, temperature, the supersaturation of the solution, and its molar volume), here the supersaturation  $S$  is defined as the ratio of the solution concentration to the equilibrium concentration in a saturated solution.

Since the surface term always increases with radius squared, and volume term being negative with radius cubed, there must be a critical radius, which corresponds to a maximum total free energy, i.e. maximally unstable nuclei. By setting  $d\Delta G/dr = 0$ , the critical radius  $r^*$  is found to be

$$r^* = -2\gamma / \Delta G_v \quad (3)$$

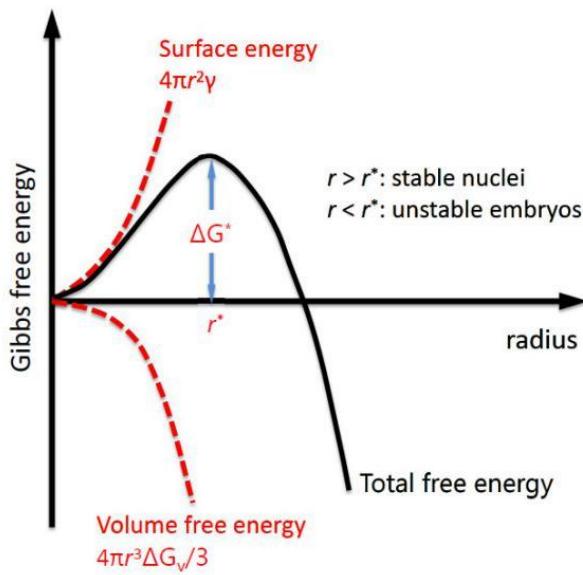
and the critical (maximum) total free energy barrier is

$$\Delta G^* = 16\pi\gamma^3 / (3\Delta G_v^2) \quad (4)$$

---

<sup>1</sup> For more details about the classical nucleation theory part, you can refer to, for example, CHAPTER 8: SOLIDIFICATION in *Physical Metallurgy* (Fourth, Revised and Enhanced Edition), Edited by. Robert W. Cahn, Peter Haasen. © 1996 Elsevier Science B.V.

<sup>2</sup> <http://www.sciencedirect.com/science/article/pii/S0022024813003278>



*Figure 15. Free energy of a cluster of atoms as a function of radius: two main components of energy change in forming a cluster are defined: upper curve is surface free energy due to solid–liquid interfacial energy  $\gamma$  which increases with radius; lower curve is volume free energy that decreases with radius as the nucleation phase is more stable than the matrix and  $\Delta G_v$  is the difference in volume free energy between matrix and nucleus; middle curve is the sum of total free energy change as a function of radius. The “critical nucleus” with a radius of  $r^*$  correlates to a free energy of  $\Delta G^*$ . Image is adapted from source [https://en.wikipedia.org/wiki/Critical\\_radius](https://en.wikipedia.org/wiki/Critical_radius)*

Equations (2) and (3) lead to the following conclusions:

The critical radius corresponds to the minimum size at which a particle can survive in solution without being redissolved into the solution, i.e., the nuclei with a size of  $r > r^*$  are stable and can grow further, while the nuclei with a size of  $r < r^*$  tend to dissolve. The particle's free energy presents a barrier for nucleation, where a critical free energy change  $\Delta G^*$  is required to obtain stable nuclei; The smaller this critical size or free energy, the easier nucleation can occur as this can be overcome by thermal energy fluctuations.

The critical nucleus size and free energy of nucleation is governed by the magnitude of surface energy and volume free energy. So any process that modifies these values will have an influence on the possible viability of the nucleation process.

**Question: List some possible factors that can influence the critical nucleus size and free energy. How can you experimentally realize them?**

### 3.2 Nucleation rate

The nucleation rate for classical nucleation theory can be determined by assuming that a single atom added to the subcritical size nucleus will result in a stable cluster.

Given the energy barrier  $\Delta G^*$ , there is a probability that thermal fluctuations can overcome the nucleation dissolutions, and the nucleation rate can be defined as the number of nuclei formed per unit time per unit volume, and can be expressed by<sup>1</sup>:

$$dN/dt = A \exp(-\Delta G^*/k_B T) \quad (5)$$

here  $A = C_0 k_B T / 3\pi\eta d^3$  is a pre-exponential factor, where  $\eta$  corresponds to the viscosity of the solution,  $d$  to the diameter of growth species, and  $C_0$  to the initial concentration.

Substituting eqs. (2) and (4) into eq. (5), we obtain:

$$dN/dt = C_0 k_B T / 3\pi\eta d^3 \exp(-16\pi\gamma^3 v_0^2 / (3k_B^3 T^3 \ln^2 S)) \quad (6)$$

This equation shows that high nucleation rates are promoted by high temperature, high initial concentration of growth species, high supersaturation (the ratio of the solution concentration to the equilibrium concentration in a saturated solution), and low viscosity.

### 3.3 Heterogeneous Nucleation on a substrate

Often, in practical synthesis or in industrial environments, the solutions can contain impurities of another solid phase materials. In the presence of such other nucleation surfaces/centers (impurities, walls, bubbles, drops, etc.), the barrier  $\Delta G^*$  for nucleation to occur is globally decreased. Unlike homogeneous nucleation, the nuclei are formed on the already existing surface of a foreign body. In this process called heterogeneous nucleation, the nuclei are no longer spherical shape but a cap forming a spherical contact angle of  $\theta$  ( $0^\circ \leq \theta \leq 180^\circ$ ) with the support.

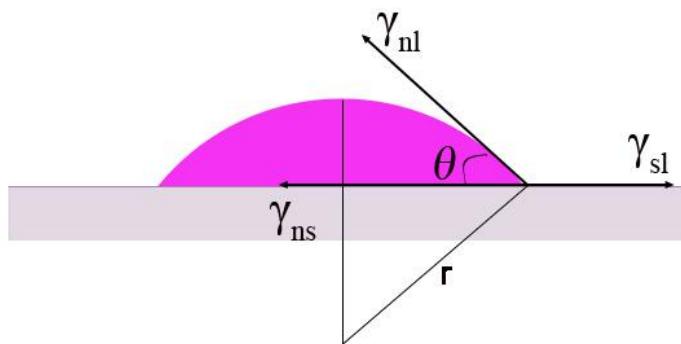


Figure 16. Geometry for heterogeneous nucleation. At the nucleus-substrate interface the nucleus (n) is a half-lens with different surface areas contacting the liquid (l) and substrate (s) interfaces. The contact angle between the nucleus and substrate is  $\theta$ .

Similar to the case of homogeneous nucleation, the free energy needed for heterogeneous nucleation ( $\Delta G_{het}$ ) can also be given in terms of surface free energy and volumetric free energy of the nucleus, but here a new surface energy term is added due to the nucleus-substrate interface:

---

<sup>1</sup> For more details about the classical nucleation theory part, you can refer to, for example, CHAPTER 8: SOLIDIFICATION in *Physical Metallurgy* (Fourth, Revised and Enhanced Edition), Edited by. Robert W. Cahn, Peter Haasen. © 1996 Elsevier Science B.V.

$$\Delta G_{\text{het}} = S_{\text{ns}}(\gamma_{\text{ns}} - \gamma_{\text{sl}}) + S_{\text{nl}}\gamma_{\text{nl}} - V\Delta G_v \quad (7)$$

$S$  and  $V$  are the surface area and volume of the nucleus. Considering the geometry shown in Figure 16,  $S_{\text{nl}} = 4\pi r^2(1-\cos\theta)$ ;  $S_{\text{ns}} = \pi r^2 \sin^2\theta$ ;  $V = 4\pi r^3(2-3\cos\theta+\cos^3\theta)/3$ , and  $\gamma_{\text{ns}} - \gamma_{\text{sl}} = -\gamma_{\text{nl}}\cos\theta$ , therefore,

$$\Delta G_{\text{het}} = (4\pi r^2\gamma_{\text{nl}} + 4\pi r^3\Delta G_v/3) * (2-3\cos\theta+\cos^3\theta)/4 \quad (8)$$

Note the first term in parenthesis is the total free energy for homogenous nucleation ( $\Delta G_{\text{hom}}$ ), so the free energy needed for heterogeneous nucleation ( $\Delta G_{\text{het}}$ ) is equal to the product of homogeneous nucleation ( $\Delta G_{\text{hom}}$ ) and a function of the contact angle ( $\theta$ ):

$$\Delta G_{\text{het}} = \Delta G_{\text{hom}} \cdot f(\theta) \quad (9)$$

$$f(\theta) = (2-3\cos\theta+\cos^3\theta)/4 \quad (10)$$

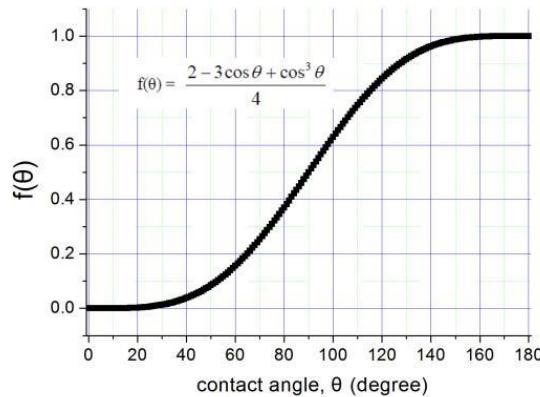


Figure 17. Curve of contact angle function. It can be seen that  $0 \leq f(\theta) \leq 1$ .

$f(\theta)$  describes the effect of wetting angle on the shape of the nucleus. The two limiting cases are easy to understand: For  $\theta = 180^\circ$ ,  $f(\theta) = 1$ ,  $\Delta G_{\text{het}} = \Delta G_{\text{hom}}$ ; no wetting of the surface, falling into the case of homogeneous nucleation; For  $\theta = 0^\circ$ ,  $f(\theta) = 0$ ,  $\Delta G_{\text{het}} = 0$ ; full wetting, no barrier for nucleation at surface. The heterogeneous nucleation can be changed by controlling the substrate wettability.



Figure 18. Contact angle and wettability; a droplet on lotus leaf; a droplet on SiNx membrane before and after oxygen plasma treatment.

Discussion: *hydrophobic, hydrophilic, superhydrophobic, superhydrophilic, and oleophobic, oleophytic, superoleophobic, superoleophytic.*<sup>1,2</sup>

<sup>1</sup> <https://www.nature.com/articles/natrevmats201736>

<sup>2</sup> <http://www.sciencedirect.com/science/article/pii/S1369702115000024>

Using the similar treatment shown in the case of homogeneous nucleation, the critical radius  $r^*$  is yielded at the maximum of  $\Delta G_{\text{het}}$ . The critical radius  $r^*$  and critical free energy  $\Delta G^*_{\text{het}}$  are given by:

$$r^*_{\text{het}} = -2\gamma/\Delta G_v = r^*_{\text{hom}} \quad (11)$$

$$\Delta G^*_{\text{het}} = \Delta G^*_{\text{hom}} \cdot f(\theta) \quad (12)$$

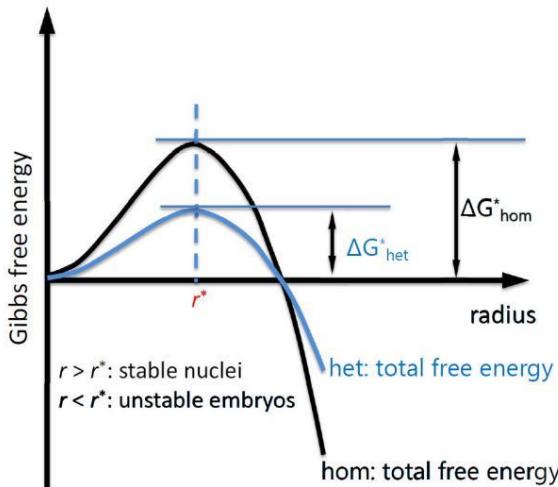


Figure 19. The comparison of free energy curves for heterogeneous nucleation and homogeneous nucleation

Important notes for the heterogeneous nucleation: The critical radius  $r^*$  is unchanged for heterogeneous nucleation and homogeneous nucleation, but being on a substrate the volume of critical nucleus can be significantly less for heterogeneous nucleation than for homogeneous nucleation. The critical free energy can be significantly lower for heterogeneous nucleation due to the wetting effect of the substrate and hence the rates in heterogeneous nucleation can be much higher.

**Question: With a significantly lower energy barrier, who does the Arrhenius dependence look for heterogeneous compared to the homogeneous case? Can you list some examples to show heterogeneous nucleation?**

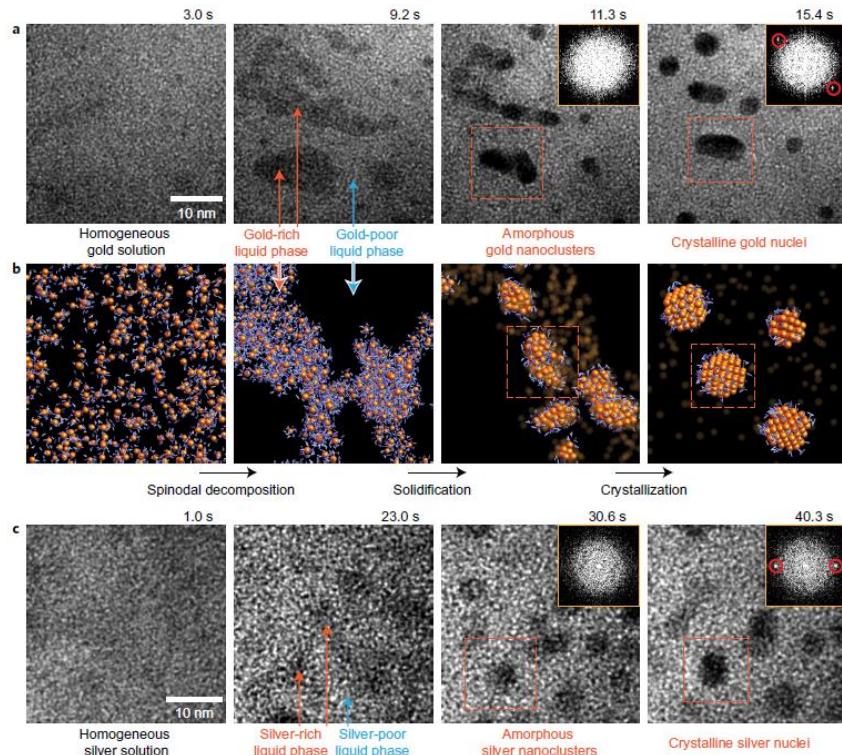
#### Example 1: Multistep nucleation of nanocrystals in aqueous solution revealed by LPEM<sup>1</sup>

The above classical nucleation theory discusses the nucleation event in a single step. This theory can be successfully used to describe a wide range of phenomena, such as the condensation of water droplets from vapor, however, the reaction process is much more complex.. Recently, by employing the LPEM method, Mirsaidov et al. directly observed unprecedented three step nucleation of gold and silver nanocrystals from supersaturated aqueous solutions: first spinodal decomposition (i.e. separation) into solute-rich and solute-

<sup>1</sup> <http://www.nature.com/nchem/journal/v9/n1/full/nchem.2618.html>

poor liquid phases, then nucleation of amorphous nanoclusters within the metal-rich liquid phase, and finally crystallization of these amorphous clusters.

The observed intermediate steps in nuclei formation has important implications not only for a better understanding of nucleation theory but also for optimizing controllable growth of crystalline and amorphous materials.



*Figure 20. Proposed three-step pathway for gold and silver nucleation in solution. a, A series of TEM images that shows the intermediate steps in nucleating gold nanocrystals from a supersaturated aqueous  $Au^0$  solution (Supplementary Movie 1)<sup>1</sup>. From 3.0 to 9.2 s, the supersaturated  $Au^0$  solution spontaneously demixes into gold-poor and gold-rich liquid phases (lighter and darker regions, respectively) via spinodal decomposition. Amorphous gold nanoclusters emerge (11.3 s) from the gold-rich phases that then crystallize (15.4 s). Insets show Fourier transforms of cropped square regions (orange) with the  $Au\{111\}$  fcc reciprocal lattice spacing circled in red. b, Schematic of the proposed steps in nucleation (gold as orange spheres, with surrounding water as blue bent lines). c, A TEM image sequence in which silver nanocrystals nucleate from aqueous  $Ag^0$  with the same intermediate steps as for gold in a. Here a supersaturated  $Ag^0$  aqueous solution demixes into silver-poor and silver-rich liquid phases (1.0–23.0 s), and the latter develops amorphous  $Ag$  nanoclusters (30.6 s) that crystallize (40.3 s). Insets show Fourier transforms of cropped square regions (orange) with the visible  $Ag\{111\}$  lattice spacing circled in red.*

### 3.4 Growth mechanisms: growth of nanoparticles in solution

After nuclei have been established, further growth can take place by different processes<sup>2,3</sup>:

(1) **LaMer mechanism.** LaMer suggested that the process of nucleation and growth can be divided into three portions: (I) A rapid increase in the concentration of free monomers in

<sup>1</sup> <http://www.nature.com/nchem/journal/v9/n1/full/nchem.2618.html#supplementary-information>

<sup>2</sup> <http://pubs.acs.org/doi/abs/10.1021/cr400544s>

<sup>3</sup> <http://onlinelibrary.wiley.com/doi/10.1002/anie.200802248/pdf>

solution, (II) the monomer undergoes “burst nucleation” which significantly reduces the concentration of free monomers in solution. The rate of this nucleation (see eq. 5) is described as “effectively infinite” and after this point, there is almost no nucleation occurring due to the low concentration of monomers after this point; (III) following nucleation, the growth occurs under the control of the diffusion of the monomers through the solution leading to adsorption on the particles’ surface.

(2) **Ostwald ripening and digestive ripening.** Ostwald ripening, which describes the change of an inhomogeneous structure over time, i.e., small crystals or sol particles dissolve, and redeposit onto larger crystals or sol particles, so the larger particles grow while the smaller ones dissolve. This is what you experience in e.g. foams, where the initial foam can be dense of small bubbles but over time has a void-filled structure of large bubbles.

### **Question: How can the Ostwald ripening effect make storing ice cream in the freezer a bad idea?**

This process was first observed by Wilhelm Ostwald in 1896. Due to the high solubility and the surface energy of smaller particles within solution, these redissolve and in turn allow the larger particles to grow even more. The Gibbs–Thomson relation (eq. 13) is responsible for the size change<sup>1,1,2</sup>.

$$C_r = C_b \exp(\Delta\mu/k_B T) \quad (13)$$

here,  $\Delta\mu (= 2\gamma v/r)$  is the chemical potential of a spherical particle with radius  $r$ , where  $v$  is the molar volume of the bulk crystal, and  $\gamma$  is the surface energy;  $C_b$  and  $C_r$  are the monomer concentrations in the bulk solution and on the particle surface.

The mathematical theory of Ostwald ripening within a closed system is described by Lifshitz and Slyozov and Wagner (LSW theory)<sup>1,3</sup> that has two main regimes:

**i. Diffusion Limited Growth:** when the diffusion is the slowest step in the growth process, the particle growth is essentially controlled by the diffusion of the monomers to the surface, and if the total mass of the system is explicitly kept conserved, the particle growth rate is reduced to the form:  $dr/dt = r^2/K$ , that can be easily solved to obtain the dependence of the particle size on time. It is given by,  $r^3 - r_0^3 = Kt$ , where  $r_0$  is the average radius of the particle at time  $t=0$ ,  $K$  is a constant. The volume grows linearly with time as monomers constantly flow in.

**ii. Reaction-Limited Growth:** If diffusion is not the rate controlling step in the growth process, then the growth rate is limited by the surface reaction of the attaching monomers, and if the conservation of mass is valid, the dependence of particle size on the growth time can be expressed as  $r^2 = K_R t$ ,  $K_R$  is a constant here. The radial growth is as the square root of time.

---

<sup>1</sup> <http://www.sciencedirect.com.proxy.findit.dtu.dk/science/article/pii/S1359645417301866>

<sup>2</sup> <http://www.sciencedirect.com/science/article/pii/S1359646204007195>

<sup>3</sup> <http://onlinelibrary.wiley.com/doi/10.1002/pssa.2211330119/pdf>

## Example 2: Observation of Single Colloidal Platinum Nanocrystal Growth Trajectories<sup>1</sup>

Using the LPEM method, Zheng et al studied the growth behavior of single colloidal platinum nanocrystals. They showed that platinum nanocrystals can grow either by monomer attachment from solution or by particle coalescence. Through the combination of these two processes, an initially broad size distribution can spontaneously narrow into a nearly monodisperse distribution. Detailed growth kinetic parameters of platinum nanoparticles were also derived from the in-situ observations. The results also showed that the growth process is a diffusion-controlled Ostwald ripening process. The direct observations of single-particle growth trajectories provide important insights into nanocrystal growth mechanisms, which are not accessible with a conventional analysis that is on the basis of an ensemble of particles.

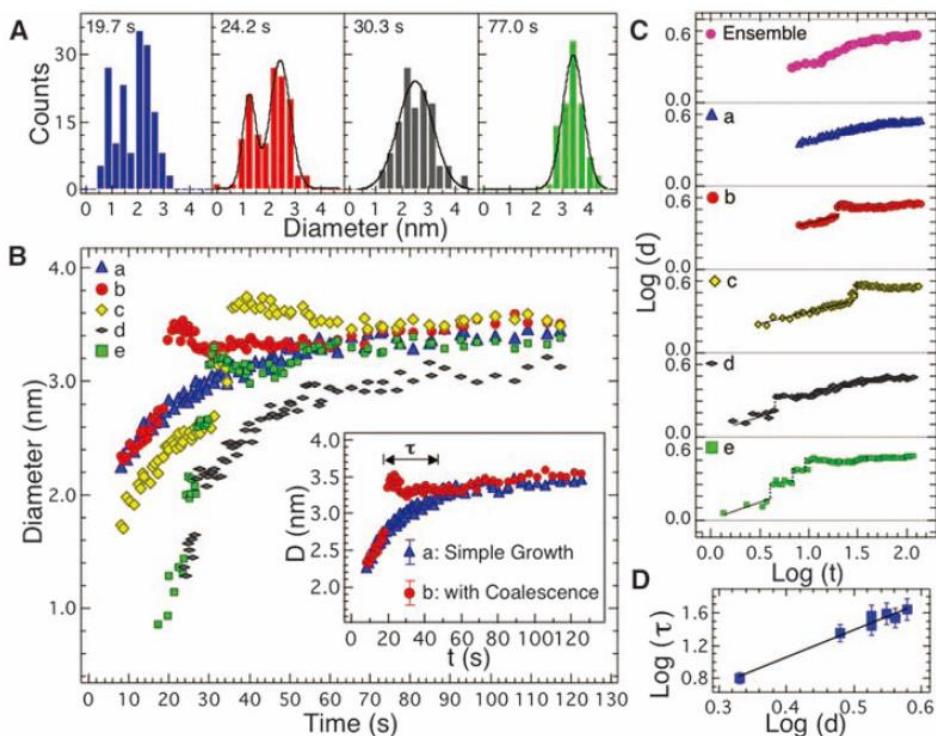


Figure 21. Growth kinetics of Pt nanoparticles.<sup>2</sup> (A) Histograms of particle size distribution at 19.7 s, 24.2 s, 30.3 s, and 77.0 s. Black curves are Gaussian fits. (B) Particle size versus growth time. These particles are highlighted in Fig. 2A. Inset shows two types of growth trajectories. A relaxation time ( $\tau$ ) was observed after a coalescence event. Error bars for particle diameter measurements are less than  $T0.18\text{ nm}$ . (C) Logarithmic relationship of particle size versus growth time for the ensemble and those individual particles in (B). Black lines are guides for the eye, and dashed lines show the coalescence events. (D) Logarithmic relationship of relaxation time versus the size of the coalesced particles. Black line shows linear fit.

Digestive ripening is effectively the inverse of Ostwald ripening. Within this case, smaller particles grow at expense of the larger ones. This process of formation is controlled once again

<sup>1</sup> <http://science.sciencemag.org/content/324/5932/1309>

<sup>2</sup> <http://science.sciencemag.org/content/324/5932/1309/tab-figures-data>

by the surface energy of the particle within solution where the larger particle redissolve and in turn smaller particles grow.<sup>1, 2</sup>

(3) **Finke-Watzky two step mechanism.** This mechanism supposes that the nucleation and growth steps can not be separated completely. Firstly a slow continuous nucleation occurs ( $A \rightarrow B$  in the image below), the newly formed nuclei will further catalyze the surface growth which is not diffusion controlled ( $A + B \rightarrow 2B$ ). Although, this method is different from classical nucleation, the nucleation step still follows a critical size described within a classical nucleation framework.

(4) **Coalescence and oriented attachment.** The two mechanisms are similar, but only differ in the orientation of the crystal lattice at the grain boundary. Within coalescence the lattice planes are randomly oriented between domains, while for oriented attachment there is a perfect alignment of the planes.<sup>3</sup> It worth noting that during the oriented attachment, Coulombic and van der Waals interactions with anisotropic polarizability can not be ruled out completely.

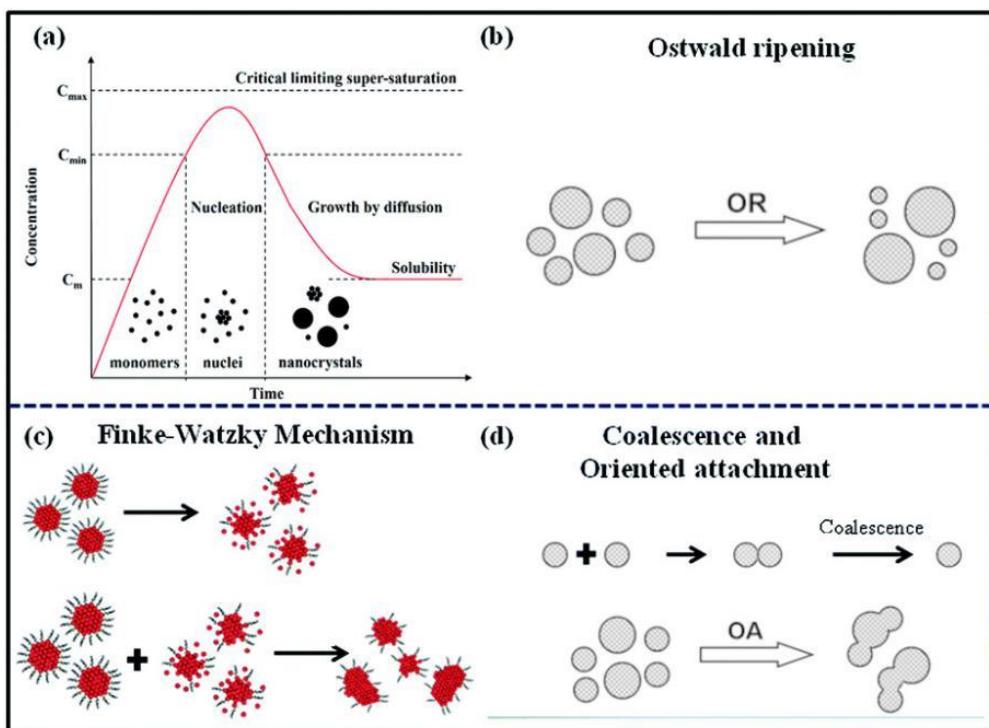


Figure 22. Different possible homogenous growth mechanisms. (a) is adapted from Chem. Rev., 2014, 114, 7610; (c) is adapted from Faraday Discuss., 2015, 181, 85

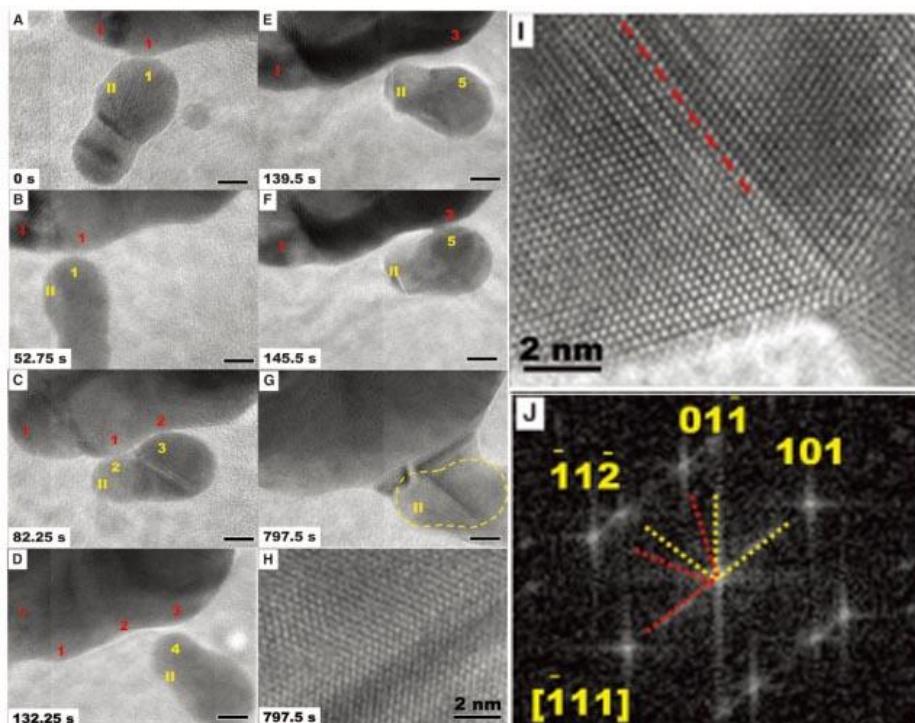
<sup>1</sup> <http://www.sciencedirect.com/science/article/pii/S1359645407003825>

<sup>2</sup> <https://link.springer.com/content/pdf/10.1134%2FS0036024417080131.pdf>

<sup>3</sup> <http://science.sciencemag.org/content/289/5480/751>

### Example 3: Direction-Specific Interactions Control Crystal Growth by Oriented Attachment<sup>1</sup>

Although the oriented attachment is recognized as an important mechanism of crystal growth from solution, the alignment process and attachment mechanism have not been fully understood yet. The development of LPEM now makes it possible. Li et al performed TEM using a fluid cell to directly observe oriented attachment of iron oxyhydroxide nanoparticles. Key evidence of oriented attachment, including the continuous rotation and interaction of particles before achieving a perfect lattice match, particle jumping to contact each other, has been confirmed with this in-situ technique.



*Figure 23. (A to G) Sequence of images from movie S2 showing typical dynamics of the attachment process (see fig. S4 for high-resolution images). The surfaces of particles I and II made transient contact at many points and orientations (points 1-1, 1-2, 2-3, and 3-4) before finally attaching and growing together (points 3-5). (H) High-resolution image of interface in (G) showing twin structure (an inclined twin plane). The yellow dashed line in (G) shows the original boundary of the attached particle. (I and J) High-resolution in situ TEM image (I) and fast Fourier transform (FFT) (J) of an interface from another OA event demonstrating formation of a (101) twin interface after attachment. The grain boundary is delineated by a dashed line in (I). Scale bars are 5 nm for (A) to (G).<sup>2</sup>*

### 3.5 Growth mechanisms: growth on a substrate

(1) **Frank-Van der Merwe growth.** This model is also known as “layer-by-layer” growth. A layer of atoms is started and completed before another layer starts to grow. The model assumes that attachment of atoms to the edge of a growing layer is far more likely than attachment to the top of the layer, i.e., new nucleation, and hence the latter rarely occurs. The simplest energetic interpretation is that the atom attaching to the growth layer edge makes two or more bonds and hence reduces the interfacial energy by much more than one connecting to

<sup>1</sup> <http://science.sciencemag.org/content/336/6084/1014>

<sup>2</sup> <http://science.sciencemag.org/content/suppl/2012/05/23/336.6084.1014.DC1>

the top surface which makes only one. In a kinetic interpretation the atom attached to the top of the surface will diffuse quickly until it encounters a new layer edge, and, now having two bonds, thereafter will have a much lower probability of moving back to the surface. Similarly, an atom bonded to the edge of the surface layer may diffuse along the edge until it can find additional bonding opportunities at “kink” sites.

(2) **Volmer–Weber growth or island growth.** In this model, nucleation of islands is favored over extended growth at layer edges. This can occur if the new precipitating species have a tendency to cluster, i.e., the new atoms bond more strongly with one another than to the surface, or if diffusive processes slow layer growth until new surface nucleation far exceeds it (resulting in increased nuclei density). The net effect is that layers are filled in as islands are nucleated and merge.

(3) **Stranski–Krastanov growth.** This model is a mixed alternative, also known as layer-plus-island growth, where growth is initiated as a layer-by-layer mode but shifts to an island mode.

The above three classical heterogenous growth models can also be understood from the point of interface/surface energy. For Frank van der Merwe growth, the interface energy is relatively low and layers form readily from nearly any size nuclei; for Volmer–Weber growth, the interface energy is relatively large. Hence discrete nuclei form and grow before a layer is completed by infilling. For Stranski–Krastanov growth, the interface energy is comparable to island interaction energy, and hence layer formation competes with nuclei formation.

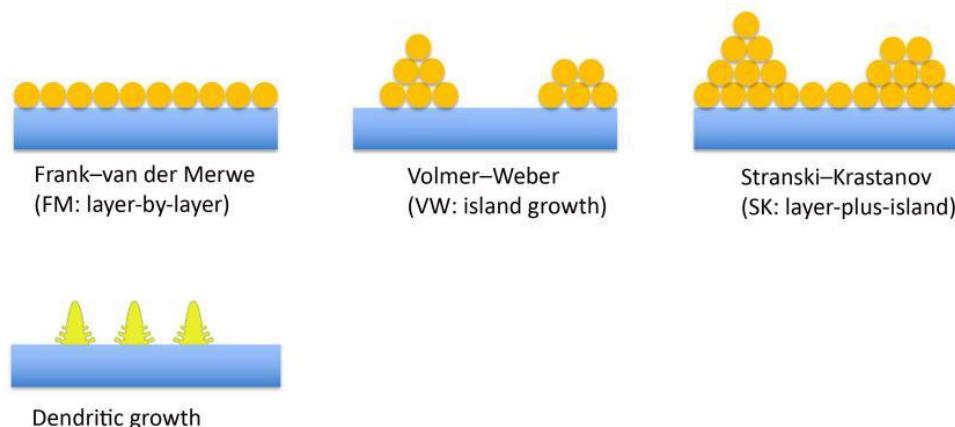


Figure 24 Different possible heterogenous growth mechanisms.

**Dendritic growth.** When the mass transportation limits the rate at which a crystal can grow, it will often greatly affect the shape of the growing crystals. Besides the three typical heterogenous growth models, dendritic growth via the balance of liquid and surface diffusion is also possible. Generally, the fluctuation or instability of crystal growth can be described by Mullins–Sekerka theory.<sup>1,2,3,4</sup>

<sup>1</sup> <http://aip.scitation.org/doi/abs/10.1063/1.1713333>

<sup>2</sup> <https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.52.1>

<sup>3</sup> <http://jes.ecsdl.org/content/138/11/3305.abstract>

<sup>4</sup> <https://mogadalai.wordpress.com/2008/10/09/morphological-instabilities-during-growth-linear-stability-analyses/>

Metal dendrites are harmful for the development of high energy batteries, such as lithium-metal, sodium-metal batteries. Firstly, the formation and growth of dendrites during electrochemical reaction can penetrate through the separator and cause cell short circuit, imposing on the wide-ranging safety concerns. Secondly, the dendrites could also lead to an electrical detachment of Li or Na from the current collector and become “dead Li or Na”, significantly shortening the cycle life of the batteries. Thirdly, due to the intrinsic high-reactivity of metal dendrites in the organic electrolyte, corrosive reactions occur on the dendrites/electrolyte interface, leading to an instable interface and continuous depletion of the electrolyte. Therefore, characterizing the microstructures of metal dendrites and chemical composition of solid electrolyte interface<sup>1</sup>, observing the detail formation process of dendrites<sup>2,3,4</sup>, and developing suitable approaches to prevent the dendrites generation<sup>5</sup> are crucial for the battery performance improvement.

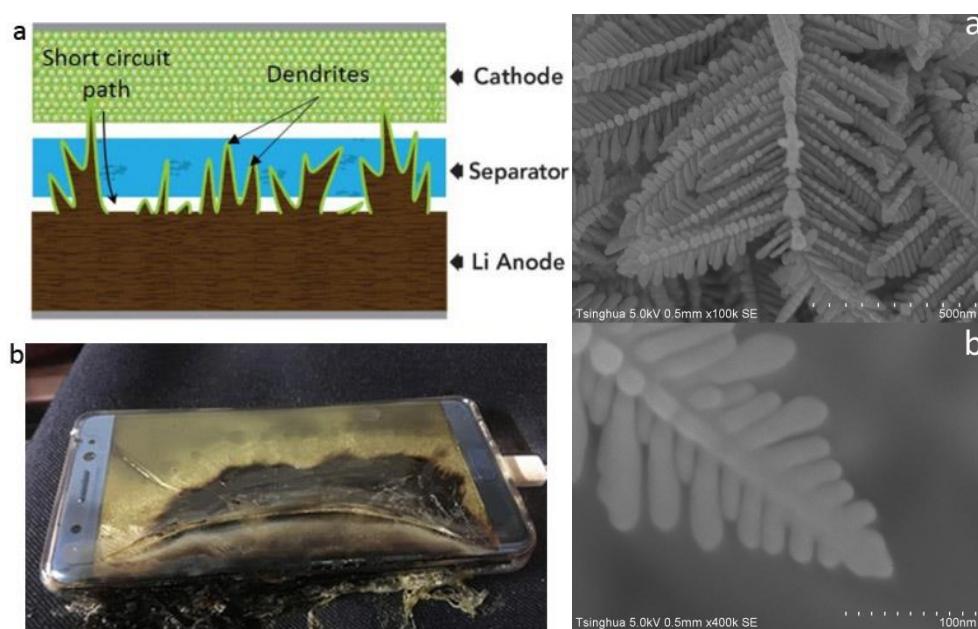


Figure 22 (left) Dendrites are the enemy of lithium-ion batteries<sup>6</sup>; (right) typical morphology of Ag dendrites.

<sup>1</sup> <http://science.sciencemag.org/content/358/6362/506>

<sup>2</sup> <http://pubs.acs.org/doi/pdfplus/10.1021/acs.nanolett.5b02479>

<sup>3</sup> <https://www.youtube.com/watch?v=gljoBkGIKmk>

<sup>4</sup> [https://www.youtube.com/watch?v=tDuaQ4Am\\_c](https://www.youtube.com/watch?v=tDuaQ4Am_c)

<sup>5</sup> <http://pubs.acs.org/doi/pdfplus/10.1021/acs.chemrev.7b00115>

<sup>6</sup> <https://deepresource.wordpress.com/2017/08/28/efficiency-breakthrough-lithium-ion-batteries/>