AFM study of the interaction between vaccinia topoisomerase IB and DNA

L.H.F.M. Holtzer

Master Thesis

April 2005

Delft University of Technology
Kavli Institute of Nanoscience
Section of Molecular Biophysics
Supervision: Dr. N. Dekker, Prof. Dr. C. Dekker
Daily guidance: Dr. F. Moreno-Herrero

TU Delft

Abstract

Type IB topoisomerases (TopIB) change the degree of DNA supercoiling by inducing a single-strand break and subsequently religating it, which causes supercoils to be removed. Topoisomerase from vaccinia virus is the smallest type IB topoisomerase known (36 kDa) and acts as a model system, also because of its sequence specificity. We have used Atomic Force Microscopy to study the interaction between vaccinia TopIB and DNA and we obtained static and dynamic images of experiments. Vaccinia TopIB binds non-covalently to DNA in a sequence-independent manner. We found that TopIB forms filaments on DNA after the formation of an intramolecular node on a DNA molecule. The formation of filaments is a highly cooperative process (ω=5181$\pm$621). We propose a model that describes this process in terms of cooperativity theory. In the presence of divalent (cat)ions and at high TopIB concentrations, DNA is aggregated by TopIB. This unexpected aggregation can be explained as an extension of the proposed model. The influence of monovalent ions was much less dramatic. Control experiments showed clean DNA molecules without the enzyme and that TopIB is a monomer in solution.

The dynamics of the TopIB-DNA interaction was explored by imaging in liquid environment. Imaging speeds up to 9 $^{s}/_{\text{frame}}$ were obtained. We found that TopIB is very loosely bound to DNA and that the off-rate is influenced by tip-sample interaction. The movement of several TopIB molecules bound to DNA was analyzed and diffusion-like behaviour was observed. The obtained diffusion constant for bound TopIB was smaller compared to the diffusion constant of a TopIB diffusing on a mica surface while not bound to DNA.

Contents

Abstract i

Contents iii

List of Symbols v

1 Introduction 1
- 1.1 Topoisomerases . . . 1
- 1.2 Atomic Force Microscope . . . 4
- 1.3 Goal of the project . . . 7

2 Theory 9
- 2.1 Cooperative binding . . . 9
- 2.2 Diffusion of molecules . . . 15

3 Materials and Methods 17
- 3.1 DNA preparation . . . 17
- 3.2 Vaccinia Topoisomerase IB . . . 18
- 3.3 Sample preparation . . . 19
- 3.4 Atomic Force Microscope . . . 20
- 3.5 Data analysis . . . 22

4 TopIB aggregates DNA in a cooperative fashion 27
- 4.1 Binding of TopIB to DNA shows cooperative behaviour . . . 27
- 4.2 TopIB aggregates DNA in the presence of divalent ions . . . 30
- 4.3 Control experiments . . . 34

5 Dynamics of TopIB-DNA interaction 39
- 5.1 Characterization of TopIB-DNA interaction . . . 39
- 5.2 Lifetime of TopIB bound to DNA . . . 41
- 5.3 1D-diffusion of TopIB on DNA . . . 42

6 Discussion 47
- 6.1 Cooperativity . . . 47
- 6.2 Aggregation of DNA by TopIB . . . 51
- 6.3 Experimental methods . . . 54
- 6.4 Concluding remarks . . . 56

7 Conclusions **59**

Bibliography **61**

List of Symbols

Symbol	Meaning	Units
b_1	Probability of picking the left end of a bound protein	[-]
b_n	Probability of picking the right end of a bound protein	[-]
B	Average total number of bound proteins per DNA	[-]
B_{dc}	Average total number of d.c. bound proteins per DNA	[-]
B_{isol}	Average total number of isolated bound proteins per DNA	[-]
B_{sc}	Average total number of s.c. bound proteins per DNA	[-]
c	Filament length	[-]
$\bar{C}$	Average filament length	[-]
D	Diffusion constant	$[\mathrm{m^2 s^{-1}}]$
e	Elementary charge	[C]
f	Probability of picking a free base pair	[-]
f_r	Resonance frequency	$[\mathrm{s^{-1}}]$
F	Force	[N]
F_c	Probability distribution function of bound proteins in filaments	[-]
F_L	Lateral force	[N]
F_N	Normal force	[N]
g	Gap size between two bound proteins	[bp]
k	Cantilever spring constant	[N/m]
k	Boltzmann constant	[J/K]
k_{off}	Off-rate	$[\mathrm{s^{-1}}]$
K, K_{DNA}, K_{dimer}	Intrinsic association constants	$[\mathrm{M^{-1}}]$
l_{fil}	Filament length	[m]
l_i	Location of protein on DNA at time i	[m]
L	Free protein concentration	[M]
L_T	Total protein concentration	[M]
n	Part of DNA covered by one protein	[bp]
N	Concentration of base pairs	[M]
N	Number of frames a protein is bound	[-]
$P(\cdot\|\cdot)$	Conditional probability	[-]
P_c	Normalized probability distribution of filament lengths	[-]
P'_c	Probability distribution of filament lengths	[-]
P_g	Probability distribution of gap sizes	[-]

Symbol	Meaning	Units
r_i	Location of protein on surface at time i	[m]
R_c	Radius of the AFM tip	[m]
R_m	Radius of a molecule	[m]
$\bar{s}$	Average number of free binding sites per gap	[-]
$\bar{s}_{dc}$	Average number of free d.c. binding sites per gap	[-]
$\bar{s}_{isol}$	Average number of free isolated binding sites per gap	[-]
$\bar{s}_{sc}$	Average number of free s.c. binding sites per gap	[-]
t	Time	[s]
T	Temperature	[K]
V	Potential energy	[J]
W	Apparent molecule width	[m]
W_{conv}	Tip-sample convolution	[m]
W_{DNA}	Width of DNA molecule	[m]
z	Distance	[m]
z_i	Charge of ion i	[C]
ϵ	Dielectric constant	[F/m]
ϵ_0	Permittivity of free space	[F/m]
λ_D	Debye length	[nm]
ν	Binding density of proteins	[-]
ρ_i^∞	Concentration of ions in the bulk	[M]
ω	Cooperativity parameter	[-]

1. Introduction

1.1 Topoisomerases

1.1.1 Function and types

Topoisomerases are enzymes that remove or introduce supercoils in DNA. They do this by cleaving, manipulating and religating DNA strands. Supercoils are formed in DNA when twist is added to or subtracted from the DNA double helix. Processes in cells (for example DNA transcription or DNA replication) can generate supercoiling of DNA. Figure 1.1 shows the difference between a relaxed DNA molecule and a molecule that is partly positively and partly negatively supercoiled. These supercoils present a problem for processes like DNA replication (the two strands need to be separated) and the DNA can even be damaged if supercoils are not released. Because transcription and replication is very important for the survival of the cell, topoisomerases are also very important. This is the reason why topoisomerases are an interesting target for anti-cancer drugs. Inhibiting topoisomerase activity in cancer cells will induce cell apoptosis, which constitutes a potential cancer therapy.

There are two families of topoisomerases [1]. Type I topoisomerases cleave one of the two strands of a dsDNA[1] molecule and allow for supercoil release by single-strand passage or rotation. Type II topoisomerases create a double-stranded gap through which another duplex can pass, either from the same DNA molecule or from another one [1]. Both subfamilies can be divided in an A-type and a B-type subfamily. Type IA topoisomerases

[1] dsDNA = double stranded DNA

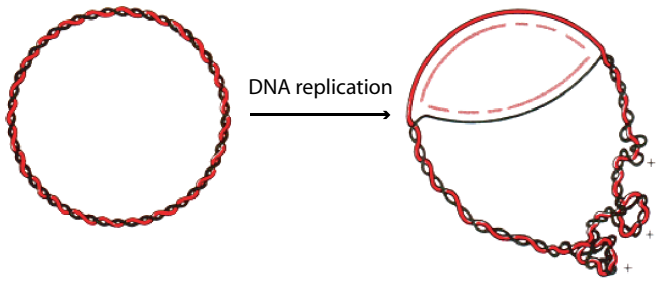

Figure 1.1: Positive and negative supercoils are induced during DNA replication. These supercoils need to be removed by topoisomerases to allow the replication fork to proceed.

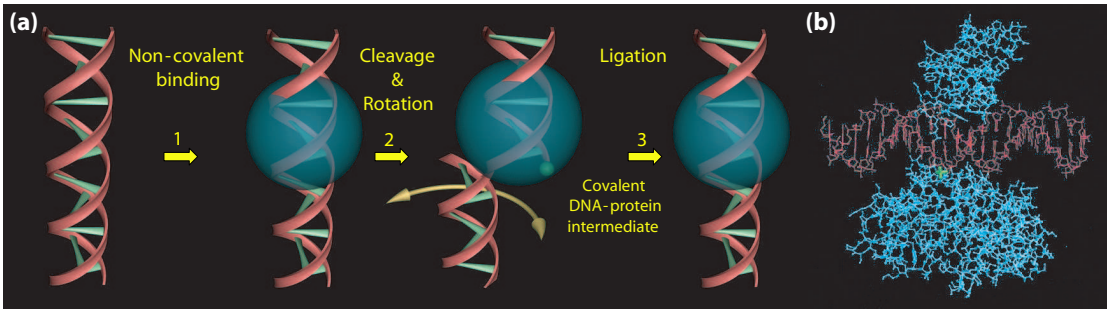

Figure 1.2: **a)** Scheme showing the different steps in supercoil removal by type IB topoisomerases. **b)** Suggested model of a type IB topoisomerase non-covalently bound to a DNA molecule (red=DNA molecule, blue=TopIB) [5]. The protein forms a C-shaped clamp around the DNA as was predicted earlier [6]. The TopIB amino acids that catalyze cleavage can interact directly with the DNA phosphate that will be cleaved.

bind covalently to a 5' phosphate while type IB topoisomerases covalently bind to a 3' phosphate. In practice type IA topoisomerases remove only negative supercoils and they ligate DNA after the DNA has rotated once, thereby changing the linking number by one [2]. They also need Mg^{2+} for DNA relaxation activity. Type IB topoisomerases can both relax positive and negative supercoils and the rotation after cleavage is not limited to one turn. Mg^{2+} or ATP is not required by type IB topoisomerases. The type II topoisomerases all belonged to the same subfamily until in 1997 the first type IIB topoisomerase was identified which has almost no similarities to type IIA topoisomerases [3]. Nowadays it is known that both type IIA and type IIB topoisomerases are widely found [4]. Our research will focus on the type IB topoisomerases.

1.1.2 Type IB topoisomerases

There are three classes of type IB topoisomerases (TopIB): the eukaryotic topoisomerases I, the poxvirus topoisomerases and the prokaryotic topoisomerase from *Methanopyrus kandleri* [1]. Type IB topoisomerases remove either positive or negative supercoils. This is done in several steps (Fig. 1.2a). First, TopIB binds non-covalently to DNA (1), then it cleaves one strand of the dsDNA and forms a covalent DNA-(3'-phosphotyrosyl)-protein intermediate. This allows the DNA to rotate around the created nick (2). Finally the DNA is religated (3) and TopIB will dissociate from the DNA [7]. The cellular eukaryotic topoisomerases have sizes from 765 to 1019 amino acids [8]. Type IB topoisomerase from vaccinia virus was used in this research and consists of 314 amino acids (36 kDa). TopIB from vaccinia virus is the smallest TopIB known and consists only of the minimal functional unit. Therefore they can be treated as a model system for all type IB topoisomerases. Vaccinia TopIB consists of two structural domains that have been crystallized individually: a N-terminal domain and a C-terminal catalytic domain, connected by a hinge [9, 5]. The crystal structure of TopIB non-covalently bound to DNA has not been determined yet but a model was suggested (Fig. 1.2b) [5].

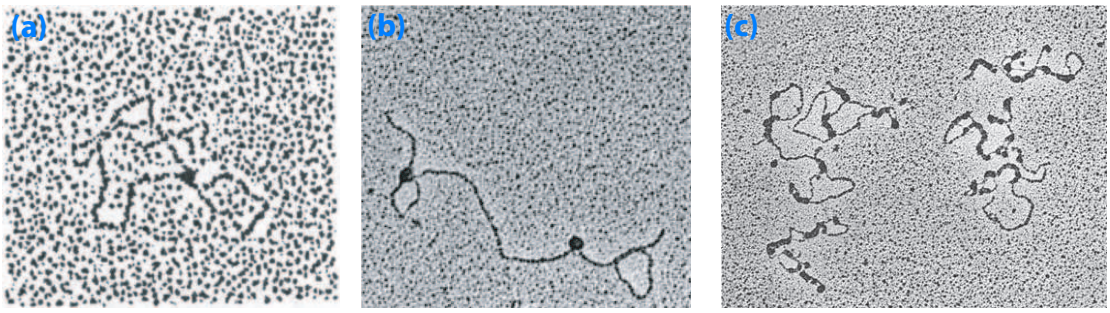

Figure 1.3: EM images of intramolecular nodes on DNA formed by calf thymus TopIB [12] **(a)** and vaccinia TopIB [7] **(b)**. **c)** Vaccinia TopIB zips up the DNA molecule by forming filament-like structures [7].

Another advantage of the use of vaccinia TopIB for studying type IB topoisomerases is that it prefers to cleave DNA at a specific 5'(C/T)CCTT↓ pentamer site in dsDNA (the ↓ arrow indicates the location of the nicking site). In other type IB topoisomerases this specificity was not found. Recently it has been shown that vaccinia TopIB can also cleave DNA when the specific sequence is not present [10]. The specific pentamer is also preferred for the non-covalent binding of TopIB to DNA although it can also bind to DNA without the pentamer with a seven to tenfold lower affinity [11]. This suggests that TopIB can bind to DNA non-specifically (for example through interaction with the DNA backbone) and diffuse linearly along the DNA to locate the preferred pentamer. Linear diffusion would certainly increase the TopIB efficiency of finding the cleavage site because the fraction of 5'(C/T)CCTT↓ pentamers in a natural DNA molecule is small (0.2% assuming that each base has an equal probability to be found in DNA). Combined with a hopping mechanism it would be even more efficient. When vaccinia TopIB is non-covalently bound to DNA, it covers 25 bp as shown with DNase I footprinting [11].

Besides changing the number of supercoils in DNA molecules, other properties and functions have been observed for type IB topoisomerases. Zechiedrich and Osheroff showed using electron microscopy (EM) that calf thymus topoisomerase I can form intramolecular nodes in dsDNA molecules (Fig. 1.3a) [12]. Later it was shown that vaccinia topoisomerase IB has similar capabilities (Fig. 1.3b) [7]. Because vaccinia topoisomerase is a monomer in solution [13, 14] these nodes are formed by two TopIB proteins that are already bound to a dsDNA molecule, forming a TopIB dimer on DNA. When an intramolecular node is formed on DNA, TopIB can zip up the DNA to form filament-like structures, which was shown with EM (Fig. 1.3c) [7]. In 1996 it was found that eukaryotic topoisomerases catalyze the resolution of four-way Holliday junctions [15]. Several years later Sekiguchi et al. observed the same behaviour for vaccinia topoisomerase [16].

While the role of the enzyme *in vivo* is still unknown, it was suggested that the protein helps in packaging the DNA genome of the virus which has a length of 192 kbp

[8]. It was also suggested that vaccinia topoisomerase plays a role in viral recombination [8]. This is supported by findings that type IB eukaryotic topoisomerases have structural similarities to site-specific recombinases [5]. Both type IB topoisomerases and tyrosine recombinases can catalyze similar reactions. Furthermore RNase activity was detected for both enzymes [17]. The origin and function of this activity is still unclear. Shuman suggested that this RNase activity actually has a functional significance. TopIB could for example play a role in RNA recombination [18]. On the other hand, this RNase activity could just be a result of evolution. An hypothesis is that an early protein which catalyzed certain transfer reactions might have existed in an *RNA world*. After transition to a world with interplay between RNA and DNA it might have used its functions in transfer reactions in RNA-DNA hybrids. Later it might have evolved in such a way that it could be able to process DNA in the absence of RNA [19]. The question of how the vaccinia virus has incorporated the topoisomerase still remains unanswered. Likely it was obtained from a eukaryotic source and consequently through time and evolution the virus might have removed portions of the enzyme which were not needed by the virus [20].

1.2 Atomic Force Microscope

The research presented here was carried out using the Atomic Force Microscope (AFM) as the main technique. Therefore a brief description of the AFM is given below.

The Atomic Force Microscope (or Scanning Force Microscope (SFM)) belongs to the Scanning Probe Microscopes (SPM) family. Apart from the AFM, other members of this family are the STM (Scanning Tunneling Microscope) and the SNOM (Scanning Near-Field Optical Microscope). These microscopes operate following the same principle. A very sharp probe scans a surface while keeping the interaction between the tip and the surface constant. This interaction depends strongly on the distance between the tip and the surface. When the interaction changes during the scanning of a sample this change can be reversed by adjusting the height of the sample using a piezo tube. Combining all these adjustments results in a height profile of the sample. The STM uses the tunneling current between tip and sample as the control parameter. The SNOM uses the evanescent waves coming from the surface of a sample as the control parameter. For the AFM it is the force between tip and sample that matters. The AFM generates a topographic image of a surface in a similar way as a blind person can read Braille characters by probing a sheet of paper with his/her fingers.

The AFM was invented in 1986 by Binnig and Quate [21]. It consists of three parts: the software, the electronic and the mechanical part. Figure 1.4 shows a schematic overview of the three parts and the connections between them. The sample is located in the mechanical part and is placed on a piezo tube as shown. A cantilever with a very sharp tip is located above the sample at a certain distance. The cantilever undergoes a certain degree of bending because of the interaction between tip and sample. This bending is measured using a laser beam that reflects off the surface of the cantilever. When

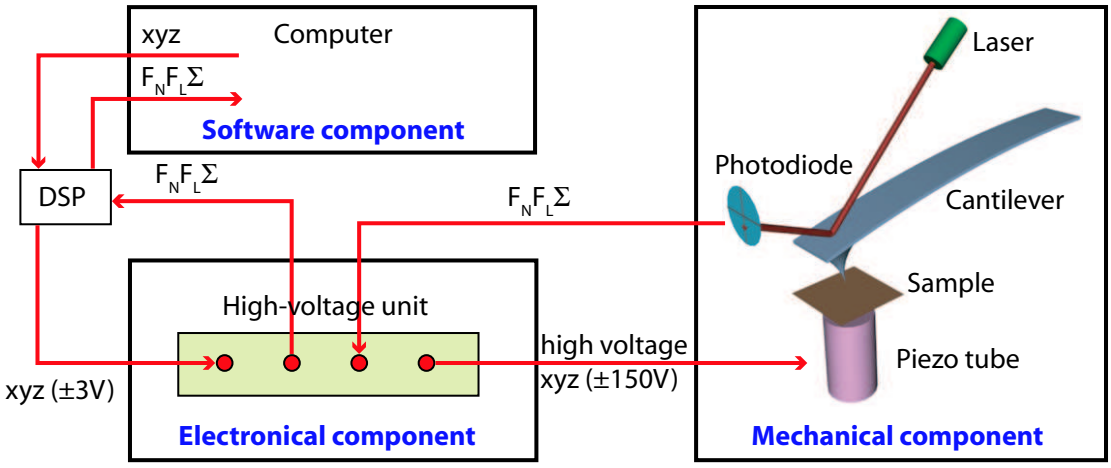

Figure 1.4: Schematic overview of the three main parts of the AFM. The software component consists of a computer which functions as the user interface. The electronical component consists of a unit that can apply high voltages to the piezo tube. The mechanical component consists of the piezo, the sample and the cantilever. Red arrows indicate which parameters are transferred between the parts.

the interaction between tip and sample changes, the deflection changes accordingly. The reflected laser beam is captured by a photodiode which can accurately measure the position of the laser spot. A horizontal displacement of the laser spot corresponds with a change in lateral force (F_L), a vertical displacement corresponds with a change in the normal force (F_N). The intensity of the laser spot on the photodiode is indicated by Σ. These analog signals (F_L, F_N and Σ) are digitalized in the Digital Signal Processor (DSP) for further processing by the software. The main functions of the software are the following: first, it serves as a user interface to the AFM; secondly, it generates scans; and finally it controls the feedback mechanism that is necessary to keep the interaction between tip and sample constant. This is done by changing the z-position of the sample. The necessary z-movement of the sample is recorded by the software and converted into an image of the surface. The movement of the sample in the z-direction, but also in the x- and y-direction (which is necessary for scanning the surface), is controlled by the software and applied by the electronical part which applies high voltages to the piezo tube. In order to move the sample several microns $\sim 100\,\mathrm{V}$ must be generated by the electronical part with a low noise level ($\sim 10\,\mathrm{mV}$).

The behaviour of the cantilever close to a surface depends on several interaction forces. The potential (energy) of the interaction between AFM tip and the surface is shown for air (Fig. 1.5a) and liquid (Fig. 1.5b). The force between the tip and sample is determined by $F = -\frac{\partial V}{\partial z}$ and is negative (attractive) for long distances and positive (repulsive) for short distances [22]. Long range interactions (micrometer scale) are caused by magnetic forces (in this research magnetic forces are not present) and electrostatic forces. Attractive Van der Waals interactions are present and cause medium

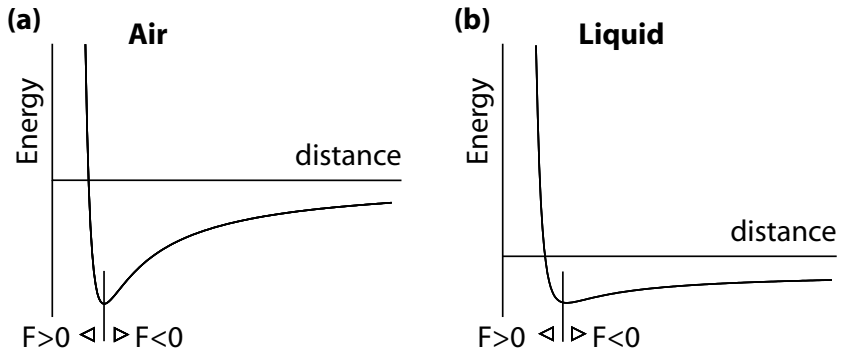

Figure 1.5: Qualitative graphs showing the potential of the interaction between the AFM tip and a surface depending on the distance between the two. This is shown for air **(a)** and liquid **(b)**. A minimum value exists for the potential resulting in an area with negative force (attraction) and positive force (repulsion).

range interactions (tens of nanometers scale) resulting in negative forces between tip and sample. Also the tip can chemically bind (depending on the surface) to the surface which also results in an attractive force. On the short range scale (nanometer scale) the interaction is dominated by contact forces. The electron clouds will repel each other and there will be also ionic repulsion. In this region a clear difference between air and liquid exists. In air the surface is covered with a very thin layer of water and this will result in (short range) capillary forces (attractive) which explains the deep well in the potential diagram (Fig. 1.5a). In a liquid environment these capillary forces are not present (Fig. 1.5b).

The AFM can operate in several modes of which *contact* mode and *dynamic* mode (often called tapping mode) are the most widely used. In contact mode the AFM scans the surface by keeping the force (and consequently the height) between tip and surface constant. To obtain images with high resolution the tip needs to be very close to the sample. Because the tip constantly exerts a force on the sample this can result in damage of the sample especially when biomolecules are imaged. In addition the molecules may be moved by the lateral forces that exist in contact mode. Operating the AFM in dynamic mode reduces these problems. In this mode the cantilever is oscillating close to or at its resonance frequency and the feedback system tries to keep the amplitude of the oscillation constant. During each period of the oscillation the tip is only close to the sample for a very short time. Therefore only a small force is applied on the sample and not much damage is done. The force that is applied on the sample also depends on the used cantilever. A wide range of different cantilevers exist with different spring constants k (ranging from $0.03\,^{\mathrm{N}}/_{\mathrm{m}}$ to $100\,^{\mathrm{N}}/_{\mathrm{m}}$). For experiments with biological molecules usually soft cantilevers are used which have low spring constants and consequently apply less force on the sample compared to cantilevers with high k.

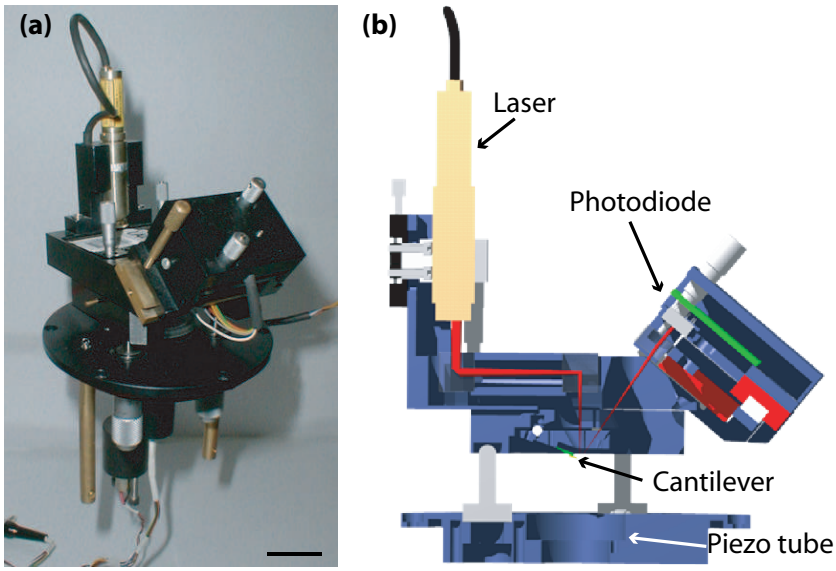

Figure 1.6: a) Picture of the AFM head. Scale bar size is 2 cm. **b)** Cross section of the AFM head showing the path of the laser beam and the location of the cantilever and the photodiode.

The AFM head used in this experiment is shown in figure 1.6a. A cross section of the AFM head with the optical system visible is shown in figure 1.6b. The beam from the laser is reflected by several mirrors. These mirrors can be adjusted to position the laser beam on the cantilever. The cantilever is mounted directly above the piezo tube on which the sample is placed. The reflected laser beam reaches the photodiode through a lens which can be adjusted to center the laser beam on the photo diode. To reduce noise in the image the setup is placed in a isolated box. This prevents sound from inducing vibrations in the cantilever. The AFM is placed onto a steel plate which hangs inside the box and is connected by a set of elastic cords. This reduces mechanical vibrations.

1.3 Goal of the project

In this project several questions were addressed, mainly focused on the binding of TopIB to DNA and the associated dynamics. Reproduction of filament-like structures formed by TopIB and DNA and the quantitative analysis of this binding was one of the goals. An important issue was the cooperativity of this process. It was already shown that TopIB can form filament-like structures on DNA [7]. This was done however with EM and it is known that the EM sample preparation can introduce artifacts [23]. AFM does not have these drawbacks and is therefore a better tool to investigate the cooperative behaviour of the enzyme. Also the effect of ions present in solution was investigated and a model was derived to explain the observed behaviour. In addition this research was helpful in finding the right conditions for measurements on vaccinia TopIB in a magnetic tweezers setup.

1. Introduction

AFM operated in a liquid environment is a useful tool to study the dynamics of biological processes. By binding DNA molecules loosely onto the surface TopIB molecules can interact with the DNA and this can be visualized with the AFM. The dynamics of the binding of TopIB to DNA and the formation filaments were investigated. By following TopIB bound to DNA the linear diffusion of TopIB along DNA was investigated.

2. Theory

2.1 Cooperative binding

In our experiments the cooperativity of the binding of TopIB to DNA was investigated. The theory used in the analysis is described in this section.

Proteins that interact non-specifically with DNA (having no preference for a certain base sequence) can bind in two ways, in a non-cooperative or in a cooperative manner. A theoretical treatment of these types of protein-DNA interaction was described by McGhee & Von Hippel [24]. A brief description is given below.

DNA is treated as an infinitely long 1D lattice with N identical units (base pairs for dsDNA or bases for ssDNA[1]). A protein that binds covers n base pairs. There are three different types of binding events (figure 2.1 shows this for a protein which covers 3 base pairs; base pairs are indicated by black triangles).

- A protein can bind to an isolated binding site (Fig. 2.1a). This binding has an intrinsic association constant K (M^{-1}).
- It can bind next to an already bound protein on one side, while on the other side there is no bound protein (singly contiguous (s.c.), Fig. 2.1b). This reaction

[1] ssDNA = single stranded DNA

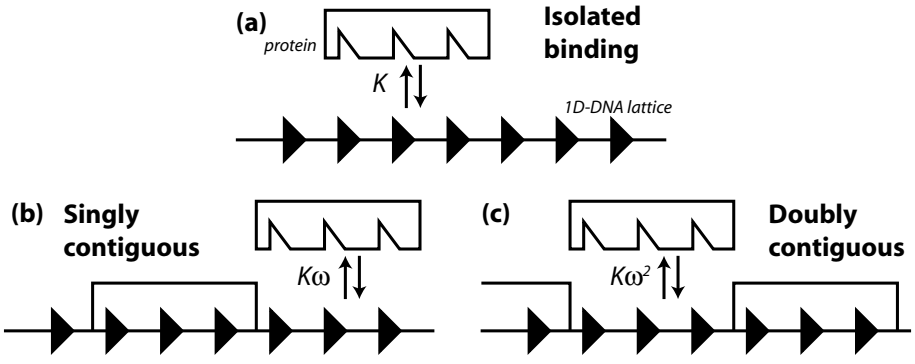

Figure 2.1: Schematic drawing of three possible types of binding of a protein to a 1D-DNA lattice: **a)** Isolated, **b)** singly continuous and **c)** doubly contiguous. Each base pair is indicated by a black triangle and in this case the protein covers 3 base pairs.

has an intrinsic association constant $K\omega$ (M^{-1}), in which ω is the (dimensionless) cooperativity parameter.

- A third type of binding consists of a protein binding between two already bound proteins in such a way that there are no empty base pairs between the proteins after binding of the protein (doubly contiguous (d.c.), Fig. 2.1c). This reaction has an intrinsic association constant $K\omega^2$ (M^{-1}). Doubly contiguous binding is a second order effect because the probability to find a gap with size n is low (for ω=1000, n=5 this probability is around 0.1% depending on percentage of DNA that is covered by proteins).

The (dimensionless) cooperativity parameter ω indicates the relative affinity of a protein for a contiguous, as opposed to an isolated, binding site. For a non-cooperative process $\omega = 1$. A cooperative process will have $\omega > 1$, which indicates that a protein binds preferentially next to an already bound protein, because then the binding rate $K\omega$ is larger than the inherent DNA binding rate K. The opposite is also possible, $\omega < 1$ indicates anti-cooperative behavior. Usually a process is called cooperative when $\omega \geq 10n$ [25]. An example of a cooperative protein is the gene 32 protein which binds cooperatively to poly(rA) (ω=1.2·10^3, n=7.5 and K=9·10^3 M^{-1}) [26]. Other examples are the recA protein which binds cooperatively to ssDNA (ω=50±10, n=5±0.7 and K=5.1·10^4 M^{-1}) [27] and the T4 UvsX protein which also forms filaments on ssDNA (ω=100, n=4.0±0.3 and K=4.3·10^4 M^{-1}) [28].

To derive an equation that relates the binding parameters of interest K, n and ω to parameters one can experimentally measure such as the total[2] protein concentration and the average fraction of DNA covered by the protein θ (also called the fractional DNA saturation), several aspects have to be considered. First, for each type of binding the number of binding sites needs to be calculated (Fig. 2.2). For a free DNA molecule the number of free binding sites equals $N - (n - 1)$, because a protein which covers n bases cannot bind to the last $n - 1$ bases of the DNA molecule (Fig. 2.2a). When a protein binds to a completely free DNA molecule it can eliminate as much as $2n - 1$ binding sites, because the protein itself already occupies n sites and also prevents another protein from binding to $n - 1$ sites to the left of the bound protein. When one or more proteins are already bound the number of free binding sites depends completely on where the protein(s) are bound. If a protein binds to a gap g (in units of bp[3]) which is exactly n bases long, lying between two already bound proteins, only one free binding site is eliminated.

The distribution of gap sizes is described by the probability distribution P_g and is a function of n, ω and ν. ν represents the binding density of the proteins (in units of

[2]It is important to note the difference between *free* and *total* protein concentration. Total concentration includes both proteins in solution and proteins bound to the DNA, while free concentration only includes the protein in solution.

[3]bp = base pairs

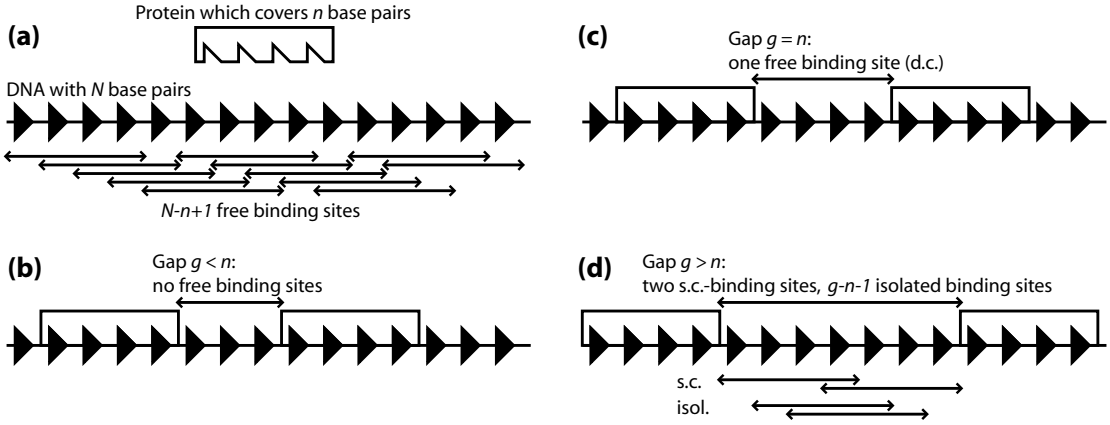

Figure 2.2: Calculation of the number of binding sites for each type of binding. **a)** A free DNA molecule has $N-(n-1)$ free binding sites, **b)** there are no free binding sites for $g<n$, **c)** if $g=n$ there is one free binding site (d.c.), **d)** for $g>n$ there are two s.c.-binding sites and $g-n-1$ isolated binding sites.

moles of bound protein per mole of base pairs). For a completely free DNA molecule $\nu=0$, and $\nu=\frac{1}{n}$ for a DNA molecule that is completely covered by proteins with size n. ν is related to the average DNA coverage θ by

$$\nu=\frac{\theta}{n} \tag{2.1}$$

To calculate P_g it is necessary to define how the size of a gap is determined. We count the number of free bases in a gap by starting at the right end of a bound protein and continuing until the left end of a next bound protein is reached. Using this counting convention we can define three conditional probabilities:

- $P(f|b_n)$ = the probability of finding a free base pair f next to the right end of a bound protein b_n.
- $P(f|f)$ = the probability that a free base pair has another free basepair to the right of it.
- $P(b_1|f)$ = the probability to find the left end of a bound protein b_1 next to a free base pair f.

All these conditional probabilities are a function of n, ω and ν. Using these probabilities, P_g can be written as

$$P_g=P(f|b_n)\cdot[P(f|f)]^{g-1}\cdot P(b_1|f) \tag{2.2}$$

It is also important to notice that in a gap of length g there are no free binding sites if $g<n$ (Fig. 2.2b); one doubly contiguous binding site if $g=n$ (Fig. 2.2c); two singly

contiguous binding sites if $g \geq n+1$ (Fig. 2.2d); and finally, $g-n-1$ isolated binding sites per gap if $g \geq n+2$ (Fig. 2.2d). Using these relations and multiplying it with P_g results in $\bar{s}$, the average number of free binding sites per gap given a certain gap size distribution P_g. $\bar{s}$ for each different type of binding can be calculated as shown in [24]:

$$\bar{s}_{isol} = \sum_{g=n+2}^{\infty} (g-n-1)P_g = \frac{P(f|b_n) \cdot [P(f|f)]^{n+1}}{P(b_1|f)} \tag{2.3a}$$

$$\bar{s}_{sc} = \sum_{g=n+1}^{\infty} 2 \cdot P_g \qquad = 2P(f|b_n) \cdot [P(f|f)]^n \tag{2.3b}$$

$$\bar{s}_{dc} = 1 \cdot P_n \qquad = P(f|b_n) \cdot [P(f|f)]^{n-1} \cdot P(b_1|f) \tag{2.3c}$$

where P_n is the probability to find a gap $g = n$ on the DNA molecule, $\bar{s}_{isol}$ the average number of free isolated binding sites per *gap*, $\bar{s}_{sc}$ the average number of s.c. binding sites per *gap* and $\bar{s}_{dc}$ the average number of d.c. binding sites per *gap*.

The average *total* number of bound proteins per DNA is defined as B and therefore $\nu = \frac{B}{N}$. These bound proteins can be divided in three groups: proteins that are bound in an isolated manner (B_{isol}); proteins that are singly contiguously bound (B_{sc}) and proteins that are doubly contiguously bound (B_{dc}). We define:

$$B = B_{isol} + B_{sc} + B_{dc} \tag{2.4}$$

For B bound proteins there are $B-1$ gaps between those proteins for $g \geq 0$. The free part of DNA that exists from the beginning of the DNA molecule till the first bound protein is counted as a gap and also the part between the last bound protein and the end of the DNA molecule. Therefore there are $B+1 \simeq B$ total gaps on the DNA for $g \geq 0$. Multiplying equations 2.3a-c (which calculate the average number of binding sites *per gap*) by B results in the average number of each type of free binding site present on the *entire DNA molecule*.

In general we can write the binding of a protein to a free binding site as follows:

$$\text{Free binding site} + \text{Free protein} \underset{}{\overset{\text{K}}{\rightleftharpoons}} \text{Bound protein} \tag{2.5}$$

The associated mass-action equation is:

$$\frac{[\text{Bound protein}]}{[\text{Free protein}]} = K \cdot [\text{Free binding sites}] \tag{2.6}$$

Both sides can be divided by the concentration of DNA and the resulting equation can

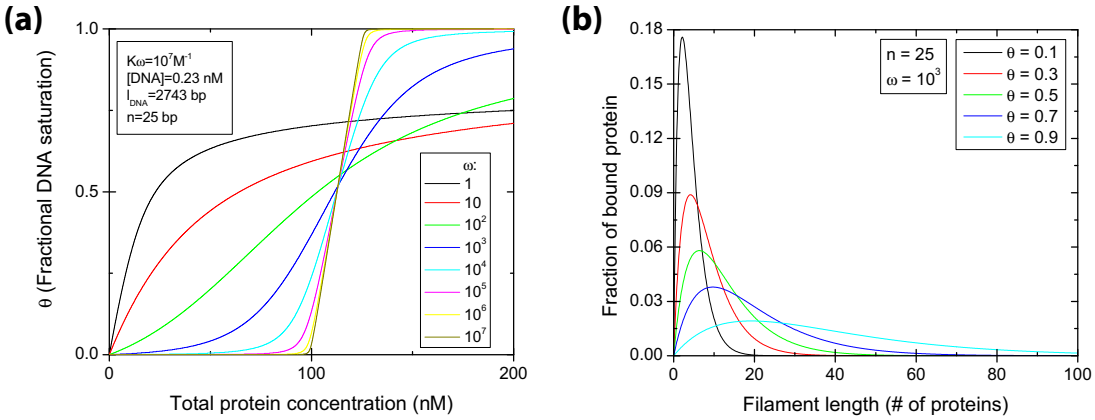

Figure 2.3: a) Fractional DNA saturation versus total protein concentration for various values of K and ω. **b)** Fraction of protein that is bound to DNA in a filament of a certain length versus filament length for various values of fractional DNA saturation.

be written for each type of binding (isolated, s.c. or d.c.):

$$\frac{B_{isol}}{L} = K \cdot (\text{average number of free isolated binding sites per DNA}) \tag{2.7a}$$

$$\frac{B_{sc}}{L} = K\omega \cdot (\text{average number of free s.c. binding sites per DNA}) \tag{2.7b}$$

$$\frac{B_{dc}}{L} = K\omega^2 \cdot (\text{average number of free d.c. binding sites per DNA}) \tag{2.7c}$$

where L is the free protein concentration. Now we substitute $(B \cdot \bar{s}_{isol})$, $(B \cdot \bar{s}_{sc})$ and $(B \cdot \bar{s}_{dc})$ for the terms in parentheses in equations 2.7a-c. Substituting equations 2.7 into equation 2.4 results in the overall binding equation [24]:

$$\frac{\nu}{L} = K(1-n\nu)\left[\frac{(2\omega-1)(1-n\nu)+\nu-R}{2(\omega-1)(1-n\nu)}\right]^{n-1}\left[\frac{1-(n+1)\nu+R}{2(1-n\nu)}\right]^{2} \tag{2.8}$$

where

$$R = \sqrt{[1-(n+1)\nu]^2 + 4\omega\nu(1-n\nu)} \tag{2.9}$$

is used for simplification. The quantities that can be experimentally obtained are the fractional DNA saturation $\theta = \nu \cdot n$ and the total protein concentration L_T which can be substituted in equations 2.8 and 2.9 by using

$$L = L_T - \nu N \tag{2.10}$$

Figure 2.3a shows several theoretical curves plotting θ versus the total protein concentration for different values of ω. The difference between non-cooperative binding and cooperative binding is very clear from the graph. For ω=1 (non-cooperative) and low protein concentrations θ increases almost linearly with increasing protein concentration

and saturates at high protein concentrations. For ω>1000 the curves are completely different. At low protein concentrations θ remains zero and the transition from empty DNA to completely covered DNA happens in a narrow protein concentration range resulting in complete coverage of the DNA molecule at high protein concentrations. The shape of the curves becomes clearly sigmoidal when $\omega \geq 10n$ [25]. The shape of the curves in this regime is largely defined by the ratio $\frac{\omega}{n}$ [24].

2.1.1 Filament length distribution

An experimentally useful approach to determine the parameters n, K and ω is measuring the filament length distribution. The filament length c is defined as the number of proteins that are bound next to each other without a gap between them. This number can range from 1 to $\frac{N}{n}$. The distribution of filament lengths depends on the values of n, K, and ω during the binding reaction [25]. Using AFM imaging this distribution can be measured and therefore an estimation can be obtained for these parameters. The average filament length ($\bar{C}$) is the average number of proteins bound per DNA molecule (B) divided by the average number of filaments per DNA molecule. Each filament ends with a protein that has a free binding site at its right and therefore the average number of filaments per DNA molecule equals

$$B - B[P(b_1|b_n)] \tag{2.11}$$

where

$$P(b_1|b_n) = \frac{1-(n-2\omega+1)\nu - R}{2\nu(\omega-1)} \tag{2.12}$$

For $\bar{C}$ this results in

$$\bar{C}(n,\omega,\nu) = \frac{B}{B - B \cdot P(b_1|b_n)} = \frac{1}{1-P(b_1|b_n)} = \frac{2\nu(\omega-1)}{(n-1)\nu - 1 + R(n,\omega,\nu)} \tag{2.13}$$

$\bar{C}$ does not have any information about the distribution of filaments lengths. Also in our experiments we obtained data for three data sets resulting in only three values for $\bar{C}$ which is not enough for fitting a line through it.

A more practical approach is to calculate the distribution of filament lengths at a given DNA saturation θ. θ can be determined experimentally and therefore it is not necessary to know the initial total protein concentration. This presents the advantage that results of this theory are completely independent of errors in the determination of the initial total protein concentration. The probability P'_c to find a filament with length c next to a free binding position on the DNA equals $P(f|f)$ for $c = 0$ and

$$P'_c = P(b_1|f) \cdot [P(b_1|b_n)]^{c-1} \cdot P(f|b_n) \tag{2.14}$$

for $c > 0$, in which $P(b_1|f)$ indicates the probability of finding the begin of a filament, $[P(b_1|b_n)]^{c-1}$ the probability of finding c proteins bound next to each other and $P(f|b_n)$

is the probability of finding the end of a filament. Because filaments of size 0 are not taken into account P_c' needs to be renormalized. This results in

$$P_c = [P(b_1|b_n)]^{c-1}[1 - P(b_1|b_n)] \tag{2.15}$$

and P_c is the normalized probability (for $c > 1$) to find a filament of length c. To calculate the fraction (F_c) of all bound proteins that are in a filament with length c, P_c has to be multiplied by the filament length and normalized:

$$F_c = \frac{cP_c}{\sum_{c=1}^{\infty} cP_c} = c[1 - P(b_1|b_n)]^2[P(b_1|b_n)]^{c-1} \tag{2.16}$$

Combining equations 2.12 and 2.16 results in a final expression for the

$$F_c = c\left[1 - \frac{1 - (n - 2\omega + 1)\nu - R}{2\nu(\omega - 1)}\right]^2\left[\frac{1 - (n - 2\omega + 1)\nu - R}{2\nu(\omega - 1)}\right]^{c-1} \tag{2.17}$$

Figure 2.3b shows several theoretical curves of F_c versus c for several values of θ, indicating that for large θ there is a wide distribution of filament lengths. This means that almost all bound proteins are found in large filaments, while for small θ almost all bound proteins are found in short filaments.

In this theory DNA was treated as a molecule of infinite length. In real experiments the DNA length is finite. The theory of filament length distribution presented in subsection 2.1.1 breaks down when theoretical filament lengths exceed the actual length of the DNA used in experiments. The effect of this approximation on the experiments will be discussed in section 6.1.2.

2.2 Diffusion of molecules

In a liquid environment molecules are constantly colliding with water molecules which are constantly moving because of their thermal energy. The resulting random movement of molecules in solution is called *Brownian motion*. Due to this motion molecules such as proteins will be at different positions in time but on average the movement of the protein is zero. The mean-square displacement $\langle(\Delta l_n)^2\rangle$ of the protein in a certain amount of time can be calculated with

$$\langle(\Delta l_n)^2\rangle = \frac{\sum_{i=1}^{N-n}(l_i - l_{i+n})^2}{N - n} \tag{2.18}$$

in which N is the total number of measured positions of the molecule which is equal to the number of frames in the movie that shows the molecule. The time between each measurement is fixed. l_i and l_{i+n} are the measured positions of the molecule with a time difference of n frames.

The diffusion constant D is an indication of the size of the excursion of a molecule. The relation of D to the mean-square displacement of a protein that can diffuse only in one dimension is

$$\langle(\Delta l_n)^2\rangle = 2Dt \tag{2.19}$$

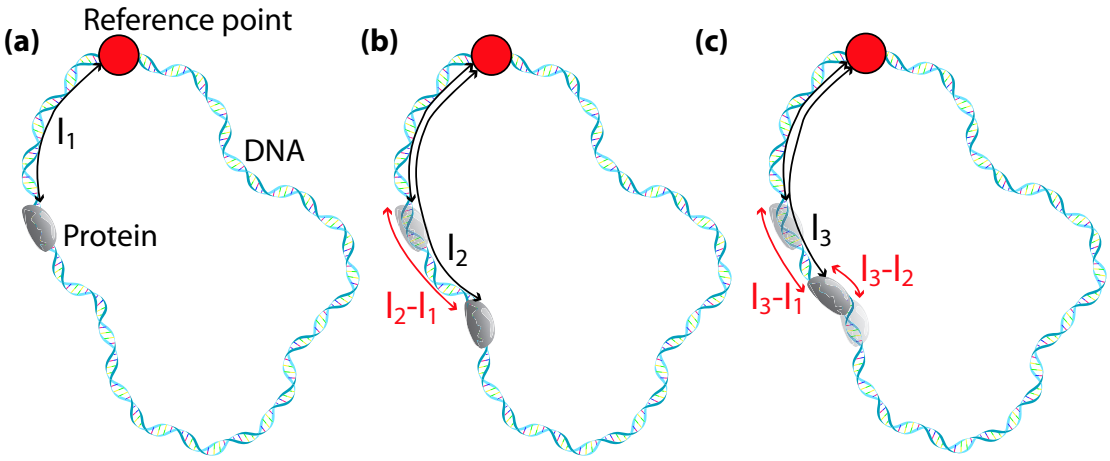

Figure 2.4: Schematic drawing of a protein that is diffusing along a DNA molecule. Between each of the consecutive images the protein has moved to another position on the DNA.

Figure 2.4 shows an example of how linear diffusion of a protein along a DNA molecule is measured. For a series of consecutive snapshots a fixed point on the DNA is located which will function as a reference point. In the first frame (Fig. 2.4a) the position of the protein is measured relative to the reference point along the contour of the DNA molecule. This distance is denoted as l_1. This procedure is repeated in the second frame (Fig. 2.4b) resulting in l_2. Then $l_2 - l_1$ is calculated for use in equation 2.18 for $n = 1$. Again for the third frame (Fig. 2.4c) the distance from the protein to the reference point is measured (l_3) and $l_3 - l_2$ is calculated for equation 2.18 for $n = 1$. For $n = 2$ the value for $l_3 - l_1$ is used. This procedure is repeated for all frames where the protein is still bound.

For a molecule that can move in two dimensions (for example on a mica surface) equation 2.19 changes into

$$\langle(\Delta r_n)^2\rangle = \langle(\Delta x_n)^2\rangle + \langle(\Delta y_n)^2\rangle = 4Dt \tag{2.20}$$

in which

$$\langle(\Delta x_n)^2\rangle = \frac{\sum_{i=1}^{N-n}(x_i - x_{i+n})^2}{N-n} \tag{2.21a}$$

$$\langle(\Delta y_n)^2\rangle = \frac{\sum_{i=1}^{N-n}(y_i - y_{i+n})^2}{N-n} \tag{2.21b}$$

x and y indicate the coordinates of the molecule that is diffusing on a 2D surface. AFM is a powerful tool to investigate diffusion of proteins on a DNA molecule and it was used to examine the diffusion of RNA polymerase [29].

3. Materials and Methods

3.1 DNA preparation

In all the experiments the vector *pGEM®-3Z*[1] (2743 bp) was used in circular or in linear form. The theoretical length of the DNA is 933 nm assuming a rise per base pair of 0.34 nm compatible with B-form DNA [30]. There are several reasons to use this DNA. First, because of the relative short size of the DNA there are not many intramolecular crossings which simplifies the analysis of cooperativity and aggregation. Secondly, the DNA length is close to 1 µm and this is a convenient size for imaging an area of 2 µm x 2 µm on the surface. Using this image size we can adjust the DNA concentration to get 5-10 DNA molecules in one AFM image with a good resolution ($4\,^{\mathrm{nm}}/_{\mathrm{pixel}}$). Finally, there is no noticeable creep during imaging and the image of the DNA has a FWHH [2] of 10-15 nm with the used cantilevers.

The stock solution of circular DNA is mostly supercoiled although there is a small population of nicked molecules probably caused by degradation during the purification process (Fig. 3.1a lane 2). For most experiments the stock needs to be processed further. For practical reasons linear DNA was used for the aggregation experiments. Linear DNA was produced by adding the restriction enzyme *Bam*HI to the stock solution of circular DNA. In figure 3.1a an image from an agarose gel shows a clear band corresponding to a linear 2.7 kbp DNA molecule (lane 3). After purification[3] the linearized DNA molecules were imaged with the AFM, a typical image is shown in figure 3.1b. The contour length of 87 DNA molecules was determined by tracing using a homemade software package[4]. The result is shown in figure 3.1c. The average contour length obtained from this analysis is 919 ± 69 nm which agrees nicely with the theoretical value.

For cooperativity experiments relaxed nicked circular DNA was used in order to reduce the aggregation process and favor the formation of intramolecular TopIB-DNA filaments. As mentioned above the plasmid DNA from stock is negatively supercoiled. To remove the supercoils the nicking enzyme *N.Bst*NB I[5] was added. The recognition site for this enzyme (5'...GAGTC...3') was found twice in the used DNA. As a result the DNA is nicked on 2 positions and the supercoils are removed (figure 3.1a lane 4).

[1] Promega Biotech: http://www.promega.com/
[2] FWHH=Full Width at Half Height
[3] Purification was done with a UltraClean^{TM} PCR Clean-up^{TM} Kit from Mo Bio Labs.
[4] Written in IDL language by John van Noort and Thijn van der Heijden
[5] New England Biolabs: http://www.neb.com/

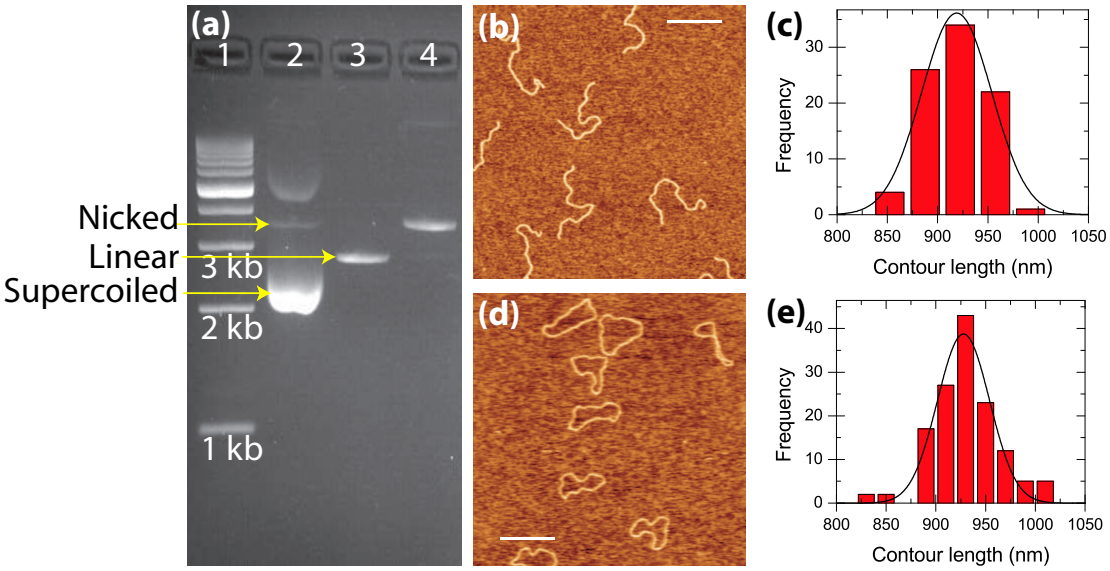

Figure 3.1: DNA used in experiments. **a)** Gel from agarose gel electrophoresis with lane 1 showing a 1 kbp ladder; lane 2 shows the DNA from the stock solution with the bright band indicating supercoiled DNA; lane 3 shows a band around 2.7 kbp for linear DNA; lane 4 shows the relaxed circular plasmid. **b)** Typical AFM image of linear DNA adsorbed on mica using $MgCl_2$ (13 mM). **c)** Distribution of measured DNA contour length showing an average of 919±69 nm. **d)** Typical AFM image of relaxed circular DNA adsorbed on mica using $MgCl_2$ (13 mM). **e)** Distribution of measured DNA contour length showing an average of 928±52 nm. Scale bar size of AFM images is 400 nm.

After purification the DNA molecules were imaged with AFM. A typical image is shown in figure 3.1d. The contour length of 136 DNA molecules was determined. The result is shown in figure 3.1e. The average contour length obtained from this analysis is 928 ± 52 nm which agrees nicely with the theoretical value.

3.2 Vaccinia Topoisomerase IB

Wild-type vaccinia topoisomerase IB enzymes where purified by the laboratory of S. Shuman as described previously [14]. TopIB was aliquoted and stored at $-80\,^{\circ}C$ in a 10 mM Phosphate Buffer (PB) at pH 7.4. For each day of experiments one aliquot was thawed, diluted to the concentration needed and kept on ice. After one day of experiments the thawed aliquot was discarded.

3.3 Sample preparation

3.3.1 Binding reaction

Sample preparation consisted of two steps: the dilution of TopIB and the binding reaction between TopIB and DNA.

TopIB was diluted to 5-100 nM by adding PB (10 mM, pH 7.4) or Tris·HCl buffer (10 mM, pH 7.4 or 8.0). PB was used in air experiments, except in the cases where TopIB was combined with $MnCl_2$ or $CaCl_2$ because otherwise residue was formed. In these cases Tris·HCl buffer (pH 7.4) was used. This buffer was also used in the liquid experiments, because the imaging buffer contained also Tris·HCl (pH 8.0).

The binding reaction consisted of 0.22 nM DNA (circular or linear, depending on the experiment), 10 mM buffer (Tris·HCl or PB) and 2.5-40 nM Topoisomerase IB. Depending on the experiment 13 mM $MgCl_2$ was added. This was needed to deposit the sample onto a mica surface. The total volume of the sample was 6 µl. After mixing, the sample was kept at 37°C for 15 minutes and subsequently it was deposited onto a mica surface.

3.3.2 Sample deposition on mica

For imaging the sample with the AFM the molecules need to be adsorbed onto a flat surface. Mica was used as a surface because the material consists of very thin atomically flat layers. Both the mica surface and the DNA are negatively charged in water (pH 7.4-8.0). In order to fix the DNA on the surface the mica needs to be positively charged. We used two methods to do this. Most of the time $MgCl_2$ was added to the binding reaction. The Mg^{2+}-ions replace the negative charges on the mica surface and as a result the DNA molecules are attracted to it. This method results in *equilibrium* trapping [31]. This means that DNA molecules are allowed to equilibrate on the 2D mica surface before they are finally captured in a particular conformation. In this conformation the distribution of two consecutive DNA segments of length l is governed by thermal fluctuations following a Gaussian distribution. Instead of Mg^{2+} it is also possible to use other divalent ions for this process by adding for example $CaCl_2$ or $MnCl_2$ to the solution.

Treating the mica with polylysine (PLL) is another method to positively charge the surface. This is done by placing 10 µl of 0,01% PLL-solution on the mica surface, rinsing it with water after a few seconds and drying it. This method results in *kinetic* trapping [31]. In the presence of PLL the DNA molecules do not equilibrate on the mica surface and the resulting conformation resembles a projection of the 3D conformation of the molecule on a 2D surface. Figure 3.2a shows equilibrated molecules which look less compact than the kinetically trapped molecules in figure 3.2b. In both images the contour length of the DNA and the image size are identical. Usually equilibrated DNA is easier to analyze. After the sample has been placed on the surface for about one minute, it needs to be rinsed with a few milliliters of MilliQ water and dried.

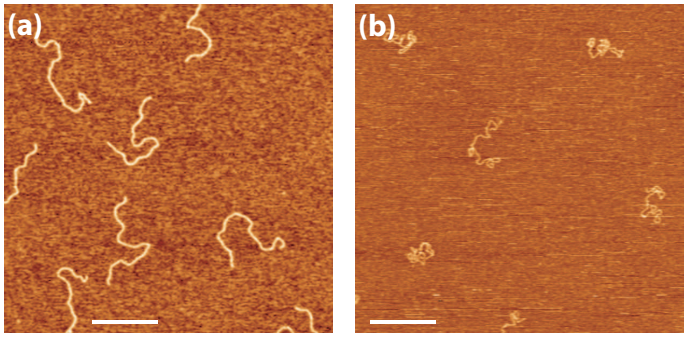

Figure 3.2: AFM images of equilibrated **(a)** and kinetically trapped **(b)** DNA molecules. Scale bar size is 400 nm.

3.3.3 Imaging buffer

Imaging in liquid requires a careful selection of the imaging buffer. In our experiments the *imaging buffer* contained a Tris·HCl buffer, a certain concentration $MgCl_2$ and a certain concentration KCl. For imaging in liquid the DNA molecules need to be loosely bound to the surface in such a way that they are still visible with AFM and are able to move reasonably free over the surface. An increase in $MgCl_2$ creates stronger binding of DNA to the surface. An increase in KCl increases the screening which reduces the attachment of the DNA to the surface. The optimal values were found to be 3 mM $MgCl_2$ and 75 mM KCl in Tris·HCl buffer (10 mM, pH 8.0).

3.4 Atomic Force Microscope

A commercially available AFM (Atomic Force Microscope) from Nanotec Electrónica[6] was used. The AFM was operated in dynamic mode. In this mode the cantilever oscillates at or close to its resonance frequency. For imaging in air soft[7] cantilevers were used (Fig. 3.3a, *Olympus OMCL-RC800PSA*, k=0.39 N/m). The amplitude of the cantilever oscillation depends on the interaction between the surface and the tip and the amplitude is used as the parameter for the feedback system. When the sample is brought closer to the cantilever the amplitude is reduced because of interaction with the surface (Fig. 3.3b). The amplitude is a measure for the distance between tip and sample and this distance can be determined to sub-nanometer precision. The AFM was usually operated at small amplitude-reduction. The normal force (Fig. 3.3b) is on average very small but this is mainly because most of the time the tip is not in contact with the sample. When closer to the sample the time that the tip is in contact with the sample increases and also the force it applies to the sample. It is interesting to note that when approaching the sample at a certain point the cantilever snaps into contact with the sample (discontinuity in the normal force and amplitude) and is actually attracted to the sample

[6] http://www.nanotec.es/
[7] with a low spring constant k

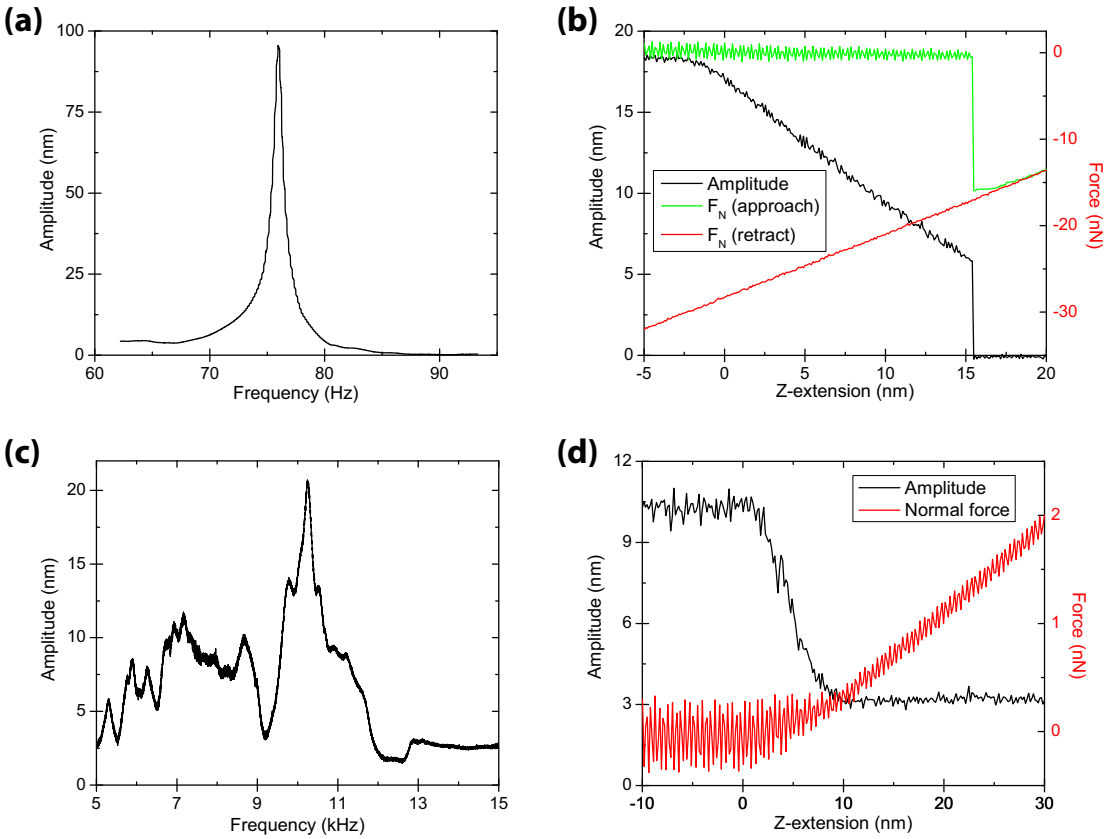

Figure 3.3: Typical characteristics of AFM cantilevers used in air and liquid. **a)** Amplitude of a cantilever in air vs. drive frequency; the peak indicates the resonance frequency. **b)** Cantilever amplitude and normal force between tip and sample vs. the z-piezo extension. The amplitude reduces when the distance between tip and sample is reduced (z-piezo extension is increased). The green curve indicates the normal force when the tip approaches the sample; the red curves indicates the normal force when the tip is retracted from the sample. **c,d)** Similar graphs for a cantilever in liquid. For liquid there is no difference in the normal force for approaching or retracting the tip, so only one curve (red) is plotted.

due to capillary forces. When the tip is retracted again it stays in contact with the surface due to the capillary forces and a relatively high force (in this case $\sim 70\,\mathrm{nN}$) is needed to overcome these capillary forces. Therefore hysteresis is observed between the approaching and retracting curve.

In liquid softer cantilevers were used (*Olympus OMCL-TR400PSA*, k=0.08 $\mathrm{N/m}$) to prevent damaging the biological sample and usually the amplitude of the cantilever is smaller (Fig. 3.3c) and also the resonance frequency because of damping by the water. Often the frequency spectrum presents more than one peak. This is due to acoustic vibrations in the liquid cell. It is difficult to choose a good peak and in practice a peak

around 11 kHz was chosen, because from experience it was known that this frequency usually resulted in the best possible image quality in liquid. Also there is more noise in the amplitude signal because water molecules are constantly bumping into the cantilever. Because no capillary forces exist in liquid (figure 1.5) there is no attracting force when the tip approaches the sample and no snap-in of the cantilever occurs. Also the normal force is identical for the approaching and retracting curve.

3.5 Data analysis

3.5.1 Maximum-likelihood estimation

A common method to analyze large data sets is to represent them in a histogram. To make a histogram all values are grouped and put into bins of a certain size. Obviously the data is changed in this process and a fit of a model to the data will depend on the choice of the bin size. While for large data sets this problem plays a minor role, for small data sets it is often difficult, if not impossible, to choose a satisfying bin size. A powerful method to unambiguously estimate a parameter from a series of measurements is to look at the likelihood to measure a given data set for a certain function and its parameters. Maximizing the likelihood results in the best estimation for the parameters of the model. The maximum-likelihood (ML) [32] method was used for estimating the cooperativity parameter ω. Let's consider a data set $x_1, x_2, x_3, \ldots, x_N$ with n independent measurements. Each measurement has a certain probability to be measured based on the underlying model with parameter ω. This probability is given by the probability density function $P(x_i;\omega)$. The likelihood function is then defined as

$$L(x_1, x_2, x_3, \ldots, x_n) = \prod P(x_i;\omega) \tag{3.1}$$

Combining equations 2.15 and 3.1 gives the likelihood function for the cooperativity model:

$$L(c_1, c_2, c_3, \ldots, c_N) = \prod_{k=1}^{N} [1 - Q(\omega)] Q(\omega)^{c_k - 1} \tag{3.2}$$

in which c_k are the measurements of the filament length and $Q(\omega) = P(b_1|b_n)$. Maximizing this function with respect to ω will give the best estimator for this parameter. First, all unnecessary terms were removed from the product:

$$L = [1 - Q(\omega)]^N \cdot \prod_{k=1}^{N} Q(\omega)^{c_k - 1} \tag{3.3}$$

The product can be transformed into a sum by taking the logarithm of the likelihood:

$$\ln L = N \ln [1 - Q(\omega)] + \sum_{k=1}^{N} \ln Q(\omega)^{c_k - 1} \tag{3.4}$$

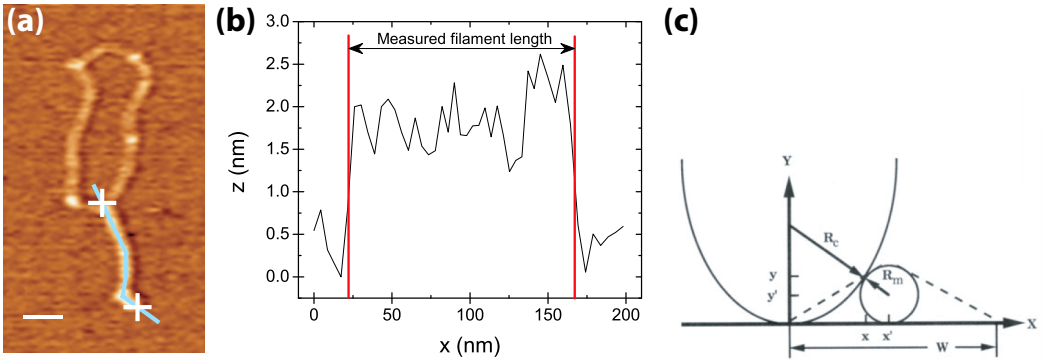

Figure 3.4: **a)** Typical AFM image of a topo-filament on a DNA molecule. Scale bar size is 50 nm. **b)** Height profile of the filament following the line drawn in figure a. **c)** Schematic drawing of tip-sample convolution.

Rewriting gives:

$$\ln L = N \ln [1 - Q(\omega)] - N \ln Q(\omega) + \left(\sum_{k=1}^{N} [\ln Q(\omega)] c_k \right) \tag{3.5}$$

To maximize this equation with respect to ω the derivative is taken and set to zero. Using the Maple software package[8], $Q(\omega) = P(b_1|b_n)$ and equation 2.12 results in an analytical solution for the maximum likelihood estimator for ω ($\hat{\omega}$):

$$\hat{\omega} = \frac{N\theta n \sum_{k=1}^{N} c_k + \theta N^2 - \theta n \sum_{k=1}^{N} c_k - \theta N \sum_{k=1}^{N} c_k - nN \sum_{k=1}^{N} c_k + n \left(\sum_{k=1}^{N} c_k \right)^2}{\theta N^2} \tag{3.6}$$

With this equation $\hat{\omega}$ can be determined for each data set. θ is experimentally obtained from the data set, N is the number of data points in the data set, n is the number of base pairs the protein covers and c_k indicates kth measured filament length from the data set. The uncertainty in $\hat{\omega}$ corresponds to the places where $\ln L$ is 0.5 below its maximum [33]. Because $\ln L$ is not symmetric around the peak value, the uncertainty is also asymmetric.

3.5.2 Measuring filament lengths

Figure 3.4a shows a typical AFM image of a circular DNA molecule with a TopIB filament. The length of this filament was measured by tracing the filament by hand. This results in a height profile (Fig. 3.4b). The length of the filament l_{fil} is measured by taking the begin and the end of the filament as the positions where the height of

[8] http://www.maplesoft.com/

the profile is half the maximum value (Fig. 3.4b). The tip of the AFM is not infinitely sharp, it has a certain radius R_c, which results in tip-sample convolution as is shown in figure 3.4c [34]. The apparent width of the molecule W is given by

$$W = \frac{4(R_c + R_m)\sqrt{R_m(R_c - R_m)}}{R_c}, R_c > R_m \tag{3.7}$$

in which R_m is the true radius of the molecule. The tip-sample convolution can be estimated by measuring a DNA molecule. The width of a DNA molecule is known, so the contribution of W_{conv} to the apparent length of the molecule can be calculated with

$$W_{conv} = W - W_{DNA} \tag{3.8}$$

The true filament length can then be calculated by subtracting W_{conv} from the measured l_{fil}.

In some cases, especially at higher TopIB concentrations, individual DNA molecules had a blob associated to them. For these events it is impossible to measure the filament length directly. Instead, the length of the DNA that was not covered with TopIB was measured. This value is then subtracted from the average DNA contour length (section 3.1). The resulting value was divided by two to account for the fact that two DNA duplexes are present in a filament. These events were only used in the calculation of the average DNA saturation θ and not used in the filament-length distribution, because it is not known if the blob consists of one filament or more. For TopIB-DNA aggregates that consist of more than one DNA molecule a similar procedure was used. The total length of free DNA in an aggregate was measured. Again this value was subtracted from the total length of the DNA molecules in the aggregate and divided by two. The filament length for each molecule in the aggregate was then calculated by dividing the total filament length by the number of DNA molecules in the aggregate.

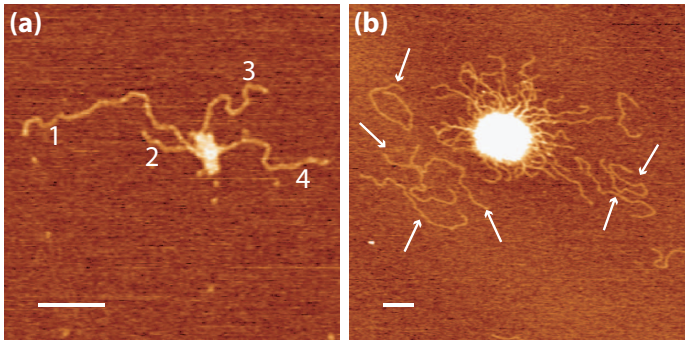

Figure 3.5: Method to count DNA molecules in aggregates. Scale bar size is 200 nm. **a)** Counting the number of DNA 'legs' and dividing the number by 2. **b)** Sometimes DNA molecules are found very close to an aggregate (indicated by the white arrows), these molecules are counted as if they were part of the aggregate.

3.5.3 Measuring the size of the DNA-TopIB aggregates

The number of DNA molecules in a DNA-TopIB aggregate was measured by counting the DNA molecules by hand. For small aggregates (1-5 DNA molecules) this was not a problem. For larger aggregates it was more difficult to distinguish all the DNA molecules. Aggregates usually have a big 'blob' in the center which makes the counting of DNA molecules even more difficult. To estimate the number of DNA molecules the number of visible DNA ends was counted and divided by two (a linear DNA molecule obviously has two ends, an example is shown in figure 3.5a). Then also the DNA molecules that are very close (< 100 nm) to the aggregate were counted as being part of the aggregate (Fig. 3.5b). This was done because these DNA molecules were probably removed from the aggregate by the rinsing process. This is a reasonable assumption because the probability of a DNA molecule being deposited next to an aggregate of 10+ DNA molecules are slim.

4. TopIB aggregates DNA in a cooperative fashion

4.1 Binding of TopIB to DNA shows cooperative behaviour

We first investigated the formation of intramolecular nodes and filament-like structures by mixing TopIB and circular DNA at different ratios. As expected TopIB formed filament-like structures upon binding to circular nicked DNA. Figure 4.1 shows several typical AFM images of these TopIB-DNA filament-like structures. Images shown were taken in air, thus they are snapshots of configurations molecules adopt in solution. Three different stages in the formation of the filaments can be seen. First (1) a few TopIB molecules are bound individually to DNA. Then (2) two bound TopIB molecules come

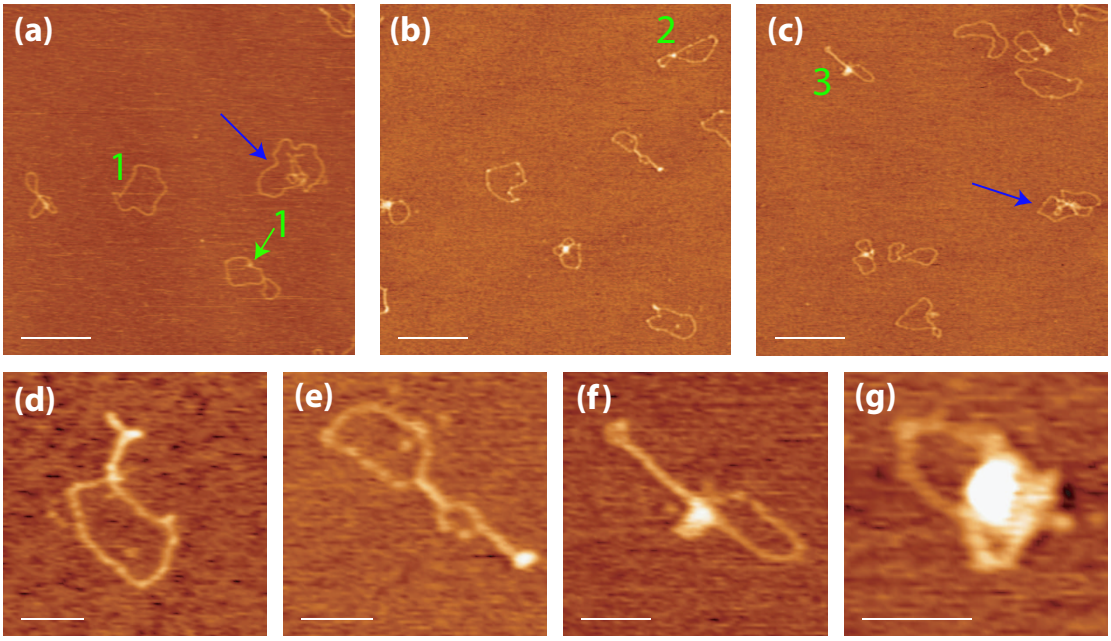

Figure 4.1: AFM images taken in air of TopIB-DNA filaments, [DNA]=0.22 nM. Three different stages in the formation of the filaments are indicated by green numbers. The blue arrows indicate aggregation of 2 DNA molecules by TopIB. **a)** [TopIB]=7.6 nM. **b,c)** [TopIB]=13.4 nM. **d,e,f)** Close-up images of circular DNA containing TopIB-DNA filaments. **g)** Close-up image of a cluster associated to DNA. Scale bar size is 400 nm for **a-c** and 100 nm for **d-g**.

close enough to form a TopIB-DNA node which in this case is called an *intra*molecular node. The formation of an intramolecular node requires two TopIB, bound to DNA, to form a dimer. Finally (3) more TopIB molecules extend the node to form a filament.

Interestingly two or more DNA molecules could aggregate in the presence of TopIB molecules (blue arrow). In this case two TopIB molecules bound to different DNA molecules form an *inter*molecular TopIB-DNA node. This node can also extend into filament-like structures resulting in aggregates. Figure 4.1g shows another manifestation of the process which can occur in a single DNA molecule. It appears as a cluster of TopIB/DNA associated to a DNA molecule. These type of events are found when higher [TopIB]/[DNA]-ratios are used. This behaviour is addressed further in section 4.2. The fact that some DNA molecules are completely free of TopIB, while others already have formed filaments is a strong indication of cooperative behaviour of TopIB-DNA binding. If the binding of TopIB to DNA was non-cooperative the sample should have been much more homogenous. This cooperative behaviour was analyzed by using the theory presented in section 2.1. Although this theory was developed for cooperative binding of proteins to a 1D DNA lattice, it can be adapted to describe the filament formation observed in TopIB-DNA interaction. The first step in filament formation in the presented theory is the binding of a protein to a free binding site on the DNA with a binding constant K. In the TopIB case the formation of an intramolecular TopIB-DNA node will be considered the first step:

$$\mathrm{TopIB}_{unbound} + \mathrm{TopIB}_{unbound} + \mathrm{DNA} \xrightarrow{K} \text{intramolecular node}$$

If binding is cooperative a filament will be formed from this node requiring two TopIB proteins (to form a dimer) for each extension of the filament. The binding constant of the extension of a node or filament by a TopIB dimer equals $K\omega$.

To quantify the cooperativity the length of filaments was measured for several hundreds of DNA molecules at different [TopIB]/[DNA]-ratios using the method described in subsection 3.5.2. The measured filament length was corrected for tip-sample convolution. The average total filament length per DNA molecule was calculated and the DNA saturation θ was obtained for each data set. This was done by dividing the average total filament length per DNA by the maximum possible filament length per DNA molecule which equals half the contour length of a DNA molecule (= 464 nm, considering the fact that always two duplexes are needed to form a filament). The obtained filament lengths were then divided by n (=25 bp) to get the number of proteins involved in the formation a filament. The distribution of filament lengths for three different data sets is shown in figure 4.2a and table 4.1 summarizes the result. DNA coverage by TopIB was measured as described in 3.5.2 and with the ML-method a value for ω was obtained (table 4.1). From all the molecules that have filaments there is a small part where the filaments led to cluster formation. This percentage is indicated in parenthesis in the legend of figure 4.2. The higher this percentage, the higher the overestimation of θ will be.

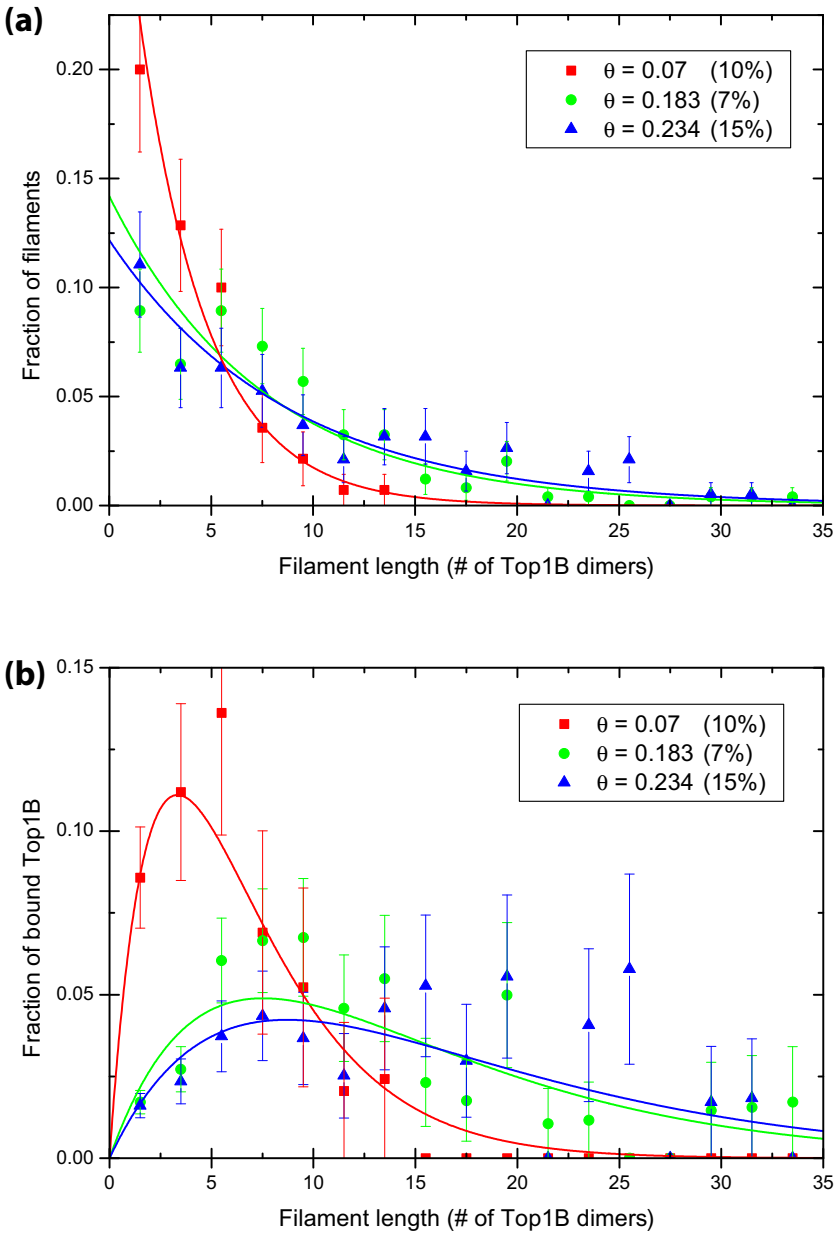

Figure 4.2: **a)** Filament length distribution for three different data sets. For each set ω was estimated using the ML-method described in subsection 3.5.1 resulting in ω=3667^{+1025}_{-775} for θ=0.07, ω=6318^{+1273}_{-1031} for θ=0.18 and ω=6183^{+1439}_{-1133} for θ=0.23. (In parenthesis the percentage of molecules with clusters is shown as an indication of the uncertainty in θ. For each experiment equation 2.15 is plotted using the obtained ω. **b)** Distribution of bound TopIB for each filament length obtained by multiplying the distributions in figure **a** by the filament length (x-axis value). For each experiment equation 2.16 is plotted using the obtained ω.

Set	# of filaments	# of DNA	θ	% clustered	ω	- error	+ error
1	70	107	0.07	10%	3667	-775	+1025
2	123	116	0.183	7%	6318	-1031	+1273
3	95	98	0.234	15%	6183	-1133	+1439

Table 4.1: Calculated values for the cooperativity parameter ω for three different data sets.

From these three experiments the weighted average for ω was calculated using

$$\bar{\omega} = \sqrt{\frac{\sum_i \omega_i^2/\sigma_i^2}{\sum_i 1/\sigma_i^2}} \qquad \sigma(\bar{\omega}) = \sqrt{\frac{1}{\sum_i 1/\sigma_i^2}} \tag{4.1}$$

which results in ω=5181±621. Although the error is around 12% the obtained result indicates that the binding of TopIB to DNA is a strongly cooperative process. Figure 4.2b shows the distribution of the bound TopIB over all the filament lengths for each experiment together with a plot of equation 2.16 using experimentally obtained parameters. At higher filament lengths (>15 TopIB dimers) the fits deviate from the data. This can be explained by the method by which the histograms are created. Each bin from the histogram in figure 4.2a is multiplied by its x-axis value (filament length) and renormalized. Because the data sets are relatively small a few large filaments that are about the same size will produce a relatively large bar in the histogram and this bar has therefore a large error associated with it.

To estimate the binding constant K of the formation of a TopIB-DNA node the DNA coverage θ was measured for different TopIB-to-DNA ratios. Figure 4.3 shows the obtained data for 4 different TopIB concentrations[2]. Equation 2.8 was used for fitting the data. Because θ cannot be written explicitly as a function of K the fitting was done using a self-written Labview[3] program using the least squares fitting method which resulted in K=44·10^3 M^{-1}. Another fit was done for ω=3667 which corresponds to the data set obtained at low [TopIB] (θ=0.07). The difference between the two curves is very small although the difference in ω is almost 30%. The product $K\omega$=2·10^8 M^{-1} is constant for the two curves.

4.2 TopIB aggregates DNA in the presence of divalent ions

In the previous section it was already shown that TopIB-DNA aggregates are formed at higher TopIB-to-DNA ratios. To investigate this behaviour several experiments were done using linear DNA instead of circular. The advantage of linear DNA is that intermolecular interactions are more likely to happen (it is not constrained by its circular

[1] The concentration of TopIB dimers is half the concentration of TopIB molecules.

[2] The value for [TopIB]=0 was not determined experimentally, but obviously when no TopIB is present in solution no filaments can be formed

[3] http://www.ni.com/labview/

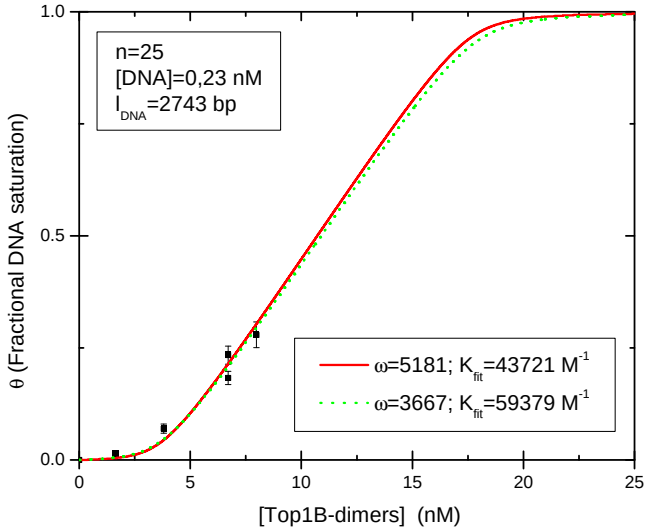

Figure 4.3: Fractional DNA saturation θ plotted vs. the total concentration of TopIB dimers[1]. Equation 2.8 is plotted using the values in the figure. Two different fits are shown: the red and the green curve show fits for the complete data set for different ω.

shape) and for big aggregates it is easier to count the number of DNA molecules in it.

Several stages of the interaction of TopIB and DNA were imaged with AFM by mixing TopIB and DNA at different ratios. The results are shown in figure 4.4. At low TopIB concentrations figure 4.4a shows DNA molecules with single TopIB molecules associated to it, which appear as dots on the DNA. This image resembles the images of the experiments with circular DNA. Increasing the TopIB-concentration resulted in the formation of TopIB-DNA nodes (Fig. 4.4b). Cooperative binding behaviour can also be seen. Increasing the TopIB concentration even further resulted in larger DNA-TopIB aggregates which differ a lot in size, ranging from small aggregates containing only 2 DNA molecules to very large aggregates containing over 60 DNA molecules (Fig. 4.4c,d). The size of the aggregates and the distribution over the surface was very inhomogeneous. For example, individual DNA molecules can be found in one place, while an aggregate containing 10 DNA molecules can be can be found within 4 µm of the individual molecules. Note also that the TopIB concentration is the same in figures 4.4c and 4.4d.

To quantify the aggregation of DNA molecules by TopIB the fraction of individual DNA molecules in a sample was measured for each TopIB concentration. Approximately 1300 DNA molecules were analyzed and the results are shown in figure 4.5. The data shows a clear trend and it appears that the fraction of individual DNA molecules de-

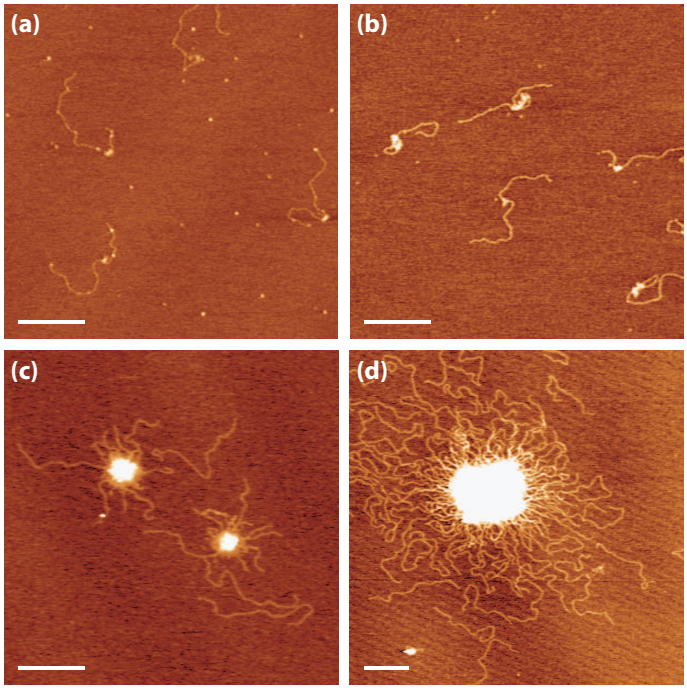

Figure 4.4: AFM images taken in air of TopIB bound to linear DNA at several TopIB concentrations, [DNA]=0.22 nM. Experiments were done in PB (10 mM), but using Tris·HCl (10 mM) as buffer yielded similar results. Scale bar size is 400 nm. **a)** [TopIB]=4 nM, individual TopIB are bound to DNA. **b)** [TopIB]=7 nM, intramolecular TopIB-DNA nodes are formed, cooperativity is also observed. **c)** [TopIB]=45 nM, DNA molecules are aggregated by TopIB molecules. The distribution of the DNA molecules over the aggregates is very inhomogeneous. **d)** [TopIB]=45 nM, a very large aggregate containing tens of DNA molecules.

creases almost linearly with increasing TopIB concentration and reaches a plateau where 20% of the DNA molecules is still not part of an aggregate. The error bars are determined in the following way. The data points in the plot are percentages and DNA molecules have only two options. They can be in an aggregate (probability p) or not (probability $1-p$). The error in a point can then be calculated using

$$error = \sqrt{\frac{p(1-p)}{N}} \tag{4.2}$$

in which N is the number of DNA molecules analyzed for the data point.

In order to study the influence of divalent (cat)ions on the aggregation process, several experimental parameters were varied. Vaccinia Topoisomerase IB does not need Mg^{2+} to relax DNA [14]. Still it could be possible that Mg^{2+} is crucial for the aggregation process. Therefore $MgCl_2$ was replaced by $MnCl_2$ or $CaCl_2$ during the binding reaction. Both Mn^{2+} and Ca^{2+} do not inhibit DNA relaxation by TopIB [14]. Experiments showed that DNA molecules can be deposited onto a mica surface using either $MnCl_2$ or $CaCl_2$

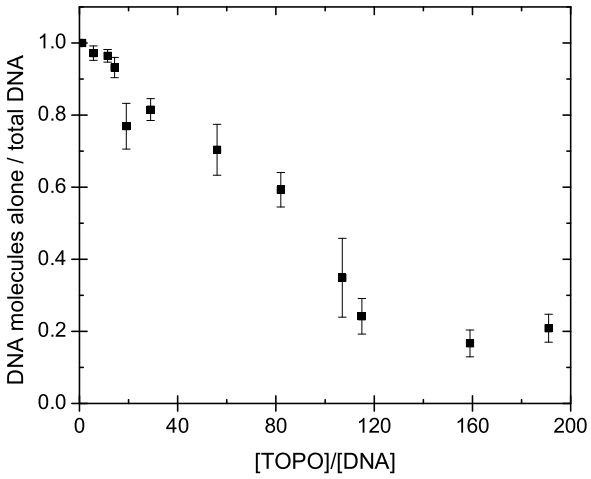

Figure 4.5: Fraction of DNA molecules that are not part of an aggregate plotted versus the TopIB-to-DNA ratio. Over 1300 DNA molecules were analyzed ([DNA]=0.22 nM).

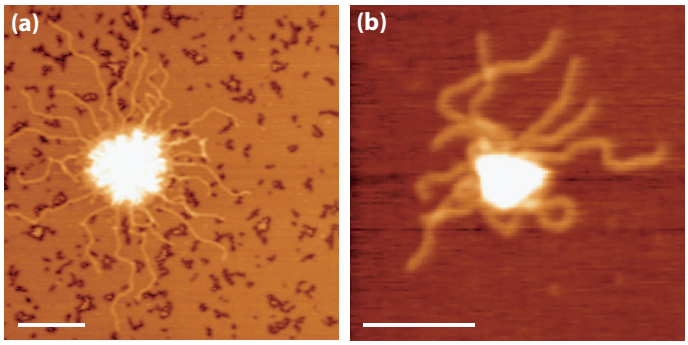

Figure 4.6: AFM images taken in air of TopIB-DNA aggregates formed with other divalent ions present during the binding reaction. Concentrations of TopIB and DNA were similar as in previous experiments ([TopIB]=13 nM, [DNA]=0.22 nM). Scale bar size is 200 nm. **a)** Molecules deposited using 13 mM $MnCl_2$ in the binding reaction. **b)** Molecules deposited using 13 mM $CaCl_2$ in the binding reaction.

instead of $MgCl_2$ (images not shown here). Figure 4.6 shows the effect that other divalent ions have on the aggregation process. Both images show the presence of aggregates and this confirmed that Mg^{2+} is not crucial to this process but divalent ions in general.

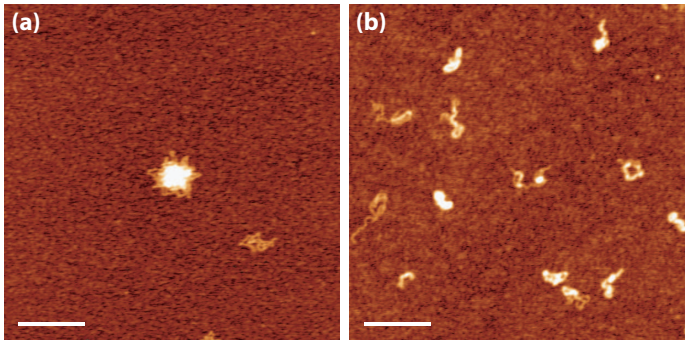

Figure 4.7: AFM images of DNA-TopIB interaction taken in air. Molecules were deposited using 0.01% PLL and therefore they are kinetically trapped onto the surface. [DNA]=0.22 nM, [TopIB]=27 nM. Scale bar size is 400 nm. **a)** Adding $MgCl_2$ (13 nM) to the binding reaction resulted in the formation of aggregates similar to earlier experiments. **b)** No divalent ions in the binding reaction resulted in DNA molecules that appear compacted by TopIB. No aggregates were found after analyzing 100 DNA molecules. The observed behaviour is clearly cooperative.

4.3 Control experiments

Control experiments were done to investigate the effect of the ions in the solution during the binding reaction and to confirm that both TopIB and DNA do not aggregate by themselves. The results of these experiments are described in the following subsections, while for DNA this was already shown in section 3.1..

4.3.1 In the absence of divalent ions no aggregates are found

The effect of divalent ions was investigated by removing them from the binding reaction. Subsection 3.3.2 describes a method to deposit DNA molecules onto a mica surface without the use of divalent ions. By using this method the effect of divalent ions on the aggregation process can be determined. DNA molecules have a different appearance when polylysine is used to deposit them (Fig. 3.2). This presents no problem for the observation of DNA-TopIB aggregates. Experiments were done at high TopIB concentration ([TopIB]=27 nM) and linear DNA was used to promote the formation of aggregates. Two experiments, one containing 13 mM $MgCl_2$ in the binding reaction, the other in the absence of $MgCl_2$, were done depositing the molecules on a PLL-coated surface. Experimental conditions were identical. In the presence of $MgCl_2$ in the binding reaction aggregates were also found when the sample was deposited onto PLL-coated mica (Fig. 4.7b). This rules out the possibility that the aggregates were an artifact of the deposition method. The difference in appearance of the aggregate can be attributed to the deposition method. It is more difficult to distinguish DNA coming out of the central part of the aggregate. In the absence of divalent ions in the binding reaction the result was dramatically different (Fig. 4.7a). No aggregates were observed, instead individual DNA molecules were found with TopIB associated to them. Part of the DNA was com-

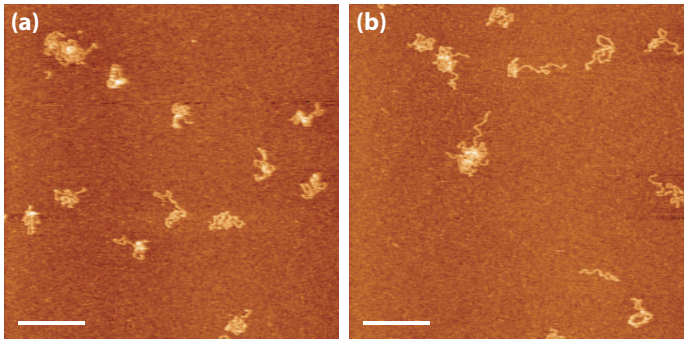

Figure 4.8: AFM images taken in air of TopIB-DNA interaction ([DNA]=0.22 nM, [TopIB]=22 nM). 25 mM NaCl was present during the binding reaction. No large aggregates were found, while in a similar sample with 13 mM $MgCl_2$ instead of NaCl more than 60% of the DNA was part of aggregates. Scale bar size is 400 nm.

pacted and showed the formation of filaments. Some DNA molecules seemed unaffected by TopIB while others were completely compacted and covered by the enzyme. This again strongly indicates that the binding is cooperative.

The effect of monovalent ions on the aggregation was also investigated. At high TopIB concentrations for which $MgCl_2$ experiments showed lots of aggregates, NaCl (25 mM or 50 mM) was added and the sample was deposited using PLL-coated mica. Results are shown in figure 4.8. After analyzing over 100 DNA molecules no large aggregates were found. Occassionaly two DNA molecules were found to be linked together by TopIB molecules of which an example can be seen in figure 4.8b. Besides that, the results were very similar to the results obtained without adding any monovalent or divalent ions. Although in the latter case no ions are added there are still ions present. Both the phosphate buffer and Tris·HCl (which is there because DNA is diluted in it) are a source of positive monovalent ions and therefore there are always some 'background' positive ions with a concentration of 5-15 mM. Possible explanations for the aggregation of DNA molecules by TopIB will be discussed in section 6.4.

4.3.2 In the absence of DNA no TopIB clusters are found

Bulk methods showed that vaccinia topoisomerase IB is a monomer in solution [13]. To confirm that TopIB was also a monomer under the conditions of the binding reaction AFM images of TopIB were taken in liquid. The imaging buffer was similar to the buffer of the binding reaction and the TopIB concentration was quite high (27 nM). Figure 4.9a shows a typical AFM image of TopIB molecules on a mica surface. The height distribution is shown in figure 4.9b. A sharp peak was found and fitting a Gaussian resulted in an average height of 3.6±1.8 nm which agrees nicely with the height you would expect for a circular protein of 36 kDa. This also confirms that TopIB is monomer and not a dimer in solution. A dimer would have resulted in a broader peak in the height

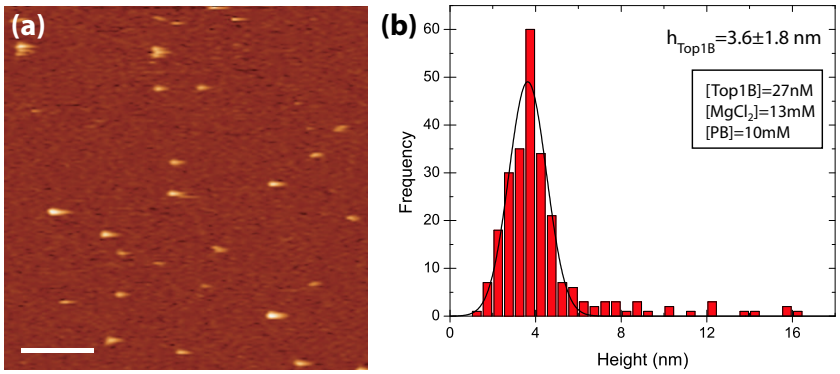

Figure 4.9: TopIB is a monomer in solution. **a)** AFM image of TopIB molecules imaged in a liquid environment. Individual dots can be seen which represent individual TopIB-enzymes. Scale bar size is 150 nm, [TopIB]=27 nM. **b)** Analysis of 246 TopIB molecules is plotted in a histogram. The data is sharply peaked around 3.6 nm, indicating that TopIB is indeed a monomer.

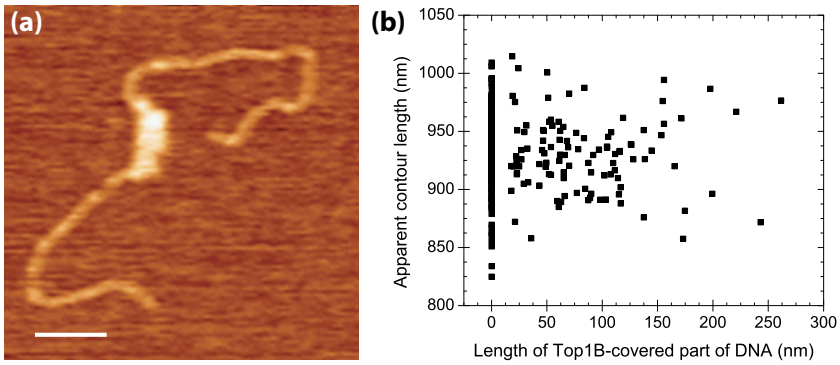

Figure 4.10: a) Typical AFM image of TopIB bound to linear dsDNA. Scale bar size is 100 nm, [TopIB]=6.7 nM, [DNA]=0.22 nM. **b)** Apparent contour length of DNA measured for 108 DNA molecules with TopIB bound. The large number of points (172) at a coverage of 0 nm indicates the spread in apparent contour length for clean linear dsDNA.

distribution because the height of the dimer will depend more on its orientation on the surface. Results were confirmed in air where no TopIB aggregates were found in the absence of DNA.

4.3.3 Bound TopIB does not change the contour length of DNA

Proteins can change the contour length of DNA. For example, the RecA protein elongates a DNA molecule by 50% when it is completely covered [35]. To determine if this is also the case for the binding of individual TopIB molecules to DNA, the apparent contour length of DNA molecules was measured. This was done at low TopIB concentrations to prevent aggregation. No significant change in contour length was detected (Fig.

4.10b). The measured contour of clean linear dsDNA always has a Gaussian distribution. The measured contour length of DNA with TopIB bound always stayed within this distribution for all DNA molecules.

5. Dynamics of TopIB-DNA interaction

5.1 Characterization of TopIB-DNA interaction

The dynamics of the interaction between TopIB and DNA was studied with the AFM in a liquid environment. A sample containing circular nicked DNA and TopIB was prepared and imaged in a buffer (3 mM $MgCl_2$, 75 mM KCl and 10 mM Tris·HCl, pH 8.0). Figure 5.1 shows four consecutive images taken at an image rate of 35 $^{s}/_{frame}$ extracted from a movie (total duration=21 min). In this sequence a TopIB molecule binds to the DNA (1) and remains bound for two frames (2,3). After this, it again unbinds from the DNA (4). The direction of the TopIB molecule movement was not influenced by the tip of the AFM which was scanning from left to right. The TopIB-molecule moved in the opposite direction. The images show almost no TopIB on the surface because the sample was diluted 160-170 times because of the addition of 1 ml imaging buffer, which resulted in a sub-nanomolar TopIB concentration. To increase the probability of TopIB-DNA interaction, TopIB was added to the imaging buffer. Images showed more TopIB on the surface (Fig. 5.2). In this sequence (total duration=62 min), TopIB was seen to move along the surface and also leave and enter the imaged surface area.

Filament-like structures that were observed in air experiments were also reproduced in liquid (Fig. 5.3). The imaging buffer contained 13 mM $MgCl_2$, no KCl and 10 mM Tris·HCl. Compared with the experiment shown above the imaging buffer contained 4 times as much $MgCl_2$, which was necessary to fix the filaments to the surface. When

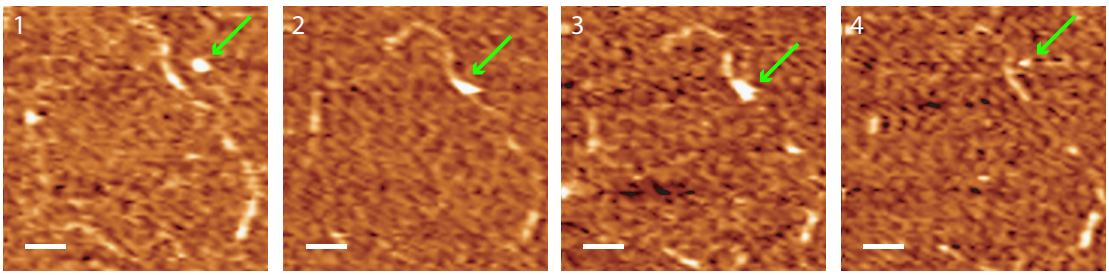

Figure 5.1: Four consecutive frames showing a TopIB molecule (indicated by the green arrow) binding to DNA on a mica surface. The molecule is bound on frames 2 and 3. Frame rate = 36 $^{s}/_{frame}$, no TopIB in imaging buffer. Conditions during binding reaction: [DNA]=0.22 nM, [TopIB]=13 nM, [$MgCl_2$]=13 mM, scale bar size is 50 nm.

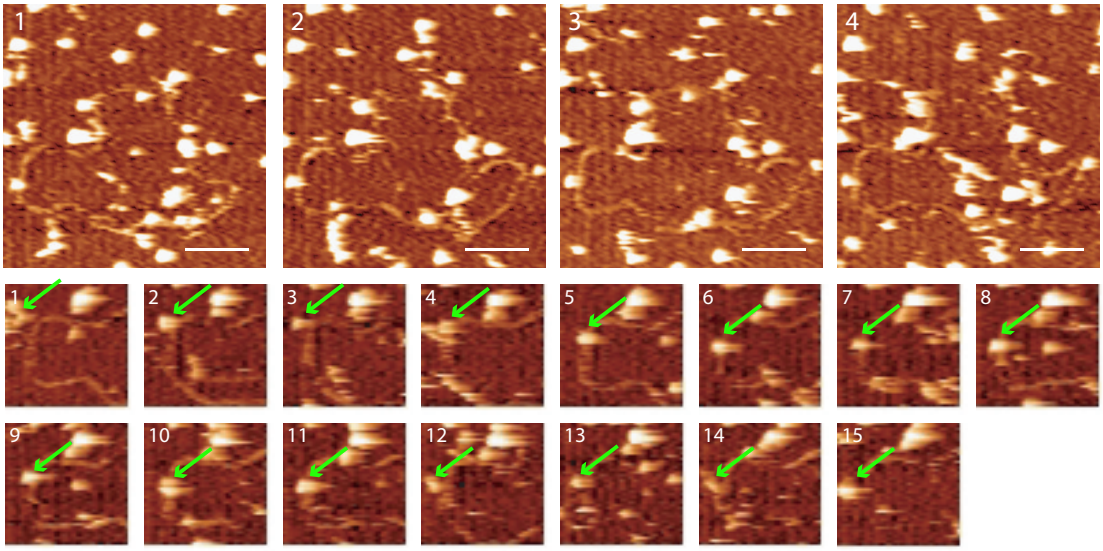

Figure 5.2: Four consecutive frames showing TopIB activity. Scale bar size is 100 nm. The 15 small frames show a TopIB molecule binding to DNA, moving along the DNA and unbinding after several frames. The imaging buffer contained 5 nM TopIB, frame rate=33 s/frame.

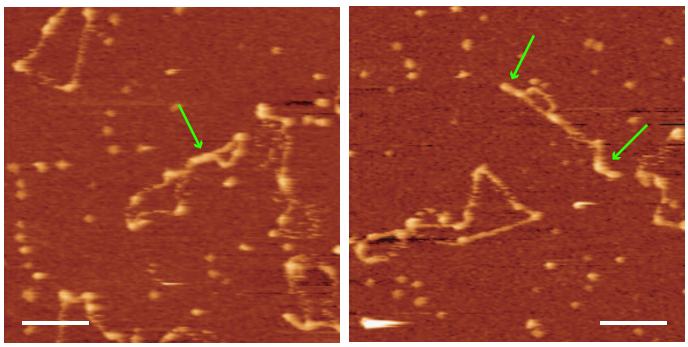

Figure 5.3: AFM images in liquid showing filaments (indicated by green arrows) of TopIB on DNA molecules. Imaging buffer contained 13 mM $MgCl_2$, scale bar size is 200 nm.

a similar experiment with 3 mM $MgCl_2$ and 75 mM KCl in the imaging buffer was performed, no filaments were found but DNA was still visible along with TopIB from the imaging buffer. While this experiment confirmed the presence of TopIB filaments on DNA it was impossible to see any dynamics because the TopIB molecules (and therefore also the filaments) were well fixed to the surface due to the high concentration of Mg^{2+}-ions.

5.2 Lifetime of TopIB bound to DNA

To estimate the off-rate of the binding of TopIB to DNA we measured the lifetime of bound TopIB to DNA. The lifetime is defined as the time that a TopIB is bound to DNA and it can be measured as follows. We counted for how many frames a TopIB was bound to DNA for a complete movie. The data was plotted in a histogram showing how many TopIB molecules were still bound after n frames. For example the TopIB molecule in figure 5.1 was bound for two frames and would therefore be counted in both bin 1 and bin 2 (using a bin size of 1 frame). The binding of TopIB to DNA can be described by the following reaction:

$$\text{DNA} + \text{TopIB} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{DNA:TopIB} \tag{5.1}$$

The change in number of bound proteins (N_{bound}) is proportional to the number of bound proteins:

$$\frac{dN_{bound}}{dt} = -k_{off} \cdot N_{bound} \tag{5.2}$$

Solving the differential equation results in

$$N_{bound} = N_0 \cdot e^{-k_{off}t} \tag{5.3}$$

Thus, an exponential decay is expected for the lifetime of the DNA:TopIB complex. The lifetime was measured at three different frame rates (Fig. 5.4). Fitting of equation 5.3 resulted in three different values for the off-rate k_{off} (table 5.1). For a frame rate

Frame rate (s)	# of events	k_{off} (frame^{-1})	k_{off} (s^{-1})
9	73	1.77±0.05	15.9
20	111	1.71±0.08	34.2
33	206	1.27±0.04	41.9

Table 5.1: Calculated off-rate for TopIB-DNA interaction at three different frame rates.

of 20 s/frame (green) and 9 s/frame (blue) a TopIB protein remains bound on average for 1.7 frames. Both data sets were taken from the same sample. For a frame rate of 33 s/frame we measured k_{off}=1.27frame^{-1}. Unfortunately, the measured values for k_{off} were approximately the same for the three AFM frame rates. This suggests that the measured k_{off} is a measurement of the chance that the AFM will kick a bound TopIB off the DNA. This also explains why the off-rate is lower for the data set taken at 33 s/frame. This data set was taken from another sample at different imaging conditions[1] and the force applied on the sample was apparently larger. The higher the force that the cantilever applies on the TopIB, the lower the value for k_{off} will be. Although the 'real'

[1] Conditions such as cantilever amplitude, force applied on the surface and feedback parameters can differ from cantilever to cantilever.

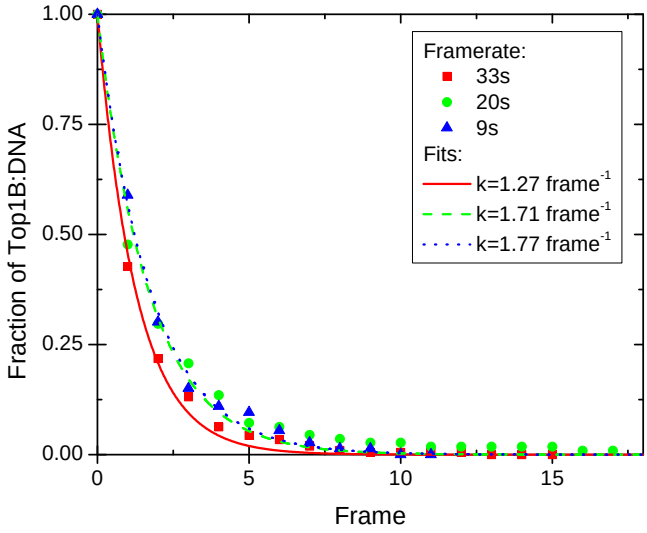

Figure 5.4: Lifetime of TopIB bound to DNA measured at three different frame rates. The fit of an exponential decay shows that the rate constant is not proportional to the frame rate. Data for 20 s and 9 s frame rate was taken from the same sample.

off-rate could not be determined in this experiment a lower limit could be estimated. The data shows that for the three frame rates the measured k_{off} is still influenced by the AFM and therefore the actual k_{off} must be higher than 41.9 s^{-1}, lying more in the minutes range than in the seconds range. Fluorescence spectroscopy studies done by Kwon et al. showed an off rate of 3.0-5.0 s^{-1} for the TopIB Y274F mutant binding non-covalently to short DNA oligos containing the preferred pentamer sequence [36].

5.3 1D-diffusion of TopIB on DNA

TopIB showed 1D-diffusive behaviour along a DNA molecule when imaged in buffer. Sekiguchi and Shuman already suggested in 1994 that TopIB might bind non-specifically to DNA and use a diffusion mechanism to locate the recognition site [11]. To date this has never been visualized in any way. Our previous data (Fig. 5.2) already suggested that TopIB might be able to diffuse on DNA after binding to it.

In another series of experiments (with a frame rate of 20 s/frame, taken from a 42 min movie) two TopIB molecules were analyzed. Both molecules were bound for nine consecutive frames and in their proximity a suitable reference point on the DNA was found (a part of the DNA that did not move during these nine frames). The TopIB molecules were found to be moving in both directions on the DNA and also opposite to the

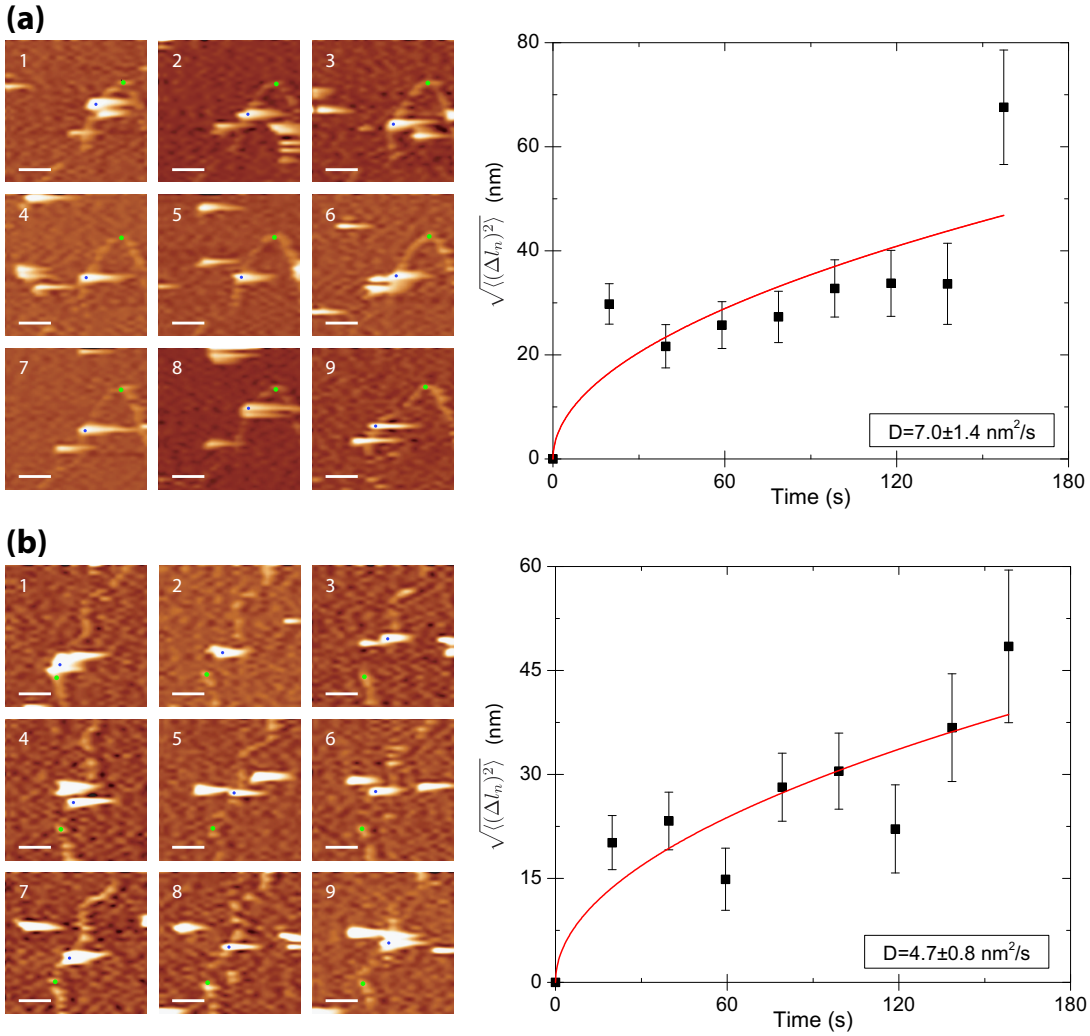

Figure 5.5: Two series of 9 consecutive AFM images showing a TopIB molecule (indicated by a blue dot) bound to DNA and diffusing along the DNA. The fixed reference point is indicated in each frame by a green dot. The 2 molecules have a diffusion constant of **a)** $7.0\pm1.4\,\mathrm{nm^2 s^{-1}}$ and **b)** $4.7\pm0.8\,\mathrm{nm^2 s^{-1}}$. Frame rate = $20\,\mathrm{s/frame}$, scale bar size is 50 nm.

AFM scanning direction (which is always from up to down and from left to right). This means that the movement of the TopIB was not induced by the interaction with the AFM tip, which is an important control. To determine if the TopIB movement can be described by a diffusive behaviour it was necessary to analyze the position of the DNA and the TopIB in the movies more quantitatively. Using the theory presented in section 2.2 a 1D-diffusion constant (D) was found for both molecules. For the first molecule D=$7.0\pm1.4\,\mathrm{nm^2 s^{-1}}$ and for the second D=$4.7\pm0.8\,\mathrm{nm^2 s^{-1}}$. These values are in agreement (order of magnitude) with similar experiments using RNA polymerase [29]. The error

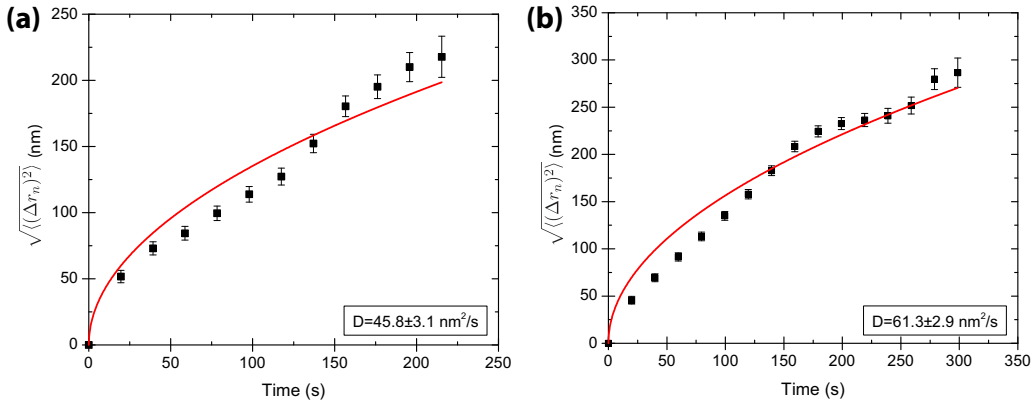

Figure 5.6: 2D-diffusion of two TopIB molecules moving over a mica surface not bound to DNA. The values of D obtained by fitting of equation 2.20 to the data results in higher values of D than for DNA-bound TopIB.

bars in the diffusion graphs were determined in the following manner. The position of the TopIB has an error of 1 pixel (5.5 nm). For the calculation of $\Delta l = l_i - l_{i+n}$ along the contour of the DNA two independent measurements are needed: the position of a TopIB and the position of the reference point. The error in Δl is therefore the sum of the individual errors in l_i and l_{i+n} resulting in ± 11 nm. Because all the $\Delta l = l_i - l_{i+n}$ are averaged the error can be calculated by $\frac{11}{\sqrt{N-n}}$.

Besides bound molecules also several TopIB molecules were analyzed which were not bound to DNA, but which were moving on the mica surface. These molecules were imaged under the same experimental conditions as the molecules shown in figure 5.5. The 2D-diffusion of the two molecules is plotted in figure 5.6. Fitting the 2D-diffusion equation (eq. 2.20) yielded D=45.8±3.1 nm^{2}s^{-1} for one molecule and D=61.3±2.9 nm^{2}s^{-1} for another. These values are clearly larger than the 1D-diffusion constants obtained for bound TopIB. This indicates that good imaging conditions were employed where TopIB can freely move over the mica surface which means that the interaction of TopIB with the surface was small. Therefore the binding of TopIB to DNA was specific. Our results support the idea that TopIB can diffuse along the DNA.

Another interesting set of images was made at a different frame rate and is an example of the difficulties that arose during analysis (Fig. 5.7, taken from a 31 min movie). In this movie a TopIB molecule was bound to the DNA in many frames (Fig. 5.7a). In figure 5.7b a TopIB molecule is bound to DNA and in the 7th frame it transfers from one DNA duplex to the other. Two problems arose during the analysis of these frames. As was shown in section 5.2 there were not many TopIB molecules that stayed bound to DNA for a long time (>10 frames). When looking at the molecules that were bound for a long time a second problem arises, namely the fact that it was difficult to find a good

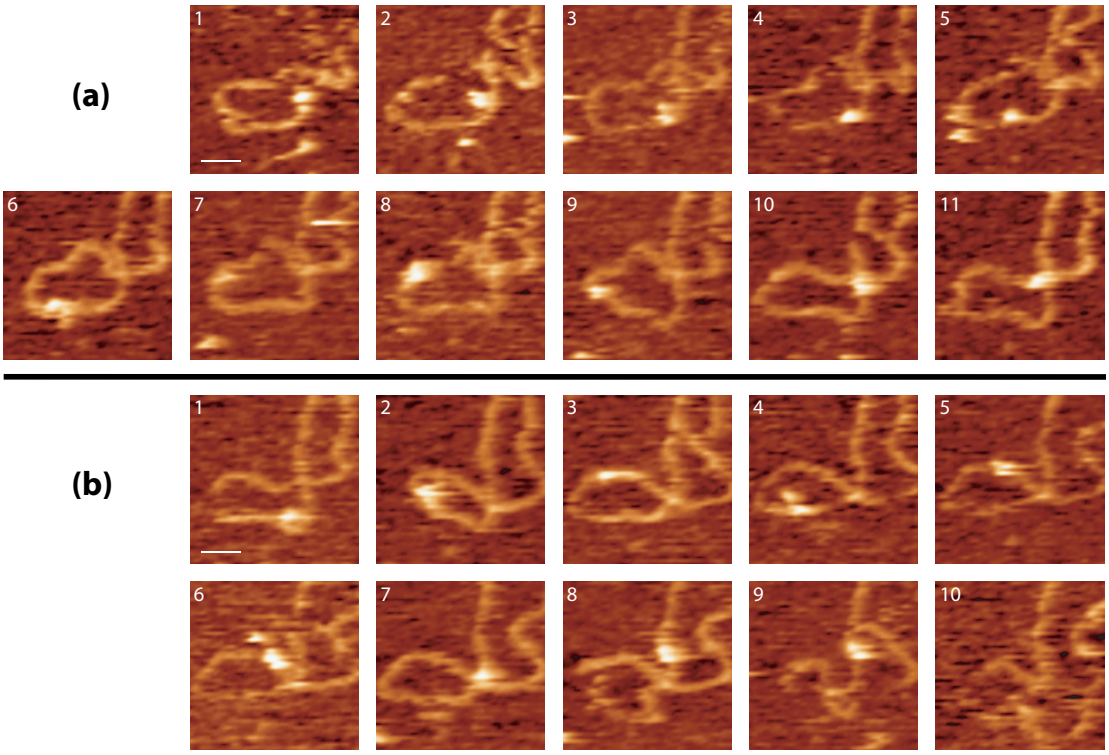

Figure 5.7: a) 11 consecutive frames which show a TopIB molecule interacting with a DNA molecule. In 10 frames the molecule is bound to DNA. **b)** 10 consecutive frames showing a TopIB molecule that binds to a DNA duplex and transfers to another duplex of the same molecule. Scale bar size is 50 nm.

reference point on the DNA which fixed during the whole movie. As a result there were very few TopIB molecules that could be analyzed and unfortunately it was not possible to analyze the images from figure 5.7 unambiguously.

During these experiments the obtained frame rates were close to the maximum possible frame rate available with this AFM. Increasing the scanning speed was possible but this resulted in images with poor quality in which DNA and single TopIB molecules could not be distinguished anymore.

6. Discussion

6.1 Cooperativity

6.1.1 Model for cooperative binding

We propose a model that describes and quantifies the cooperative binding of TopIB to DNA (Fig. 6.1). The formation of filaments of TopIB on DNA requires several steps. The first step in this process is the non-covalent binding of TopIB to DNA at a random position (1). This step has an association constant K_{DNA}. This bound TopIB can then diffuse linearly along the DNA. In the meantime another TopIB can bind non-covalently to DNA at a random position (2), resulting in two bound TopIB proteins that can diffuse along the DNA. Due to brownian motion of the DNA molecule the two bound TopIB proteins have a certain probability to approach each other and interact. The result can be an intramolecular node on the DNA formed by a TopIB dimer (3). The formation of a TopIB dimer from two TopIB proteins that are bound to DNA has a association constant K_{dimer}. The process in which an empty DNA molecule (1) transforms into a DNA molecule with a TopIB dimer bound (3) has an overall association constant K which depends on the two binding constants K_{DNA} and K_{dimer}. More TopIB proteins can bind in the meantime and extend the existing intramolecular node (4). The extension of the node by one dimer (there are now two consecutive dimers bound to the DNA) has a association constant of $K\omega$. Depending on the concentration of TopIB the filament will extend even more (5). It is also possible that another intramolecular node is formed on the DNA molecule. This can result in complex structures which are difficult to analyze

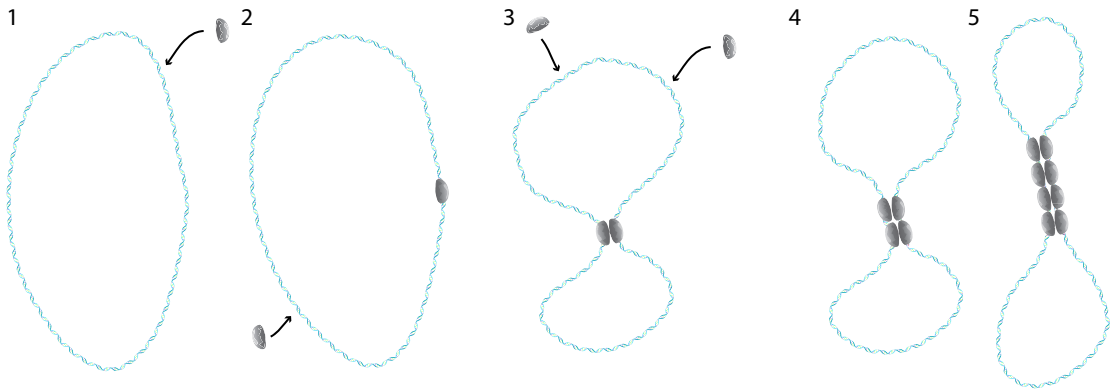

Figure 6.1: Schematic view of the proposed model for cooperative binding of TopIB to DNA.

(for example figure 4.1g). In the presence of divalent ions the formation of a node can also happen between two DNA molecules. This is called an intermolecular node instead of an intramolecular node. The formation of intermolecular nodes results in aggregation of DNA molecules by TopIB, discussed extensively in subsection 6.2. The process in which DNA and TopIB transform from state 3 to state 4 (Fig. 6.1) is still unknown, but several possibilities exist. Bound TopIB might diffuse individually along the DNA molecule until they 'hit' the existing intramolecular node, stay there and form a dimer with another TopIB that is also diffusing along the DNA. Another possibility could be that first a TopIB dimer is formed (another intramolecular node) which diffuses until it binds to the other intramolecular node.

Analysis of the filament formation was done using the cooperativity theory presented in section 2.1. This theory applies for filaments that are formed on dsDNA and it was assumed that this theory could be extended for the case of type IB topoisomerase binding to DNA. In the theory the binding of a single molecule to DNA is the first step in filament formation. In the TopIB case the formation of a TopIB dimer on DNA is the equivalent step. As was shown before TopIB is a monomer in solution (subsection 4.3.2). If TopIB were a dimer the formation of one intramolecular TopIB-DNA node would automatically increase the probability of another dimer binding next to it. This is because the first dimer has brought the two DNA duplexes close to each other and the next dimer does not need to do this anymore and cooperativity would be an intrinsic property of the binding of dimers. Our analysis method however cannot distinguish this process (TopIB dimers binding to DNA) from the process described in our model (TopIB monomers binding to DNA and subsequently forming a dimer).

If the binding of TopIB to DNA were not cooperative an increasing number of individual nodes would be found instead of long filaments. Nevertheless, also for the non-cooperative case it is still easier for two TopIB molecules to form a node next to an already existing node because simply the two parts of the DNA are already closer to each other. As is shown in figure 4.3 our data shows that the coverage of DNA by TopIB also follows cooperative behaviour and for a non-cooperative process this curve would be completely different. This proves that the formation of filaments actually is a cooperative process.

However, we showed that DNA gets aggregated (because of the formation of more intramolecular nodes and intermolecular nodes) at higher TopIB concentrations and determination of the fractional DNA saturation θ becomes more and more difficult. The fractional DNA coverage θ has been measured for several concentrations of TopIB. However, as can be seen in figure 4.3, only data points below $\theta = 0.25$ have been measured. The only method to determine θ at high [TopIB] is to measure the contour length of the DNA that is not covered with TopIB. We assumed that within the aggregate the DNA was fully covered with TopIB. There is no way to check this assumption because the inside of an aggregate is inaccessible with AFM. Therefore this assumption is probably an overestimation of the coverage of the DNA. Consequently the values of θ

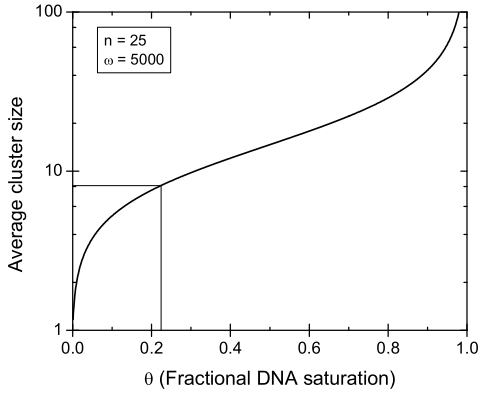

Figure 6.2: Average filament length of bound TopIB dimers plotted versus fractional DNA saturation with ω and n close to experimentally observed values. The horizontal and vertical lines point to the highest observed value for θ in our experiments.

in figure 4.3 are probably slightly overestimated, especially the data points with a high percentage of molecules with aggregates.

6.1.2 Validity of the model

In the theory presented in 2.1 on the cooperativity of the binding of proteins to DNA the DNA was considered to be infinitely long. In our experiments DNA had a length of 2743 bp and the maximum number of TopIB dimers that can bind to one DNA molecule in a filamentous structure is 54 using a footprint of $n = 25$ [11]. Obviously, the theory breaks down when measured filament lengths approach this number. In figure 6.2 the average filament length using equation 2.13 is plotted and it shows that for $\omega \approx 5000$ and $\theta \approx 0.23$ (which are similar to values found experimentally) the average filament length is below 10 proteins. Figure 4.2 showed that even the largest filament found is still well below the maximum possible filament length for one DNA molecule. Therefore in this research the assumption of DNA to be infinitely long is valid.

6.1.3 Dynamics of filament formation

Shuman et al. showed with EM that type IB topoisomerases can zip up DNA molecules [7]. However preparation of samples for EM imaging requires fixation using glutaraldehyde and it is known that this treatment can introduce artifacts and prevents dissociation of proteins from the DNA [23]. In our AFM analysis no fixation is needed and obtained images are a 2D snapshot of the 3D molecules in solution. Molecules imaged with AFM resemble their natural state better than EM-imaged molecules.

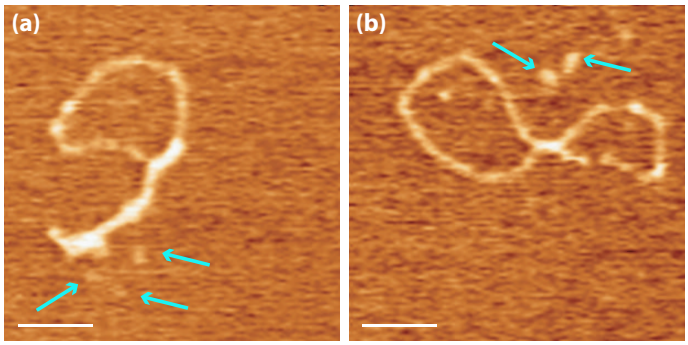

Figure 6.3: Two AFM images taken in air showing TopIB filaments on a DNA molecule. The blue arrows indicate TopIB molecules that presumably were part of the filament but were removed during the rinsing and drying process. Scale bar size is 100 nm.

The results from AFM imaging shown in this research suggest that the filament formation is a dynamic process. When an imaging buffer with low concentrations of Mg^{2+}-ions and also K^{+}-ions was used no filaments were observed. Probably the filaments were present just after deposition of the DNA on the surface, but they are probably disappearing in the time (several minutes) between sample deposition and AFM imaging. This can be explained as follows. At low ion concentrations in the imaging buffer, TopIB molecules can bind and unbind from the DNA while loosely attached to the surface. When a TopIB molecule unbinds from an intramolecular node in DNA this means that these two parts of DNA are not held together anymore. If this happens these parts of the DNA regain there ability to move independent from each other. Because the DNA is loosely bound to the surface it can equilibrate and the two DNA parts that were initially connected through two TopIB molecules will move away from each other through diffusion. Due to the presence of the surface the DNA cannot move as freely as in solution and it is unlikely that the two DNA parts will come close to each other again. It is even more unlikely that at that moment there will be two TopIB molecules bound on these parts, able to form a new TopIB dimer on DNA. After a while all filaments will have disappeared in this manner and this will result in DNA molecules without any nodes. This is also supported by images taken in air (Fig. 6.3) which show TopIB-DNA filaments. Images also showed TopIB molecules lying very close to a filament or an aggregate. This suggests that TopIB is only weakly bound to the DNA when present in a filament and that it can bind and unbind constantly. This also supports the idea that one of the functions of TopIB is to compact the DNA. By forming these filaments the volume (radius of gyration) that the DNA occupies in solution is reduced. By having the TopIB only weakly bound in these filaments it is easy to remove the filaments and 'open up' the DNA when required.

6.1.4 Height of filaments

It can be argued that the images of DNA molecules with TopIB filaments associated to them are actually images of supercoiled DNA. For example the height of some of the filaments is approximately twice the height of a single dsDNA molecule. There are two reasons why the observed DNA cannot be supercoiled. First, the plasmid DNA used in the experiments has two nicks. So if TopIB would induce supercoils the DNA will immediately relax because of the nicks. Nonetheless it is still possible that two TopIB proteins form a intramolecular node on the DNA, creating two DNA loops. One of these loops could not include a nick and supercoils might be induced in this area. It remains unlikely that supercoils were present in the DNA because supercoils will also be relaxed by the topoisomerase present in the solution. Considering the incubation time (15 minutes) and the relatively high TopIB concentration every supercoil that might be created will be relaxed ultimately.

The height of the filaments in air is lower (it is only 1-2 nm) than what is expected from the TopIB height measurements done in liquid (section 4.3.2) especially because a filament should consist of two dsDNA held together by two TopIB monomers. Height measurements of TopIB in air showed an average height of 3.2±1.3 Å which is only 10% of the measured height in liquid. The reduced height is a combination of several effects [37]. The same height reduction is found for DNA molecules imaged in air. It has been suggested that the reduced height is due to the presence of a salt layer in which DNA molecules are embedded. This layer introduces an offset in the measured height of any molecule on the surface and as a consequence the measured height is smaller than the actual height. However, the effect of this layer (which is only several angstroms thick) is not enough to account for the observed height difference. Another cause of the reduced height could be drying artifacts. It is known that vaccinia topoisomerase IB consists of two parts that are connected by a hinge [5]. After drying the protein can adopt a conformation completely different from the natural form. Finally, it is possible that the hinge breaks during drying and that topoisomerase is split into two smaller parts. Also the TopIB might be deformed by the interaction with the AFM tip reducing the apparent height of the protein.

6.2 Aggregation of DNA by TopIB

6.2.1 Aggregation is an extension of cooperative binding

The observed aggregation of DNA molecules by TopIB can be explained with the same process that causes the formation of filaments in circular DNA. For circular DNA intramolecular TopIB interaction is favored, for linear DNA intermolecular TopIB interaction happens more often. Because of the shape of circular DNA (it has a smaller radius of gyration) it will interact more often with itself than linear DNA. The formation of an intermolecular node results in a small aggregate (of two DNA molecules). Therefore the formation of larger aggregates is also expected in samples with circular DNA and

higher TopIB concentrations as was shown (figure 4.4c). The unexpected aggregation behaviour observed at higher TopIB concentration using $MgCl_2$ cannot be explained as an artifact due to deposition on mica but as a consequence of the cooperative behaviour of TopIB and the formation of filaments. The reasons for this are as follows. The aggregation was never observed when only DNA or TopIB was imaged. To investigate the effect of drying the sample an experiment in liquid was performed. Using similar conditions in the binding reaction (although using Tris·HCl instead of PB) the sample was deposited on mica and rinsed with water. Without drying, the imaging buffer was added and AFM images showed the same aggregation behaviour as observed in the air experiments. Also the possibility that the phosphate buffer caused the aggregation was ruled out in this experiment because Tris·HCl was used.

It was shown that the use of $MgCl_2$ promotes aggregation of DNA molecules with TopIB. Unfortunately there is not enough experimental data to develop a model that describes and explains the role of charges in the aggregation behaviour in a detailed manner. However, it seems that when divalent ions are present in solution the electrostatic repulsion between DNA molecules is reduced. There are two effects that can possibly explain this reduction: the reduction of the *Debye length* and the screening of the surface charge. The *Debye length* (λ_D) is a measure for the length over which electrostatic interactions between molecules play a role [38]. It is defined as

$$\lambda_D = \left(\sum_i \frac{\rho_i^\infty e^2 z_i^2}{\epsilon \epsilon_0 k T} \right)^{-\frac{1}{2}} \tag{6.1}$$

in which the sum is over all different ions (numbered by i) in solution, ρ_i^∞ is the concentration of ions in the bulk, e the charge of an electron, z the charge of ion i, ϵ is the dielectric constant, ϵ_0 the permittivity of free space, k the Boltzmann constant and T the temperature. Equation 6.1 shows that the presence of ions in a solution reduces the Debye length and as a consequence electrostatic repulsion between molecules with the same charge has a shorter range. Putting in numbers for different ions into equation 6.1 results in

$$\lambda_D = \begin{cases} \frac{0.304}{\sqrt{[\mathrm{NaCl}]}}\mathrm{nm} & \text{for a 1:1 electrolyte} \\ \frac{0.176}{\sqrt{[\mathrm{MgCl_2}]}}\mathrm{nm} & \text{for a 2:1 or 1:2 electrolyte} \end{cases} \tag{6.2}$$

which means that increasing the concentration of ions in solution decreases λ_D. Divalent ions reduce the λ_D more than monovalent ions. For example λ_D for $[MgCl_2]$=13 mM is equivalent to λ_D for [NaCl]=39 mM. In subsection 4.3.1 it was shown that adding an equivalent amount of monovalent ions to the solution instead of divalent cations increased the formation of aggregates a little bit. Also the aggregates found were definitely much smaller and there were only very few of them found. So the reduction of the Debye length by itself cannot explain by itself the aggregation in the presence of divalent cations, although it shows that electrostatics plays a role in the process.

The presence of divalent cations in a solution has a very dramatic effect on the reduction of the surface potential (ψ_0) of molecules or surfaces and thereby reducing the

repulsion between two molecules (or surfaces) with the same charge [38]. Even at concentrations of a few mM the concentration of divalent cations at the surface will be very high (for example, for [$CaCl_2$]=3 mM and [NaCl]=100 mM the concentration of Ca^{2+} at a negatively charged surface (surface charge density=-0.2 Cm^{-2}) is around 7 M) and the divalent ions can sometimes even bind to the negatively charged surface [38]. When adding 13 mM of $MgCl_2$ to a solution which contains 10 mM NaCl the surface potential will be 16 times reduced. To get the same reduction using only NaCl the concentration NaCl needs to be increased to 56 M, which is of course impossible. From this it can be seen that the surface charge of a DNA is changed dramatically in the presence of divalent cations which leads to less repulsion between DNA molecules, consequently increasing the probability of having two DNA molecules in close proximity. This can explain the observed aggregation. In the presence of divalent cations it is not only possible for TopIB to form intramolecular nodes on DNA but also to form intermolecular nodes which manifest themselves as aggregates. The underlying process for aggregation is actually the same process that forms intramolecular filaments in circular DNA.

6.2.2 Analysis of aggregate size

The analysis of the number of DNA molecules in an aggregate brings some difficulties. At low TopIB concentrations analysis of the data presents no problems and imaging and analyzing hundred or more DNA molecules is straightforward. But when more aggregates are formed, DNA molecules are not homogeneously spread on the surface. And because a large number of aggregates is needed to get results without large errors, the area that has to be scanned with the AFM increases. In the extreme case an area of 25 μm^2 can be completely empty except for one aggregate containing 50 DNA molecules. Another difficulty is the counting of the number of DNA molecules in an aggregate. While the ends of the DNA molecules can be counted, it is difficult to estimate how many DNA molecules are completely inside the aggregate. Also during sample deposition on the surface and subsequent rinsing it is possible that some DNA molecules are removed from the aggregate. When a large aggregate is on the surface it can have DNA molecules that are not close to the surface at all. If these molecules are not very well bound to the rest of the aggregate they will be torn off by the forces that exist during the rinsing process. Therefore the amount of DNA molecules in a big aggregate was probably underestimated. This also explains why the number of free DNA molecules did not go down to zero at high TopIB concentrations (Fig. 4.5). The individual DNA molecules that were found were probably there as a result of rinsing. In solution the number of free DNA probably approaches zero because if enough TopIB is available all DNA molecules should finally be covered with it and be part of an aggregate.

6.3 Experimental methods

6.3.1 Tip-sample and sample-surface interaction

In the AFM experiments we have tried to minimize the perturbation of the samples and the occurring biological processes as much as possible. However during imaging the biological process is still always affected by interaction with the tip and with the surface. During imaging with the AFM the AFM tip exerts a certain force on the sample. In all experiments the AFM was operated in dynamic mode. Forces between tip and sample are much smaller in this mode than in contact mode. For imaging in air the force between tip and sample usually does not affect the sample since consecutive images showed no changes in the surface. The molecules are fixed on the surface and dried. When the force between tip and sample is increased during imaging most likely the tip will become more blunt and the image quality will go down.

Two ways of depositing DNA molecules on the surface have been used in this research, either using Mg^{2+}-ions or polylysine coated mica. The interaction between sample and surface is smaller when using Mg^{2+}-ions, because DNA molecules can equilibrate on the surface. This is usually the preferred method to deposit DNA molecules [34]. However often a background structure is visible on the mica surface after depositing a sample using $MgCl_2$ (Fig. 4.6a). The mica surface is covered with a salt layer which in this case does not cover the surface entirely and has some holes. This type of structure was also found when a solution of only $MgCl_2$ was placed onto mica. This confirmed that the effect was not caused by the presence of TopIB.

Because we found that TopIB-DNA interaction was influenced by the presence of Mg^{2+}-ions and other divalent cations we tried to avoid the presence of divalent ions during the TopIB-DNA binding reaction. Mica discs were treated with a solution of $MgCl_2$ for a certain period (ranging from one minute to one day) in order to replace the potassium ions in the mica surface. This procedure is similar to the one presented by Vesenka et al. [39]. However no DNA molecules were found on the surface after putting a sample with DNA on it. Similar results were found when we added NaCl to the binding reaction, which should result in equilibrium trapping [31]. Unfortunately not many DNA molecules were found on the surface after deposition. Ultimately polylysine was chosen to study TopIB-DNA interaction in the absence $MgCl_2$. Unfortunately, when this deposition method is used DNA molecules are kinetically trapped on the surface.

For imaging in liquid the interaction between the tip and the sample has a larger effect than for imaging in air. Molecules are loosely bound to the surface and they can be moved easily when a force is applied by the tip. Also the biological process can be disturbed by the tip. Therefore it is very important to keep the force between tip and sample as low as possible, but obviously some interaction is needed to get reasonably sharp images. A large force will result in damage of the sample and the molecules will be moved by the AFM cantilever. The force can be divided into two components: the

normal and the *lateral* force. The normal force can easily be estimated by looking at the signal that is coming from the photodiode of the AFM. When imaging the free amplitude of the oscillating cantilever is reduced because of interaction with the sample (Fig. 3.3). Because the spring constant of the used cantilever is known (k=0.08 $^{\mathrm{N}}/_{\mathrm{m}}$) the reduction in amplitude can easily be converted into a force that is applied on the sample. The normal force usually lies somewhere between 50 and 100 pN. The lateral force depends for example on the imaging speed, on the angle between the surface and the tip ($\sim 15°$) and also on the amount of time that the tip is interacting with the sample. A reasonable estimate for the lateral force is 5-10% of the normal force resulting in a force of 3-10 pN. A TopIB molecule which occupies on average 4 pixels on an image is probed approximately 44 times each image (1 ms per pixel and a resonance frequency of 11 kHz). For proteins that are weakly bound to a DNA molecule the AFM can influence the binding and remove the protein from the DNA. This was also observed for TopIB (section 5.2).

The surface plays also an important role in imaging in liquid, because it affects dynamic processes that take place on the surface. It can affect the movement of the molecules because they are attracted to the surface. In section 5.1 it was shown that the ion concentrations in the imaging buffer are such that TopIB can still move on the surface and that it also can detach from the surface into the bulk and vice versa. Also parts of the DNA are mobile and in these regions TopIB can certainly interact with DNA. The results from section 5.2 shows that the lifetime of TopIB bound to DNA shows an exponential behaviour which indicates that there is interaction between TopIB and DNA even while the DNA is adsorbed to the surface. Because TopIB molecules are attracted to the surface they tend to move only in two dimensions. This might result in a longer observed binding time to DNA. On the other hand access to the DNA is also more difficult because of the surface and therefore the bond between TopIB and DNA might be weaker than in bulk experiments. Another proof that TopIB can interact with DNA on the surface was shown in section 5.3. Several TopIB molecules bound to DNA were found to diffuse more slowly than their counterparts that were not bound to DNA. We propose a future experiment to prove that TopIB can interact with surface-bound DNA. In this experiment TopIB is added to a sample where supercoiled DNA molecules are loosely bound to a mica disc. In the meantime the sample is scanned with the AFM and it should be possible to see the supercoils being removed from DNA by TopIB in real time.

6.3.2 Technical limitations on imaging speed

To image dynamic processes the AFM has to be able to acquire images as fast as possible. However we will show that the maximum frame rate is limited by the hardware that is used. In this section the maximum frame rate for the Nanotec AFM will be estimated. The cantilever has a fixed resonance frequency f_r and to determine the amplitude at least two periods of the wave are needed for each point. The minimum time t per image

can be calculated with

$$t > \frac{2nN^2}{f_r} \tag{6.3}$$

in which n is the number of periods of the wave needed for amplitude calculation, N is the number of pixels in one direction. The piezo z-scanner also has a certain resonance frequency f_z which limits the speed at which it can move the sample. For the z-scanner the minimum time t per image follows from

$$t > \frac{2pN^2}{\lambda f_z} \tag{6.4}$$

in which p is the pixel size in nm and λ the apparent width of features on the surface which is typically 10 nm. The Nanotec z-scanner has a f_z of a few kHz.

In table 6.1 the minimum time per image is calculated for several typical image sizes.

Image size (nm)	Pixels	Cantilever limit (s)	Z-scanner limit (s)	Limiting factor
1000	256	29	17	Cantilever
500	128	7,3	4,3	Cantilever
250	64	1,8	1,1	Cantilever

Table 6.1: Minimum time per image at certain imaging conditions. In all cases the cantilever limits the rate of AFM imaging.

The cantilever is the limiting factor in these conditions, although only a 1.7x increase in speed can be obtained by using cantilevers with a higher resonance frequency. To increase the speed even more the speed of the z-scanner needs to be increased also.

6.4 Concluding remarks

The formation of filaments and aggregates confirms that vaccinia TopIB has more properties than changing the supercoiling density of DNA molecules. Although the role of TopIB in the virus is still unknown the observed filament formation can play a role in DNA compaction in the virus. This role might actually be more important for the virus than the relaxation of supercoils. For DNA replication the virus can use the infrastructure and machinery of the infected host cell and it does not need a topoisomerase of its own. TopIB is only loosely bound in the filaments and this suggests that the amount of compaction can be changed according to the needs of the virus. An interesting experiment in liquid would be to fix the filaments on the surface and consequently change the imaging buffer to conditions in such a way that TopIB-DNA dynamics are visible and TopIB can unbind from the filament. By changing the ion concentration also the amount of compaction changes especially when divalent ions are added. Although divalent cations can enhance the compaction of DNA by TopIB, the big aggregates found at high TopIB concentrations probably will not be found in the vaccinia virus. Still

the aggregates are a result of a process that also happens at lower TopIB and Mg^{2+} concentrations. It would also be interesting to repeat these experiments with human type IB topoisomerase. If aggregation and/or filament formation was also found in samples with human TopIB this would give some insight into the origins of the type IB topoisomerases.

Previous research already suggested that TopIB might diffuse along DNA when it tries to locate its preferred cleavage site. The data shown here supports this idea, but still a hopping mechanism could not be ruled out. In a hopping mechanism the TopIB would bind to DNA for a short time. After dissociation the TopIB diffuses in the solution but still stays in the vicinity of the DNA. After a while it binds again and so on. The reduced diffusion constant (compared to diffusion of unbound TopIB on the mica surface) could be explained by the fact that TopIB is bound to DNA for short periods which inhibit the movement of TopIB during these periods. Effectively, this leads to a lower diffusion constant compared to free diffusion on the mica surface. To rule out one of the two mechanisms a faster experimental setup is needed in order to get a higher sample frequency for the position of the TopIB on the DNA. Also a better method for finding a fixed point on the DNA has to be used. A gold bead for example could be attached to DNA to provide a clearly visible mark on the DNA. This idea is currently being pursued in our research group. In addition a gold bead would allow to determine the distance between this bead and the bound TopIB more accurately.

7. Conclusions

Using AFM in air, it was shown that TopIB forms filament-like structures on DNA molecules. A model that describes the formation of these filaments was proposed. The filament starts with the formation of an intramolecular node by two bound TopIB. The extension of this node into a filament can be described by cooperativity theory. Analysis of several data sets showed a cooperativity parameter ω=5181±621 which indicates that the formation of filaments is highly cooperative.

At high TopIB concentrations and in the presence of divalent (cat)ions DNA molecules were aggregated by TopIB. In the absence of divalent ions no aggregates were found. Aggregation was caused by the formation of intermolecular nodes on DNA by TopIB. The influence of divalent cations on the formation of intermolecular is explained in terms of electrostatic interaction. Divalent cations cause screening of the charge of the DNA molecules which reduces the repulsion between them. This was supported by the fact that addition of monovalent ions instead of divalent ions did not result in large aggregates. Control experiments showed that both DNA and TopIB did not aggregate in the absence of each other. Also both the filaments and the aggregates were found when samples were imaged in a liquid environment. The binding of individual TopIB proteins did not change the apparent contour length of DNA.

The dynamics of the TopIB-DNA interaction was investigated by AFM in liquid using an imaging buffer in which DNA and TopIB are only loosely bound to the surface. Imaging rates of 9 $^{s}/_{frame}$ were obtained for small surface areas. TopIB molecules were found to bind and unbind to DNA molecules in consecutive frames. The lifetime of a TopIB bound to DNA was found to be influenced by the AFM tip and an off-rate of $1.7\,\mathrm{frame}^{-1}$ was found. Experiments also suggested that TopIB is only weakly bound in a filament. Several movies containing TopIB-DNA complexes were analyzed and data suggests that TopIB can diffuse linearly along DNA. For TopIB molecules bound to DNA a 1D-diffusion constant ranging from $4.7\pm0.8\,\mathrm{nm}^2\mathrm{s}^{-1}$ to $7.0\pm1.4\,\mathrm{nm}^2\mathrm{s}^{-1}$ was found while unbound TopIB on the surface had a larger 2D-diffusion constant ranging from $45.8\pm3.1\,\mathrm{nm}^2\mathrm{s}^{-1}$ to $61.3\pm2.9\,\mathrm{nm}^2\mathrm{s}^{-1}$.

Three experiments are undergoing. First, for diffusion measurements a method to create a clear reference point on DNA is being developed; secondly, an experiment in which the removal of supercoils from DNA by TopIB is visualized by AFM is being done; finally, experiments with human TopIB instead of vaccinia TopIB are planned.

Bibliography

[1] Champoux, J. J. Dna topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, 70:369–413, 2001.

[2] Dekker, N. H., Rybenkov, V. V., Duguet, M., Crisona, N. J., Cozzarelli, N. R., Bensimon, D., and Croquette, V. The mechanism of type ia topoisomerases. *Proc Natl Acad Sci U S A*, 99(19):12126–31, 2002.

[3] Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A., and Forterre, P. An atypical topoisomerase ii from archaea with implications for meiotic recombination. *Nature*, 386(6623):414–7, 1997.

[4] Wang, J. C. Cellular roles of dna topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol*, 3(6):430–40, 2002.

[5] Cheng, C., Kussie, P., Pavletich, N., and Shuman, S. Conservation of structure and mechanism between eukaryotic topoisomerase i and site-specific recombinases. *Cell*, 92(6):841–50, 1998.

[6] Sekiguchi, J. and Shuman, S. Vaccinia topoisomerase binds circumferentially to dna. *J Biol Chem*, 269(50):31731–4, 1994.

[7] Shuman, S., Bear, D. G., and Sekiguchi, J. Intramolecular synapsis of duplex dna by vaccinia topoisomerase. *Embo J*, 16(21):6584–9, 1997.

[8] Shuman, S. Vaccinia virus dna topoisomerase: a model eukaryotic type ib enzyme. *Biochim Biophys Acta*, 1400(1-3):321–37, 1998.

[9] Sharma, A., Hanai, R., and Mondragon, A. Crystal structure of the amino-terminal fragment of vaccinia virus dna topoisomerase i at 1.6 a resolution. *Structure*, 2(8):767–77, 1994.

[10] Koster, D. Personal communication.

[11] Sekiguchi, J. and Shuman, S. Requirements for noncovalent binding of vaccinia topoisomerase i to duplex dna. *Nucleic Acids Res*, 22(24):5360–5, 1994.

[12] Zechiedrich, E. L. and Osheroff, N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with dna crossovers. *Embo J*, 9(13):4555–62, 1990.

[13] Shaffer, R. and Traktman, P. Vaccinia virus encapsidates a novel topoisomerase with the properties of a eucaryotic type i enzyme. *J Biol Chem*, 262(19):9309–15, 1987.

[14] Shuman, S., Golder, M., and Moss, B. Characterization of vaccinia virus dna topoisomerase i expressed in escherichia coli. *J Biol Chem*, 263(31):16401–7, 1988.

[15] Sekiguchi, J., Seeman, N. C., and Shuman, S. Resolution of holliday junctions by eukaryotic dna topoisomerase i. *Proc Natl Acad Sci U S A*, 93(2):785–9, 1996.

[16] Sekiguchi, J., Cheng, C., and Shuman, S. Resolution of a holliday junction by vaccinia topoisomerase requires a spacer dna segment 3' of the ccctt/ cleavage sites. *Nucleic Acids Res*, 28(14):2658–63, 2000.

[17] Sekiguchi, J. and Shuman, S. Site-specific ribonuclease activity of eukaryotic dna topoisomerase i. *Mol Cell*, 1(1):89–97, 1997.

[18] Shuman, S. Polynucleotide ligase activity of eukaryotic topoisomerase i. *Mol Cell*, 1(5):741–8, 1998.

[19] Sherratt, D. J. and Wigley, D. B. Conserved themes but novel activities in recombinases and topoisomerases. *Cell*, 93(2):149–52, 1998.

[20] Redinbo, M. R., Champoux, J. J., and Hol, W. G. Structural insights into the function of type ib topoisomerases. *Curr Opin Struct Biol*, 9(1):29–36, 1999.

[21] Binnig, G., Quate, C. F., and Gerber, C. Atomic force microscope. *Physical Review Letters*, 56(9):930–933, 1986.

[22] García, R. and Pérez, R. Dynamic atomic force microscopy methods. *Surface Science Reports*, 47(6-8):197–301, 2002.

[23] Dame, R. T. and Goosen, N. Hu: promoting or counteracting dna compaction? *FEBS Lett*, 529(2-3):151–6, 2002.

[24] McGhee, J. D. and von Hippel, P. H. Theoretical aspects of dna-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J Mol Biol*, 86(2):469–89, 1974.

[25] Kowalczykowski, S. C., Paul, L. S., Lonberg, N., Newport, J. W., McSwiggen, J. A., and von Hippel, P. H. Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters. *Biochemistry*, 25(6):1226–40, 1986.

[26] Lonberg, N., Kowalczykowski, S. C., Paul, L. S., and von Hippel, P. H. Interactions of bacteriophage t4-coded gene 32 protein with nucleic acids. iii. binding properties of two specific proteolytic digestion products of the protein (g32p*i and g32p*iii). *J Mol Biol*, 145(1):123–38, 1981.

[27] Menetski, J. P. and Kowalczykowski, S. C. Interaction of reca protein with single-stranded dna. quantitative aspects of binding affinity modulation by nucleotide cofactors. *J Mol Biol*, 181(2):281–95, 1985.

[28] Ando, R. A. and Morrical, S. W. Single-stranded dna binding properties of the uvsx recombinase of bacteriophage t4: binding parameters and effects of nucleotides. *J Mol Biol*, 283(4):785–96, 1998.

[29] Guthold, M., Zhu, X., Rivetti, C., Yang, G., Thomson, N. H., Kasas, S., Hansma, H. G., Smith, B., Hansma, P. K., and Bustamante, C. Direct observation of one-dimensional diffusion and transcription by escherichia coli rna polymerase. *Biophys J*, 77(4):2284–94, 1999.

[30] Stryer, L. *Biochemistry*. W.H. Freeman and Company, New York, 6th edition, 1999.

[31] Rivetti, C., Guthold, M., and Bustamante, C. Scanning force microscopy of dna deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J Mol Biol*, 264(5):919–32, 1996.

[32] Chatfield, C. *Statistics for technology*. Chapman & Hall, 3rd edition, 1996.

[33] Brock, I. Statistical methods of data analysis, 2000.

[34] Bustamante, C., Vesenka, J., Tang, C. L., Rees, W., Guthold, M., and Keller, R. Circular dna molecules imaged in air by scanning force microscopy. *Biochemistry*, 31(1):22–6, 1992.

[35] Stasiak, A., Di Capua, E., and Koller, T. Elongation of duplex dna by reca protein. *J Mol Biol*, 151(3):557–64, 1981.

[36] Kwon, K. and Stivers, J. T. Fluorescence spectroscopy studies of vaccinia type ib dna topoisomerase. closing of the enzyme clamp is faster than dna cleavage. *J Biol Chem*, 277(1):345–52, 2002.

[37] Moreno-Herrero, F., Colchero, J., and Baro, A. M. Dna height in scanning force microscopy. *Ultramicroscopy*, 96(2):167–74, 2003.

[38] Israelachvili, J. *Intermolecular & Surface Forces*. Academic Press Limited, London, 2nd edition, 1991.

[39] Vesenka, J., Guthold, M., Tang, C. L., Keller, D., Delaine, E., and Bustamante, C. Substrate preparation for reliable imaging of dna molecules with the scanning force microscope. *Ultramicroscopy*, 42-44 (Pt B):1243–9, 1992.