

Research in the field of single biopolymers using scanning probe microscopy

THÈSE N° 6372 (2015)

PRÉSENTÉE LE 18 SEPTEMBRE 2015
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE PHYSIQUE DE LA MATIÈRE VIVANTE
PROGRAMME DOCTORAL EN PHYSIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Andrey MIKHAYLOV

acceptée sur proposition du jury:

Prof. M. Q. Tran, président du jury
Prof. G. Dietler, Dr S. Sekatski, directeurs de thèse
Prof. P. De Los Rios, rapporteur
Dr D. Klinov, rapporteur
Prof. O. Marti, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

Sapere aude.
— Horace

To my daughter...

Acknowledgements

Here I would like to express my acknowledgements for all who helped me during the work on this thesis. First of all, I have to thank Switzerland, the country, which gave me a priceless opportunity to live and work here. I remember the day, when I saw it first time in summer 2007 and fell in love with it at first sight. After that I was constantly dreaming of getting a position here. I thank Switzerland for providing me with a possibility to learn different French, German and Italian.

I thank kindly my thesis directors, Giovanni Dietler and Sergey Sekatskii. It was a very good chance to meet and work with such good scientists and people. They impress me every day by their infinite knowledge. When I show Giovanni some data, he has always an idea of how to process and present it. When I show Sergey an equation, he knows every particular solution of it.

I would like to show gratitude to all the people I was working with. This intercultural and multi-language team was constantly inspiring me with good jokes and new ideas. I thank especially the users of my program, Jeong Jae Sun, Aleksandre Japaridze, Simone Ruggeri, Dusan Vobornik, Andrea Cerreta, Urszula Cendrowska and many others. They helped me with the testing of the software and provided reports of their experience, found bugs and desirable features. Without them I could not create a complete product. I thank Fabrizio Benedetti for teaching me how to prepare samples and process the data and for his programming ideas. I thank Andrzej Kulik for fruitful and wise scientific and not only discussions. I thank Michel Kessous for his help with every kind of software/hardware issue. I thank Christine Vuichoud for perfect organization. I cannot imagine a case when Michel or Christine forget to do something or do it in not a thorough manner. I thank our mechanics and electronics engineers for precise fabrication of any setup.

Of course, I heartily thank my family for supporting me during these years. My mother visited me many times, so that I do not feel that great distance of several thousand kilometers from my native city. Finally, I cannot imagine my life here without my lovely wife, who constantly helping me with everything and giving me power to overcome any obstacle, and my daughter, who inspires me every day with her curiosity.

Lausanne, le 25 août 2014

A. M.

Abstract

Current thesis is devoted to the investigation of biopolymers using AFM, a well established and widely used technique. To begin with, we studied processing of AFM images and its automation. Our software product, *DNA Trace*, permit one to quickly extract scientifically meaningful data from a huge number of AFM scans. Then, this program enabled us to perform a study on how polymers, with DNA as a model, change their conformation from three to two dimensions upon their deposition onto the surface. DNA molecules suit well for such an investigation as they are studied in detail, easy to deposit onto mica surface, have relatively large persistence length to be traced on the images and, the most important, relax from 3D to 2D, the timescale of this process being in order of tens seconds to minutes.

Besides, we introduced so-called *concentration function* to describe numerically such transition and adapted formulas of the theory of diffusion to polymers. We depicted this process as a diffusional propagation of relaxation of short segments of polymer. The results we have obtained are very promising and hopefully will lead to the better understanding of the theory of polymers, as the latter are widely used in all areas of industry.

Then, this work describes the possibility to increase the sensitivity of cantilever deflection detection. We etched a diffraction grating on the backside of cantilever and used diffractive rather than mirror reflection of a laser beam, which gave geometrical advantage. We also used position-sensitive duo-lateral diode in this setup and finally got the increase of sensitivity of about 5 - 10 times compared to common detection scheme.

Finally, a big chapter of this thesis is devoted to force spectroscopy, the technique, which permit one to measure forces of interaction between biomolecules and thus obtain kinetic parameters of interaction. We studied barnase - barstar pair and then flaviviral proteins - putative cell receptors interaction. The results obtaining by the latter study will hopefully lead to the model of flaviviral - cell membrane fusion.

Key words: atomic force microscope (AFM), DNA relaxation on the surface, polymer 3D-2D transition.

Résumé

Cette thèse est dédiée à l'étude de biopolymères en utilisant le microscope de force atomique (AFM), technique connue qui est bien établie et largement utilisée. Tout d'abord, nous avons étudié le traitement des images d'AFM et son automatisation. Notre logiciel, *DNA Trace*, nous permet d'extraire les données significatives d'un grand nombre d'image AFM et ce très rapidement. Puis, en utilisant ce logiciel, nous avons étudié le changement de conformation de polymères passant de trois à deux dimensions lors du dépôt sur la surface, prenant comme model de polymères les molécules de l'ADN. Cette molécule convient bien pour cette recherche, car elle est bien étudiée, est facile à déposer sur la surface de mica, à tracer grâce à sa grande taille et, le plus important, est capable de se relâcher de 3D à 2D avec une échelle de temps allant de la dizaines de secondes aux minutes. En plus, nous avons introduit la *fonction de concentration* pour transcrire cette transition de manière numérique et adapter les formules de la théorie de diffusion aux polymères. Nous avons décrit ce processus comme la propagation diffusive de la relaxation de segments courts du polymère. Les résultats obtenus sont prometteurs et nous espérons qu'ils aideront les chercheurs à mieux comprendre la théorie des polymères, largement dans tous les domaines de l'industrie.

L'autre chapitre de cette partie traite de la possibilité d'augmenter la sensibilité de détection de la déflexion du cantilever. Pour le faire, nous avons gravé une grille de diffraction sur la surface arrière et utilisé la réflexion du faisceau de laser diffractive au lieu de la réflexion spéculaire. Cette technique nous a donné un avantage géométrique et, combinée avec l'utilisation de la photodiode duo-latérale sensible à la position, a permis d'augmenter de la sensibilité de 5 à 10 fois.

Enfin, le dernier chapitre de spectroscopie de force, traite de la technique permettant de mesurer les forces d'interaction entre des biomolécules et d'en obtenir des paramètres cinétiques. Nous avons étudié la paire de protéines barnase - barstar ainsi que les combinaisons entre des protéines flavivirales et les récepteurs cellulaires putatives.

Mots clefs : microscope à force atomique, relaxation de l'ADN sur la surface, 3D-2D transition, spectroscopie de force, interaction entre deux protéins.

Contents

Acknowledgements	v
Abstract (English/Français)	vii
List of figures	xii
List of tables	xiv
1 DNA Trace	5
1.1 Introduction	5
1.2 Flattening	7
1.3 Tracing algorithms	11
1.4 Types of objects	14
1.5 Results and Discussion	15
1.6 Current applications	17
1.7 Conclusions	17
2 Diffraction grating on a cantilever	19
2.1 Photodiode	21
2.2 Results and Discussion	23
2.3 Conclusions	27
3 DNA relaxation	31
3.1 Introduction	31
3.2 Experiment	34
3.2.1 Materials and methods	34
3.2.2 Scaling exponent	36
3.2.3 Diffusion of relaxation	37
3.2.4 Persistence length	39
3.2.5 Error assessment	40
3.3 Conclusions	45
4 Concentration function	47
4.1 Motivation	47
4.2 Theoretical model	48

Contents

4.2.1	Equation of diffusion	48
4.2.2	Concentration function	49
4.2.3	Normalization	53
4.3	Results and Discussion	55
4.3.1	Simulation	59
4.4	Conclusions	61
5	Force spectroscopy	65
5.1	Introduction	65
5.1.1	Force - loading rate dependence	67
5.1.2	Spring constant calibration	69
5.1.3	Surface treatment and protein deposition	70
5.1.4	Event selection	71
5.2	Barnase-barstar interaction	72
5.2.1	Scientific interest for the protein pair	72
5.2.2	Experimental setup	73
5.2.3	Results	73
5.3	Viral proteins	76
5.3.1	Introduction	76
5.3.2	Flavivirus	76
5.3.3	Flaviviral membrane - cell interaction	77
5.3.4	Results	78
5.3.5	Discussion	79
5.4	Conclusions	82
A	Used acronyms	91
A	Listings of the code	93
A.1	Smart Fit	93
A.2	Flood Fill	96
A.3	Put path	98
	Bibliography	112
	Curriculum Vitae	

List of Figures

1	Real profile of the sample compared to the apparent one	2
1.1	Screenshot of <i>DNA trace</i> with a short molecule of double-stranded DNA.	7
1.2	Flattening artifacts and their elimination using <i>DNA Trace</i>	8
1.3	Example of the work of tracing algorithm.	13
1.4	Visualization of Semi-automatic further tracing algorithm.	14
1.5	Possible types of objects of <i>DNA Trace</i>	15
1.6	Screenshots of output of <i>DNA trace</i>	17
2.1	Scheme of AFM photodiode	20
2.2	Schematic of detection of the tip position using a grating on the backside of it.	21
2.3	Laser beam detection at two different positions.	22
2.4	SEM images of the backside of CSC38 AFM lever with a diffraction grating.	24
2.5	Force curves recorded in diffractive and mirror reflection modes.	25
2.6	Picture of a position sensing detector used in our setup.	25
2.7	Circuit diagram of an amplifier used to read out the signal from position sensing detector.	26
2.8	SEM image of the ultra-short cantilever with a grating etched on the backside of it.	27
2.9	Geometrical improvement of the angular sensitivity by a spherical surface.	28
3.1	Images showing DNA relaxation with deposition time from 10 to 300 seconds.	35
3.2	$\langle R_{ee}^2(s) \rangle$ vs. s for relaxation time from 10 seconds to 5 minutes.	36
3.3	Local scaling exponent as a function of contour length.	37
3.4	Square distance up to which molecule is relaxed.	39
3.5	Dimensionality of the system as a function of relaxation time.	40
3.6	2D projections of the simulated polymer molecules.	42
3.7	Histogram of the values of the persistence length for different sets of molecules.	43
3.8	The values of $\langle l_p \rangle (N)$ and $\sigma(N)$ for the sets with different number of molecules.	43
3.9	Values of l_p calculated on the sets with different number of molecules.	44
3.10	Scaling of $\sigma(N)$ from fig. 3.8.	44
4.1	Sketch of a molecule showing calculation of concentration function.	50
4.2	Sketch showing calculation of concentration function for multiple molecules.	51
4.3	Concentration function for DNA molecules with different deposition time.	56
4.4	Concentration function for samples with different deposition time and its fit.	57

List of Figures

4.5	Goodness of fit for different models and relaxation time.	58
4.6	Concentration function of the contours of the simulated polymer molecules. . .	63
5.1	Scheme of measuring forces between the surface and the cantilever.	66
5.2	Typical retract force curve with a specific interaction rupture event.	67
5.3	EM image of an AFM tip with several cantilevers	68
5.4	Typical force curve of a single I27 molecule stretching.	69
5.5	Force histogram for selected loading rate range.	74
5.6	Barnase - barstar force - loading rate dependence.	75
5.7	Dissociation energy landscape in absence and presence of the applied force. . .	76
5.8	Force - loading rate dependence for protein E with putative cell receptors. . . .	78
5.9	Interaction between protein E and LBP and corresponding monoclonal antibodies. .	84
5.10	Examples of force curves of viral - cell protein - protein interaction.	85
5.11	Examples of force curves of viral - cell protein - protein interaction (continued). .	86
5.12	Histograms of the distributions of heights and lengths of force plateaus.	87
5.13	Extension force curve of a polymer.	88
5.14	Different regimes of force - extension dependence.	89

List of Tables

2.1	Incident and reflected angles along with the sensitivity increase coefficient. . .	26
3.1	Dependence of the scaling exponent on the dimensionality of the system. . . .	34
3.2	Dependence of dimensionality of the system on relaxation time.	40
4.1	Fitting parameters of the concentration function of the samples with different relaxation time.	59
4.2	End-to-end distance of the molecules and its relation to the parameters D and T .	59
4.3	Parameters of the fit of the concentration function data for the simulated samples.	60
5.1	Kinetic parameters of the interaction of viral protein with putative cell receptors.	80

Introduction

My thesis is dedicated to scanning probe microscopy and to atomic force microscopy (AFM) as its major part. This method is so well known nowadays, that barely needs a long introduction and its importance is doubtless. In this section I would like to give a brief overview of the technique.

The predecessor of AFM was scanning tunneling microscopy (STM), developed in 1981 by Gerd Binnig and Heinrich Rohrer at IBM Zürich. The idea behind this technique was to use a sharp conductive "needle" to scan conductive surface of the sample and measure the local tunneling current between the probe and the sample. This invention was one of the most important key points in the history of science and brought Nobel Prize in Physics to its inventors in 1986. Good STM devices achieve 10^{-1} nm lateral resolution (XY directions) and 10^{-2} nm perpendicular resolution (Z direction) [6]. Such a device permits one to visualize individual atoms and observe not only the periodicity of the lattice, but its single defects as well. STM can work under different conditions including gas, liquid, ultra-high vacuum and temperatures varying from fractions of kelvin to several hundred degrees Celsius. Nevertheless, STM is based on the tunneling current effect, so its important limitation is that the investigated sample has to be conductive. Such a condition seriously limits the use of STM for studying of biological samples.

After the introduction of scanning tunneling microscope, AFM was its logical successor. Patented in 1986 [11], it started to gain popularity quickly. AFM is free from the limitation to work only on conductive samples, as AFM does not rely on the tunneling current but rather on the interaction force between the sample and the probe. This is a very important technique in many areas including biophysics. The image obtained with such a microscope is not just a 2D image but rather a 3D topography profile with nanometric, if not angstrom resolution [129]. For example, resolution sufficient to observe chemical structure of a single molecule of pentacene was shown by Gross et al. [58].

One of the key aspects of high resolution of scanning probe techniques is the sharpness of the probe. In case of AFM a typical probe is called *cantilever* and consist of a lever with a sharp tip at the end mounted on a chip. If one assume that scanning is done in a contact mode and both

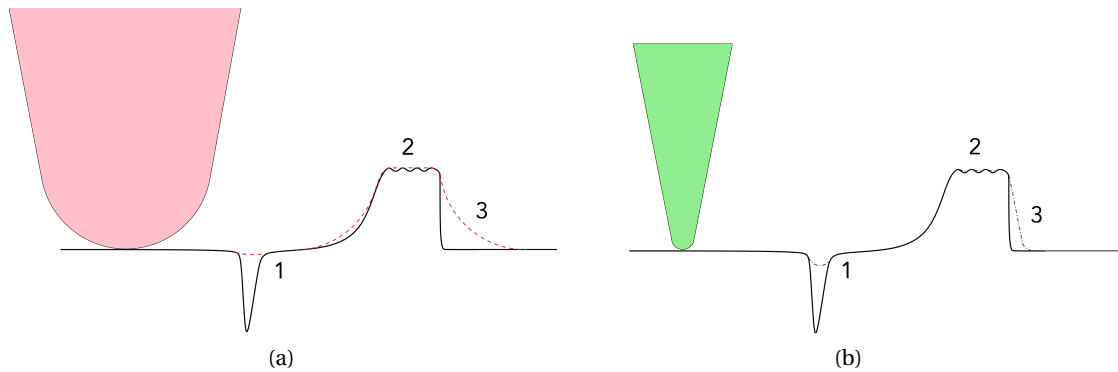


Figure 1 – Real profile of the sample (thick line) compared to the apparent one, recorded by an AFM device (dashed line). (a): relatively blunt tip with radius of curvature of the apex larger than the fine structure of the sample does not permit to resolve certain features of the sample. In particular, the depth of the hole is underestimated (1), fine periodicity (2) is not resolved and the apparent shape of an object (3) deviates from the real one and is sufficiently wider. (a): sharper tip produces better results in cases (2) and (3), but is still not able to resolve fully the real profile of the hole (1).

the tip and the sample are rigid enough, the apparent image of the sample recorded by AFM is represented by the convolution of the real profile with the shape of the end of the probe (see fig. 1). There are many consequences that prohibit one to obtain the actual information about the sample. First of all, if one wants to investigate the fine structure, the typical size of details has to be comparable to the radius of curvature of the tip R . Indeed, scanning of a sample with periodicity much lower than R (fig. 1, (2)) does not reveal this periodicity and records only considerably larger details. On the contrary, using of a sharp tip (fig. 1b) permits one to record the profile, which is much closer to the real profile of the surface. Of course, if there is even finer structure, it will not be resolved.

There exist different modes of obtaining images. The first one is called contact mode, in which the tip is dragged across the surface. The topographical profile of the sample is obtained by measuring the deflection of the cantilever directly or keeping the deflection constant using a feedback loop. Another possibility is to use a dynamic mode, in which the tip is oscillating and frequency or amplitude of oscillations can be used as feedback channel. This mode is typically the most appropriate for scanning soft biological samples, as the interaction force in this case is relatively small, which permits one to scan the sample without damaging it.

AFM is widely used for scanning of biological samples including DNA. It can even resolve its helical periodicity and permit one to calculate the pitch of the double helix [22, 93]. Besides, one can extract huge amount of information from the data coming from an AFM device. We will discuss in details processing of scans and extraction of different parameters from them in

chapter 1. Moreover, AFM permit one not only to visualize the topographical profile of the sample, but also the forces of interaction of nanoscale objects, for example, antigen-antibody or other proteins. This topic will be described in chapter 5.

In chapter 2 we are going to present the investigation, which can permit to increase the resolution of AFM tip deflection detection. We have etched a diffraction grating on the backside of the cantilever and have successfully shown, that using an appropriate AFM setup, one can obtain a significant improvement of the resolution and minimum detectable angle of the deflection of AFM probe.

Chapter 3 discusses the physics of the transition of a polymer from three to two dimensions upon the deposition of it onto the surface. We used molecules of DNA as a sample, as they are widely studied and can be easily observed using AFM. We have shown that this transition process is time-dependent and can be observed using AFM. We also present the images and statistical distributions of DNA molecules in 3D-2D quasi-projection, 2D equilibrium and intermediate states.

Chapter 4 develops this study further and describes the analogy between the contour of a polymer and the trajectory of a particle undergoing Brownian motion and applies the laws of diffusion to the polymer theory. We believe that the results presented in this chapter will be helpful for deeper understanding of the behavior of the polymer in two and three dimensions, as well as for the description of the transition process from three to two dimensions.

1 DNA Trace

Science is what we understand
well enough to explain to a computer.
Art is everything else we do.
Donald Knuth

1.1 Introduction

AFM is a brilliant tool that can give us amazing images. Scientists appreciate especially its use for biological samples like DNA and proteins, which can be proven by many papers [135, 39, 13, 132, 2]. Some of these studies show qualitative results, for example, binding of some proteins on the DNA which can be easily visualized by AFM. Others require numerical calculations of some values, for instance, persistence length, which is related to the rigidity the molecule and its elastic parameters [33, 8].

AFM itself is a well-automated device. It permits us to obtain relatively easily images with the rate of 5-10 images per hour, not speaking about ultrafast techniques with video frame rates, and can scan over big areas automatically, producing tens and even hundreds of images. One can calculate and average different topological parameters, which are statistically robust. To do so, one has to process all of these images, applying some filters to the images (for example, flattening) and selecting interesting objects on them. We will use the term *tracing* which means the extraction of the coordinates of the objects.

Such a processing requires usually a large amount of handwork, although some techniques for automation of object recognition were already developed before [106, 89, 42]. The main drawbacks of such techniques for tracing of the molecules usually are: 1) the lack of the ease of use, 2) small level of automation or 3) the possibility to use them only in a narrow range of specialized applications. This drew us to the idea of developing a new tool for automatic image

processing. We have created *DNA Trace*, a user-friendly multi-platform software designed to work with raw images from most common commercial Atomic Force Microscopes [92].

The first and the main goal of the software was to determine the coordinates of the objects on AFM images, mostly of DNA molecules. This routine extraction of the coordinates is common for the field and is necessary, when one wants to characterize quantitatively the sample and to report the values of the persistence length, periodicity of height (which can be obtained by analyzing the Fourier transform of the profile of the height along the object), angular and height distributions and many other values. As most of them need to be statistically averaged, the error of their determination typically decreases as $\sim 1/\sqrt{N}$, where N is the number of samples in the set. This explains, why the researchers often want to obtain and treat hundreds, if not thousands of molecules for every sample. Automatic scanning is relatively well implemented in many devices, especially the ones having low drift scanners, while the following processing of the obtained images still requires a lot of handwork.

Our software combines several filters for images and a set of fully automatic, semi-automatic and manual algorithms for tracing (digitizing of the coordinates) of long polymer molecules. *DNA Trace* is written entirely in Java SE[®] platform and can be run on every operating system for which a *Java Virtual Machine* exists, including Windows[®], GNU/Linux[®] and Mac OS X[®]. It implements a convenient graphical user interface and looks like a native application (fig. 1.1).

DNA Trace was written by me, based on the existing software *DNA Statistics* (which was formerly written by me as well). Of course, the help from the colleagues, their usage and bug reports and proposals of new features was really appreciated. The ideas of the algorithms to use were based on the tracing procedures described before [133], on my own and my colleagues' experience of image processing and common programming paradigms. I chose Java as an efficient, modern multi-platform environment and mainly consulted with [14] to create clean and robust code.

The software can directly open raw files from most of the commercial microscopes. This means that the user does not have to perform any calibration, all the parameters such as scanner properties (range, offset etc.), image mode and many others are extracted immediately and there are no losses of precision due to conversion. Currently the list of the supported microscope vendors is the following: *Bruker*, *Veeco*, *NT-MDT*, *Aist-NT*, *Park systems*, *Nanonis*, *Asylum Research*, *JPK*. Files from the devices of other vendors will be added in future.

DNA Trace can open common image formats (JPEG, GIF, PNG, BMP, TIFF, etc.) as well, provided that the user specifies pixel-to-nanometer XY and Z calibration factors by clicking on the scale bar.

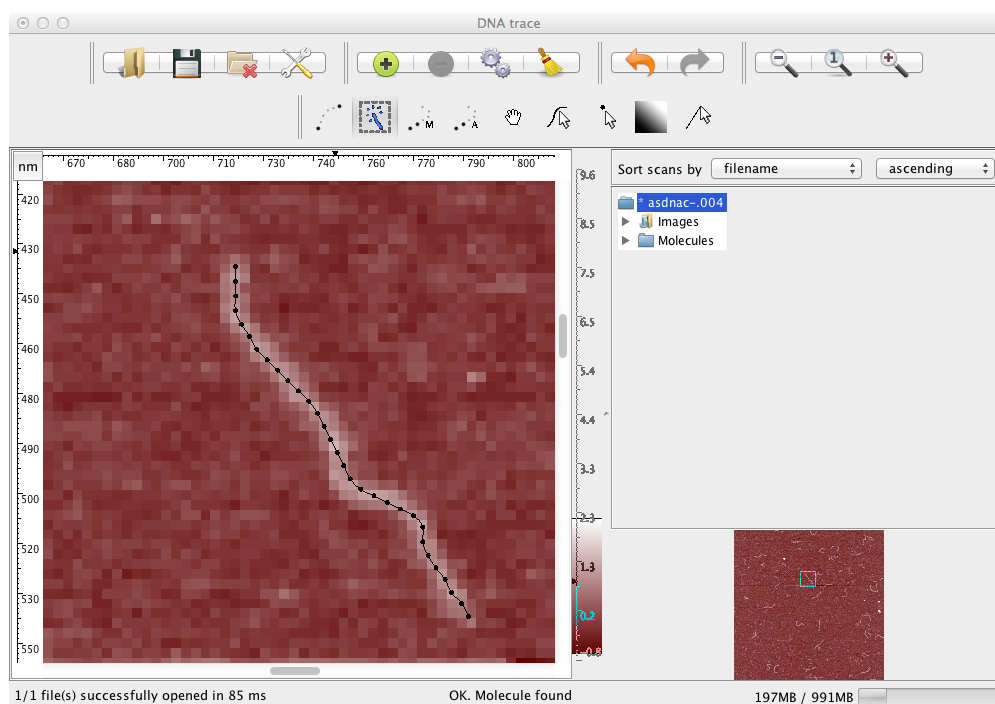


Figure 1.1 – Screenshot of *DNA trace* with a short molecule of double-stranded DNA. Such clear image of short molecules can be processed completely automatically.

When an image is opened and modified (filters applied, molecules traced, etc.), it can be saved in a proprietary *DNT* format with open specification (see the description of it below). This type of file can be opened with the program to continue the work with the image or to export different information.

1.2 Flattening

It is well known that most of SPM images require flattening [87]. For example, a typical substrate for AFM, mica, is supposed to be flat, but on the raw image from the microscope different lines are often shifted with respect to each other (fig. 1.2a). This effect appears due to the line-by-line scanning nature and thermal drift of the piezo scanner.

Other pronounced effects can be the inclination of the sample surface of the substrate and higher-order curvatures due to the non-linearity of the scanner. Even if the scanner is well calibrated using a special grating and some hardware correction is applied to every image, the characteristics of the piezoelectrical parts of the scanner change as a function of the temperature and time. The problem typically becomes more pronounced with the increase of the size of the scanned area.

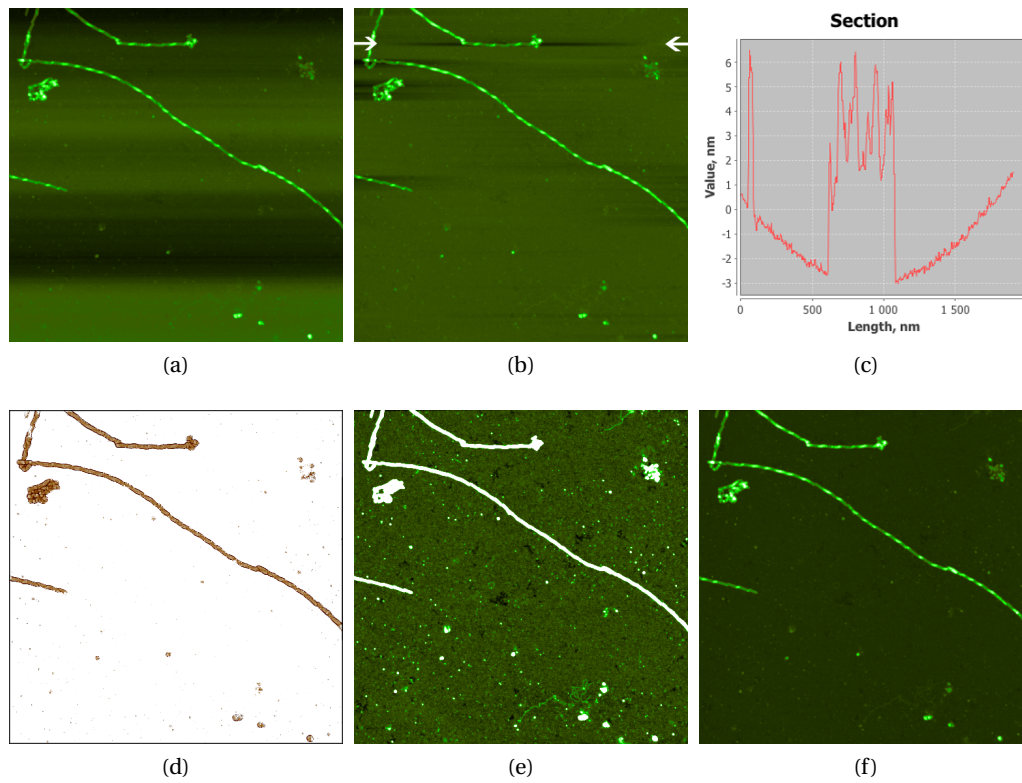


Figure 1.2 – Flattening artifacts and their elimination using *DNA Trace*. (a): a raw image of an amyloid filament from the microscope (*Veeco Nanoscope III Multimode AFM*); (b): standard line-by-line flattening using 2^{nd} -order curve subtraction gives strong deviation from a flat background; (c): a line of an image, showing a strong flattening artifact (marked with white arrows on the image (b)). Background, which is supposed to be flat, is substantially curved, due to this we have about 3 nm of underestimation of the height of certain parts of the filaments; (d): contours of objects determined by the software; (e): smart flattening algorithm used by *DNA Trace* shows absolute flat background with no artifacts; the height of the sample is correctly determined in this case; (f) the same image with the contrast showing the features on the filaments. Images (a), (b), (d), (e) and (f) are $2 \times 2 \mu\text{m}$ size.

Usually researchers use the least-square subtraction of a polynomial surface from the whole image or a polynomial curve from every line, the order can vary from 0 (constant offset), 1 (a plane or a line, respectively) to 3 or more[75]. This technique works well for the images, where the background occupies the major part and the objects are relatively low (for example, most of the images of DNA molecules).

Nevertheless, some images flattened using this technique are unacceptable for the precise measurements of the height of the objects (fig. 1.2b) [73]. Subtraction of the curve should take into consideration only points of the background, otherwise the parts of the image, belonging to the objects, introduce deviation, which can be easily seen by making a section along the line, for which an object takes a substantial part (fig. 1.2c). This section should show flat background with zero average, but instead it is curved, which leads to the underestimation of the height of the substrate for about 3 nm in this case.

DNA Trace includes a fuzzy-logic algorithm, which automatically detects objects on a non- or wrong-flattened image, exclude them and subtract polynomial curves, fitting only background parts of an image (fig. 1.2e). The final image shows a flat background with zero average and only random noise, coming from the device. This feature can substantially improve many images, obtained using SPM [53].

This flattening process is performed in the following way. First, the user or the software itself selects an area, which will be used as a background. Then a polynomial function is subtracted from every line, minimizing the root mean square deviation (RMSD) between the pixels in the area and the function. In many common software solutions this area is equal to the whole image, which gives the source of the aforementioned flattening artifacts and discrepancy. Ideally, such an area should include only the parts corresponding to the flat background and exclude "peaks", corresponding to the deposited sample, or "holes". The main problem is to find objects on the image before it is flattened.

We have noticed, that if one applies the Discrete Laplace operator to the image ¹, the contours of the objects become substantially clearer and easier to detect. Namely, the background value becomes close to zero, while the borders of objects show the value, which is greater than the noise of the background (it can be easily seen on fig. 1.2d). Moreover, the Discrete Laplace

¹Discrete Laplace operator is an analog to continuous Laplace operator $\Delta f(x, y) = \nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$. Discrete Laplace operator is applied to an image by calculating the convolution of 2D-array of the pixels with some

of *Discrete Laplacian* kernels. We used a kernel with diagonals: $D_{xy}^2 = \begin{bmatrix} 0.5 & 1 & 0.5 \\ 1 & -6 & 1 \\ 0.5 & 1 & 0.5 \end{bmatrix}$

operator is almost insensitive to the smooth change of the height of pixels due to the scanning artifacts, but is sensitive enough to the sharp gradients corresponding to the objects on the surface. In other words, Discrete Laplace operator applied to the same image before its flattening and after it produces very close results. These facts permit us to use the following algorithm:

1. Create a copy A of the original image I :

$$A = \text{Copy_Image}();$$

2. Perform Laplacian transform on this copy (fig. 1.2d):

$$A = \text{Laplacian_Transform}(A);$$

3. Calculate $RMSD$ of A :

$$RMSD = \text{Root_Mean_Square_Deviation}(A);$$

4. Choose a set of points $S \subset A$ on the image; in current version we choose angle points:

$$S = \{(0, 0), (0, \text{height}), (\text{width}, 0), (\text{width}, \text{height})\};$$

5. Select an area using *Flood fill*² algorithm with starting points $(x_i, y_i) \in S$ and the following criteria that pixels are connected: $p < 3 * RMSD$, where p is the value of the pixel in the copy A (the listing of the implementation of the algorithm is given in appendix A.2);

$$\text{Selection} = \text{Flood_Fill}(A, S, p < 3 * RMSD);$$

6. Fit the original image line by line using least squares method with a polynomial of n^{th} order (we have found $n = 3$ as a sufficient value) minimizing the sum of square deviations between pixels in the original image and the polynomial function; only pixels which belong to the *Selection* are considered. If p_{xy} is the value of the pixel of I with coordinates x and y , fitting of the i^{th} line can be described as the following:

$$\text{For } j = 1 \text{ to } \text{width} \text{ do } p_{ji}^{\text{new}} = p_{ji} - f(p_{ji}),$$

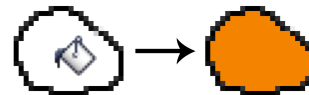
where $f(x)$ is a polynomial of n^{th} order, calculated by least squares method over all pixels of line i that belong to S :

$$\sum_{x=1, p_{xi} \in S}^{\text{width}} (f(p_{xi}) - p_{xi})^2 = \min.$$

The listing of the part of the code concerning this algorithm can be found in appendix A.1. The described flattening procedure is fully automatic and can be applied upon opening of an image. For the scans where the sample occupies the major part of the image or the noise is high, this procedure can produce undesirable result. In this case the user can select the area of background manually (using rectangular selection or magic wand tool) or use classic flattening.

2

Flood fill is the algorithm that iterates over the area connected to a given node in a multi-dimensional array. The criterium that the nodes are connected is given by a particular function f . This algorithm is widely used in many image processing programs, for example, *bucket fill* tool in *Windows* build-in *Paint* editor. In this case the pixels are the nodes of two-dimensional array and f applied to a pixel returns true if its color is equal to the one of an initial pixel (e.g. the one which the user clicks at). See the enclosed image for the explanation.



1.3 Tracing algorithms

In this section the terms *step*, *SectionWidth*, *UseGaussianFit* and *scale* mean some parameters which can be adjusted by the user. They are explained below.

Once the image is opened and flattened, the user wants to trace the objects on it, i.e. extract the coordinates of long structures. Each of them is finally represented by a sequence of points with a constant distance between them (*step*). The software has a broad choice of the algorithms for tracing. They can be described as the following:

1. Fully automatic tracing of all molecules on an image: the user has to press one button. It works well for clear images of relatively short and isolated molecules.
2. Automatic tracing of each molecule: the user has to click on one point of each molecule and the program finds the rest of it. The algorithm suits for relatively short molecules with zero or few crossings.
3. Automated searching of the part of the molecule between two points: the user has to do two consequent clicks along the molecule and the program traces the molecule between them. This algorithm is the most used and can be applied for every image. The clearer the image is, the fewer clicks the user has to perform. For most of images it is enough that the user points to the ends of the molecule and to the crossings.
4. Semi-automatic further tracing: if the molecule is already partly traced, the program is going forward along it while the user presses the button.
5. Manual tracing: user manually shows the direction of the molecule to the program.

The algorithms can be combined in order to achieve the fastest and the most precise tracing. For each point the program records the coordinates, the height and the width. The algorithms are adjustable and show high stability of molecule's tracing under different conditions. For most scans of polymer it takes the user from several seconds to several minutes to process an image, depending on its quality and the number of the molecules on the surface.

Let us look deeper into some of these algorithms. The most useful and precise is normally the one number 3. It is the most complex and most accurate algorithm, which works in almost all cases. It can be described in the following way:

1. Let us assume that the image is flattened and mean value of every background pixel in every line is close to zero.

2. At the first step the user has chosen two points A and B manually; they have coordinates (x_a, y_a) and (x_b, y_b) respectively.
3. The software takes into consideration only some portion of the image to reduce the calculation overload; in our case we choose a rectangle which center coincides with the middle of the segment AB, $width = 2 * |x_a - x_b|$ and $height = 2 * |y_a - y_b|$.
4. We make a connected graph the area in this rectangle using the following rule. All pixels are the vertices in the graph. For every two adjacent (horizontally, vertically and diagonally) pixels we add an edge connecting two corresponding vertices. We associate a weight w to every edge using the following formula: $w = \exp(-\frac{p1+p2}{scale})$, where $p1$ and $p2$ are the values of height of the pixels and $scale$ is some parameter (it will be discussed later in this text). Therefore the weight of every edge is strictly positive. This weight is multiplied by $\sqrt{2}$ for diagonal edges.
5. The Dijkstra's algorithm³ is used to solve the graph (i.e. to find the "shortest" path between vertices A and B, where the length of the path is equal to the sum of the weights of the edges of the path).
6. At the end all the vertices belonging to the solved "shortest" path are unified into a sequence; then this sequence is renormalized so that all the adjacent points have the same distance equal to $step$ between them.
7. The complexity of Dijkstra's algorithm is $O(E + V \log V)$ [47], where E is the number of edges and V is the number of vertices. If the rectangle has $N \times N$ pixels, $E = N^2$ and $V = 8N^2$ the resulting complexity is $O(N^2 \log(8N^2)) = O(N^2 \log N)$, which permits us to solve the graph with $N = 1000$ within fractions of a second. Typical AFM image counts 512×512 pixels and rarely exceeds 2048×2048 pixels, so the performance of the algorithm is not an issue. The listing of the part of the code is presented in appendix A.3.

The user has to set $scale$ to the typical height of the object. In this case the algorithm gives smooth tracing and small number of errors. The physical sense of this algorithm is to find the path which goes through relatively high points (with height close to $scale$) keeping the length of the contour minimum. The example of the work of the aforementioned tracing algorithm and its dependence on $scale$ parameter is presented on fig. 1.3. At latter modifications the algorithm was improved to take into consideration the edges between points separated by two pixels as well, which slightly improved the smoothness of the tracing.

The algorithm number 4 has the following steps (see fig. 1.4):

1. In the beginning the molecule has to have at least two points; if it is not the case, the user has to select them manually (first one is chosen exactly at the point where the user

³A simple and efficient algorithm for finding the shortest path in a graph with non-negative edge costs [32].

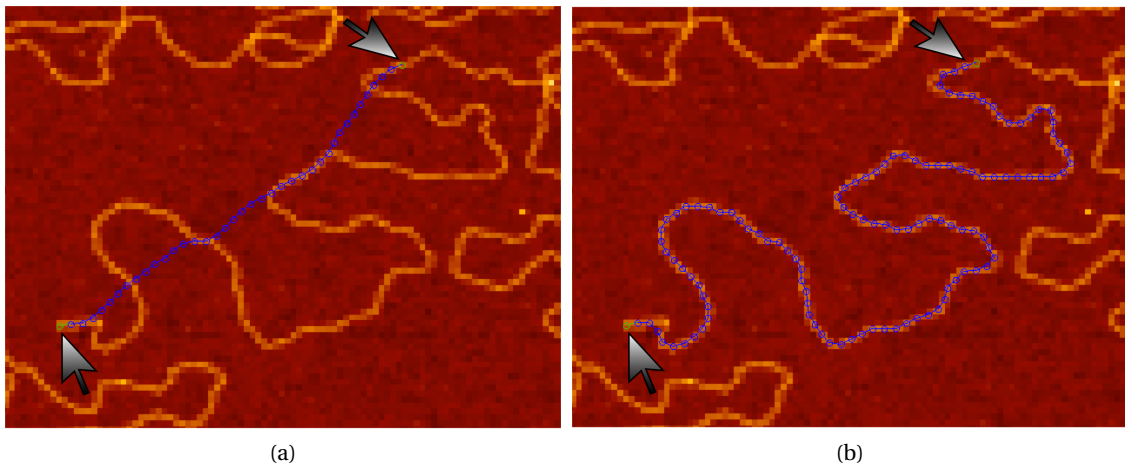


Figure 1.3 – Example of the work of tracing algorithm that follows the molecules between consequent user clicks and dependence of its quality on the adjustable parameter *scale*. On the AFM image of DNA molecules you can see the points where the user has performed clicks (showed with arrows) and the contour of the molecule found by the program. (a): *scale* = 5 nm, which is substantially higher than the apparent height of the object (about 0.5 nm in this case) gives too high priority to the shortest path and leads to wrong tracing; (b): *scale* = 0.5 nm permits the algorithm to produce an accurate contour of the molecule.

clicks the first time, the second one is chosen so that it lies in between the first one and the user's second click and is *step* away from the first one).

2. Let us call the last point of the molecule P_2 and the previous one P_1 .
3. A point P_3 is chosen so that P_1 , P_2 and P_3 lie on the same line and $|P_1P_2| = |P_2P_3|$.
4. A line a is chosen so that $P_3 \in a$ and $a \perp P_1P_2$.
5. A section along the line is taken so that its middle is P_3 and its length is *SectionWidth* in pixels. The values in the section correspond to the height of every pixel.
6. The section is plotted (x and y values correspond to the position of the point in the section and the height of the corresponding pixel on the image respectively) and one point P_{chosen} in it is chosen; depending of the parameter *UseGaussianFit* this point corresponds to the maximum x_0 of the Gaussian distribution function $f(x) = A * \exp\left(-\frac{(x-x_0)^2}{B}\right)$, where A and B are some fitting parameters, or to the maximum point of the section.
7. A point P_{next} ($P_{next} \in P_2P_{chosen}$ and $|P_2P_{next}| = step$) is added to the end of the molecule.

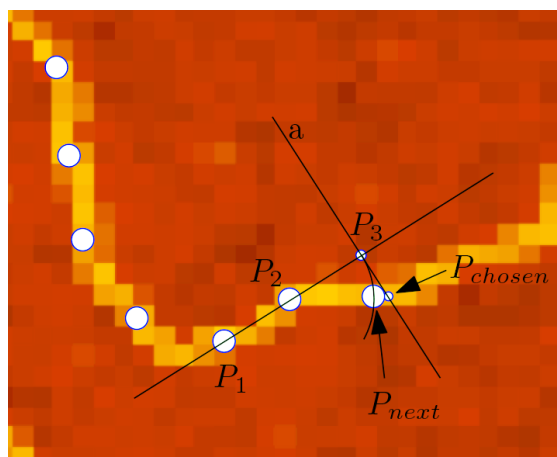


Figure 1.4 – Visualization of Semi-automatic further tracing algorithm. Points P_1 and P_2 are already existing last two points in the contour of the molecule; P_{next} is the next point determined automatically by the algorithm.

1.4 Types of objects

DNA Trace can work with different types of objects presented on a scan (Figure 1.5). They can be divided into four groups:

1. Linear objects (Figure 1.5a). This type includes long polymer molecules as DNA, protein, amyloid filaments and other lengthy objects. This type is the most common for processing by the program.
2. Circular molecules (Figure 1.5b). Plasmid DNA molecules, for example, can be included in this subtype of linear objects.
3. Branched structures. Such objects represent several linear strands unified in a tree-like conformation.
4. Non-linear shapes (Figure 1.5c). Different objects, which are non-linear, such as blobs, spherical objects, proteins and aggregates, are described by this type. The program has possibility to find such objects automatically.

A linear object is stored in memory as a contour, represented by a set of points with certain step between them (the step can be adjusted manually and should be normally set to one-two pixels). Having the coordinates of these points, one can iterate along the contour and obtain the height profile along the object, its length, end-to-end distance and correlation function. Circular objects are represented by the same structure, except for the last point, which is connected to the first one. Branched structures are stored as sets of linear objects. Non-linear shape is represented by a set of pixels of the edge of the object. To calculate the surface, the volume or the distribution of the height of the points of the objects, the software uses

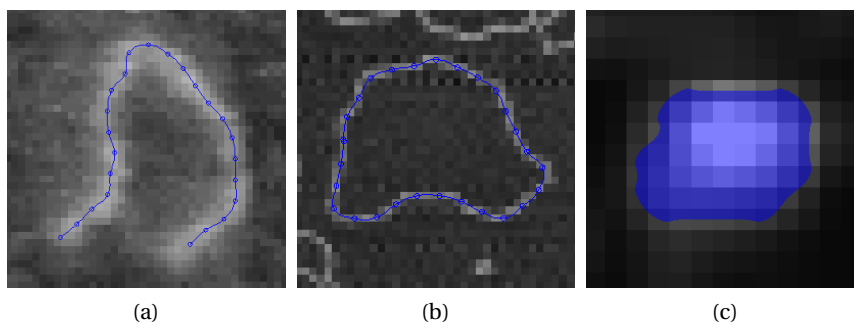


Figure 1.5 – Possible types of objects of *DNA Trace*. (a): linear molecule (DNA); (b): circular molecule (DNA) (c): non-linear shape (a diamond nanocrystal). Images have sizes 50×50 , 100×100 and 40×40 nm.

an algorithm that iterates over the pixels enclosed in a polygon. We have used an efficient Polygon Fill Algorithm from [43] and implemented it in Java. However, for common applications with AFM images not exceeding 512×512 pixels the simplest algorithm (iterate over all of the pixels of the scan and test each pixel if it is inside the polygon) would work fine as well.

All the objects found on the image can be saved, along with the raster itself, into a proprietary file format *DNT*. The file itself is a ZIP-archive with several entries inside it. The first one is a file named *scan.xml*. Inside it there is stored all the information about the traced objects and about the pixels-nanometers correspondence in the XML format. Then there is one entry per each raster in the image (there are commonly several of them: height, amplitude, phase channels, *etc.*). These entries are in the binary format and every one of them consists of a two-dimensional array of eight bit floating point numbers. The values in the arrays are in the corresponding units (typically *nm* for the height channel and *V* for the other ones). Finally, there is a text entry named *scan.parameters*, where all the additional information is stored (scan rate, x and y offsets, setpoint, captured date, scan angle, *etc.*) in the textual format.

1.5 Results and Discussion

Having traced the molecules on the image, the user can get different information out of the files. The user can either export the coordinates of the molecules to the text file or make a plot and save it to open in an external program or calculate all the needed characteristics directly in *DNA Trace*.

One of the most interesting characteristics is the persistence length. To calculate this quantity

one uses the bond-bond correlation function [61]:

$$\langle \cos\theta(s) \rangle = \exp\left(-\frac{(d-1)s}{2l_p}\right), \quad (1.1)$$

where s is the contour length, d is the dimensionality of the system, θ is the angle between the tangents to two points on the molecule separated by the distance and l_p is the intrinsic persistence length. The averaging is done over all possible pairs of points of the molecule, for which s is constant, and then over all of the molecules.

DNA Trace calculates $\langle \cos\theta(s) \rangle$, plots it versus s and makes a fit with the function mentioned before (eq. (1.1)) or any other. From this the user immediately gets the value of l_p (fig. 1.6a).

If other polymer physics quantities are investigated, it is interesting to calculate the average end-to-end distance $\langle R_{ee}^2(s) \rangle$. For the calculation of the scaling exponent or the fractal dimension of the molecule, one can plot $\langle R_{ee}^2(s) \rangle$ versus s . According to the polymer physics theory, the end-to-end distance will be then expressed in the following way [28]:

$$\langle R_{ee}^2(s) \rangle \underset{s \rightarrow \infty}{=} A s^{2\nu}, \quad (1.2)$$

where A is some constant and ν is the scaling exponent. The fractal dimension will be simply: $d_f = 1/\nu$. Plotted on a double logarithmic scale, this plot can give an information of the scaling exponent as the slope of the local tangent to the curve. Moreover, the probability distribution for $\langle R_{ee}^2(s) \rangle$ at fixed s can be plotted as well [127].

When transformed to

$$\frac{1}{2} \frac{d \ln \langle R_{ee}^2(s) \rangle}{d \ln s} = \nu(s), \quad (1.3)$$

the scaling behavior of the polymer with finite s can be studied.

Another important possibility is that the program determines the length of the molecules, the quantity, important for the characterization of the sample [108]. Using this quantity, *DNA Trace* reports the values of the lengths as a mean value, as a histogram or as raw data (fig. 1.6b).

Although the height of the samples can be distorted when obtained by SPM, it is interesting to get this information from the sample as well. For example, some objects show periodicity along the molecule or molecule array [1, 93, 53, 35]. In this case *DNA Trace* can help to plot a height profile along the molecule. Applying the Fourier transform to it, one can get the period.

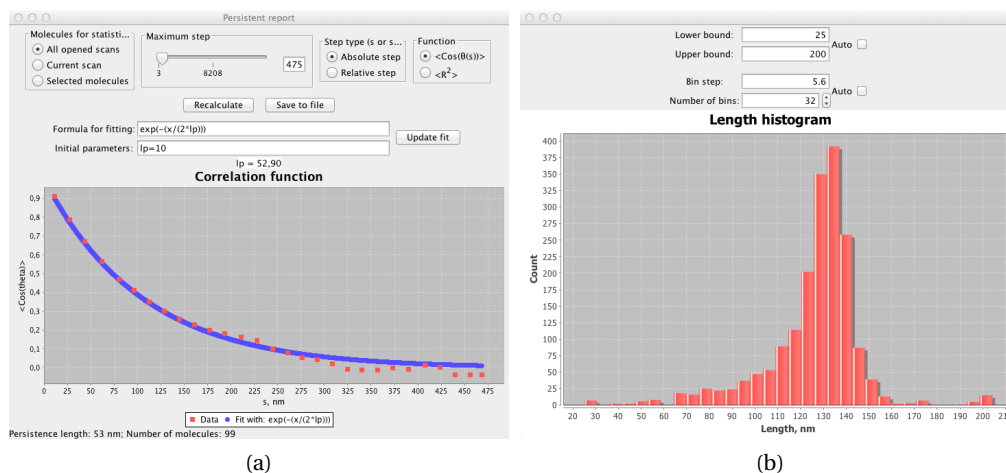


Figure 1.6 – Screenshots of output of *DNA trace*. (a): Correlation function $\langle \cos\theta(s) \rangle$ versus s with a fit $\langle \cos\theta(s) \rangle = \exp(-\frac{s}{2l_p})$ gives $l_p = 52.9$ nm; (b) histogram of the contour length of the set of DNA molecules.

1.6 Current applications

DNA Trace is widely used in our lab and permits us to quickly extract data from the images [21, 22]. We use it to characterize the height of the sample, for example, to calculate periodicity or conformation of the molecules, to extract background information to check the quality of modification of the surface, to flatten images or apply some filters before presenting them. The main usage, of course, remains the measuring of the persistence length of the molecules.

With the use of *DNA trace* software we are able to make analysis of amyloids formed from Huntingtin Protein with different mutations. This analysis starts from check if the images obtained with AFM are properly flattened. Then, analysis based on *DNA trace* consists of measuring the length and height of structures (average value for each amyloid) and their shape. We also using the function of plotting the histograms form the table of traced molecules to see statistical distributions of results. Other articles containing data obtained with this program are prepared to be submitted.

1.7 Conclusions

In this chapter we have described the software *DNA Trace* intended for quick processing of images obtained by AFM. We have succeeded to create a versatile tool, which makes the routine extraction of meaningful data from AFM scans fast, easy and precise. First of all, it works directly with the files obtained with most of common microscope types, so that the user does not need to open the images with original software first to convert them. This feature

permits one to use *DNA Trace* as a browser for the files recorded by different setups.

Then, the software incorporates a set of filters that process raster images contained in these files. Of the most importance are *flattening* algorithms. Flattening needs to be applied to AFM images in most of cases, and *DNA Trace* can be adjusted to apply a selected filter to an image automatically upon its opening. Besides common algorithms such as subtraction of a n^{th} order polynomial line-by-line or a n^{th} order polynomial surface from the whole raster that are typical from the field, *DNA Trace* offers so-called *smart flattening*, which determines the pixels of the background only and fits the image line-by-line with a polynomial, considering only background pixels. Such a filter permits one to avoid well-known flattening artifacts and get an image with absolutely flat background.

Main advantage of the software, in our opinion, is the feature of tracing of objects of different type on the surface. Among them, the most useful are the algorithms permitting to find contours of linear objects, for example, DNA molecules. These contours are stored in files of proprietary format *DNT*, which permits to reopen images with *DNA Trace* and calculate data on the objects on them without the need to re-trace molecules. Once all the contours are known, the software can extract information of different type from the sample, such as length, height, angular distributions, end-to-end distances, persistence length and so on. This information helps the researchers to characterize their samples not only qualitatively, but also quantitatively and do it in a quick and precise manner.

Finally, the software is cross-platform, user-friendly and easy to use. The significant amount of data collected by it not only in our group, but by our colleagues in other research institutions show the importance of *DNA Trace* and the need of its further development. We propose the implementation of particle analysis and a tool, which can measure the mapping of some objects with respect to the position on linear molecules (for example, that can be used to determine exact position of the proteins, which bind specifically to some sites of DNA) as the mostly desirable features.

2 Diffraction grating on a cantilever

Measure what is measurable,
and make measurable what is not so.
— Galileo Galilei

The resolution and sensitivity of AFM technique is pushed nowadays towards its extreme limits. One can study the topography of different samples with a resolution in the order of nanometers or even angstroms [58, 129]. Interaction forces can be measured with piconewton sensitivity using force spectroscopy technique [110] (see chapter 5 for further information). To reach such limits of resolution and measurement precision, one has to determine the position of the tip, *i.e.* bending of the cantilever as accurate as possible. In this chapter we will discuss the detection of the cantilever bending and the ways to improve it.

A typical AFM cantilever bending detection system consists normally of a laser and a four-quadrant photodiode measuring the position of the reflected beam (see fig. 2.1a). It permits to detect the bending of the lever in two axes using the following principle. A laser beam reflected from the backside of a cantilever falls onto a photodiode, which has four independently quadrants. The diode can be adjusted, so that when the cantilever is out of contact with the surface, the reflected beam is in the center of the diode and the signal from all of the four zones is equal (see fig. 2.1b). When the cantilever is bent in one or two directions, the laser beam is shifted. The signals $A+B-(C+D)$ and $A+C-(B+D)$ give us vertical and horizontal displacement of the laser spot respectively and permit us to measure the deflection and torsion of the lever. From deflection, the force acting perpendicularly to the lever can be determined. This technique is well established nowadays, is relatively easy to implement and gives sufficient sensitivity for most of AFM applications.

However, the sensitivity of a classical detection scheme can be increased, which is of great interest, especially if one keeps in mind the sub-nanometric precision of the movement of piezo scanner. We have shown the possibility to improve the angular sensitivity (see the de-

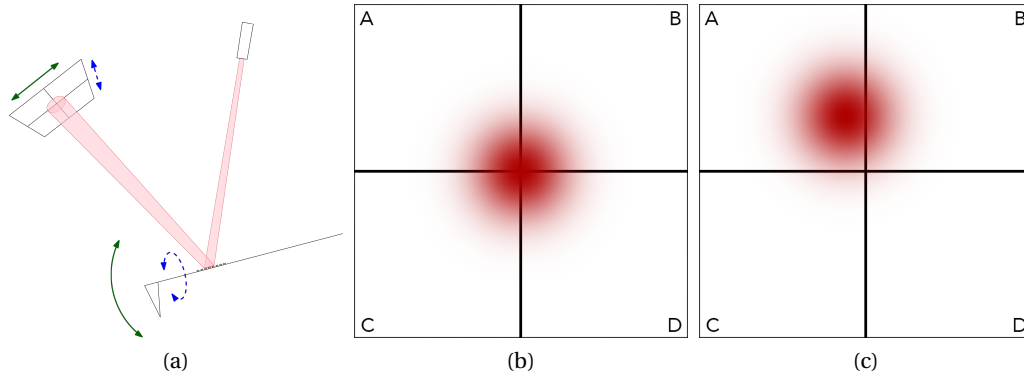


Figure 2.1 – Scheme of cantilever deflection measuring system. (a): principle of working: when the lever is bent, the laser beam spot is displaced in one direction (shown with the green solid line), while when it is twisted, the spot moves in the other direction (blue dashed line). (b): four-quadrant photodiode with a laser beam in the center of it. The signal from A,B,C and D zones is equal. (c): when the cantilever is bent in one or two directions, the laser beam is shifted from the center of the diode. The signal $A+B-(C+D)$ gives vertical displacement and $A+C-(B+D)$ gives the horizontal one.

scription of this term below) significantly using the following technique. If a beam is reflected from the backside of the cantilever, the incident angle of the laser beam with respect to the normal to the grating surface, θ_0 , is equal to the reflected angle θ , so $d\theta = d\theta_0 = d\phi$, where ϕ is the angle between the backside of the lever and the surface. The goal is now to increase the ratio $k = d\theta/d\theta_0$.

If one etches a diffraction grating on the same backside and observes diffractive reflection rather than the mirror one, θ_0 and θ are related with the following equation:

$$l(\sin\theta - \sin\theta_0) = n\lambda, \tag{2.1}$$

where n is the order of diffraction, l is the period of the grating and λ is the wavelength of the laser. Developing eq. (2.1), one obtains:

$$k = \frac{d\theta}{d\theta_0} = \frac{\cos\theta_0}{\cos\theta}. \tag{2.2}$$

Now k can be as large as possible, if θ is close to 90° . In a real setup it is easy to achieve k of ten-twenty units with $\theta \approx 70^\circ - 80^\circ$. We would like to stress attention to the difference of the following terms used widely in this chapter. The (geometrical) angular sensitivity represents the displacement of a laser beam along the surface of the detector compared to the bending of the cantilever. It is proportional to k and can be sufficiently increased by the use of a grating on the backside of the lever. However, the total sensitivity, which corresponds to the change of the output signal of the detector compared to the bending of the probe, is not necessarily

proportional to the angular one. Thus we call *angular sensitivity* $k = \frac{d\theta}{d\theta_0}$ and *total sensitivity* $\alpha = \frac{dS}{d\theta_0}$, where S is the output signal from the sensor. These terms are not general and are chosen for this chapter for the explanatory purposes. The difference between them will be discussed next.

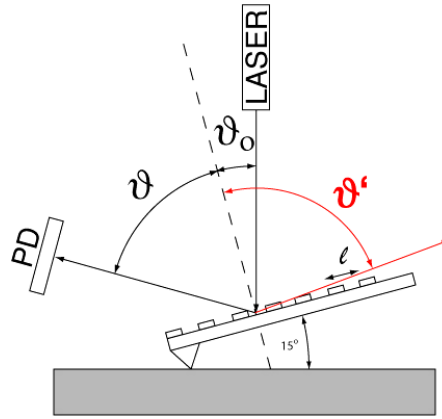


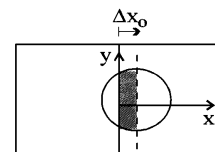
Figure 2.2 – Proposed technique of detection of the tip position by the use of light reflected from grating etched on the backside of an AFM cantilever. Image taken from [113].

2.1 Photodiode

Regardless of theoretical infinitely large k , the practical sensitivity of the deflection detection system depend on the photodiode as well. In case of an aforementioned split detector, its sensitivity is inversely proportional to the angular divergence of the beam [52]. More precisely, if the laser power is equal to P , distance from the lever to the detector is L and the size of a beam spot on it is a (which is much less than L), the total sensitivity $\alpha = dS/d\theta$ (S is the output signal from the detector) can be expressed as $\alpha = 2k\sqrt{2\pi} \frac{e\eta P}{h\nu} \frac{L}{a}$, where k is the angular sensitivity, which is equal to unit for a classical mirror reflection detection scheme, e is the electronic charge, η is the quantum efficiency, h is Planck's constant and ν is the frequency of the light [52]¹.

The formula shows the dependence $\alpha \sim 1/\beta$, where $\beta = a/L$ is the angular divergence of the beam. If we use reflection from aforementioned grating, the geometrical angular sensitivity is amplified by the factor k , but the angular divergence $\beta \sim \frac{L}{a} \frac{d\theta}{d\theta_0} = k \frac{L}{a}$ is also amplified by the same factor (see fig. 2.3, (4)), so the advantage cancels out and the total sensitivity remains the same: $\alpha' \sim kPL/ka = PL/a \sim \alpha$. The aforementioned formula underlines an important fact:

¹ The derivation of the formula comes from the following principle. If the laser beam spot (which is assumed to be a Gaussian beam) moves along the detector, the total output signal is increased by twice the signal from the part Δx (see the enclosed figure). Integration of the part of a Gaussian beam inside this area gives the aforementioned formula. More complete derivation can be found in [52].



Chapter 2. Diffraction grating on a cantilever

the total sensitivity has a fundamental physical limit - the diffractive divergence of the laser beam (if it would be not the case, we could infinitely increase both the angular and the total sensitivity just by placing the detector far enough from the cantilever).

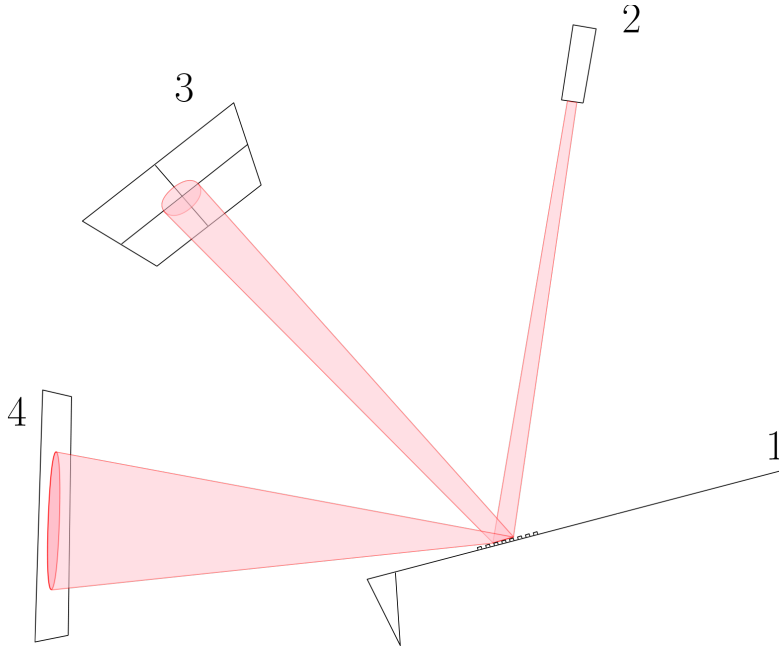


Figure 2.3 – Schematic of the detection of a laser beam (2) reflected from a grating etched on the backside of the cantilever (1) in mirror reflection mode (3) and diffraction reflection mode (4). In the first case a four-section split photodiode is utilized, in the latter one – a duo-lateral position-sensitive photodiode.

In order to circumvent the just mentioned limitation of the detection scheme, we utilized a duo-lateral diode to keep the advantage in the sensitivity. First of all, its theoretical *minimum detectable angle* is smaller and sensitivity bigger by the factor of $\sqrt{2\pi}$ compared to a split detector [52]. But the main advantage is that its sensitivity does not have the dependence on the angular divergence of the beam, admitting that the size of the beam is sufficiently smaller than the size of the detector. These detectors permit us to measure the deflection of a cantilever in two modes: mirror and diffractive reflection, with the total sensitivity increased by the geometrical factor of k in the latter case.

Another important advantage of the proposed technique is the following. In most of AFM setups the surface of the backside of a cantilever is almost parallel to the surface of the sample, so the parasite light reflected from the latter falls onto the detector along with the beam corresponding to the reflection from the cantilever. The positions of these two spots lie close to each other. In the case of the most common AFM cantilevers this problem appears to be minor, as the width of the lever is comparable or bigger than the laser spot on it (typical scale of both the laser spot and width of the lever is in order of several tens micrometers).

Consequently, the aforementioned effect does not play a big role in the detection. But for some applications, including high-speed AFM imaging, one has to use tips with high resonance frequency, and as a consequence with high spring constant. Such cantilevers are typically very narrow and have width down to 0.5 - 1 micron, so the parasite reflection of the laser beam from the sample cannot be neglected anymore. This reflection can seriously disturb the measurement of the deflection of the cantilever. In the scheme proposed by us the detector is placed at the angle far from mirror reflection (for example, 18° and 70° respectively in one of our experiments), so that the beams reflected from the sample and cantilever surface do not interfere with each other.

2.2 Results and Discussion

We have performed a set of experiments using an AFM tip with a grating etched on the back-side of it. The first one was an amplitude diffraction grating of a period of $1\ \mu\text{m}$ etched by the Focused Ion Beam (FIB) (figs. 2.4a and 2.4b). In this grating 500 nm aluminium strips were interleaved with 500 nm aluminium-free grooves. The tip used was aluminium coated silicon levers CSC38 from MicroMash, Tartu, Estonia, with the length of $250\ \mu\text{m}$, width $35\ \mu\text{m}$ and thickness of $1\ \mu\text{m}$. The thickness of the coating was 30 nm and the nominal spring constant $0.08\ \text{N/m}$. As duo-lateral detectors are not widely used in AFM setups and due to different detection angle, we had to prepare custom cantilever holder and electronics.

We used CW He-Ne laser beam of 5 mW in air, which was focused by a lens with a focal distance of 15 mm to a spot of about 20 - 40 microns. First we used a two-quadrant photodiode as a detector to prove the possibility the proposed approach. A 0.17 mm thick cover glass used as a sample was mounted onto a NIS-70 scanner (Nanonis, Israel). We set $\theta_0 \approx 18^\circ$, which gave $\theta \approx 70^\circ$ for diffracted reflection of the order $n = 1$. The incidence angle was easy to control, by observing the reflection of the order $n = -1$. In our case auto-collimation condition, *i.e.* $\theta_0 = -\theta$, is satisfied, when the incidence angle $\theta_0 \approx 18.4^\circ$. Therefore we chose the incidence angle, so that diffractively reflected beam with $n = -1$ coincided with the incident one.

We applied triangular signal to the piezo-scanner corresponding to the periodical engagement/disengagement of the surface towards the tip and recorded force curves read out by the split photodiode with the diameter of 25 mm in two different positions corresponding to mirror and grating reflection (the setup was chosen such that the power of these two signals were close to each other and reached about 0.5 - 0.8 mW). Both scanner control and signal recording was performed using a custom LabVIEW program. The results presented on fig. 2.5 show the increase of the amplitude of the signal of about 2.2 times, which is not far from the theoretical $\cos 18^\circ / \cos 70^\circ \approx 2.8$. This proves the concept of the increase of the angular sensitivity by using diffractive reflection mode. There were no problems to increase k further by using θ up to 80° , but we do not present quantitative results with such angles, as the parameters of diffraction

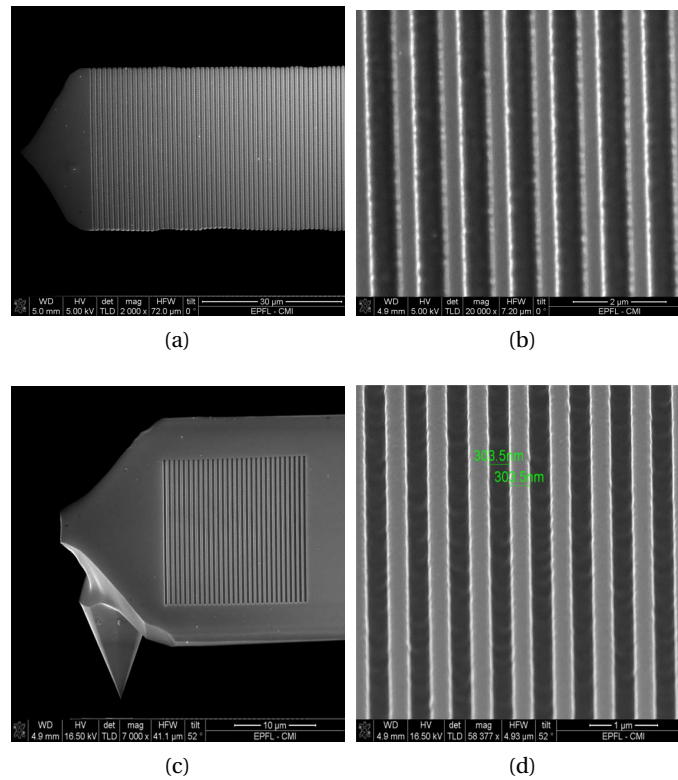


Figure 2.4 – SEM images of the backside of CSC38 AFM lever with a diffraction grating. (a),(b): 1 micron-period; (c),(d): 600 nm-period. The grating was produced by focused ion beam; technical limitations of this particular device prevented us to etch a grating on the whole surface of the lever in the latter 600 nm-period case as it required the highest resolution. Image taken from [113].

and mirror reflection beams differed too much and their direct comparison would not be appropriate. Please note that the aforementioned experiment does not claim the increase of the total sensitivity but only of the angular one (as we adjusted the photodiode to keep the size of the laser beam the same, the noise was multiplied by the same geometrical factor k).

In the next series of experiments we switched to high linearity position sensing detector (duo-lateral photodiode, UDL Sensors Inc., USA) with the size of an active area 5×1 mm (see fig. 2.6). The diode was connected to a custom-made amplifier and the difference of voltage between its two outputs was measured (see fig. 2.7). We used the aforementioned grating with a period of 1 micron along with another one with a period of 600 nm produced by the same technique (figs. 2.4c and 2.4d). The latter gave us the most interesting results when illuminated with a He-Ne laser. In such a case the grating can reflect only one single order of diffraction with $n = -1$ if the incident angle θ_0 is greater than 3.1° . This order of diffraction corresponds to the reflection to the same side as the incident beam. We have successfully observed force curves recorded with duo-lateral photodiode for the angles representing case

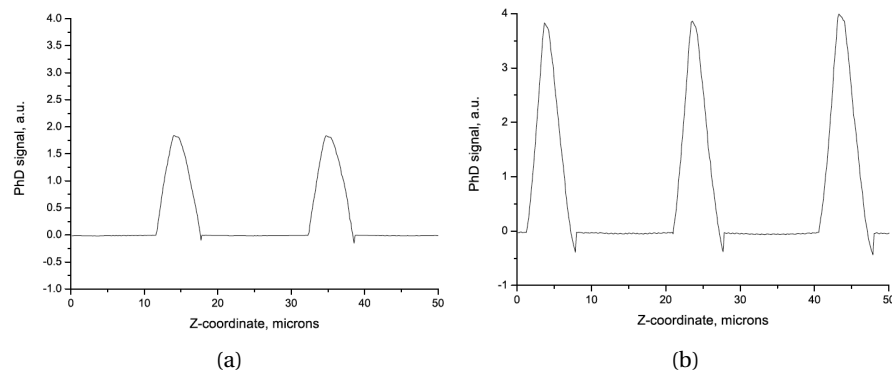


Figure 2.5 – Force curves recorded by the photodiode placed in two different positions: (a) - mirror reflection, (b) - diffractive reflection.

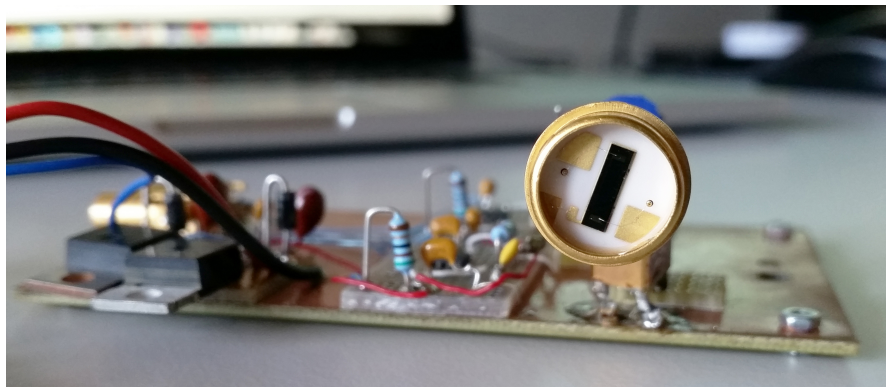


Figure 2.6 – Picture of a position sensing detector (UDL Sensors Inc., USA, active area 5×1 mm) used in our setup.

1 and 2 in Table 2.1. The power of the reflected laser beam reached 1.5 mW and almost the whole surface of the detector was illuminated when the distance between it and the cantilever was about 30 mm. We successfully registered cantilever deflection and recorded force curves with this setup, which appears to be promising for the further experimental application of the proposed deflection detection scheme.

Finally, we etched grating with a period of 720 nm on the backside of an ultra-short cantilever designed for high-speed AFM (USNMCB, NanoWorld AG, Neuchatel, Switzerland; resonance frequency 5 MHz, size $10 \times 5 \mu\text{m}$). An EM image of the tip with the grating is presented on fig. 2.8. We successfully registered diffractive reflection of He-Ne laser beam with the power reaching 0.4 mW (the laser was focused onto the end of the tip by a lens with the focal distance of 15 mm). Using the first diffraction order and the incident angle of 6° , one obtains a reasonable angle $\theta = 80^\circ$, which permits one to isolate easily the cantilever deflection signal from the parasite light reflected from the sample. The angular sensitivity increases by a factor

Chapter 2. Diffraction grating on a cantilever

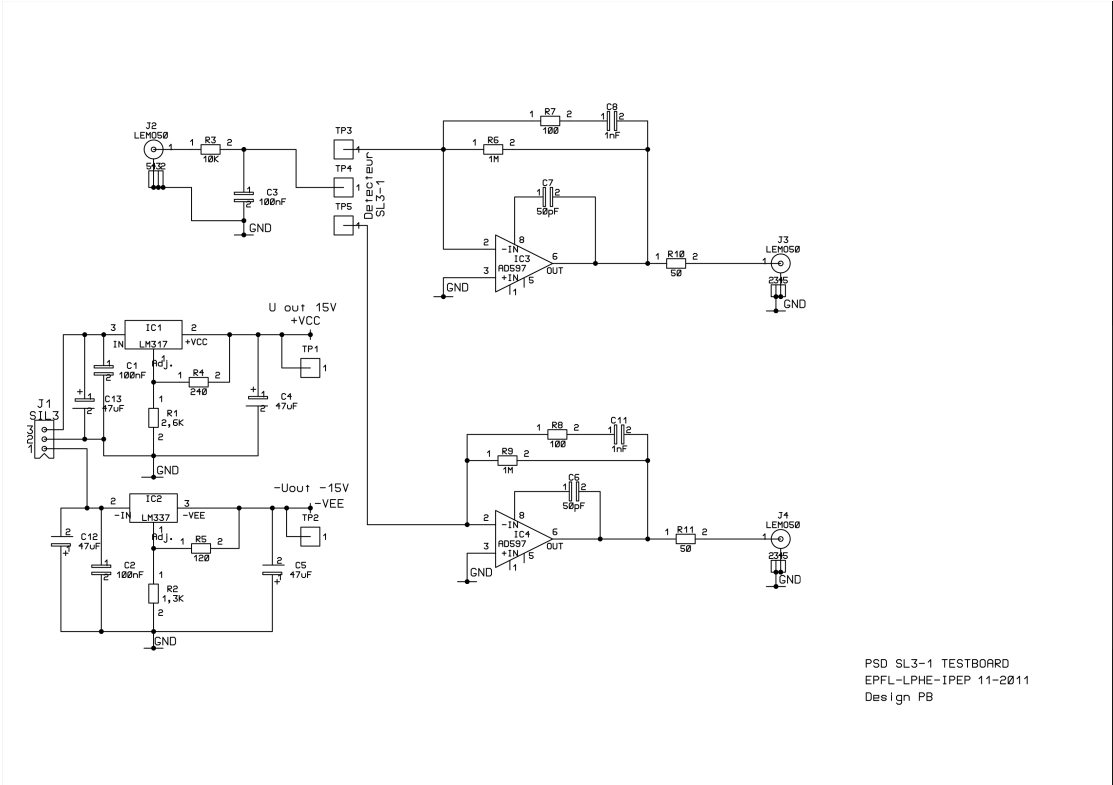


Figure 2.7 – Circuit diagram of an amplifier used to read out the signal from position sensing detector.

Case	$\theta_0, ^\circ$	$\theta, ^\circ$	$k = \frac{\cos\theta_0}{\cos\theta}$
0	3.1	90	∞
1	3.3	85	11.5
2	4	80	5.7

Table 2.1 – Incident and reflected angles along with the sensitivity increase coefficient k for He-Ne laser, grating with a period of 600 nm and $n = -1$.

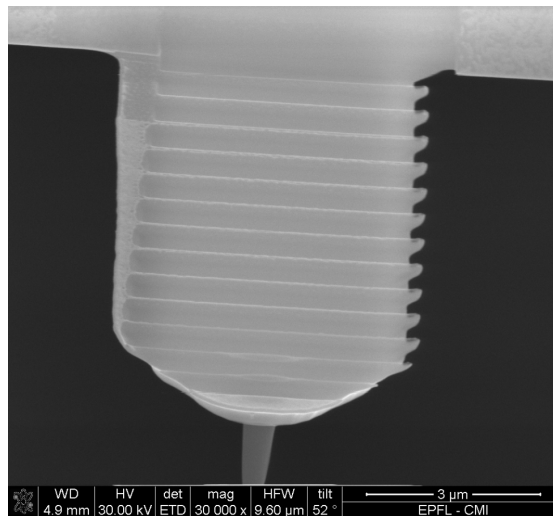


Figure 2.8 – SEM image of the ultra-short cantilever (USNMCB, 5 MHz, NanoWorld AG, Neuchâtel, Switzerland) with a 720 nm-period grating etched by focused ion beam. This particular type of cantilevers is specially designed to be used in high speed AFM setups. Image taken from [113].

of $k = 5.7$, which can also bring, combined with duo-lateral detection diode, the increase of the total sensitivity to a ultra-fast AFM device.

2.3 Conclusions

In this chapter we described the idea of the increasing of the sensitivity of cantilever deflection and force measurement in AFM device and its implementation. There exist a patent stating the possibility to use such an approach [69], but up to our knowledge there existed no implementation of it. We would like to mention certain articles exploring the usage of diffraction grating in combination with AFM as well [138, 124, 62, 31]. These studies mention the possibility to form a grating on the sample or on an AFM tip and observe the diffracted light from it while scanning of the sample. However, the grating is used only as an interferometer in this case: the mirror-reflected or passing through laser beam interferes with the diffractively reflected or transmitted beam and the intensity change is recorded. These papers do not discuss the effect of the increase of the sensitivity of measuring of the cantilever deflection or acting force.

We have already discussed the difference between the angular and total sensitivity of the deflection detection system. The first one is pure geometrical and is equal to unit for a mirror reflection from the cantilever. It can be significantly increased by the use of a diffraction grating on the backside of a cantilever (theoretically, it can go up to infinity for $\theta = 90^\circ$). We would like to mention other geometrical possibilities to increase angular sensitivity k . For example, reflection from a spherical or cylindrical surface can rise k ratio as well [52, 137]. In this case

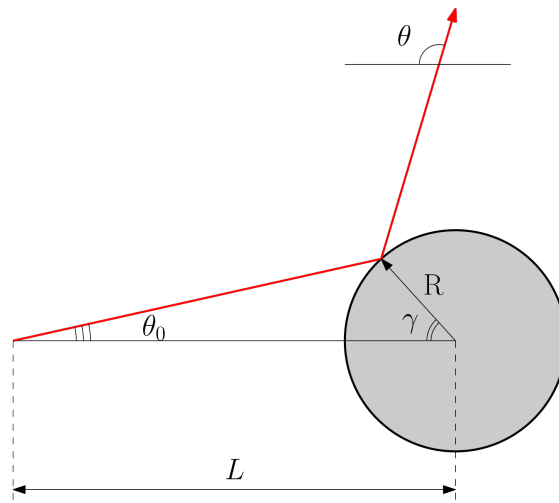


Figure 2.9 – Geometrical improvement of the angular sensitivity by the use of reflection from a spherical surface.

$\theta = \theta_0 + 2\gamma$, where $\gamma = \sin^{-1}(L \sin \theta_0 / R) - \theta_0$ (see fig. 2.9 for a geometrical explanation). This leads to $k = \frac{d\theta}{d\theta_0} = 2L/R \cos(\theta_0) / \sqrt{1 - (L \sin \theta_0 / R)^2} - 1$ and converges to infinity, when $\sin \theta_0$ is close to R/L (*i.e.* the beam is tangent to the sphere). Note that both of described schemes lead to the rise of both angular sensitivity and angular divergence of the beam by the same factor k .

The total sensitivity is the most important factor that describes the increase of the output signal from the detector with respect to the change of the deflection of the lever or the force acting to the tip. Although it is normally proportional to the angular sensitivity, it can also be inversely proportional to the angular divergence of the laser beam, if a split photodiode is used. The final goal of this study was to improve the total sensitivity of the detection scheme.

We have obtained approach-retract force curves and shown an effective increase of angular sensitivity by the factor of 2-12 times. As described before, it does not lead to the increase of the total sensitivity when a common split photodetector is used. To utilize the geometrical advantage of the proposed scheme, we used duo-lateral photodiode, which does not have the dependence of the measuring sensitivity on the angular divergence of the beam and permitted us to have the increase of the total sensitivity close to the advantage in angular sensitivity. This advantage can be especially useful when applying to force spectroscopy technique, which is widely used by many researchers including our laboratory (see chapter 5 of the present work). Such technique require great precision in measuring the exact position of the tip. We suppose that another good alternative to split detector or duo- (tetra-) lateral photodiode would be CMOS active-pixel image or CCD sensor. Such pixel arrays are widely used nowadays in digital cameras and cell phones. The technology is well established and low-noise and relatively low-cost sensors could permit one to detect the position of the reflected laser beam with high

precision.

We have also successfully proven the concept of applying this approach to high-speed AFM. Ultra-small cantilever with a grating etched on it permitted us to measure its position placing the detector so that the parasite beam reflected from the surface of the sample do not get into it. Such approach can significantly simplify the setup of an AFM using these narrow levers and bring better Z position sensitivity.

3 DNA relaxation

If a man insisted always on being serious,
and never allowed himself a bit of fun and relaxation,
he would go mad or become unstable without knowing it.
— Herodotus

3.1 Introduction

The physics of linear polymers and of linear DNA in particular is well described [46, 28]. DNA can be easily imaged using AFM for what it has to be deposited on a surface. During the deposition molecules undergo conformation changes. Polymer molecules are confined to two dimensions, but their configuration may vary. Some papers report, that the molecules appear to be projected onto the surface from the three-dimensional quasi-equilibrium [19, 127], while the others show images where polymers are relaxed on the surface and adopt two-dimensional quasi-equilibrium [136, 89]. It is also known that DNA molecules deposited on the surface can have 2D or 3D conformation depending on deposition conditions [105].

Indeed, after a polymer comes in contact with the surface, it can be fixed completely or have a possibility to detach from or move along the surface. This process depends on the interaction between the surface and the polymer and the timescale of the observation. Below we will show, that the process, when polymer molecules are not strongly fixed and change their configuration on the surface, can be relatively slow and happen in the timescale from seconds to minutes, so it is observable by common experimental AFM conditions.

While all molecules on the surface formally are in the two-dimensional space, the conformation of the molecules can vary. If the polymer is confined to two dimensions, but can move along a very thin layer on the surface, the molecules can reach the "pure" two-dimensional state after some long time. To be precise, there cannot be any self-intersection in a polymer

chain in the pure two-dimensional state, which is rarely observable in practice for long polymers. Nevertheless, statistically rare chain self-intersections can be neglected and the polymer can be described in terms of two-dimensional statistics. If the polymer was not confined, but was rather immediately projected onto the surface from volume, it will physically be in two dimensions, but will rather exhibit most of the properties, corresponding to three-dimensional structures.

There exist two common ways to deposit DNA molecules, which have negative charge in solution onto the surface of mica, which is also negatively charged [105]. The first one is to treat surface with APTES, which results in a reversal of the charge of the surface and leads to the adsorption of DNA molecules. It commonly leads to the strong interaction between deposited molecules and the surface [82], so they appear to have 3D-2D projection configuration [127]. Although we have observed the change of the conformation of DNA molecules deposited onto APTES-treated mica in some experiments, the deposition time was in order of hours in that case and up to now we do not have enough experimental data to study this process in details.

The other way to treat mica surface consists of using divalent ions (mostly Mg^{2+}) [128, 18]. This results in weaker binding, so molecules are not completely fixed and can adopt a two-dimensional conformation [136]. This method is proved to be more interesting for us, as it enabled us to observe the change of the conformation of DNA molecules within more practical timescale of 10–300 seconds.

To characterize the conformation of the polymer molecules on the surface, we will first of all recall eq. (1.2). This equation describes the scaling of polymer end-to-end distance with respect to the increase of its length. For the simplest polymer theory, freely-jointed chain with N completely independent segments \vec{r}_i ($\langle \vec{r}_i \vec{r}_j \rangle = 0$ when $i \neq j$) of the length l_0 each and the total contour length $L = Nl_0$, the mean squared end-to-end distance $\langle R_{ee}^2 \rangle$ can be calculated in the following way:

$$\begin{aligned} \langle R_{ee}^2 \rangle &= \langle \overline{R_{ee}^2} \rangle = \left\langle \left(\sum_{i=1}^N \vec{r}_i \right)^2 \right\rangle = \left\langle \sum_{i=1}^N \vec{r}_i^2 + \sum_{i \neq j} \vec{r}_i \vec{r}_j \right\rangle = \\ &= \left\langle \sum_{i=1}^N \vec{r}_i^2 \right\rangle + \left\langle \sum_{i \neq j} \vec{r}_i \vec{r}_j \right\rangle = \left\langle \sum_{i=1}^N l_0^2 \right\rangle = Nl_0^2 = Ll_0, \quad (3.1) \end{aligned}$$

so $R_{ee} \sim L^{1/2} = L^\nu$ and the scaling exponent $\nu = 1/2$.

Nevertheless, real polymers have chains with finite diameter, which somehow interact with themselves, at least due to the fact that two segments cannot occupy the same space. This effect is called *excluded volume*. Flory proposed a simple but approximate theory, that takes

into account the excluded volume effect and predicts the scaling exponent ν . In brief, it can be described in the following terms [28]. If a polymer is in a good solvent (*i.e.* pairs of monomers repel each other), its repulsive energy can be estimated as

$$F_r \sim \nu_{ex} \frac{N^2}{R^3}, \quad (3.2)$$

where ν_{ex} is the excluded volume of each monomer, N - the number of monomers and R - end-to-end distance of the chain. In the same time, the entropy of the system is

$$S_N(R) = k_B \log P(R, N) \sim -\frac{R^2}{Nb^2}, \quad (3.3)$$

where $P(R, N)$ is the probability of the polymer to have certain R with current N and b is the length of a monomer. Combining eqs. (3.2) and (3.3), one gets the following expression for the total free energy:

$$F(R) = F_0 + c_1 \frac{R^2}{Nb^2} + c_2 \nu_{ex} \frac{N^2}{R^d}, \quad (3.4)$$

where F_0 , c_1 and c_2 are some dimensional constants, that do not depend on R . Minimizing free energy $F(R)$, one gets the following:

$$\frac{dF}{dR} = \frac{2c_1 R}{Nb^2} - \frac{c_2 \nu_{ex} d N^2}{R^{d+1}} = 0, \quad (3.5)$$

$$R^{d+2} = \frac{c_2 \nu_{ex} b^2 N^3}{2c_1}, \quad (3.6)$$

$$R \sim N^{3/(d+2)}. \quad (3.7)$$

Finally, this theory gives the following dependence of ν on the dimensionality of the system d [28]:

$$\nu = \begin{cases} \frac{3}{2+d}, & \text{if } d \leq 3, \\ \frac{1}{2}, & \text{if } d > 3. \end{cases} \quad (3.8)$$

The special case for $d > 3$ is excluded from the aforementioned calculations, because the excluded volume effects cannot make the polymer more swollen, than the Gaussian freely-jointed chain, so $\nu = 1/2$ in this case. More complex calculations result in a value of ν presented on table 3.1, which are very close to the ones predicted by Flory (eq. (3.8)).

As one can see, the scaling exponent is closely related to the dimensionality of the system, this fact pushed us to choose it as one of the measures of the conformation of the molecules.

d	≥ 4	3	2	1
ν	0.5	0.588	0.75	1

Table 3.1 – Dependence of the scaling exponent on the dimensionality of the system.

The exponent ν was already determined previously [127, 136], so we decided to extract the dimensionality of the system from our data by calculating the end-to-end distance of the molecules, the scaling exponent, and applying this to the formula

$$d = \frac{3}{\nu} - 2. \quad (3.9)$$

We have to stress here, that the aforementioned scaling behavior is observed only for sufficiently long polymer molecules. Any physical polymer has some finite bending stiffness, so it can be represented as a rigid rod on short length scales and $\nu = 1$ in this case. Then there typically is a range of s , at which the molecule no longer behaves as a stiff rod, but the excluded volume effects are still not present [85]. On these length scales the molecule is not long enough for the monomers feel the presence of each other and remains in the so-called *intermediate Gaussian regime* with $\nu = 0.5$, independent on the dimensionality of the system. This intermediate range of s depend both on the polymer and on the value of d . We do not analyze this regime in details in the current Thesis.

Another characteristics, which depends on the dimensionality, is the correlation function (see eq. (1.1)). The intrinsic persistence length of the polymer l_p is a physical characteristic of the molecules and depends on the stiffness of the chain. One can show, that the same molecules will exhibit different correlation function in 2D and 3D spaces. Nevertheless, the values of d and l_p are correlated in eq. (1.1), so one cannot extract them separately. Our idea was to use the value of $l_p = 50$ nm for DNA, which was measured by many different methods previously [105, 8], and calculate the value of d .

3.2 Experiment

3.2.1 Materials and methods

λ -DNA (solution λ -phage DNA, 48502 bp long, which was cleaved by restriction enzymes to give a mixture of total contour length L ranging from 1503 bp up to the maximum 48502 bp) dissolved in water with 1 mM Mg^{2+} was deposited on freshly cleaved mica. After certain relaxation time t_0 varying from 10 seconds to 5 minutes, the sample was rinsed with UHQ water and dried with air, thus preventing the molecules from further relaxation. Images were obtained using Veeco Nanoscope III Multimode AFM in tapping mode in air with Bruker TAP150A Si probes (fig. 3.1).

We have tried different concentration of binding molecules (Mg^{2+}) varying from 0.5 mM to 10 mM. We have observed, that the rate of the relaxation process depend on the concentration of the salt (higher concentration leads to slower relaxation), but have not described this effect quantitatively. The molecules appear to be attached strongly with 10 mM concentration, because the molecules cannot reach 2D-equilibrium after 5 minutes deposition time. We have chosen the concentration of 1 mM to have, on one hand, the possibility to observe an unrelaxed state after 10 seconds, and, on the other hand, a relaxed one after 5 minutes.

Time shorter than 10 seconds cannot be well controlled, as the researcher first has to deposit the solution of DNA molecules with Mg^{2+} onto the surface, then wait the deposition time and, finally, rinse the sample with water and dry with air thus fixing the polymer on the surface. This procedure is not feasible in less than several seconds and the error of deposition time calculation is 2-3 seconds.

One can neither wait considerably longer than several minutes. The sample is drying out and there are different side effects. Nevertheless, we have never seen any difference in the conformation of the molecules after 5 minutes of deposition and results show clear two-dimensional behavior of the molecules (see below).

After the scans were obtained, the coordinates of the molecules were extracted by tracing the scans with our software *DNA Trace* (chapter 1, [92]). For each set of data at least several tens of images were obtained and processed.

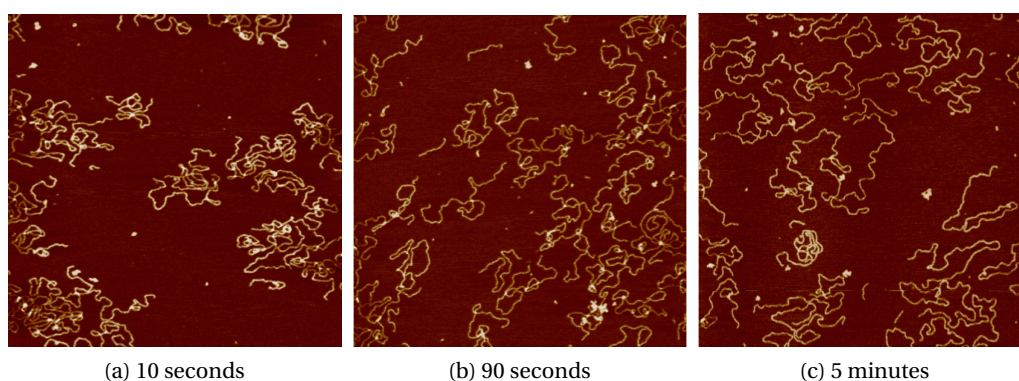


Figure 3.1 – Examples of AFM images showing DNA relaxation as a function of the deposition time ranging from 10 to 300 seconds. One can see, that the molecules become less compact with the increase of time. Molecules were deposited onto the mica surface using 1mM $MgCl_2$. The size of each image is $5 \mu\text{m} \times 5 \mu\text{m}$.

3.2.2 Scaling exponent

We have plotted the average square of end-to-end distance $\langle R_{ee}^2(s) \rangle$ versus contour length s (fig. 3.2) and fitted it with eq. (1.2). It is clear from the graph, that $\langle R_{ee}^2(s) \rangle$ is increasing for any s with the increase of relaxation time. We have calculated the dimensionality of the system by fitting the data with eq. (1.2) and then eq. (3.8). From the results presented in table 3.2 one can see, that the dimensionality decreases from circa 3 to 2 with the increase of relaxation time.

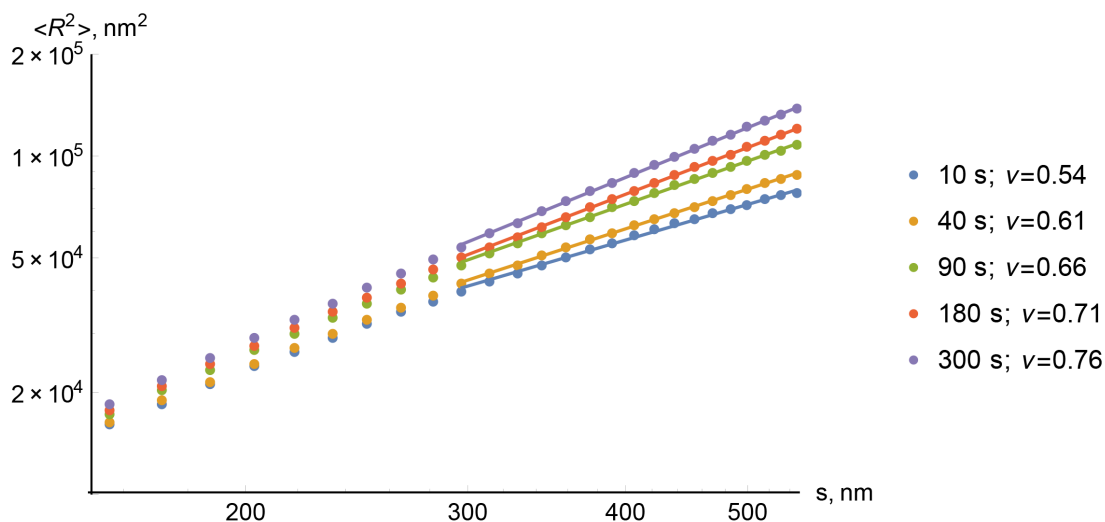


Figure 3.2 – Mean squared end-to-end distance versus contour length for relaxation time varying from 10 seconds to 5 minutes. The data were plotted in log-log scale and linear fit $\log \langle R_{ee}^2(s) \rangle = \log(C) + 2\nu \log(s)$ was applied. ν increases from 0.54 to 0.76 for the samples with relaxation time from 10 seconds to 5 minutes respectively.

We have plotted the *local scaling exponent* (fig. 3.3). For short lengths it is always close to 1, which corresponds to a straight rod. This is caused by the fact that molecules have finite persistence length. With the increase of s , scaling exponent has a tendency to decrease to the value of 0.6 for a short relaxation time and 0.75 for a longer one (fig. 3.3). This fact is in a good agreement with polymer theory, which predicts ν equal to 0.589 for the three-dimensional state and 0.75 for the two-dimensional one.

Also one can see, that the curves exhibit the tendency of ν to decrease with further increase of s rather than to reach a plateau. It can be attributed to the fact, that it takes certain time for the molecule to relax on the surface. If one observes a small area of an image, the parts of the molecules inside this area have to run a short distance to make a transition from the projected three-dimensional state to the two-dimensional equilibrium state. With the increase of the observation area, the segments have to move more and more to make the same transition. We propose that this motion on the surface is a kind of diffusional process, so one can expect a dependence $s \propto \sqrt{t}$, where s is the size of the area and t is the time needed to make a transition.

Assuming the finite experiment duration, on large enough length scale the molecules will always appear to obey the three-dimensional statistics. We have found that the time to relax under these deposition conditions (mica, 1mM Mg^{2+}) on the scale of 500 nm is around 5 minutes, while after 10 seconds the molecules adopt two-dimensional state only for s up to 100-150 nm.

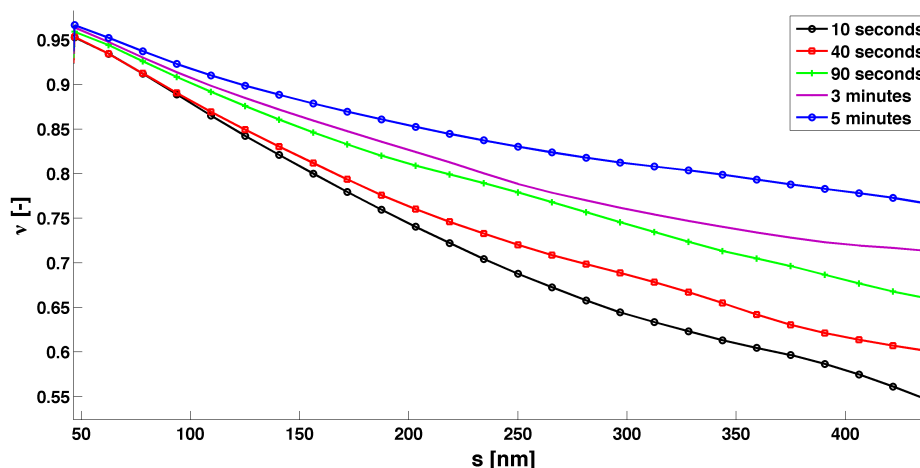


Figure 3.3 – Local scaling exponent as a function of contour length, calculated using the relation eq. (1.3). One can see from the graph, that for 10 seconds relaxation time ν is close to the theoretical value for 3D of 0.589, while for 5 minutes relaxation it is close to 0.75, corresponding to 2D-case. See section 3.2.5 for the experimental errors of the presented data.

3.2.3 Diffusion of relaxation

The relaxation process of the molecule can be thought of as a kind of diffusion of segments of molecules along the surface. If one follows only short parts of the molecules, they adopt 2D equilibrium in very short time. For example, during any accessible relaxation time, parts of the molecules that are shorter than *persistence length* are completely relaxed.

On the contrary, for the parts with the contour length more than 1 μm , even in 5 minutes they still have conformation of 3D-2D quasiprojection. This pushes us to think that the 2D equilibrium state propagation is kind of a diffusion process as well.

Although it is difficult to define the contour length s , up to which the molecule is relaxed (it is not a sharp but rather smooth transition, fig. 3.3), we assumed that this transition is reached in the point s , where scaling exponent ν crosses the value of 0.75. We have plotted this value of s^2 versus t (fig. 3.4) and we have seen a quasi-reasonable linear agreement, except for a small

deviation for short relaxation time.

Two dimensional diffusion process can be described in the following terms:

$$s^2 = 4Dt, \tag{3.10}$$

where D is a diffusion coefficient. We have fitted the data on fig. 3.4 with eq. (3.10) and the value of effective diffusion coefficient $D = 260 \pm 30 \text{ nm}^2/\text{s}$.

The value of effective diffusion coefficient is several orders of magnitude smaller than the one of the diffusion of a free DNA molecule in solution (for example, [105] report $D = 5.5 \text{ cm}^2/\text{s}$ for molecules of about $1 \text{ }\mu\text{m}$ length). Indeed, a free molecule in solution undergo relatively fast Brownian motion, while a molecule deposited on the surface is partly fixed (with higher salt concentration it does not relax during the observation time, section 3.2.1). We can propose two possible mechanisms of the transition:

1. Molecules are completely confined to two dimensions and can slowly move along the surface. This model resembles a droplet of oil on the table, which after certain time can cover the whole surface of the table, yet the process is relatively slow;
2. Molecules cannot move along the surface; instead, part of a molecule can detach from the surface with certain probability, then move in three dimensions along the surface and attach to the surface again. This can be modeled as a coil of a rope, which lies on a shaking sticky surface; after certain time this rope will no longer be a coil but rather a relaxed rope.

Real process can be a mixture of both proposed models. It seems quite natural that the rate of this process is much slower than the rate with which a single part of the polymer can move along or close to the surface.

We would like to note that the process of the diffusion of the polymer in two dimensions has already been studied by Maier and Rädler [86]. They have deposited DNA molecules on a cationic bilipid layer and watched their displacement in time. From these data they have calculated the diffusion coefficient with respect to the length of DNA molecules. In our model we consider short fragments of DNA molecules to move, namely the fragments of persistence length which can be imagined as rigid parts. Such length of about $50 \text{ nm} \approx 150$ basepairs should be used to compare data of Maier and Rädler to our results. For the molecules of such length their article attests the diffusion coefficient to be about $0.7 \text{ }\mu\text{m}^2/\text{s}$ which is roughly 3000 times bigger than the aforementioned diffusion coefficient of relaxation. This result is not very surprising for two reasons. First, we compare the free diffusion of short molecules of DNA to the motion of the parts of DNA molecules which have the same length, but are connected

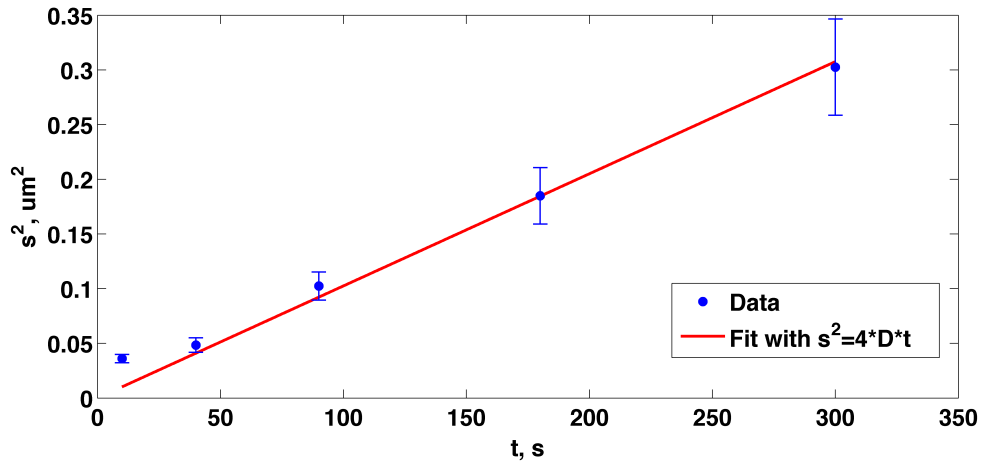


Figure 3.4 – Squared distance up to which molecule is relaxed (s for which v on fig. 3.3 reaches the value of 0.75) versus relaxation time and its fit with $s^2 = 4Dt$, $D = 260 \pm 30 \text{ nm}^2/\text{s}$. Y axis error bars are presented on the graph, x axis error is estimated to be constant and equal to about 2-3 seconds.

to similar parts at both ends. It introduces new constraints and decreases the freedom of the motion. Second, our surface is rigid mica modified by divalent ions, which has relatively strong adsorption of DNA, while Maier and Rädler exploit cationic lipid bilayer, where single lipids can have diffusive Brownian motion [4, 25].

The time-lapse motion of DNA molecules on the surface of mica treated with divalent ions was already directly studied by high-speed AFM techniques [96, 5, 83]. In the cited studies authors observe the segments of molecules attached to the surface moving along it in time and the timescale of such motion (one - several seconds to several minutes) is in qualitative agreement with our experiments.

3.2.4 Persistence length

A common way to calculate the *persistence length* of the polymer is to calculate the *correlation function* (eq. (1.1)). For two-dimensional images obtained using AFM one can assume that $d = 2$ and then calculate the persistence length.

Instead, we have determined from the decay constant a of the correlation function $\langle \cos\theta(s) \rangle = e^{-s/a}$ the dimensionality of the system. From $a = l_p$ for $d = 3$ and $a = 2l_p$ for $d = 2$, we can expect the dimensionality:

$$d = 4 - \frac{a}{l_p}. \quad (3.11)$$

Chapter 3. DNA relaxation

The data presented on fig. 3.5 and table 3.2 show the same result for the two ways of calculation of d : the dimensionality of the system is decreasing from circa 3 to 2 with increasing relaxation time.

t, s	10	40	90	180	300
d_v	3.6 ± 0.5	3.0 ± 0.4	2.6 ± 0.2	2.21 ± 0.15	1.94 ± 0.05
d_{l_p}	3.4 ± 0.3	3.2 ± 0.3	2.3 ± 0.2	2.29 ± 0.10	2.00 ± 0.05

Table 3.2 – Dependence of dimensionality of the system d on relaxation time t , calculated from the scaling exponent using eq. (3.9) (d_v) and correlation function eq. (3.11) (d_{l_p}).

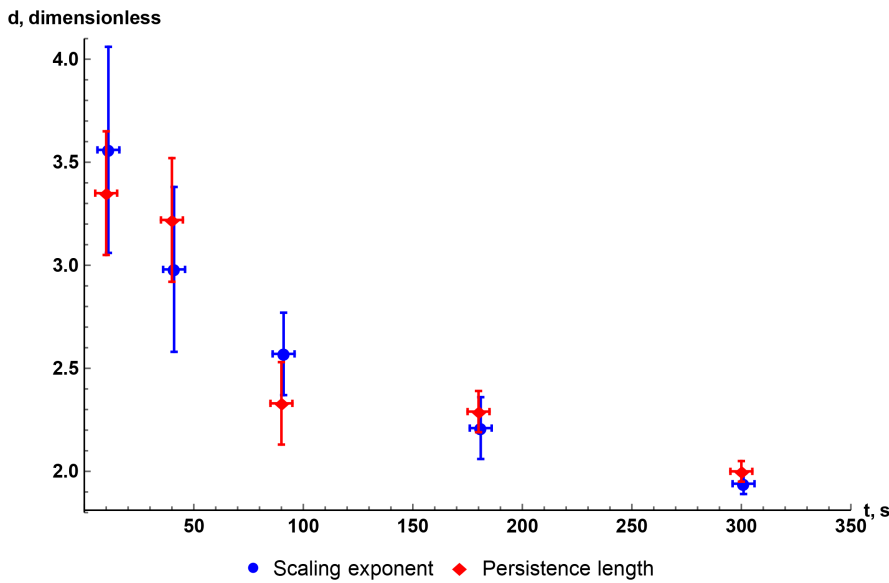


Figure 3.5 – Dimensionality of the system obtained from scaling exponent data (eq. (1.2) and eq. (3.9)) and persistence length data (eq. (3.11), where $l_p = 50$ nm). This figure visualizes the data presented in table 3.2.

The values of d calculated with two aforementioned techniques coincide up to the values of the experimental errors. Of course, $d > 3$ does not make any physical sense in our case and should be attributed to experimental error, which is relatively high for this region (see next chapter for the clarifications). We do not propose, up to now, any quantitative model of the dependence of d on the relaxation time, but the results clearly show qualitative transition from three- to two-dimensional conformation.

3.2.5 Error assessment

As soon as we perform any physical experiment, we deal with a finite error that prohibits us to obtain absolutely exact values. Any value measured in this Thesis has confidence only up to a certain digit. In this section we would like to explore the topic of errors that are typical for

AFM measurements and that correspond to our studies.

There is a common classification that divides errors into *systematic* and *random* ones. The systematic error depends on the experimental setup and conditions and always occurs, even if the experiment is repeated many times. It can have either constant value or vary in time. In the latter case it is called *drift*: the error changes in time and have a linear or non-linear trend, but it is not random. A classical example of it can be thermal drift of a piezo scanner used in AFM setups. Researches face often the problem, that even very slight change of temperature causing its gradient on different parts of AFM leads to the scanning artifacts: objects appear to be distorted and their geometrical characteristics can be measured imprecisely. Even if the drift is normally easy to detect by observing of a gradual change of the mean of a measured value in time, it is difficult to estimate its effect on the measurement. Constant error can cause even more difficulties to the researcher, as it is difficult to detect. Common sources of a constant error may include inherent non-linearity of the device, its aging or principle of measurements, etc. In case of AFM, a typical issue is the non-linearity of the scanner: the same object scanned in the middle of the frame and in the corner of it can exhibit different apparent size. The discrepancy varies from scanner to scanner and can reach the level of several percent. Of course, this issue is well-known by the producers of AFM devices and one should not underestimate their effort to linearize the scanners both on the hardware and the software levels, but it never can be done up to hundred percent due to the obvious physical reasons.

The random error is caused by unpredictable fluctuations and is always present in the experiment to some extent. The random error is typically easier to estimate and to decrease, if one can repeat the experiment many times. In this case the random error decreases as $\sim 1/\sqrt{N}$, where N is the number of measurements. However, one can see, that to get random error decreased by one order of magnitude, one has to perform the same experiment a hundred times more. A common practice is to increase the number of measurements, until the value of the random error becomes comparable to the value of the systematic one.

If we look at the formulas describing the statistics of the polymer, for example, eqs. (1.1) and (1.2), we could notice the averaging sign $\langle \dots \rangle$. It means that we have to average our functions over a big set of molecules to calculate precise values and decrease random error. But in this case random error comes not only from the measurement procedure, these values are intrinsically *stochastic*. To demonstrate this, we have simulated $N_{total} = 10\,000$ polymer molecules in three dimensions. Each molecule was represented as a chain of 100 segments, so that the length of each segment $l_0 = 1$ nm and the angle between every two consecutive segments is a normally distributed random value with zero mean and mean deviation $\sigma = \sqrt{2l_0/l_p}$, where $l_p = 50$ nm is the persistence length. Such a procedure produces the contours that are consistent with the *WLC* model (see below). If we project these molecules

onto a 2D-plane and calculate the average end-to-end distance of the projected contours $\sqrt{\langle R_{ee}^2 \rangle}$, we will get the value of about 59 nm. But if we calculate $\sqrt{R_{ee}^2}$ for the projection of single molecules, we will get a broad distribution from 0.6 to 94 nanometers (see fig. 3.6).

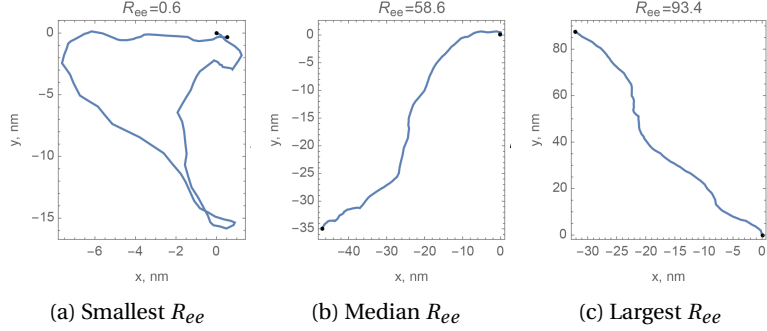


Figure 3.6 – 2D projections of the simulated contours of semi-flexible polymer molecules and the value of the projected end-to-end distances R_{ee} . Three contours with the smallest (a), the median (b) and the largest (c) R_{ee} are presented.

The same issue is present if one calculates the persistence length of the molecules and averages the correlation function $\langle \cos\theta(s) \rangle = \exp(-\frac{s}{(4-d)l_p})$. It could be rewritten as $\langle l_p \rangle = -\frac{s}{(4-d)\ln\langle \cos\theta(s) \rangle}$. This value is precise only if averaged over a large set of molecules and does not have a physical meaning, if calculated for only one molecule. We have performed a statistical analysis of the value of the persistence length for the same aforementioned simulated polymer contours in the following way. We have selected certain number of samples N of all of N_{total} simulated molecules, and calculated the value of l_p with the formula eq. (1.1). Then we iterated over all possible combinations of N molecules and calculated l_p for every of these combinations. Finally, we plotted the histogram of these values l_p for this particular N and fitted it with the normal distribution function¹; thus we got fitting parameters $\sigma(N)$ and $\langle l_p(N) \rangle$. An example of such a histogram and its fit for $N = 300$ is presented on fig. 3.7. When we vary N in the range $[1; N_{total}]$, we get fig. 3.8, where the red line represents $\langle l_p(N) \rangle$ and the blue lines represent its variation $\langle l_p(N) \rangle \pm \sigma(N)$.

We performed similar analysis for the persistence length of a real sample (namely, aforementioned molecules of DNA after 5 minutes relaxation). Values of l_p for different sets of molecules and different N are presented on fig. 3.9. Although the direct quantitative comparison between the data on figs. 3.8 and 3.9 is not appropriate due to different length of the polymer molecules and N_{total} , qualitatively the data shows the same result: with the increase of N the value σ decreases, the distribution of l_p values becomes more and more narrow with the average at $l_p = 50$ nm. One can also see, that the standard deviation scales as $\sigma \sim 1/\sqrt{N}$ (see fig. 3.10),

¹To be precise, the persistence length distribution cannot be normal, as $\langle l_p \rangle = 50$ nm and $l_p \in (0; +\infty)$, so that the distribution cannot be symmetric; however, this distribution is approximated well by a Gaussian normal distribution, which permits us to extract the mean value of l_p and its standard deviation σ .

which is consistent with the scaling of random error.

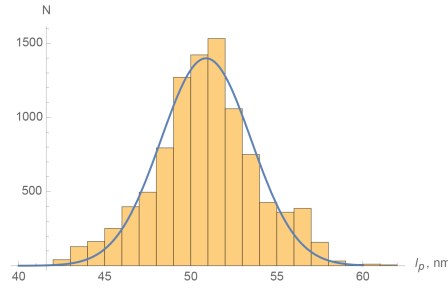


Figure 3.7 – Histogram of the values of the persistence length for different sets containing $N = 300$ molecules each. Blue line: its fit with the normal distribution function. The molecules are simulated.

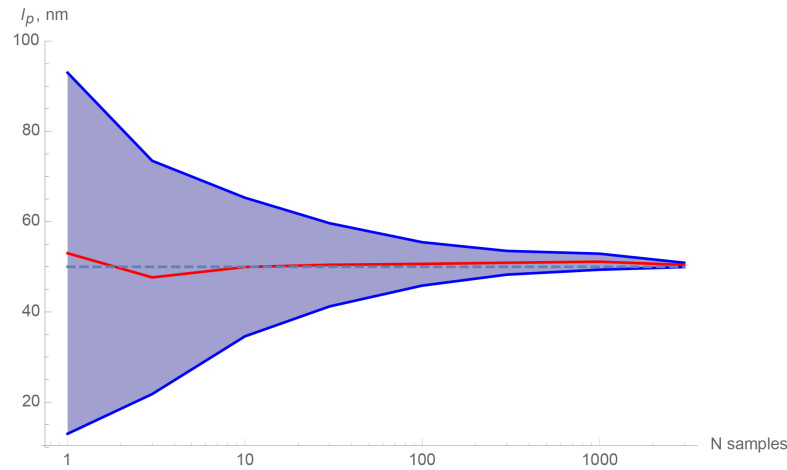


Figure 3.8 – The values of the average $\langle l_p \rangle (N)$ (red line) and the standard deviation $\sigma(N)$ (blue line) for the distributions of l_p for sets with different number of molecules in a set N . One can see, that the values converge to $l_p = 50$ nm. The molecules are simulated.

Another issue with the error of the calculation of end-to-end distance and correlation function dependence on the contour length is the following. In eqs. (1.1) and (1.2) we average over all possible contours having the same s in the set of AFM images corresponding to the current relaxation time. If we traced m contours with n segments each with the step of s_0 nm between consecutive points, the number of possible contours l with the length of $s = i s_0$ (i denotes some integer, $i \in [1; n]$) will be the following:

$$l = m(n + 1 - i). \tag{3.12}$$

Nevertheless, to estimate the random error, we should count only completely independent samples (which means that corresponding contours cannot have intersecting parts). Their

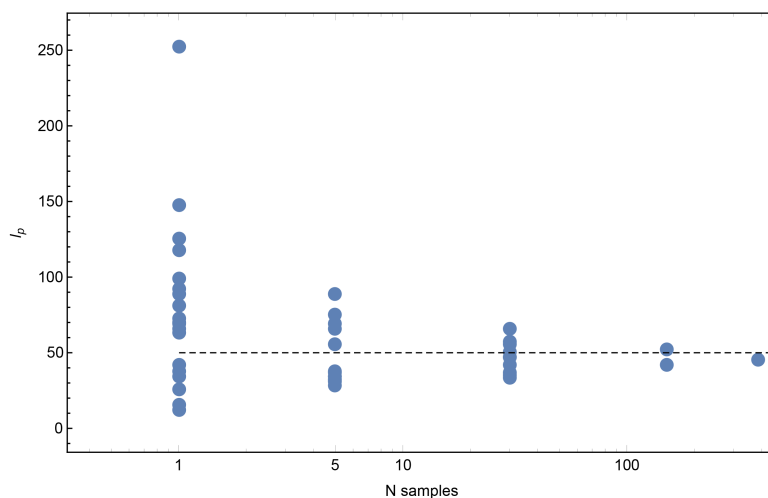


Figure 3.9 – Values of l_p calculated on the sets with different number of molecules N . The sample is aforementioned DNA molecules after 5 minutes relaxation on mica surface.

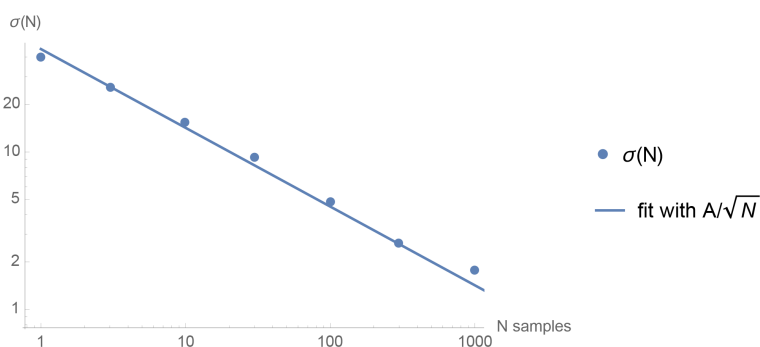


Figure 3.10 – Scaling of the standard deviation $\sigma_{l_p}(N)$ from fig. 3.8 and its fit with the inverse square root model.

number l' will be sufficiently lower for larger i :

$$l' = m \lfloor n/i \rfloor = m \left\lfloor \frac{L/s_0}{s/s_0} \right\rfloor = m \left\lfloor \frac{L}{s} \right\rfloor, \quad (3.13)$$

where $\lfloor x \rfloor$ means floor rounding function and $L = ns_0$ is the total length of the contours. Finally, the expression of the relative error for the calculation of the average end-to-end distance is the following:

$$\sigma(s) \simeq \frac{\Delta \langle R_{ee}^2 \rangle}{\langle R_{ee}^2 \rangle} \simeq \frac{1}{\sqrt{l'}} = \frac{1}{\sqrt{m \lfloor \frac{L}{s} \rfloor}}. \quad (3.14)$$

One can see from eq. (3.14), that the relative random error increases significantly with s . For example, for $m = 100$ traced contours and $s > L$ we get $\sigma \simeq 10\%$, as there is only one independent sample per each molecule contour. The aforementioned formula is also valid for the

absolute error of the correlation function (the relative error does not make sense for $s \gg l_p$ in this case, as the function tends to zero).

The aforementioned arguments show, that in some cases the error of the determination of some physical values comes not only from the imprecise measurements, but also from the stochastic nature of this value. Even if we assume that our AFM scanner is ideal, completely linear and properly calibrated and the lengths of the contours of DNA is measured *exactly*, we still have some finite precision of the determination of l_p (as well as R_{ee} and all other averaged functions of a polymer), which is caused by the intrinsic random variation of some stochastic value. Nevertheless, the value of this type of the random error is easy to measure by calculating the standard deviation of the value.

In our AFM experiments we estimate the value of systematic error of the determination of the length of about 3–5%. This includes the thermal drift, non-linearity of the scanner and its miscalibration. We normally try to repeat the experiments, so that the number of molecules on the images is sufficient to have random error in the same range. Typical number of molecules to fulfill this condition is in between one hundred and several hundreds for each set of conditions. However, the random error of the determination of relaxation time can be sufficiently higher, especially for the times of 10 and 40 seconds, because every operation (deposition of DNA, rinsing, drying) takes typically several seconds. This leads to significant errors of the measuring of $\nu(s)$ presented on fig. 3.3; for example, we estimate the total error of $\nu(400 \text{ nm})$ for 10 and 40 seconds relaxation as 10% and 7% respectively.

3.3 Conclusions

We have studied the conformational change of the polymer deposited onto the surface from the projection of the three-dimensional state to the two-dimensional equilibrium. We have deposited DNA molecules on mica and let them relax on the surface during deposition time varying from 10 seconds to 5 minutes. Results exhibit an apparent time-dependent relaxation process. On fig. 3.1 one can clearly see the transition from a compact state to a more relaxed one, which is proven by the calculation of scaling exponent (fig. 3.2). We got ν equal to 0.54 and 0.76 for 10 seconds and 5 minutes relaxation respectively, which can be attributed to the fact, that molecules are projected onto the surface ($\nu = 0.589$ is scaling exponent for three-dimensional state) and then relax on it if they are not fixed strong enough ($\nu = 0.75$ corresponds to two-dimensional state).

We have estimated the dimensionality of the system by two different functions: the end-to-end distance (eq. (1.2)) and the correlation function (eq. (1.1)). The results presented on fig. 3.5 show the transition from $d \sim 3$ to $d = 2$ with the increase of deposition time for both of the aforementioned techniques. One can also see, that the experimental error is sufficiently larger

Chapter 3. DNA relaxation

for smaller relaxation times and becomes relatively low for $t = 5$ minutes. Another approach to determine the dimensionality of the system will be described in the next chapter.

In this chapter we have also given an overview of the experimental errors, typical for the field of quantitative biopolymer measurements by AFM. As the random error decreases as an inverse square root of the number of measurements, one has to compensate between the precision of the measurements and reasonable time to run the experiment. As a rule of thumb, one usually increases the number of measurements, until the random error reaches the level of the systematic one. Typical number of samples to estimate correctly the persistence length is in the range of several hundreds molecules.

We have shown that the length scale, at which the molecules can be considered to be in two-dimensional state, increases with time (fig. 3.3). Moreover, this scale is proportional to the square root of the relaxation time and can be fitted well with the equation of two-dimensional diffusion $s = \sqrt{4Dt}$. We got an effective diffusion coefficient $D = 260 \pm 30 \text{ nm}^2/\text{s}$, which is much smaller than the diffusion coefficient for a free diffusion in solution. This value is also about three orders of magnitude lower than the one measured for the 2D diffusion of DNA molecules along bilipid layer surface. We attribute this discrepancy to the fact, that in our case the molecules are fixed or trapped by relatively strong interaction with the mica surface. We have proposed two models of the relaxation of the molecules and we assume, that this diffusion coefficient attributes to slow motion along the surface.

4 Concentration function

There is no branch of mathematics,
however abstract,
which may not some day be applied
to phenomena of the real world.
— Nikolay Ivanovich Lobachevsky

4.1 Motivation

We have already discussed the importance of the polymer theory in chapter 3. In this chapter we are going to introduce new functions to develop further on this theory and describe the conformation of a polymer. Actually, if one traces a polymer, its contour looks similar to a trajectory of a particle under the diffusion process. Indeed, one of the simplest models of polymers, freely-jointed chain [107], is substantially similar to the random walk model of diffusion [119, 55]. This analogy is well described by de Gennes [28]. It pushes us to consider the application of the well-elaborated theory of diffusion to polymers and DNA molecules in particular in order to study the dynamics of their transition from 3D to 2D conformation on the surface. Of course, this analogy is not perfect and complete (in particular, it allows neither for the effects of the excluded volume and self-avoiding of intersections, which can be significant for long chains, nor for the finite bending rigidity and the persistence length, which plays the role on short length scales), but we hope that it could be useful and predict the dynamical polymer behavior in non-equilibrium state. Below we will show, that this proposed technique permits one to distinguish between 2D and 3D state of a polymer on the surface without any prior knowledge of the persistence length of it.

4.2 Theoretical model

4.2.1 Equation of diffusion

Diffusional processes are governed by Fick's laws. The first Fick's law relates the diffusion flux J_x through the section perpendicular to the axis x with the concentration of the particles C :

$$J_x = -D \frac{\partial C}{\partial x}, \quad (4.1)$$

where D is the diffusion coefficient measured in $\text{length}^2 \text{time}^{-1}$, or m^2/s . In our work we will focus on the second Fick's law, which relates the distribution of concentration of particles throughout the space and its change in time t :

$$\frac{\partial C}{\partial t} = D \Delta C, \quad (4.2)$$

where Δ is the Laplace operator¹. As the media is isotropic, no angular dependence can be expected and in polar coordinates the Laplacians become

$$\Delta c(r, t) = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial c(r, t)}{\partial r} \right) \quad (4.3)$$

and

$$\Delta C(R, t) = \frac{1}{R^2} \frac{\partial}{\partial R} \left(R^2 \frac{\partial C(R, t)}{\partial R} \right), \quad (4.4)$$

for two- and three-dimensional case respectively. Here and later we use lower case r with $c(r)$ for 2D case and capital R with $C(R)$ for 3D case.

There exist many angle-independent solutions of eq. (4.2). First of all, we would like to mention stationary solutions, when $\partial C / \partial t = 0$. In this case eq. (4.2) becomes $\Delta C = 0$. The only stationary solution (except for the trivial one $C \equiv \text{const}$) in two dimensions is:

$$c(r) = c_0 \ln \frac{r_0}{r}, \quad (4.5)$$

where r_0 is the maximal attainable distance and c_0 is the normalization constant. Equation (4.5) describes the concentration, which is infinite at $r = 0$, zero at $r = r_0$ and negative at $r > r_0$. The latter case is not physical, because the concentration of particles is proportional to their number per surface area and cannot be negative. Nevertheless, the formula eq. (4.5) still can have a physical sense to some extent: there is an infinite point source of particles at $r = 0$; the particles spread around and are consumed at the circle $r = r_0$. So this solution is considered only for $r \in (0, r_0]$, where at every point $c(r)$ is constant in time (the particles arrive

¹ $\Delta f(x, y) = \nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \left[+ \frac{\partial^2 f}{\partial z^2} \right]$ for 2D [3D] case.

and depart with the same rates).

In three dimensions the only non-trivial solution of Fick's law has the following form:

$$C(R) = C_0 \left(\frac{1}{R} - \frac{1}{R_0} \right), \quad (4.6)$$

where, again, R_0 is the maximal attainable distance and C_0 - normalization constant. This solution describes the case similar to the aforementioned two-dimensional one, where the particles appear at $R = 0$ and disappear on the sphere $R = R_0$, so $C(R)$ is stationary and defined for $R \in (0; R_0]$.

The most common solution of Fick's law is the one, which corresponds to the case of the point source of the particles (distributed according to the Dirac delta function) at $t = 0$ and $R = 0$, which propagate in time. This case is described by the solution

$$c(r, t) = \frac{c_0}{t} \exp\left(-\frac{r^2}{4Dt}\right) \quad (4.7)$$

for 2D and by

$$C(R, t) = \frac{C_0}{t^{3/2}} \exp\left(-\frac{R^2}{4Dt}\right) \quad (4.8)$$

for 3D [101]. These solutions realize, when one injects pointwise some particles at $t = 0$ and let them diffuse freely.

There exist other solutions of the second Fick's law. For example, the function $c(r, t) = c_0 \exp(-at) Y_0(\mu r)$, where $Y_0(z)$ is the Bessel function of the second order and μ and a are some constants, also satisfy eq. (4.2) in two dimensions [101]. There are many polynomial solutions as well, which rarely have any physical applications. We will not cover all the solutions in this work and will limit ourselves to the most appropriate ones.

4.2.2 Concentration function

Now we are going to introduce *concentration function*, a function of a polymer conformation, which will be analogous to the concentration in Fick's law. Let us imagine a polymer molecule in two dimensions. We use the system of coordinates associated with one of its ends. Then we can divide the space into rectangles with the sides (dx, dy) and calculate the parts of the molecule dl (in terms of length), that are inside this rectangle. Thus we get *concentration*

Chapter 4. Concentration function

function $c(x, y)$, or $c(r)$ (see fig. 4.1), measured in m^{-1} , or in nm^{-1} more commonly:

$$c(r) = \frac{dl}{2\pi r dr}. \quad (4.9)$$

The same idea can be applied to the three-dimensional case to obtain $C(R) = \frac{dl}{4\pi R^2 dR}$ measured in nm^{-2} .

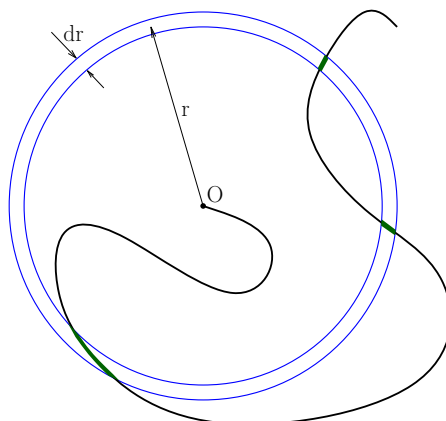


Figure 4.1 – Sketch of a molecule. Point O corresponds to the end of the molecule, the parts in green fit into the ring (O, r, dr) .

Of course, $c(r)$ and $C(R)$ have some sense, only if they are averaged over statistically significant number of molecules, the same as for persistence length, scaling exponent, end-to-end distance or other polymer distribution functions (see previous chapter). To do so, all the molecules should be virtually translated so that one of their ends is located in one point O, namely the origin of (polar) coordinates (see fig. 4.2). After the calculation of the total length all the parts of all the molecules that fit into $(O, r, r + dr)$ ring, this value has to be normalized, *i.e.* divided by the number of molecules N to get proper concentration function, which does not depend on the number of molecules in the statistical set. Please note that all the following calculations apply only to polymer molecules, which have equal contour length L .

The concentration function has an obvious analogy to the diffusion process. Indeed, we look at an image of a polymer molecule on the surface as at a picture of Brownian random walk of a (nano)particle, which begins on the same point as one of the ends of the polymer and then continues to the moment, when the length of the trajectory is equal to the length of the polymer chain. Thus the distribution of segments of polymers combined "by hand", so that one of the end of all of them coincides, can now be considered as a time-dependent diffusion of particles originating (having source) at the same point at initial moment of time. Following this analogy, the concentration function should satisfy the aforementioned Fick's law (eq. (4.2)) with properly defined D (which in this case is measured in the same standard units, namely m^2/s).

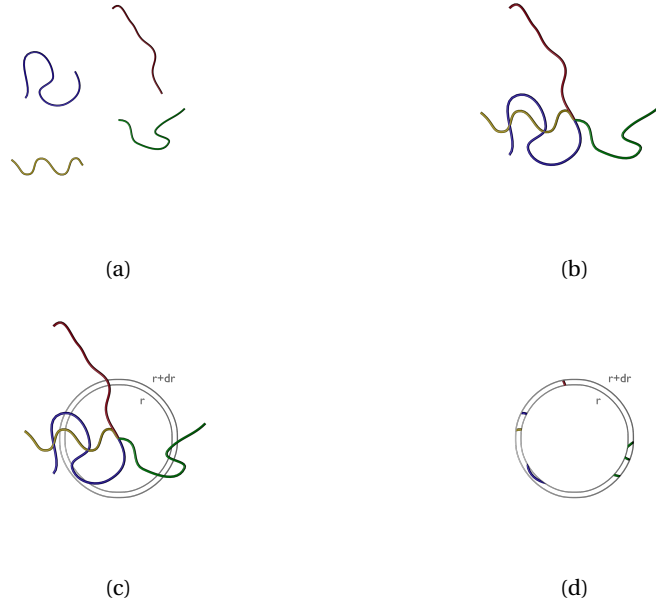


Figure 4.2 – Sketch showing calculation of concentration function $C(r)$ for multiple molecules. First all of them have to be virtually translated so that one of their ends coincide at certain point ((a),(b)), then the total length of all their segments that fit into the ring (O, r, dr) ((c),(d)) is divided by N to get proper normalization.

Nevertheless, the following difference between a diffusion trajectory of a particle and a contour of a molecule imaged by AFM still persists. Let us consider the diffusional process for $t \in [0; T]$. Any particle, that is at a certain time $t < T$ at a certain position r , is not there at $t = T$, so the information of the positions of the particles at the times prior to T is lost and we can observe finally only the distribution of particles. In our initial conditions (all the particles are at $r = 0$ at $t = 0$) this distribution is specified by eqs. (4.7) and (4.8). In case of the AFM images of the contours of molecules, for all the moments we see, where the "particle" was before: it is recorded on the image. That is why to utilize the random walk – diffusion – polymer conformation analogy for a two-dimensional case and get an equivalent concentration function, we have to integrate eq. (4.7) over time from $t = 0$ to T :

$$c(r) = \int_0^T \frac{c_0}{t} \exp\left(-\frac{r^2}{4Dt}\right) dt, \quad (4.10)$$

which can be expressed as the following, a variable change $\frac{r^2}{4Dt} = z$ being applied:

$$c(r) = c_1 \Gamma\left(0, \frac{r^2}{4DT}\right), \quad (4.11)$$

Chapter 4. Concentration function

where $\Gamma(s, x)$ is nothing else than the incomplete gamma function² [7] and c_1 is again a normalization constant. The latter can be easily calculated from the requirement $\int_0^\infty 2\pi r c(r) dr = L$. It will be discussed in the next section.

Of course, eq. (4.11) is inapplicable for smallest values of $r < l_p$, where the finite persistence length is of an uttermost importance. It is neither applicable for sufficiently large r , as it defines $c(r)$, which is strictly positive over all $r > 0$, while the actual $c(r) \equiv 0$ for $r > L$. Nonetheless, the discrepancy is not so big, because eq. (4.11) has the exponential decay asymptotic and $c(L)$ is close to zero. We will show below, that this approximation makes sense for intermediate values of r .

In 3D case the same arguments will lead to the integration of eq. (4.8):

$$C(R) = \int_0^T \frac{C_0}{t^{3/2}} \exp\left(-\frac{R^2}{4Dt}\right) dt = \frac{C_1}{R} \Gamma\left(1/2, \frac{R^2}{4DT}\right). \quad (4.12)$$

Now we have to project this function from 3D to a plane passing through the origin of coordinates. Every thin spherical layer of radius R and thickness dR containing particles with a density $C(R)$ results in the following concentration function: $c(r) = C(R)/\sqrt{1-r^2/R^2}$, where r is the radius on the plane. Spherical layers with $R < r$ do not contribute to $c(r)$, so finally the projected concentration function is defined by the following expression:

$$c_{3D \rightarrow 2D}(r) = \int_r^\infty C(R)/\sqrt{1-r^2/R^2} dR = C_1 \int_r^\infty \frac{1}{R\sqrt{1-r^2/R^2}} \Gamma\left(1/2, \frac{R^2}{4DT}\right) dR, \quad (4.13)$$

which with the same variable change as in eq. (4.10) results in

$$c_{3D \rightarrow 2D}(r) = C_1 \int_{r^2/(4DT)}^\infty \frac{1}{\sqrt{z^2 - r^2 z/(4DT)}} \Gamma(1/2, z) dz, \quad (4.14)$$

which in turn with another variable change $y = z - r^2/(8DT)$ can be rewritten as

$$c_{3D \rightarrow 2D}(r) = C_1 \int_{r^2/(8DT)}^\infty \frac{1}{\sqrt{y^2 - r^4/(64D^2T^2)}} \Gamma\left(1/2, y + \frac{r^2}{8DT}\right) dy. \quad (4.15)$$

For reasonably large values of x the incomplete gamma function has the following asymptotic:

² $\Gamma(s, x) := \int_x^\infty z^{s-1} \exp(-z) dz$

$\Gamma(\alpha, x) = x^{\alpha-1} e^{-x} (1 + O(x^{-1}))$ [56]. By using it we convert eq. (4.15) to

$$c_{3D \rightarrow 2D}(r) = C_1 e^{-r^2/(8DT)} \int_{r^2/(8DT)}^{\infty} \frac{e^{-y}}{\sqrt{y^2 - r^4/(64D^2T^2)} \sqrt{y + r^2/(8DT)}} dy. \quad (4.16)$$

The latter integral turns out to be a particular case of the example N 3.384.6 from [56]:

$$\int_u^{\infty} (x-u)^{\nu-1} (x+u)^{-\nu-1/2} e^{-\mu x} dx = \frac{1}{\sqrt{\mu}} 2^{\nu-1/2} \Gamma(\nu, 0) D_{1-2\nu}(2\sqrt{u\mu}), \quad (4.17)$$

where $D_n(x)$ is a parabolic cylinder function; for $n = 0$ this function is $D_0(x) = e^{-x^2/4}$. In our case $\nu = 1/2$ and $\mu = 1$, so eq. (4.16) with the help of eq. (4.17) results in

$$c_{3D \rightarrow 2D}(r) = C_1 e^{-r^2/(8DT)} \sqrt{\pi} D_0\left(\frac{r}{\sqrt{2DT}}\right) = c_2 \exp\left(-\frac{r^2}{4DT}\right), \quad (4.18)$$

where c_2 is some normalization constant. We can see that eqs. (4.11) and (4.18) have close, but not equal asymptotic (the former can be approximated as $c(r) \sim \frac{1}{r^2} \exp\left(-\frac{r^2}{4DT}\right)$), and we will show below, that our experiments are able to distinguish between these two functions.

4.2.3 Normalization

In this section we will cover the topic of the normalization of the aforementioned concentration functions. It does not play a big role in our model and is described for the completeness. The concentration function represents the distribution of the parts of a molecule in space, so the integral of it over all the space has to be equal to the length of the molecule L . Let us first calculate c_1 from eq. (4.11):

$$L \iint_S c(r) dr = \int_0^{\infty} c(r) 2\pi r dr = \int_0^{\infty} c_1 \Gamma\left(0, \frac{r^2}{4DT}\right) 2\pi r dr. \quad (4.19)$$

With a variable change $\frac{r^2}{4Dt} = z^2$, $r = z\sqrt{4DT}$ and using the fact that $\int_0^{\infty} \Gamma(0, x) dx = 1$:

$$\begin{aligned} L &= c_1 2\pi \int_0^{\infty} \Gamma\left(0, \frac{r^2}{4DT}\right) z\sqrt{4DT} dz \sqrt{4DT} \\ &= c_1 8\pi DT \int_0^{\infty} \Gamma(0, z^2) \frac{dz^2}{2} \\ &= c_1 4\pi DT \int_0^{\infty} \Gamma(0, x) dx. \end{aligned} \quad (4.20)$$

Chapter 4. Concentration function

The last integral can be calculated in the following way. For s satisfying the condition $\text{Re}(s) > 0$ we can integrate by parts:

$$\int_0^{\infty} \Gamma(s, x) dx = x\Gamma(s, x) \Big|_0^{\infty} - \int_0^{\infty} x\Gamma'(s, x) dx. \quad (4.21)$$

By definition of the incomplete gamma function $\Gamma'(s, x) = -x^{s-1}e^{-x}$. Thus we obtain:

$$\int_0^{\infty} \Gamma(s, x) dx = \int_0^{\infty} x^s e^{-x} dx = \Gamma(s+1). \quad (4.22)$$

Due to the existence of the limit $\lim_{x \rightarrow 0} x\Gamma(0, x) = 0$, which follows from $\lim_{x \rightarrow 0} x \ln x = 0$, the last equality is valid for $s = 0$ as well. Finally, we obtain:

$$c_1 = \frac{L}{4\pi DT}. \quad (4.23)$$

Now we will calculate the constant c_2 from eq. (4.18) by using the same technique and variable change:

$$\begin{aligned} \int_0^{\infty} c_2 \exp\left(-\frac{r^2}{4DT}\right) 2\pi r dr &= 4\pi DT c_2 \int_0^{\infty} e^{-z} dz = 4\pi DT c_2 = L \\ \Rightarrow c_2 &= \frac{L}{4\pi DT}. \end{aligned} \quad (4.24)$$

In case of a three-dimensional concentration function (eq. (4.12)) we have:

$$\begin{aligned} \iiint_V C(R) dV &= \int_0^{\infty} \frac{C_1}{R} \Gamma\left(\frac{1}{2}, \frac{R^2}{4DT}\right) 4\pi R^2 dR \\ &= C_1 4\pi 2DT \int_0^{\infty} \Gamma\left(\frac{1}{2}, z\right) dz \\ &= 8\pi C_1 DT \Gamma\left(\frac{3}{2}\right) \\ &= 8\pi C_1 DT \frac{\sqrt{\pi}}{2} = L, \end{aligned} \quad (4.25)$$

where $\Gamma(s)$ is the gamma function and $\Gamma(3/2) = \frac{1}{2}\Gamma(1/2) = \sqrt{\pi}/2$ [67]. Finally, we obtain:

$$C_1 = \frac{L}{4\pi^{3/2}DT}. \quad (4.26)$$

The same procedure can be applied to eq. (4.5) and eq. (4.6) in order to find the constants c_0

and C_0 respectively:

$$\iint_S c(r) dS = \int_0^{r_0} c_0 \ln \frac{r_0}{r} 2\pi r dr = \frac{\pi c_0 r_0^2}{2} = L, \quad (4.27)$$

$$c_0 = \frac{2L}{\pi r_0^2} \quad (4.28)$$

and

$$\begin{aligned} \iiint_V C(R) dV &= \int_0^{R_0} C_0 4\pi R^2 \left(\frac{1}{R} - \frac{1}{R_0} \right) dR = C_0 2\pi R_0^2 - C_0 \frac{4}{3} \pi R_0^2 = \\ &= \frac{2}{3} C_0 \pi R_0^2 = L, \end{aligned} \quad (4.29)$$

$$C_0 = \frac{3L}{2\pi R_0^2}. \quad (4.30)$$

If we project the stationary three dimensional solution of diffusion equation (eq. (4.6)) onto a 2D plane, we have to perform a procedure similar to the one used to obtain eq. (4.18):

$$\begin{aligned} C_{3D \rightarrow 2D}(r) &= C_0 \int_0^{R_0} \frac{\frac{1}{R} - \frac{1}{R_0}}{\sqrt{1 - \frac{r^2}{R^2}}} dR \\ &= C_0 \left(\ln \frac{r}{R_0 - \sqrt{R_0^2 - r^2}} - \frac{\sqrt{R_0^2 - r^2}}{R_0} \right). \end{aligned} \quad (4.31)$$

We would like to stress again, that we do not anticipate the stationary solutions of diffusion equation to be a correct representation for the concentration function of a polymer (and it will be shown in the next section), so the last three equations are stated for the completeness.

4.3 Results and Discussion

We calculated the concentration function for the same traced AFM images used to study relaxation of DNA on the surface that were already described in chapter 3. As the molecules have to be of the same length, we chose the fragments of 1020 ± 20 nm. We calculated concentration function for these images, the results are presented on fig. 4.3. One can see that concentration function is decreasing for small r and increasing for large r with the increase of relaxation time, as it is expected due to 2D-3D relaxation process. Indeed, as it was described before, the molecule increases its average area, so for large length scale concentration of segments

Chapter 4. Concentration function

increases in time. But the integral of concentration function over all the area, in other words, the total length of the molecule remains the same.

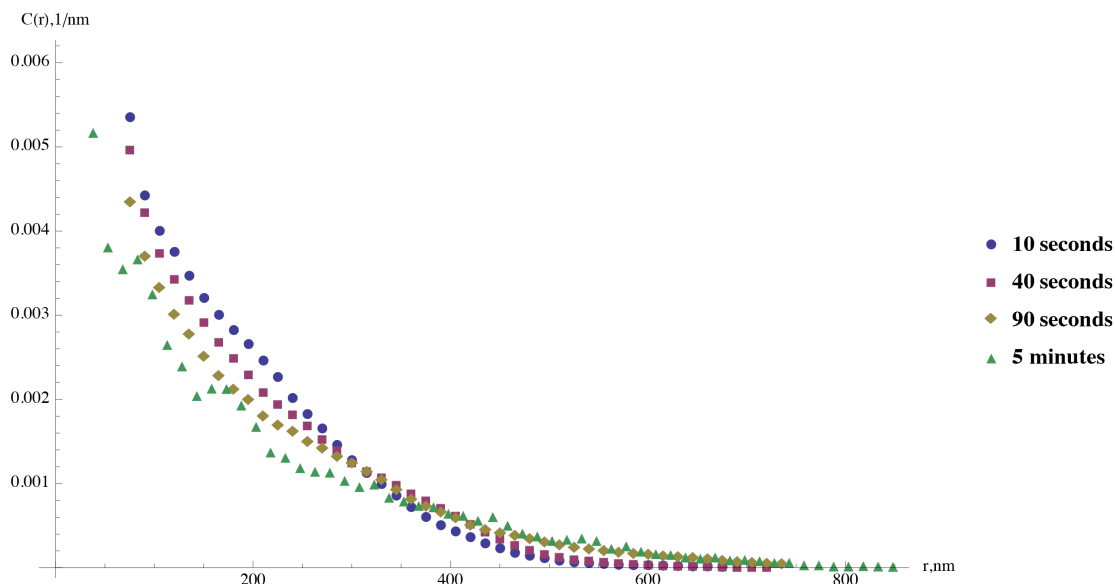


Figure 4.3 – Concentration function calculated for samples of DNA with deposition time varying from 10 seconds to 5 minutes. Preparation procedure is described in details in chapter 3.

We have fitted our data with eqs. (4.11) and (4.18) to check if our theory can be valid for relatively long molecules of $L = 1020$ nm (twenty times the persistence length). These plots are presented on fig. 4.4. If we look at the data for $t = 10$ s, we see that the model of 3D equilibrium conformation projected onto a plane (eq. (4.18)) describes the data the best (apart from the first part of the plot, which we do not consider due to the lack of stiffness effect in our models), while the 2D model described by eq. (4.11) deviates from the data. For $t = 5$ min the situation is the inverse: the latter model approximates the data significantly better. Intermediate relaxation times show results that are in between of these two models. To quantify the quality of the approximation, we used the function *DistributionFitTest*³ from Mathematica[®] software package. The values of the goodness of fit g returned by this function are presented on fig. 4.5. The data proves our hypotheses for the use of models represented by eqs. (4.11) and (4.18) to describe the molecules in three- and two-dimensional conformation respectively. Indeed, the value of g for the 3D model is close to 1 for the shortest relaxation time and decreases gradually to 0.1 for the longest relaxation time, while the 2D model gives g increasing from

³This function reports a goodness of fit value for the approximation of *data* by the specified model from the range from 0 to 1 (bad and good approximation respectively). This function performs complex calculations and takes into account not only the mean squared deviation of the data from the model, as for a commonly used R^2 criterium, but also the correlation of the error residuals. Citation from the description of the function from the Wolfram Mathematica 10.0 user manual:

"DistributionFitTest performs a goodness-of-fit hypothesis test with null H_0 that data was drawn from a population with distribution *dist* and alternative hypothesis H_a that it was not."

0 for 10 seconds relaxation time to 0.8 for 5 minutes one. We also fitted the data with the stationary solution of the diffusion equation described by eq. (4.5). As we described before, we do not anticipate this model to represent correctly the concentration function of a polymer due to the aforementioned arguments of integration of the trajectory over time. It is confirmed by the data on fig. 4.5: approximation by this model results in $g < 0.2$ for all the relaxation times.

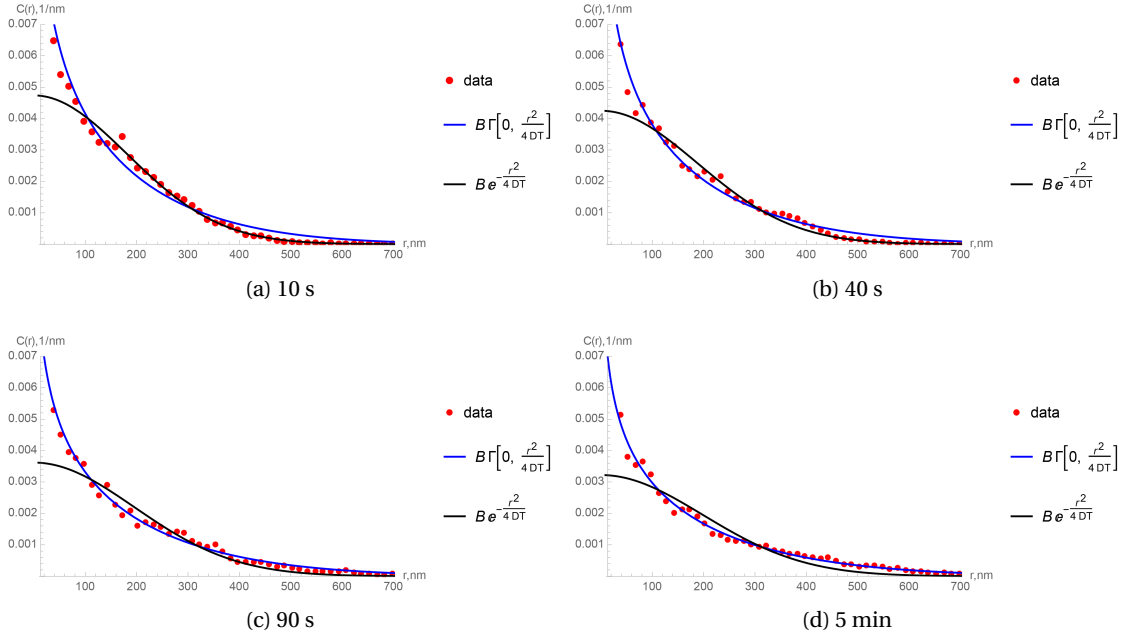


Figure 4.4 – Concentration function calculated for samples of DNA with deposition time varying from 10 seconds to 5 minutes and its fit with eq. (4.5) (dashed line) and eq. (4.11) (thick line). One can see that the latter, dynamic one approximates better the experimental data, especially in the region of relatively large r (see fig. 4.4 for further clarification).

Let us look deeply into the parameters extracted from fitting of our data with the aforementioned models presented in table 4.1. Both eqs. (4.11) and (4.12) provide us with the normalization constants c_1 and c_2 respectively, which have the same dimensionality as the concentration function (nm^{-1} in 2D), and the product DT , which have the dimensionality of nm^2 . The former ones gradually decreases in time, which is expected to hold the condition $L = \text{const}$, because the concentration function spreads for the larger and larger r (see fig. 4.3). The latter one slightly increases with the increase of the relaxation time.

We are looking for a complete description of the effective diffusion coefficient D and integration time T . First of all, we have to note that in all the equations presented in the current thesis these two parameters are not independent, but are presented as a product DT , so we cannot extract them separately. We expect that if our model describes the trajectory of a particle, which starts at the point $r = 0$ at $t = 0$, diffuses freely during the time T and after that reaches

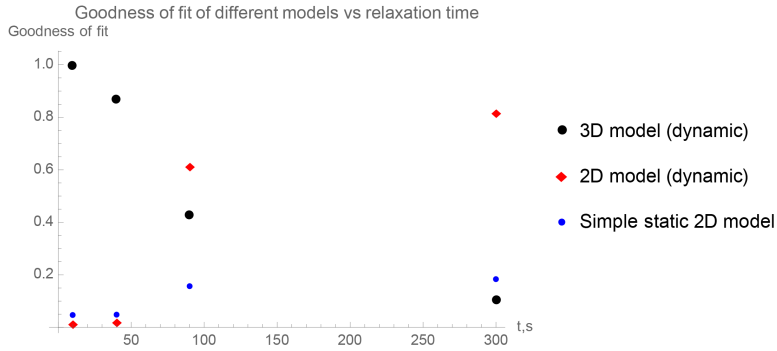


Figure 4.5 – Goodness of fit g , obtained using `DistributionFitTest` from Mathematica software for different models and relaxation time. This function returns the result from 1 to zero, if the model fits well the data or has significant statistical deviations respectively. One can see, that g for 3D-2D projection model (eq. (4.18), thick black circles) decreases from 1 to 0.1 with the increase of the relaxation time, while 2D equilibrium model (eq. (4.18), red diamonds) increases from 0 to 0.8. One can also see, that the stationary solution of the diffusion equation (eq. (4.5), thin blue circles) gives inappropriate estimation with $g < 0.2$.

the point r_0 , the mean squared displacement of this particle should correspond to the one described by the laws of diffusion: $\langle r_0^2 \rangle = 2dDT$, where d is again the dimensionality of the system. But this displacement is nothing else than the mean squared end-to-end distance of the contour of the polymer. So we expect the correspondence:

$$\sqrt{\langle R_{ee}^2 \rangle (L = 1020\text{nm})} = \sqrt{2dDT}. \quad (4.32)$$

If we look at table 4.1 and analyze the parameters presented in it, we can figure out the following facts. The normalization constant c_2 from eq. (4.12) is in a good agreement with the one predicted by eq. (4.24) for 10 seconds relaxation time, but the deviation becomes more and more significant for bigger t . The other model, eq. (4.11), on the contrary, provides a proper normalization constant c_1 for $t = 90$ seconds and 5 minutes. This result is not surprising, if we recall the goodness of fit for these models (fig. 4.5). A possible source of the statistical error in the calculation of the normalization constants is the limits of integration of the concentration function over space: our models are represented by functions that are strictly positive over all the range of r , while $c(r) \equiv 0$ for $r > L$, so the integration over the positive "tail" of $c(r)$ for $r \in [r, \infty]$ introduces slight constant error.

We have also extracted the product DT and compared it with the end-to-end distance of the molecules using eq. (4.32). To do so, we used the dimensionality of the system d calculated in the previous chapter (table 3.2, we took the average value of d calculated by the two presented techniques). One can see the results in table 4.2, which proves our choice of the models for the samples with different relaxation time. As expected, the value of mean squared end-to-end distance for the sample with $t = 10$ s is relatively close to the one calculated with eq. (4.32) for

the 3D-2D model (we have to admit the fact that experimental error is the highest in this case), while in case of $t = 5$ minutes the 2D model produces the value of \sqrt{DT} , which coincides with R_{ee} within the range of the experimental error. Certainly, the samples with intermediate relaxation times are correctly described by neither of these models, as the molecules are not in equilibrium for these t .

As it was expected, the concentration function is not represented by the stationary solutions of diffusion equation, eqs. (4.5) and (4.6). We do not present the plots of the fit of our data with these equations, because they exhibit significant deviation of the data from the model with the goodness of fit close to zero.

t_{relax}	Fit with $c_2 \exp\left(-\frac{r^2}{4DT}\right)$			Fit with $c_1 \Gamma\left(0, \frac{r^2}{4DT}\right)$		
	$c_2, 10^{-3} \text{ nm}^{-1}$	$\sqrt{4DT}, \text{ nm}$	$c_2^* = \frac{L}{4\pi DT}$	$c_1, 10^{-3} \text{ nm}^{-1}$	$\sqrt{4DT}, \text{ nm}$	$c_1^* = \frac{L}{4\pi DT}$
10 s	4.74 ± 0.09	256 ± 4	4.95	1.56 ± 0.07	493 ± 22	1.34
40 s	4.25 ± 0.09	266 ± 5	4.58	1.38 ± 0.05	521 ± 19	1.20
90 s	3.62 ± 0.10	277 ± 7	4.23	1.15 ± 0.03	555 ± 16	1.05
5 min	3.22 ± 0.11	281 ± 9	4.11	1.00 ± 0.02	579 ± 14	0.97

Table 4.1 – Parameters extracted from the fitting of the concentration function of the samples with different relaxation time using eqs. (4.11) and (4.18). The values of the normalization constants c_1^* and c_2^* are extracted from eqs. (4.23) and (4.24) respectively with $L = 1020\text{nm}$ and corresponding DT from this table.

t_{relax}	$\sqrt{\langle R_{ee}^2 \rangle} (1020\text{nm})$	d	$\sqrt{2dDT}, 3\text{D-2D model}$	$\sqrt{2dDT}, 2\text{D model}$
10 s	395 ± 20	3.5 ± 0.5	338 ± 25	652 ± 50
40 s	435 ± 15	3.1 ± 0.4	331 ± 21	648 ± 45
90 s	497 ± 15	2.5 ± 0.2	309 ± 16	620 ± 35
5 min	599 ± 10	2.00 ± 0.05	281 ± 12	579 ± 20

Table 4.2 – End-to-end distance of the molecules and its relation to the parameters D and T . The value of d is an average value from the two rows in table 3.2.

4.3.1 Simulation

We simulated contours of polymer molecules using the same technique described in the previous chapter. Each set contained 10000 molecules. The first set of samples consisted of three-dimensional polymer contours, represented by 1020 segments of $r_0 = 1 \text{ nm}$ each. The angle between the two consecutive segments is some arbitrary value, which is normally distributed with zero mean and standard deviation $\sigma = \sqrt{2r_0/l_p} = \sqrt{2/50} = 1/5$ radians. The three-dimensional concentration function $C(R)$ for these molecules is presented on fig. 4.6a along with its fit with eq. (4.12). Our model approximates well the data with fit goodness g

Chapter 4. Concentration function

equal to 0.999996. Then we have projected the same contours onto a plane and calculated two-dimensional concentration function $c(r)$. The data presented on fig. 4.6b was fitted with eqs. (4.11) and (4.18). The first model corresponding to the 2D projection of 3D equilibrium state turns out to be the most appropriate with fit goodness of 0.93, while the latter exhibits significant statistical deviations with $g < 0.0004$. Finally, we performed similar simulation of the molecules initially in two dimensions (standard deviation of the angle between the segments is now $\sqrt{r_0/l_p} = \sqrt{1/50} = \sqrt{2}/5$ to match the persistence length of 50 nm). $c(r)$ for this set of samples was again fitted with eqs. (4.11) and (4.18), but now the situation is the inverse with fit goodness of 0.005 and 0.985 respectively. We have to stress that our simulation is oversimplified and does not take into account the excluded volume effect; nevertheless, it is able to prove again the correctness of our proposed theory. It also enables us to verify our theory for a three-dimensional equilibrium case in 3D space (eq. (4.12)), which is evidently not possible by the means of AFM.

The analysis of the fitting parameters presented in fig. 4.6 reveals qualitative correspondence between the experiments *in vitro* and *in silico*. We can see that the normalization constants obtained directly from the fit are in close relation with the ones calculated using the normalization procedure with current values of DT . We also observe, that the values of end-to-end distance differ not more than 10 – 20 % from the values of $\sqrt{6DT}$ and $\sqrt{4DT}$ for 3D and 2D cases respectively.

Sample	Fit with $\frac{c_1}{R} \Gamma\left(1/2, \frac{R^2}{4DT}\right)$ (3D model)			$\sqrt{\langle R_{ee}^2 \rangle}$, nm	g
	$c_1, 10^{-3}\text{nm}^{-1}$	$\sqrt{6DT}$, nm	$c_1^* = \frac{L}{4\pi^{3/2}DT}, 10^{-3}\text{nm}^{-1}$		
3D	2.334 ± 0.008	342 ± 2	2.35	311	0.999996
3D→2D	Fit with $c_2 \exp\left(\frac{R^2}{4DT}\right)$ (3D→2D model)			254	0.93
	$c_2, 10^{-3}\text{nm}^{-1}$	$\sqrt{6DT}$, nm	$c_2^* = \frac{L}{4\pi DT}, 10^{-3}\text{nm}^{-1}$		
3D→2D	7.01 ± 0.04	256 ± 2	7.33		
2D	Fit with $c_1 \Gamma\left(0, \frac{R^2}{4DT}\right)$ (2D model)			428	0.985
	$c_1, 10^{-3}\text{nm}^{-1}$	$\sqrt{4DT}$, nm	$c_1^* = \frac{L}{4\pi DT}, 10^{-3}\text{nm}^{-1}$		
2D	1.662 ± 0.004	448 ± 1	1.62		

Table 4.3 – Parameters of the fit of the concentration function data for the simulated samples in 3D, the same samples projected onto a plane and for the ones initially in 2D with the models eqs. (4.11), (4.12) and (4.18) respectively.

4.4 Conclusions

In this chapter we have described our theoretical approach of using the diffusion – random walk – polymer physics analogy and its application to the experimental results, namely AFM images of DNA deposited onto mica surface with different relaxation time. This method can help us to analyze and predict polymer behavior on the surface and in bulk along with "more classical" techniques described in chapters 1 and 3. Our approach starts with defining *concentration function* and applying it to Fick's law of diffusion (eq. (4.2)). After that, we have noticed that the actual concentration function should be represented by a solution of eq. (4.2), integrated over time from $t = 0$ to some constant T .

We have covered three different models. The first one is based on the solution eq. (4.8) of diffusion equation in three dimensions, corresponding to the case, when all the particles are lumped in a point $r = 0$ at $t = 0$ and then diffuse in 3D space. Integration over time produces eq. (4.12). Of course, we are not able to record the contours of molecules in volume by AFM, but we have performed Monte-Carlo simulations of these contours and proved this model to be a correct representation of the concentration function in 3D.

The second model is basically the projection of the first one onto a 2D plane and is represented by eq. (4.18). We have shown that this model describes correctly the concentration function of both the contours of DNA molecules after 10 seconds relaxation on the surface and the aforementioned simulated 3D contours projected onto a plane.

The third model is analogous to the first one, except for that all the calculations are performed in two dimensions (eq. (4.7) is integrated over time and results in eq. (4.11)). This model is in good agreement with the concentration function throughout the space, and a product of the effective diffusion coefficient D and time constant T . We have not elaborated a clear description for every of the latter parameters, but we assume, that the value of $\sqrt{2dT}$, where d is the dimensionality of the system, should be equal to the mean end-to-end distance of the molecules, and this is the case in our experiments.

To prove the choice of the model in every particular case, we used the function Distribution-FitTest, which returns the value from 1 (if the model describes perfectly the data) to 0 (if the model is completely inappropriate). This function supports our intuitive judgment about the goodness of fit for every case.

All our models describe the distribution of molecules in equilibrium. That is why it is not surprising that the samples with intermediate relaxation time can be approximated by neither of the presented models and the data points lie somehow in between of the plots of eqs. (4.11)

Chapter 4. Concentration function

and (4.18). We can only see, that with the increase of the relaxation time the data tend more and more towards the 2D model and less and less to the 2D projection of the 3D model. Future research in this area should include the elaboration of the intermediate transition model of the concentration function.

To conclude this and the previous chapter, we have performed the quantitative study how linear DNA molecules (as a good model of a polymer) reaches the equilibrium on the surface of mica modified with Mg^{2+} ions and described the dynamics of the transition from 3D- to 2D-equilibrium conformations. Specially designed concentration function is sensitive enough to distinguish between different dimensions, and what has to be specially underlined - without any prior knowledge of the persistence length of the polymer molecules. We hope, that it paves the way for further fruitful application of the proposed approach to study the physics of other polymers. We have also applied our models to study in details the properties of circular DNA molecules; this work is not finished yet and the results will be published elsewhere.

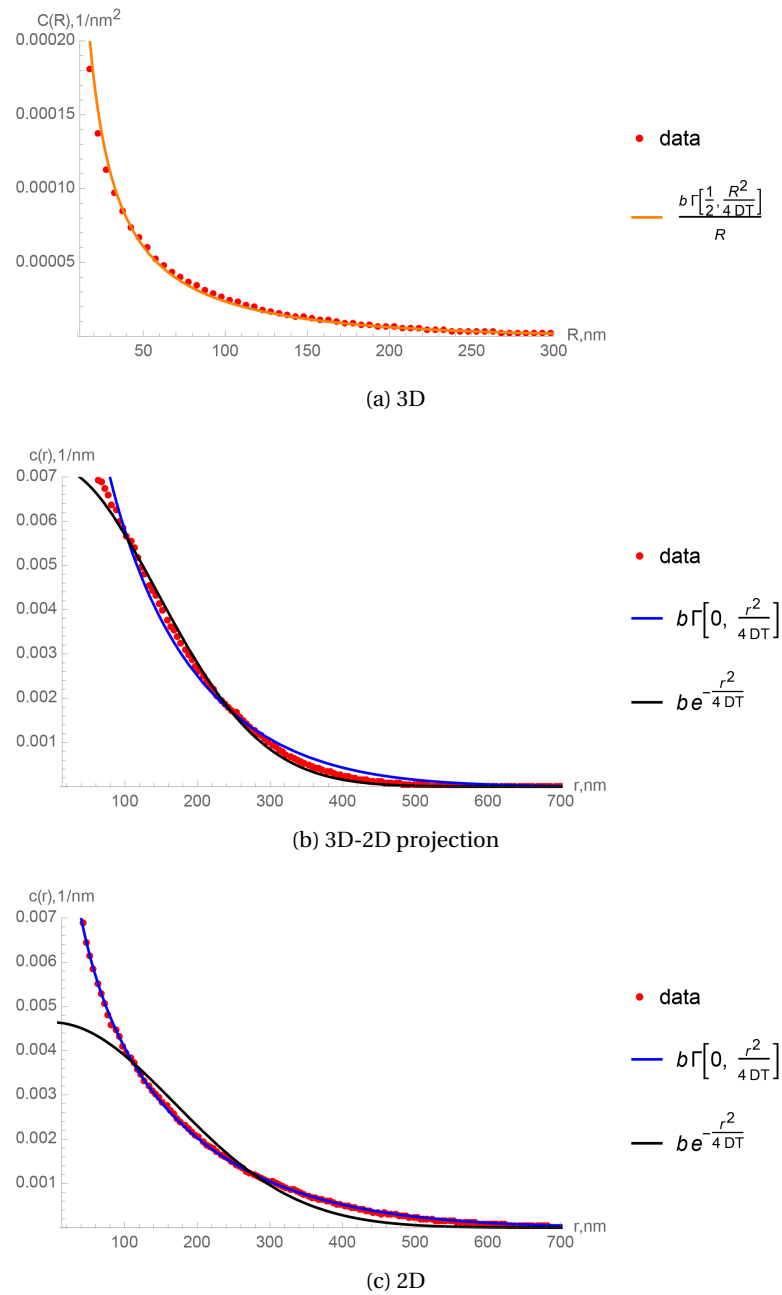


Figure 4.6 – Concentration function of the contours of the simulated polymer molecules. (a): $C(R)$ for the molecules in three dimensions and its fit with eq. (4.12), $g = 0.999996$; (b): $c(r)$ for the same contours initially simulated in three dimensions and then projected onto a 2D plane and its fits with eqs. (4.11) and (4.18), $g = 0.93$ and $g < 0.0004$ respectively; (c): $c(r)$ for the molecules in equilibrium in two dimensions and its fit with the same models, $g = 0.005$ and 0.985 respectively.

5 Force spectroscopy

Study hard what interests you the most
in the most undisciplined, irreverent
and original manner possible.
— Richard P. Feynman

5.1 Introduction

The idea of using an AFM to measure interaction forces appeared about twenty years ago [37, 77, 45, 44]. Indeed, AFM permits one to measure the bending of the cantilever, i.e. displacement of the end of the lever x , nm . If we could know a spring constant k , $[k] = N/m$, then the value of the force applied to the end of the tip is given by $F = -kx$ (fig. 5.1). In case of contact mode AFM scanning, this force is usually kept constant (fig. 5.1b). In the same time, we can measure the interaction force between two objects, for example, two proteins, by attaching one of them to the surface and the other to the tip. When it approaches the surface, two proteins can touch each other and interact. Then we start to pull the tip apart from the surface (fig. 5.1c). The cantilever start to bend, the force increases until the system is ruptured. If the proteins are attached to the surfaces strongly enough, the specific, non-covalent bond between them is broken first.

In 1994 Lee et al. [77] and Florin et al. [45] revealed their measurements of the interaction force between biotin and its specific receptors streptavidin and avidin respectively. Their experiments showed experimental evidence of the discrimination between specific and non-specific interactions. Moreover, they were able to detect interaction events between single molecules! Both of them attach biotin to the surfaces of the tip and the sample and add avidin/streptavidin to the solution, exploiting the fact that one tetrameric receptor can bind up to four ligands [130]. The first article shows symmetric technique used biotinylated BSA to attach biotin to the tip and mica surface non-covalently. Such technique shows strong attachment of biotin to the surface permitting the scientist to observe events of the rupture of

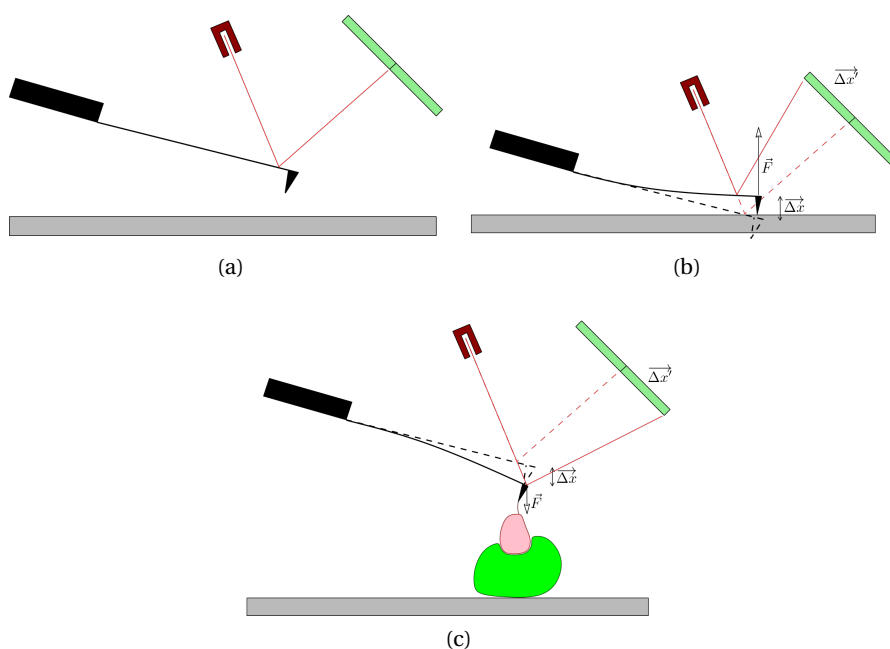


Figure 5.1 – Scheme representing the possibility of measuring forces between the surface and the cantilever. (a): AFM cantilever out of contact with the surface; (b): cantilever is pushing the surface; (c): cantilever is pulling apart a protein-protein complex. Photodiode permit us to measure the displacement of the end of the tip and to calculate the resulting interaction force using the formula $|F| = k\Delta x$, where k is the spring constant, which needs to be calibrated. Most of researchers define forces in cases (b) and (c) to have positive and negative values respectively.

the specific bond between ligand and receptor rather than protein detachment events. The second article uses the same technique to attach the protein to the tip but uses a biotinylated agarose bead as a sample. Both articles report the values of the interaction force between ligand and receptor having the value of several hundreds piconewtons compared to several tens piconewtons for non-specific interaction.

Many other interacting pairs were studied [143, 112, 131, 48, 76, 115]. Among others, these pairs include barnase – barstar [115], proteins capable of dimerisation (actin – actin [94], insulin – insulin [139]), myelin basic protein – lipid bilayers [95] and many antigen – antibody complexes [123, 68, 3]. In general, a specific bond breaking event is similar to presented on fig. 5.2. One can distinguish on this curve a non-specific interaction peak, representing adhesion of the tip to the surface, and a specific interaction event.

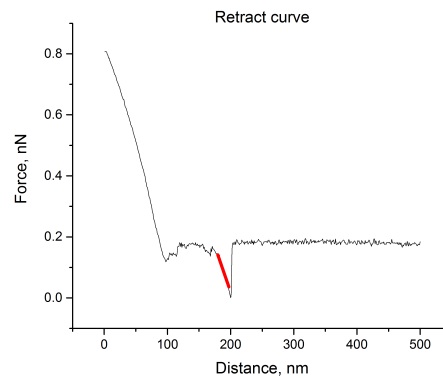


Figure 5.2 – Typical retract force curve containing an event of barnase-barstar specific interaction bond rupture. The value of the force equal to 180 pN is obtained from the height of the peak and the value of loading rate of 6 nN/s is calculated from the slope before the peak (marked with a red line).

5.1.1 Force - loading rate dependence

One could think that the specifically interacting pair should have the same interaction force in the same solution. But it turns out that the force depends on how fast you are pulling your system apart. Indeed, imagine a hook installed on the wall using double-sided adhesive tape. If you try to detach it abruptly, you will find out that you can hardly do it. But if you do it very slowly, you need to apply much less force. Finally, after a long time the same hook can detach itself even under its own weight.

This topic was firstly discussed by Bell [9] and then his model was extended by Evans et al. [40]. To describe the phenomenon, they use the term *force loading rate*, F' , N/s. This value defines the rate with which the force applied to the interacting pair increases. For any biochemical complex, if one could wait sufficiently long, the complex will dissociate without any applied force. So, at zero loading rate the rupture force is zero as well. The faster the applied force is increasing, the higher the rupture force is. For every F' the bond rupture force does not have a constant value but rather exhibit a broad distribution, so when we say "force" we mean the most probable rupture force (a discussion about difference between most probable and average bond breaking event force can be found in [114]).

Many experiments measuring the dependence of the force on the loading rate were performed both *in silico* [40, 70, 59] and *in vitro* [143, 112, 131, 48, 76, 115]. But the first ones usually exploit MD or Monte-Carlo techniques and normally cannot access the timescale longer than microseconds, so typical loading rate values are in order of 10^{12} pN/s. AFM does not permit one to go too fast. If two proteins were absolutely rigid and attached directly to the surfaces,

then the loading rate would be defined by the following formula:

$$F' = kv, \quad (5.1)$$

where v is the retract speed (the velocity with which piezoscaner retracts the cantilever) and k is the spring constant of the cantilever. In reality the proteins have their own final stiffness and are attached using some soft substance, the system can be imagined as a set of serially connected springs with an effective stiffness $1/k = 1/k_1 + 1/k_2 + \dots$, the value of which change locally over the surface of the sample. This explains why the real loading rate can vary even if cantilever stiffness and retract speed are kept constant. A typical loading rate value is in the order of 10^4 pN/s. To cover bigger range of F' , we have to change the parameters k and/or v (see eq. (5.1)). The first one depends on the cantilever used, but can be adjusted even within one experiment: many cantilever chips contain several levers (see fig. 5.3). The second one can be controlled with standard AFM software. Anyway, the range of the loading rate close to molecular dynamics simulations, unfortunately, is not accessible by the means of AFM.

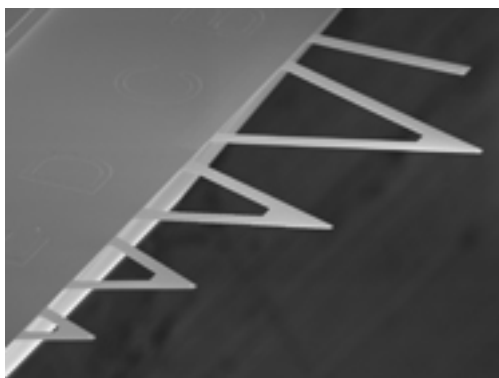


Figure 5.3 – Electron microscope image of an AFM tip (MLCT type, Bruker) showing 5 of its 6 cantilevers with nominal spring constant ranging 0.01 - 0.6 N/m. Such tip is widely used for SMDFS. Image taken from the website of the producer (www.brukerafmprobes.com).

The technique of the measuring force *vs.* loading rate dependence for the interaction of molecular scale objects using AFM is called Single Molecule Dynamic Force Spectroscopy (SMDFS). There exist so-called Bell-Evans model, which predicts the dependence of the most probable bond breaking force on the loading rate. There are several regimes in this model, but a researcher using SMDFS is usually assumed to get data being in the part, where the dependence can be described with the following equation:

$$F = \frac{k_B T}{\Delta x} \ln F' + \frac{k_B T}{\Delta x} \ln \left(\frac{\Delta x}{k_{off}^* k_B T} \right), \quad (5.2)$$

where k_B is Boltzmann constant, T is a temperature, Δx is a barrier thickness and k_{off}^* is the dissociation rate where no external force is applied (natural off-rate). This formula is valid under the assumption that the dissociation energy profile has a single potential barrier. It

permits one to extract the data about the width of the barrier and k_{off}^* from a SMDFS. Some articles report corrections to the simple Bell-Evans model as well as the dependence of the average bond rupture force on the loading rate (which is close to eq. (5.2)) [114].

Force spectroscopy is a versatile tool. The possibility of using it to probe intermolecular interactions was already discussed above. Another interesting application is to study protein unfolding [20, 51, 10]. These papers report the following results. A multimodular protein, I27, is attached to the surface and then probed with an AFM tip. During an approach-retract event there is some probability that the protein attaches to the tip as well. In this case, upon retraction cycle the protein will be stretched and thus unfolded. Experiments show that the domains of this protein exhibit abrupt unfolding events, which look like the detachment of an interacting pair described above (fig. 5.4).

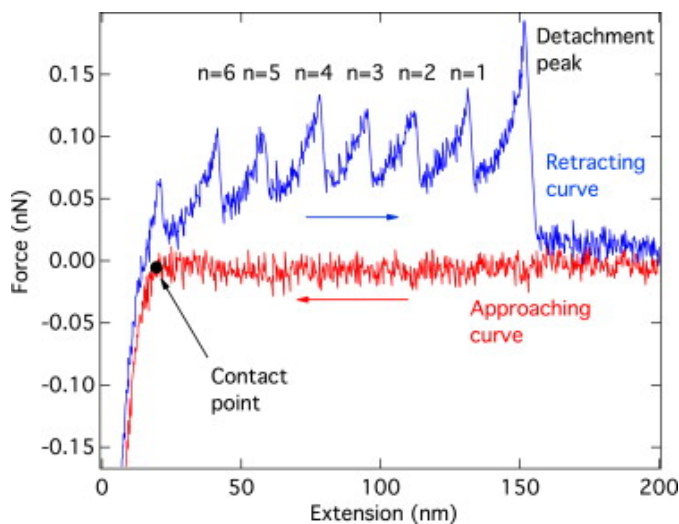


Figure 5.4 – Typical force curve of a single I27 molecule stretching. The retracting curve represents single unfolding events of the domains of the protein. Image taken from [10].

5.1.2 Spring constant calibration

An important aspect, which we have ignored up to now, is the calibration of the spring constant of the cantilever, which is required to extract the value of the force from AFM. It seems to be apparent that a cantilever is a precise nano-fabricated object with predictable and repeatable characteristics, so we could just take a nominal value of k . But it turns out that this parameter varies substantially from tip to tip (for example, Bruker gives for its DNP-10 (D) tip the value of k ranging 0.03 - 0.12 N/m; such fourfold spread is typical) and one has to use some procedure to calibrate it before the experiment. There are several techniques to do it. Probably the most used, reliable and relatively simple to implement is the thermal fluctuation analysis [81]. The idea of this method is the following: the cantilever vibrates under Brownian motion of the surrounding media. If we watch a single degree of freedom (bending in one plane), the

following relation hold on:

$$\frac{1}{2}k_B T = \frac{1}{2}k \langle x^2 \rangle, \quad (5.3)$$

where k_B is Boltzmann constant, T is the temperature and $\langle x^2 \rangle$ is an average squared displacement of the end of the tip. We can obtain power spectrum density for this noise and integrate it, applying Parseval's theorem:

$$\int |x(t)|^2 dt = \int |\tilde{x}(\omega)|^2 d\omega, \quad (5.4)$$

where $\tilde{x}(\omega)$ is a Fourier transform of the signal (displacement) $x(t)$. The theorem permits one to proceed from the integration of the signal over time to the integration of its Fourier transform over the spectrum.

Another, more precise technique described in [81] consists in obtaining a thermal noise spectrum of the cantilever oscillations and fitting its first peak with equation

$$d_{tot}(f) = \sqrt{\frac{2k_B T / (\pi k f_0 Q)}{(f^2/f_0^2 - 1)^2 + (f/(f_0 Q))^2} + d_{ds}^2}, \quad (5.5)$$

where $d_{tot}(f)$ is power spectrum density, f - frequency, f_0 - resonance frequency, Q - quality factor and d_{ds} - detection system white noise (constant) and k is spring constant. As k is one of the fitting parameters, we can get it directly from the noise power spectrum. Such technique is actually implemented in many commercial AFM intended for SMDFS, for example, Nanoscope III.

5.1.3 Surface treatment and protein deposition

As usual for AFM experiments on biological samples, the substrates have to be deposited on a flat surface. Most of SMDFS researches deal with molecule-molecule interaction, so the second interaction partner has to be deposited on an AFM tip. Mica remains one of the most used sample surfaces. It is atomically flat, relatively cheap and easy to work with. One can make the surface of mica sterile by applying an adhesive tape on it and then detach the tape. After such a cleaving process mica is separated by two layers and each of these layers are chemically clean.

There is a widely used procedure to attach covalently the proteins onto mica surface. To do it we perform the following procedure:

1. Apply 1% v/v 3-aminopropyltriethoxysilane (APTES, Sigma) solution in water for 5 min;
2. Rinse with UHQ water;

3. Treat the surface with a 1% v/v glutaraldehyde solution in water for 15 min;
4. Rinse again with UHQ water;
5. Apply a solution of a protein (typical concentration of 250 mg/L) for 15 min;
6. At the end the proteins that did not react with the surface and were loosely bound are removed by extensive washing with PBS buffer.

Typical AFM tips used for SMDFS are made of silicon nitride, Si_3N_4 . They are in general much softer than the ones used for tapping mode AFM scanning and have nominal spring constant ranging 0.02 – 0.6 N/m. We attach proteins to such tips directly using glutaraldehyde. Briefly, the algorithm of the functionalization of the tip is the same as points 3-6 of the procedure described before. In the beginning the tip can be optionally cleaned by immersing into 1% v/v H_2O_2 solution in water or some other cleaning agent and the rinsed with UHQ water. Already utilized tips can be reused for new experiments. More detailed description can be found in [24].

There is still an ongoing discussion if one has to use special linkers for SMDFS experiments. Some groups insist that their usage is obligatory and attach the proteins to the tip and the sample with benzaldehyde-PEG6-NHS linkers [38]. Nevertheless, we carried out a set of experiments where the proteins were attached using the aforementioned procedure and with special linkers. As no significant difference in loading rate - force curves was observed, we attach the proteins using our procedure without linkers, which is considerably simpler.

5.1.4 Event selection

SMDFS is a powerful technique, which permits one to get statistically robust data. A typical sample/tip set with proteins deposited can be easily used for 1-2 hours. Provided that the approach-retract cycle rate is about 1 Hz (such rate is widely used in our setup), one can get about 3000-7000 retracting force curves for only one sample/tip set. It is often enough to extract a force-loading rate dependence like presented on fig. 5.6, but how to do it quickly and efficiently? Not every force curve contain a specific interaction event and it would be a huge job to select manually such peaks and extract interaction force and force loading rate from every peak. Fortunately, some automation is well established in this area.

As one can see on a typical force curve (fig. 5.2), a specific interaction event, if it is present, appears as a peak with a pronounced triangular shape. In the beginning of the peak the force is gradually increasing (the slope of the curve at this part multiplied by spring constant gives force loading rate), then the force drops almost vertically and the following part of the curve is represented as flat background with some noise.

Such particular shape of the peaks of our interest can be used for their automatic searching on all the force curves. There exist different software solutions to do it [72, 109]. An efficient algorithm for peak recognition is first to convolute the signal with an L-shaped vector having a shape like $\{-1, -1, -1, -1, -1, -1, -1, 7\}$. Such convolution gives low value for all regions but abrupt vertical part of the peak. Setting some threshold for the convoluted value, one finds a candidate for a specific interaction peak. To separate it from the noise and non-specific peaks corresponding to the adhesion of the tip to the surface, the software usually introduces different criteria like minimal distance from the contact point to the peak, typical range of interaction force, slope of the curve before vertical part (detaching event). Adjusting these criteria one can quickly process all the curves in automatic mode with high precision. We have compared the data obtained in automatic mode of *Hooke* software [109] and compared to the one collected by manually processing of hundreds of force curves. No significant deviation was found.

5.2 Barnase-barstar interaction

5.2.1 Scientific interest for the protein pair

Ribonuclease barnase (110 amino acids) and its inhibitor barstar (89 amino acids) is much exploited in practice and nowadays can be considered as a very prospective "constructor" to create multivalent complexes used as innovative protein therapeutics. They are also considered as a model system for the studying the protein folding/refolding problem. Both of them can be easily expressed, are stable and sustain harsh environmental conditions. They were proven to restore reversibly their structure and function after denaturation. These proteins form the second most effectively interacting pair after avidin/streptavidin-biotin (dissociation constants are around 10^{-14} and 10^{-15} M respectively [90, 111, 57]). Yet there is not much force spectroscopy data concerning the interaction of the aforementioned proteins.

All arguments above make us believe that the study of barnase-barstar system serves for the better understanding of its biological and medical applications. However, there is another motivation for this study. The electrostatic contribution to the binding of this pair is still not clear. While biochemical experiments clearly show a favorable energetic contribution of an electrostatic interaction (the free energy of the binding reduces on the value of 3.1 kcal/mol for 500 nM NaCl concentration compared with the pure water case [111]), theoretical predictions differ even in sign. For example, some papers predict the absence of electrostatic contribution [116] or its unfavorable effect [78], while others clearly predict the reduce of the free energy due to the electrostatic interaction [118, 34].

5.2.2 Experimental setup

We have performed a set of force spectroscopy experiments on this pair. We have produced barnase and barstar in *E. coli* cells transformed by the corresponding plasmids (described in details in [115]). The tip and the surface were functionalized with the same procedure described in details before (see section 5.1.3), the proteins were attached covalently without special linkers. We have obtained force spectroscopy data using the AFM Nanoscope IV "Picoforce", Veeco Instruments. All experiments were performed with silicon nitride V-shaped cantilevers (Veeco) with a length of 200 μm and a nominal spring constant of 0.06 N/m (actual spring constant of every cantilever was calibrated before the measurements using thermal noise built-in calibration procedure of the microscope).

AFM images of the proteins deposited on the substrate with the procedure mentioned above show a smooth protein monolayers with single proteins seen as globules. The concentrations used result in 20-25% of successful specific interaction events for thousands of total number of approach-retract cycles.

We have processed the retract curves obtained from force spectroscopy experiments with an automatic fuzzy logic-based software (see section 5.1.4) [72]. The software permits us to extract quickly the data for the specific interaction events between the proteins from a big set of force curves (1000-1500 measurements was a typical number for the same cantilever/sample pair). A typical retract force curve containing an interaction event is presented on fig. 5.2.

We have performed a set of control experiments to prove that the interaction we observe is the specific interaction between barnase and barstar. Some of these experiments can be called "partial": non-functionalized mica and tip, functionalized tip and non-functionalized mica or the reverse. We have carried out the force spectroscopy experiments of protein pairs barnase-barnase and barnase-avidin. The last control was force spectroscopy of barnase-barstar with one of these proteins added to solution (concentration of 250 mg/L) to block receptor/ligand. We have obtained not less than 500 force curves for every control experiment. Although in all such experiments we observed events mimicking specific interaction, they appear only on less than 1% of all approach-retract cycles. Such events can be considered as noise and are almost unavoidable in the field.

5.2.3 Results

The main results of the study consists of *force F vs loading rate F'* plot. This graph is obtained using the following procedure. We have extracted points (F and F') from all specific interaction rupture events. Then we have sorted this data by F' and divided by chunks of 2 nN/s. The

middle value of the chunk gives us the value of loading rate. Then we have plotted a histogram of the forces in each of the chunks separately and fitted with Gaussian distribution. The position of the maximum of this distribution gives us the value of force (see fig. 5.5).

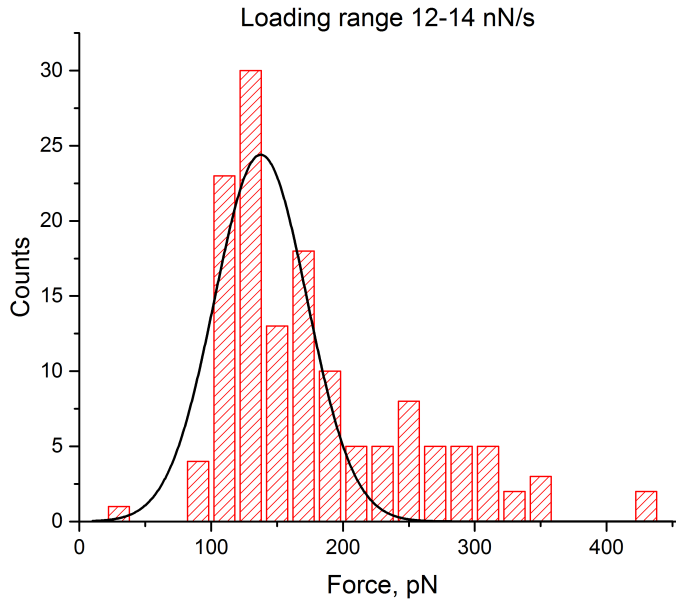


Figure 5.5 – Example of the calculation of one point of force - loading rate plot. The histogram presents the values of the force for the loading rate interval of 12-14 nN/s.

The results are presented on fig. 5.6. We used aforementioned Bell-Evans model of the dependence of the most probable single bond rupture force on the force-loading rate (see section 5.1.1).

The experimental data can be well approximated by a straight line $F = a \ln F' + b$ with $a = 33\text{pN}$ and $b = 710\text{pN}$. Using eq. (5.2) we obtain the values $\Delta x = 0.12 \text{ nm}$ and $k_{off}^* = 14 \text{ s}^{-1}$. To analyze the meaning of this data, one should compare it to the existing measurements of other receptor-ligand pairs. Most of single molecule force spectroscopy experiments show the value of Δx ranging 0.2-1 nm. Nevertheless, an observation of narrower barriers of 0.05-0.15 is not uncommon in the field and are shown, for example, for the aforementioned avidin/streptavidin-biotin pairs [30, 41, 140, 80].

Kramer's theory of thermal bond dissociation in liquids gives $k_{off}^* = \nu_0 \exp\left(-\frac{\Delta U}{k_B T}\right)$ with the typical pre-exponential factor ν_0 ranging $10^9 - 10^{10} \text{ s}^{-1}$. From this theory we can get the height of dissociation barrier as small as $0.75 - 0.85 \times 10^{-19} \text{ J}$. Biochemical data attests for the same conditions we have used $k_{off}^* = 8 \times 10^{-6} \text{ s}^{-1}$, which is considerably smaller than our data,

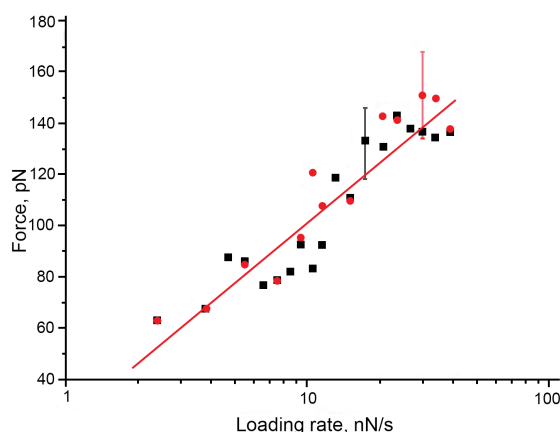


Figure 5.6 – Dependence of the specific interaction force between barnase and barstar on the force loading rate. Black squares: experiments performed in PBS buffer solution; red circles: experiments performed in PBS buffer solution with additional 650 mM NaCl. Measurements are performed at 25 °C. Typical experimental errors are shown for one point from each experiment.

and reports the value of the dissociation barrier of $1.3 - 1.5 \times 10^{-19}$ J. Such discrepancy is not uncommon in the field, other studies already reported earlier the values of off rates ranging $1 - 30 \text{ s}^{-1}$.

The main source of this discrepancy is the oversimplified model of a single-well barrier. We have already discussed above the ambiguity of experimental and theoretical data about the contribution of the electrostatic interaction to the energy landscape of barnase-barstar dissociation. We are convinced that the electrostatic interaction has favorable contribution and the barrier width is a few nanometers large. This results in the fact that even smallest observable in our experiments stretching force of a few tens piconewtons already smears out an electrostatic interaction-related broad external barrier fig. 5.7.

To prove this conclusions we have performed a set of experiments varying the concentration of NaCl from 0 to 700 mM. We have never seen any statistically robust dependence of force spectroscopy data on salt concentration. For the experiment with the highest salt concentration used we calculated the following parameters: $a=29 \text{ pN}$ and $b=700 \text{ pN}$. The discrepancy with our experiments without salt is about 10% while an experimental error is estimated to be 20%.

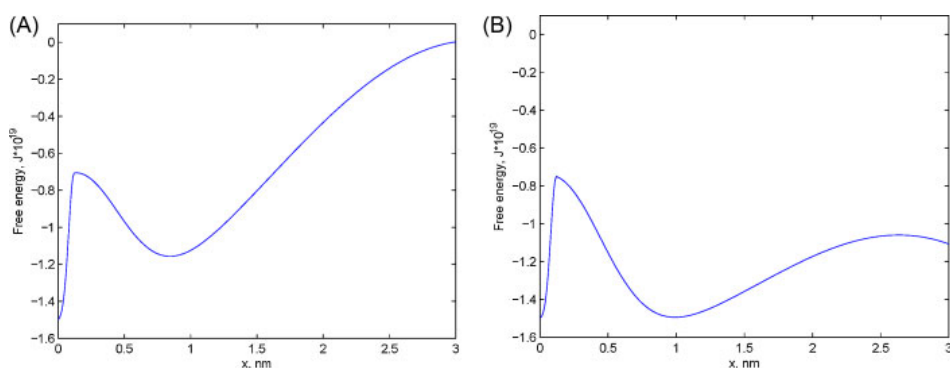


Figure 5.7 – Graph illustrating the plausible energy landscape for the barnase-barstar interaction in absence (A) and presence of the applied force $F = 35$ pN (B). Experimental values for the internal barrier width and height taken from [115] and electrostatic barrier width of 3 nm from [118] are used when preparing this graph. It is seen that already under the action of such a rather small force, the broad electrostatic interaction-related external well is not felt by a force spectroscopy method while narrow internal barrier is only slightly modified.

5.3 Viral proteins

5.3.1 Introduction

Another important part of our force spectroscopy experiments is investigation on the interaction of the proteins, which take part in the viral–cell fusion. The mechanism was widely studied in the past [23, 91, 65, 120, 50, 49, 121, 79]. The fusion of two bilayer membranes leads to the state with lower potential energy, but the system has to overcome a very high kinetic barrier, which leads to the release of large amount of energy [65].

Among all such fusion processes the one of flaviviral-cell receives the greatest interest. This process is already well studied, but many details are still not known and the whole picture is incomplete. It is proven that the penetration of the virus into the cell starts with viral ligand – cell receptor interaction and then continues with the conformational changes of flaviviral surface protein E [121, 120]. In acid environment this protein changes its configuration from dimer into trimer. Then a fusion peptide is released and interacts with the cell membrane, which leads to the fusion of the latter and the penetration of the virus into the cell.

5.3.2 Flavivirus

Flaviviruses are a group in viral family. Their name comes from Latin word "flavus" that means "yellow", as one of the most known viruses in the group is yellow fever viruses. The group also includes West Nile virus (WNV), dengue virus, tick-borne encephalitis virus (TBEV) and many others [117]. They are small RNA enveloped viruses. Most of them are pathogenic for human and animals and are of serious danger for our society. They are mainly transmitted by ticks

and mosquitoes and there is virtually no efficient treatment against many of flaviviruses, yet there exist vaccines for the prevention of the disease. Only yellow fever virus causes about 200 000 infections per year of which 30 000 patients die with about 90% of infection cases in Africa [99]. The treatment against it is mostly symptomatic. Dengue virus is even more common (50 - 100 million infections per year worldwide causing about 13 000 deaths) [98]. Tick-borne encephalitis affects mostly Russia (more than 50 000 infection cases per year) and some European countries and has mortality rate of about 1-2% [125].

Flaviviruses usually have a typical symmetrical spherical or icosahedral shape and size of about 40-65 nm. They were characterized using electron microscopy in 1962 [12] and then their scientific interest was constantly growing. Their structure and mechanism of interaction with cells was widely studied [142, 54, 84, 97]. Still many details of the infection of cells by flaviviruses are not known and the studies of the process of viral-cell membrane fusion could help to understand the whole picture and lead to the development of new anti-viral drugs.

5.3.3 Flaviviral membrane - cell interaction

One of the type of proteins that are proved to be a receptor for flavivirus is the family of integrins. It has been proved experimentally by force spectroscopy that the surface protein E of West Nile Virus interacts with integrin and the bond rupture force has been estimated for this pair. It is also proven that one can use short fragments of the surface protein E for such studies. [15, 16]. Another protein, which represent particular interest in flavivirus-cell interaction investigation, is Laminin Binding Protein (LBP or human laminin receptor, LamR). This protein serves as a cell receptor of TBEV and other flaviviruses [102].

To investigate into flaviviral-cell membrane fusion process, we have studied the interaction of TBE viral protein and its recombinant fragments (E₂₅₇₋₄₁₅ and E₁₋₂₅₄), fragments of WN viral protein E (fragments E₁₋₁₈₀ and E₅₃₋₁₂₆) with their putative cell receptors integrin ($\alpha V\beta 3$) and LBP using SMDFS and compared them to the results of immunoenzyme analysis (see [141]). Enzyme immunoassay measurements were used to complete the results obtained by SMDFS and were performed as described in [15]. Optical density was recorded at 492 nm with Uniscan spectrophotometer, LabSystem, Helsinki, Finland. The experiments were carried out at neutral and low pH values corresponding to the ones of the local environment of surface cell receptors and endosome membranes.

The setup of SMDFS and the procedure of the functionalization of tip and sample were the same as in section 5.2. The proteins were attached onto the tip and sample surfaces without special linkers. The force curves were processed automatically with fuzzy-logic software [72]. A set of control experiments was conducted for each case to prove the specificity of interaction.

5.3.4 Results

Force spectroscopy data of the interaction of the full size TBEV protein E with putative cell receptors clearly show specific interaction, varying for different receptors. The force of the interaction with LBP is larger than of the interaction with integrin for the same loading rate values (see fig. 5.8). It is consistent with our previous experiments of SMDFS of WNV protein E with LBP and $\alpha V\beta 3$ integrin. In case of protein E - LBP pair we see different slopes on force - loading rate dependency for relatively small and large values of loading rate (fig. 5.8a), which can be attributed to a complex double-well dissociation energy profile. We have calculated k_{off} and Δx using a simple Bell-Evans model eq. (5.2) for every slope of this dependence (table 5.1). The data for protein E - integrin interaction does not show two-slope dependence in experimentally accessible loading rate range (fig. 5.8b).

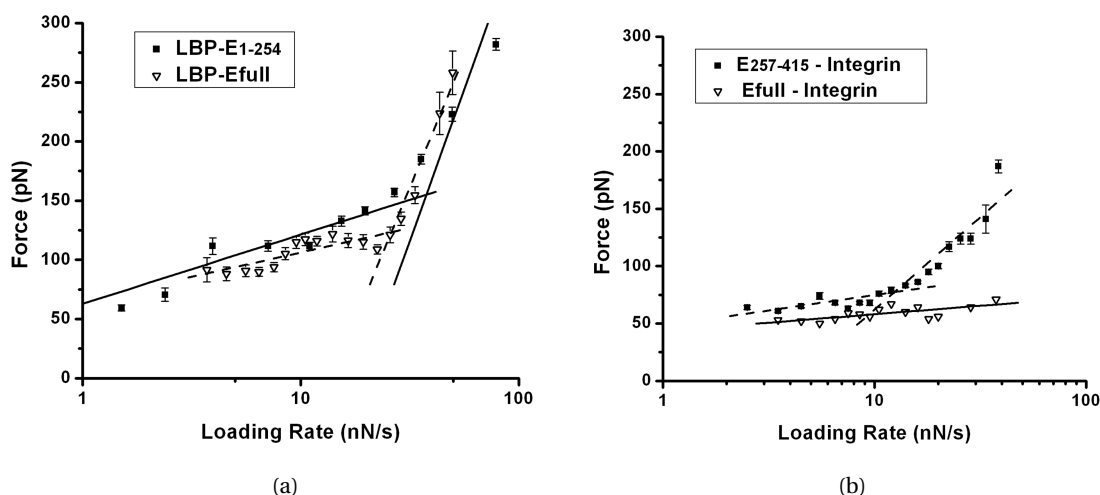


Figure 5.8 – Loading rate - force dependence curve for viral protein E and its corresponding fragments with putative cell receptors.

Next series of data describes the interaction between fragments of protein E and putative cell receptors. We have used two different recombinant fragments of protein E: E_{1-254} and $E_{257-415}$. The first one consists of domains I and II of the full-size protein, while the second one includes domain III but does not include the hydrophobic transmembrane region. We have clearly seen the specific interaction of the first fragment and LBP, as well as of the second fragment and integrin. In the same time, we have seen no specific interactions between the reverse pairs. The kinetic data of the interaction is presented on table 5.1.

We have performed a series of measurements of the interaction of protein E with putative cell receptors at different pH levels. In case of integrin - $E_{257-415}$ pair the interactions at pH=7.4 and pH=5.3 show qualitatively the same result with k_{off} smaller for the latter. On the contrary, in the SMDFS experiments of the pairs LBP - TBEV E_{1-254} and LBP - WNV E_{1-180} we observed rather unusual "flat" force curves (see fig. 5.10) at pH=5.3, while no such plateaus

were observed on the force curves obtained at pH=7.4. Other protein pairs did not show any statistically significant difference of the interactions at different pH levels.

The fact that there is no "flat" force curves if there is no LBP in the interacting pair drives us to the conclusion of some peculiarities of the behavior of LBP. To investigate further this question, we have performed the same procedure of tip-sample modification except for that we have not added the LBP counterpart protein at the final stage of AFM tip functionalization, e.g. LBP was deposited on the sample using usual procedure while the tip was functionalized but did not contain the protein. For such a setup we observed plateaus rather than gradually increasing force peaks on the force curves at both pH=5.3 and pH=7.4.

For the force curves that have plateaus we have plotted the histograms of the value of the step (force difference between plateaus on the same curve) and length of plateaus (see fig. 5.12). The curves representing the interaction of LBP - TBEV E₁₋₂₅₄ at pH=5.3 clearly show force peaks with the period of 70-80 nM and plateau lengths of about 25 nm. Such discretization of plateau length and height was less visible for the LBP - WNV E₁₋₁₈₀ pair and completely unobservable for LBP stretching experiments.

We have performed force spectroscopy and IFA experiments with mAbs recognizing the interaction sites of TBEV E protein (mAb 10H10 [103]) and LBP (mAb 8E4 [88]) to estimate the pH-induced changes of the interaction sites themselves. The results are presented in fig. 5.9. These experiments have not revealed any essential differences in the interaction of proteins for the case of viral protein E in both neutral and acid pH levels and have shown only slight changes for the case of LBP.

To exclude the effect of the surface treatment on the measurements of the force curves of LBP, we have carried out additional experiments with LBP freely (i.e. without any surface treatment) deposited onto freshly detached template-stripped gold surface at pH 5.3. They revealed similar "flat force curves" observed in about 20% of all approach-retract cycles recorded.

5.3.5 Discussion

Our experiments clearly show specific interaction between protein E of both WNV and TBEV and putative cell receptors integrin and LBP. Moreover, we have proved the existence of two different receptors on both types of protein E. For the loading rate of 10 nN/s WNV protein E show the characteristic interaction force of 105 pN with LBP and of 80 nN with integrin. The values of the interaction force for TBEV protein E with LBP and integrin at the same loading rate are 90 and 55 pN respectively. Both viral proteins E and their recombinant fragments interact specifically with the presented putative cell receptors.

Table 5.1 – Kinetic parameters of the interaction of the full size protein E and its recombinant fragments with LBP and integrin. pH=7.4, unless otherwise is stated.

Receptor	Ligand	Comments	Δx , nm	k_{off} , s ⁻¹
LBP	Protein E	Initial slope	0.23 ±0.05	1.5 ±0.05
		Final slope	0.021 ±0.05	72.0 ±30
— " —	E ₁₋₂₅₄	Initial slope	0.15 ±0.05	4.1 ±1.5
		Final slope	0.017 ±0.005	94 ±40
Integrin	Protein E	Initial slope	0.59 ±0.10	0.28 ±0.09
		Final slope	NA	NA
— " —	E ₂₅₇₋₄₁₅	Initial slope	0.30 ±0.05	5 ±3
		Final slope	0.07 ±0.02	53 ±15
		Initial slope (pH 5.3)	0.29 ±0.06	1.5 ±0.3
		Final slope (pH 5.3)	0.08 ±0.02	19 ±13

It is shown that the acid environment is the condition of the penetration of the virus into cell. We have proved the specificity of the interaction of LBP and protein E at both acid and normal pH conditions. At the same time, we have seen the change of the conformation of LBP at pH=5.3, as the force curves show "flat" regions with almost constant force instead of peaks with gradually increasing force for pH=7.4.

Some papers have already reported the phenomenon of force-induced globule-coil transition of proteins [64, 74, 27, 100, 66]. The explanation of the plateaus on the force curves is the following. If one stretches a globule with some weak force in conditions of poor solvability, it can undergo conformational changes and deform first to an elliptical form and then to a "ball - string" configuration. Such a transition can be compared to a Rayleigh instability of a column of liquid, which undergoes a sharp transition into a series of droplets [104]. This complex phenomenon was already observed for polymers, for instance for polystyrene chains in toluene and water [60].

Once a polymer molecule in good solvent is pulled, its force-extension curve shows a typical smooth shape with monotonical increase of force (fig. 5.13). Such curves are in particular typical for pulling of DNA molecules in water or a buffer solution [122, 134, 26]. This curve can be fitted well with a worm-like chain (WLC) model and permit one to extract information about the persistence length of the molecules (see section 1.5) [17]. The dependence of the force on the extension is well approximated by the following equation:

$$F = \frac{k_B T}{L_p} \left[\frac{1}{4(1 - z/L_0)^2} - \frac{1}{4} + \frac{z}{L_0} \right], \quad (5.6)$$

where k_B is Boltzmann constant, T - temperature, L_p - persistence length, L_0 - total length of

the molecule, z - extension (end-to-end distance) and F - resulting force [17].

The retract curve changes dramatically once the polymer is collapsed in poor solvent (fig. 5.13 (b)). A key region in this force curve is a flat plateau with constant force. The theory of such process is well described by Halperin and Zhulina [64]. If one starts to pull a polymer chain collapsed in a bad solvent, there will be three consecutive regimes of chain extension (fig. 5.13 (c)). First, collapsed molecule is represented by close-packed blobs of size ξ_c ($\xi_c \sim g_c^{1/2} a$, where a is the size of the monomer and $g_c \sim (\theta/\Delta T)^2$ is the number of the blobs, θ being θ -temperature and $\Delta T = \theta - T$) a sphere with radius r_c . Its shape changes upon pulling to a prolate ellipsoid with the axis of rotation of length x and two perpendicular axes of equal length d , the volume of it being constant (see fig. 5.14 a)). For weak deformations ($x \sim r_c \sim d$) the surface area S scales as $\Delta S \simeq (x - r_c)^2$. This leads to the rise of free energy equal to $\Delta A = \gamma \Delta S$, where γ represents the effective surface tension of the interface between the globule and the solvent, $\gamma \sim k_B T / \xi_c^2$. Thus the resulting force depend linearly on x :

$$f = -\frac{dA}{dx} \sim -\gamma(x - r_c). \quad (5.7)$$

For larger extensions the ellipsoid should rather be approximated as a cylinder with the length x much greater than the diameter of the base d (see fig. 5.14 b)). The lateral surface area of this cylinder is proportional to xd and the base area to d^2 , so the latter can be neglected. The volume V still remains constant and proportional to xd^2 , so $d \sim (V/x)^{1/2}$. The resulting force now is calculated as the following:

$$f = -\frac{dA}{dx} = \gamma \frac{d}{dx}(xd) \sim \gamma \frac{d(V^{1/2} x^{1/2})}{dx} \sim -\gamma V^{1/2} x^{-1/2}. \quad (5.8)$$

This region is not stable and correspond to a first-order phase transition. A model proposed for this region prescribes the coexistence of a weakly deformed globule and a stretched string of blobs of size ξ_c rather than an elongated cylinder (see fig. 5.14 c)). The system is in neutral equilibrium, so $f \sim x^0 = \text{const}$.

When the chain is elongated so that the aforementioned phase transition process is over (*i.e.* there is no more weakly deformed globule), further elongation leads to the decrease of the size of the blobs. Now ξ_c is replaced by ξ_p , which depend on the applied force: $\xi_p \sim k_B T / f$. The number of the monomers in each blob changes to g_p : $\xi_p \sim g_p^{1/2} a$, so $g_p \sim (k_B T / f a)^2$ (see fig. 5.14 d)). Finally we have the dependence

$$x \sim (N/g_p) \xi_p \sim f N a^2 / k_B T, \quad (5.9)$$

so $f \sim x$. More complex theory gives the dependence of force on loading rate.

Such characteristic force plateaus for polymers in poor solvent were observed both in experiments [60] and in computer simulations [126] and were predicted theoretically before [63]. A folded protein is often modeled as a collapsed polymer chain [36, 29]. Its chain is heterogeneous, but the strands of it interact which permit the protein to keep it compact energetically favorable shape. It is possible to imagine a slightly unfolded protein that, being pulled, behave similarly to a polymer in bad solvent nag give characteristic constant force plateaus.

Our investigation cannot propose a complete picture of globule-coil transition, which depends on the structure of the protein complex and pH level. The solubility of the protein, which plays an important role, should be studied deeper using physical-chemical approaches. Yet it clearly observes such a phenomenon involving LBP protein at pH=5.3 and relatively low force of about 50-70 pN or even less. Such values of the force seem to be achievable by different phenomena. For example, it is proven that fusion protein can form so-called jackknife structure [120]. Even surface tension effects of a particle of a radius of 1 nm reach the value of force of several hundreds piconewtons in water at room temperature.

LBP structure is well described by the means of X-ray studies [71]. It reveals a well-defined globule without disulfide bonds. Human LBP counts 295 amino acids, thus, taking into consideration the average amino acid length of 0.34 nm, the total elongated protein length can be estimated to be as large as 100 nm. In SMDFS we have measured the values of the elongation distance of 10-20 nm which is explainable using such estimation.

We propose the following model of viral-cell membrane fusion. Local changes of pH caused by small opening in the endosome membrane and a force exerted by it to LBP protein leads to an abrupt globule-coil transition. The initial opening, which could be caused by reorganization of fusion peptide, enlarges considerably in size by this explosion-like protein structure change and leads to the penetration of viral RNA into the cell cytoplasm.

5.4 Conclusions

Using our setup for SMDFS we have studied two important classes of interacting proteins: barnase-barstar pair and flaviviral membrane protein E with putative cell receptors. Both classes show efficient and highly specific interactions.

Our data for the former pair reports an existence of narrow and high (width and height of 0.12 nm and 0.8×10^{-19} J) internal barrier in the interaction energy landscape [115]. Estimating the second, "external" potential barrier caused by electrostatic interactions to be as great as 0.7×10^{-19} J, we approach closely to the value of total dissociation barrier measured before by

biochemical experiments of $1.34 - 1.53 \times 10^{-19}$ J. We explain the lack of the dependence of the interaction on the concentration by the long screening range, which is not measurable by the typical SMDFS experiment. Our data also shows the strength of this technique to separate short range interactions from the long range ones.

The latter set of experiments helps to understand the mechanism of viral-cell membrane fusion. We have shown specific interaction of putative cell receptors with WNV and TBEV protein E, as well as with its corresponding recombinant fragments. We have studied as well the conformational changes of LBP due to the change of pH and the existence of force. A model of membrane fusion based on this study was proposed by us.

The fusion of two membranes is thermodynamically favorable, but the energy landscape of such a process present a huge kinetic barrier. Our model show that LBP is able to bind viral particles with some degree of flexibility and create local membrane defects, which decrease the height of this kinetic barrier and permit viral RNA to penetrate into the cell.

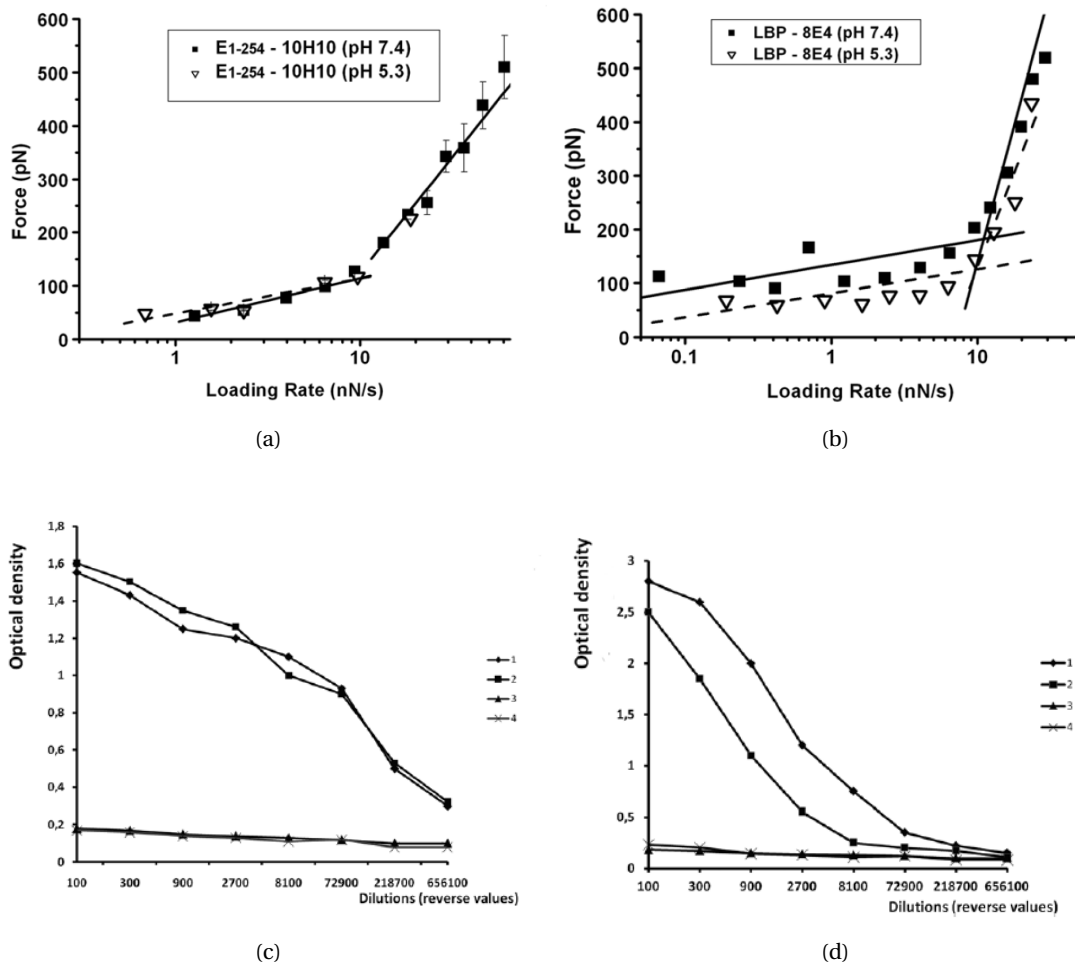


Figure 5.9 – (a),(b): Loading rate - force dependence curve for interaction of TBEV protein E and LBP with corresponding anticeptors 8E4 and 10H10. (c),(d): Corresponding immunoenzyme analysis data. Left: 1, 2 – interaction of 10H10 mAb with recombinant TBEV protein E₁₋₂₅₄; 3, 4 – interaction of 10H10 mAb with *E. coli* lysate transformed by empty PQE plasmid, (negative control). Right: 1, 2 – interaction of 8E4 mAb with recombinant protein LBP; 3, 4 – interaction of 8E4 mAb with *E. coli* lysate transformed by empty PQE plasmid (negative control). Data series 1, 3: pH 7.4; data series 2, 4: pH 5.3.

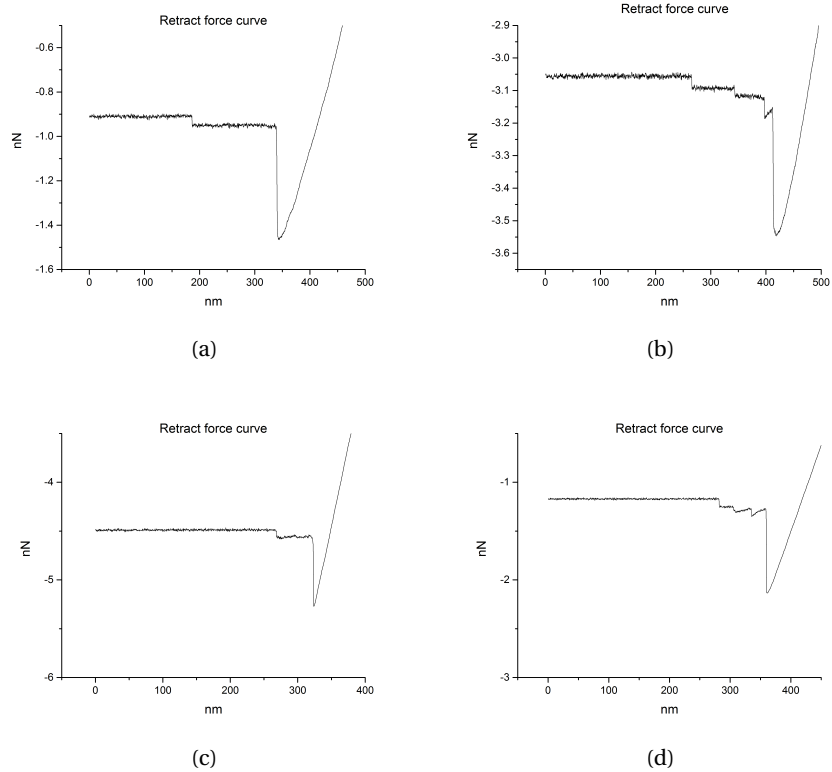


Figure 5.10 – Examples of force curves of protein-protein interaction exhibiting plateaus. (a),(b): WNV protein E₁₋₁₈₀ fragment and LBP at pH=5.3 showing one (a) or several (b) plateaus. (c),(d): WNV protein E₅₃₋₁₂₆ fragment with LBP at pH=5.4.

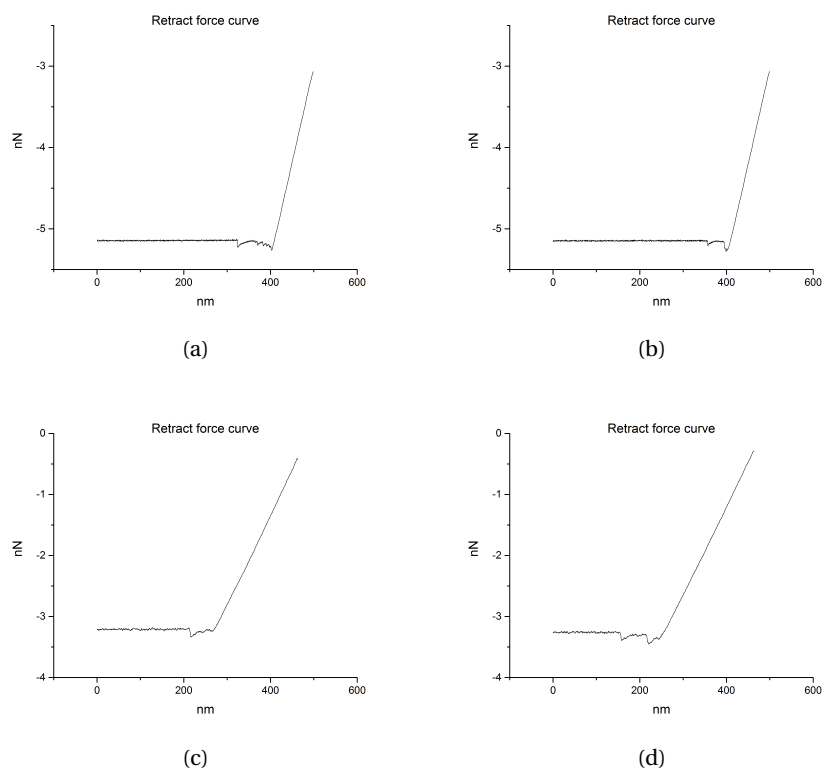


Figure 5.11 – Examples of force curves of protein-protein interaction (continued). (a), (b): WNV protein E₅₃₋₁₂₆ fragment with LBP at pH=7.4. No plateaus are observed in this case. (c),(d): protein E - integrin. One can observe that two latter force curves represent similar specific interaction events, but differ in the regions preceding them corresponding to the detachment of the tip from the surface and their non-specific interaction.

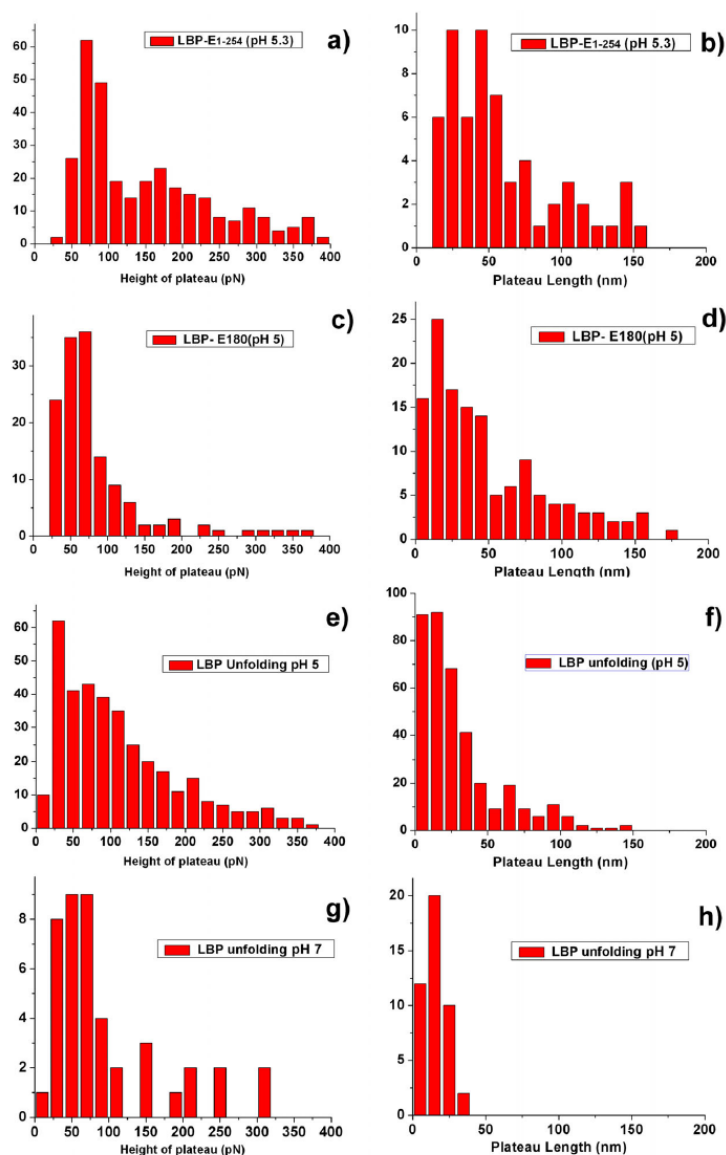


Figure 5.12 – Histograms of the distributions of characteristic heights (force values; a, c, e, g) and lengths (distances, b, d, f, h) of the plateaus on the force-distance extensions curves observed in SMDFS experiments with the protein pairs LBP-TBEV E_{1-254} , LBP-WNV E_{1-180} at pH 5.3 and pure LBP at pH 5.3 and pH 7.4.

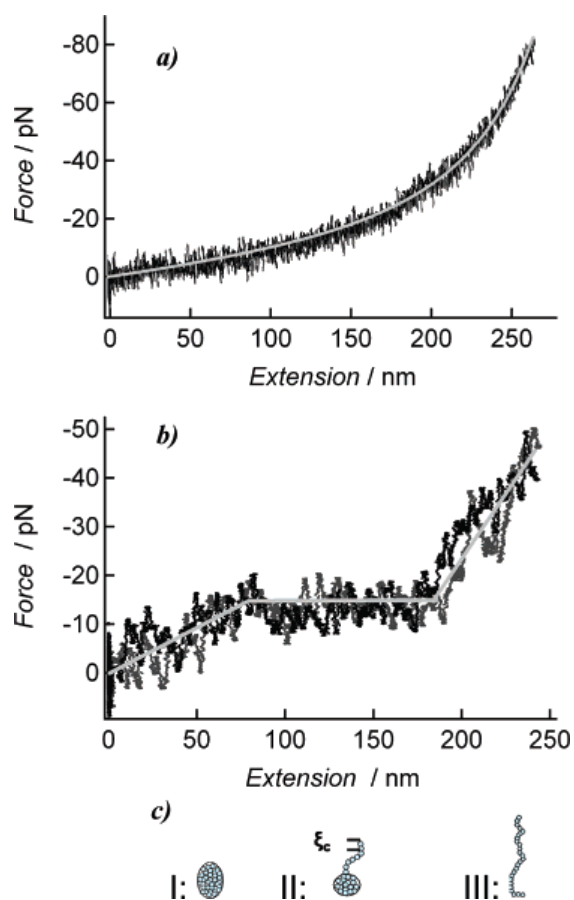


Figure 5.13 – Extension force curve of a polymer. (a) Good solvent case (polystyrene chain in toluene), gray curve: WLC fit. (b) Collapsed, poor solvent case (polystyrene chain in water). (c) Drawing illustrating the stages of the collapsed chain extension. One can see on a force curve (b) a plateau corresponding to the state of the molecule presented on (c)II . Image taken from [60].

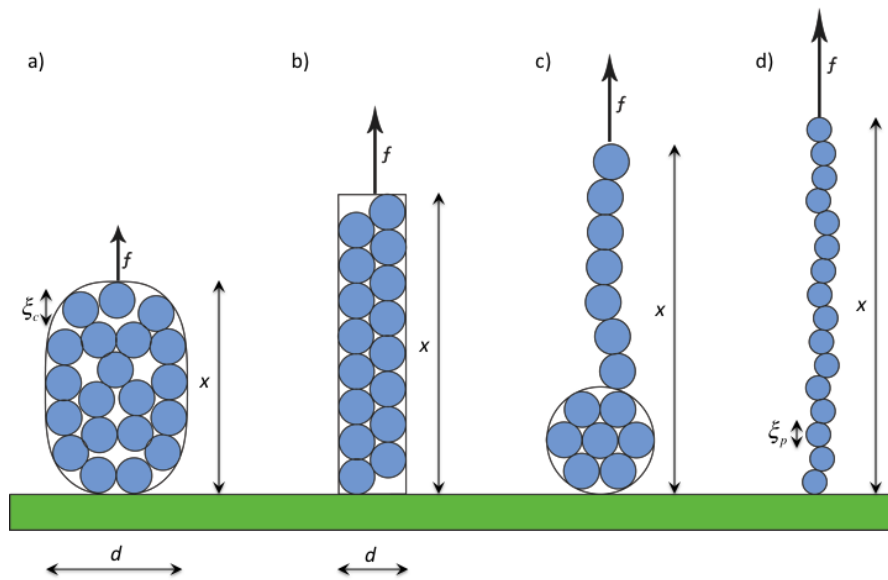


Figure 5.14 – Schematic image representing different regimes of force-extension dependence upon pulling of a collapsed polymer in bad solvent. a): small extension, globule can be presented as an ellipsoid; force is proportional to extension. b) theoretical intermediate extension regime when the polymer can be modeled as a cylinder; does not realize in practice as force decreases with extension. c) globule-coil transition regime with constant force. d) large extension regime; force is again proportional to elongation.

A Used acronyms

- AFM - Atomic Force Microscopy
- APTES - (3-Aminopropyl)triethoxysilane
- CCD - Charge-Coupled Device
- CMOS - Complimentary Metal-Oxide Semiconductor
- CW - Continious Working (laser)
- EA - Ethyl Acetate
- EM - Electron Microscope
- FIB - Focused Ion Beam
- FJC - Freely-Jointed Chain
- FRET - Fluorescent Resonance Energy Transfer
- LBP - Laminin Binding Protein
- MDA - minimum detectable angle
- WLC - Worm-Like Chain
- WNV - West Nile Virus
- PALM - Photo Activated Localization Microscopy
- RMSD - Root-mean-square deviation. If there is a set of N values $\{x_i\}$, $i = 1..N$ with the mean $\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_i$, then $RMSD(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle)^2}$.
- SMDFS - Single Molecule Dynamic Force Spectroscopy
- SNOM - Scanning Near-Field Optical Microscopy

Appendix A. Used acronyms

- SPM - Scanning Probe Microscopy
- STED - Stimulated Emission Depletion microscopy
- STORM - Stochastic Optical Reconstruction Microscopy
- TBEV - Tick Borne Encephalitis Virus
- UHQ - Ultra High Quality water, this term represents bi-distilled deionized water with resistivity about $20 \text{ M}\Omega \cdot \text{sm}$
- $\langle R_{ee}^2(s) \rangle$ - Average square of end-to-end distance
- Correlation function - $\langle \cos\theta(s) \rangle$, mean cosine of angle between tangents to the polymer at two points separated by the contour length s as a function of s

A Listings of the code

A.1 Smart Fit

Listing of the part *DNA Trace*, responsible for the aforementioned *Smart Fit* algorithm:

```
private static final float k = 6f;
/**
 * Laplacian matrix permit us to trace the borders of objects
 */
private final static float[][] matrix = {
    {-k/12, -k/6, -k/12 },
    { -k/6, k, -k/6},
    {-k/12, -k/6, -k/12}
};
/**
 * We use weighted linear least squares method to fit our data line by
 * line with polynomial;
 * weights are 1 for background or 0 for sample (actually, not zero but
 * some small value
 * to avoid singularity).
 * In other words, for every row i with width=w we are looking for a
 * function
 *  $y = a_0 + a_1 * x + a_2 * x^2 + \dots + a_n * x^n$ ,
 * where  $x = j$  (varies from 0 to w-1) and  $y$  is the value of the pixel
 * (j,i).
 * The task is linear. We build matrix X (w,n+1), diagonal matrix W(w,w)
 * and vector Y(w):
 *
 *
 *      |  $0^0 = 1$        $0^1 = 0$       ..  $0^n = 0$       |
 *      |  $1^0 = 1$        $1^1 = 1$       ..  $1^n$       |
 * X = | ..            ..            ..      ..      |
 *      |  $(j-1)^0 = 1$   $(j-1)^1 = j-1$  ..  $(j-1)^n$  |
 *      | ..            ..            ..      ..      |
```

Appendix A. Listings of the code

```

*      | (w-1)^0 = 1 (w-1)^1 = w-1 .. (w-1)^n      |,
*
*      |  $\alpha(0)$  0      .. 0      .. 0      |
*      | 0       $\alpha(1)$  .. 0      .. 0      |
*  $\bar{W}$  = | ..      ..      ..      ..      ..      |
*      | 0      0      ..  $\alpha(j)$  .. 0      |
*      | ..      ..      ..      ..      ..      |
*      | 0      0      .. 0      ..  $\alpha(w-1)$  |,
*
* where  $\alpha(j) = \{ 1, \text{ if } (i,j) \in \text{background},$ 
*                 { small  $\epsilon$ , if  $(i,j) \notin \text{background}.$ 
*
* Y - pixels in the row i.
* The goal is to find  $\beta = (a_0, a_1, \dots, a_n).$ 
*
* To do so, we solve
*  $(X^T * W * X) \beta = X^T * W * Y,$ 
* where  $X^T$  means "X transposed".
*
* More complete description:
* http://en.wikipedia.org/wiki/Linear\_least\_squares\_\(mathematics\)
*/
@Override
public float[][] process(final float[][] array,
    final boolean[][] selection, final TaskProgressWatcher watcher)
    throws InterruptedException {
    final long time = System.currentTimeMillis();
    final TaskProgressMeasure measure = new TaskProgressMeasure(1);
    if (watcher != null)
        watcher.addTaskProgressMeasure(measure);
    final boolean[][] boolArray;
    final int cl = getOrder() + 1; //number of coefficients = order + 1
    final int w = array[0].length; //width
    final int h = array.length; //height
    if (selection == null) { //If the user has not selected the
        background, find it automatically
        final Array2DProcessor laplasianProcessor = new
            MatrixImageProcessor(
                matrix);
        final float[][] laplasArray = laplasianProcessor.process(array,
            selection, watcher);
        final double rms = sdev(laplasArray);

        final FloodFiller floodFiller = new FloodFiller(new Map2D(
            laplasArray));
        final FloodFiller.FloatFunction floatFunction = new
            FloodFiller.FloatFunction() {

```

```

        @Override
        public boolean isFulfilled(final float value) {
            return Math.abs(value) < rms;
        }
    };
    // TODO Here we have to choose first point to fill
    floodFiller.floodFill(new Point(0, 0), floatFunction);
    floodFiller.floodFill(new Point(w - 1, h - 1), floatFunction);
    floodFiller.floodFill(new Point(w - 1, 0), floatFunction);
    floodFiller.floodFill(new Point(0, h - 1), floatFunction);
    //
    boolArray = floodFiller.getDetectedArray();
} else
    boolArray = selection;

final Matrix x = new DenseMatrix(w, cl);
final DenseMatrix values = new DenseMatrix(w, 1);
final DenseMatrix xt_w_x = new DenseMatrix(cl, cl);
final DenseMatrix xt_w_y = new DenseMatrix(cl, 1);
int line=0;
for (final float[] row : array) {
    for (int i = 0; i < w; i++)
        for (int k1 = 0; k1 < cl; k1++)
            x.set(i, k1, Math.pow(i, k1));
    final Matrix xt_w = new DenseMatrix(cl, w);
    x.transpose(xt_w);
    for (int i = 0; i < w; i++) {
        //If the pixel does not belong to the background,
        //set its weight to a very low value
        if (!boolArray[line][i])
            for (int j=0; j<cl; j++)
                xt_w.set(j, i, xt_w.get(j, i) * 1f/w);
        if (Float.isNaN(row[i])) {
            values.set(i,0, 0);
            for (int j=0; j<cl; j++)
                xt_w.set(j, i, 0);
        }
        else
            values.set(i, 0, row[i]);
    }
}

xt_w.mult(x, xt_w_x);
xt_w.mult(values, xt_w_y);
final DenseCholesky dc = DenseCholesky.factorize(xt_w_x);
DenseMatrix res; //It is a matrix of coefficients with one row

```

Appendix A. Listings of the code

```
try {
    try {
        res = dc.solve(xt_w_y); //try to solve equation
        // (XT*W*X) β= XT*W*Y, find β
    } catch (final MatrixNotSPDException ex) {
        final DenseLU lu = DenseLU.factorize(xt_w_x);
        res = lu.solve(xt_w_y);
    }
    final PolynomialFunction function = new PolynomialFunction(cl -
1);
    function.setCoefficients(res.getData());
    for (int i = 0; i < w; i++) {
        final double r = function.getY(i);
        row[i] -= r;
    }
}
catch (final MatrixSingularException ex) {
    MainWindow.getLogger().log(Level.FINE, "Singular matrix");
}
measure.setProgress(line * 1.0 / h);
if (Thread.currentThread().isInterrupted())
    throw new InterruptedException();
line++;
}
measure.setProgress(1);
System.out.println("Fitted in:" + (System.currentTimeMillis()-time));
return array;
}
```

A.2 Flood Fill

Flood Fill implementation used in *DNA Trace*:

```
/**
 * Performs Flood Fill algorithm on the {@code
 * data}, i.e. fills {@code detectedArray} with {@code true}
 * for the points that are "filled". {@code detectedArray} is stored
 * inside the class, so subsequent calls to floodFill skip previously
 * detected points.
 * @param startPoint a point to start
 * @param function {@code FloatFunction} used for Flood Fill:
 * if {@code function.isFulfilled(value)}, pixel is filled.
 * @return {@code true} is any pixel was affected,
 * {@code false} otherwise.
 */
public boolean floodFill(final Point startPoint,
```

```

        final FloatFunction function) {
    if (!function.isFulfilled(data[startPoint.y][startPoint.x])
        || detected[startPoint.y][startPoint.x])
        return false;
    /*
     * We do not want to use recursion, queue implementation
     * is more memory efficient.
     */
    final Queue<Point> queue = Queues.newArrayDeque();
    queue.add(startPoint);
    boolean result = false;
    while (!queue.isEmpty()) {
        final Point p = queue.poll(); //Take a point to process
        if (detected[p.y][p.x])
            continue;
        if (function.isFulfilled(data[p.y][p.x])) {
            detected[p.y][p.x] = true;
            result = true;
            int x = p.x + 1;
            final int y = p.y;
            /*
             * Go right from this point till the border of the area
             */
            while (x < width && function.isFulfilled(data[y][x])) {
                detected[y][x] = true;
                if (y < height - 1 && function.isFulfilled(data[y + 1][x]))
                    queue.add(new Point(x, y + 1)); //check the point to
                    the top
                if (y > 0 && function.isFulfilled(data[y - 1][x]))
                    queue.add(new Point(x, y - 1)); //point to the bottom
                x++;
            }
            /*
             * Go left
             */
            x = p.x - 1;
            while (x >= 0 && function.isFulfilled(data[y][x])) {
                detected[y][x] = true;
                if (y < height - 1 && function.isFulfilled(data[y + 1][x]))
                    queue.add(new Point(x, y + 1)); //point to the top
                if (y > 0 && function.isFulfilled(data[y - 1][x]))
                    queue.add(new Point(x, y - 1)); //point to the bottom
                x--;
            }
        }
    }
    return result;
}

```

```
}
```

A.3 Put path

Listing of the implementation of Dijkstra Algorithm used to trace the molecule between two consecutive user clicks in *DNA Trace*:

```
import static com.google.common.base.Preconditions.checkNotNull;
import geom.ImagePoint;

import java.util.Collection;
import java.util.Collections;
import java.util.List;
import java.util.PriorityQueue;

import message.Message;
import message.MessageManager;
import preference.SetPreference;

import com.google.common.collect.Lists;
import com.google.common.collect.Queues;

/**
 * Minimizes the total cost of the path
 * @author mikhaylo
 *
 */
public enum PixelGraphSolver {
    ;
    private static class Vertex implements Comparable<Vertex>{
        final int x;
        final int y;
        double cost = Double.POSITIVE_INFINITY;
        Vertex prev = null;

        Vertex(final int x, final int y) {
            this.x = x;
            this.y = y;
        }

        @Override
        public int compareTo(final Vertex other) {
            return Double.compare(cost, other.cost);
        }
    }
}
```

```

/**
 * Dijkstra algorithm
 * @param begin
 * @param end
 * @param map a map representing 2D pixel array
 * @return a solved path or null, if connection is not possible
 * @throws InterruptedException
 */
public static Collection<ImagePoint> solve(final ImagePoint begin, final
ImagePoint end,
    final Map2D map) throws InterruptedException {
    checkNotNull(map);
    checkNotNull(begin);
    checkNotNull(end);
    final long time = System.currentTimeMillis();
    final int ih = map.getHeight();
    final int iw = map.getWidth();
    /*
     * We use {@code PriorityQueue}, which permits one to obtain O(E+V
     logV) time complexity
     */
    final PriorityQueue<Vertex> queue = Queues.newPriorityQueue();
    final Vertex[][] vertices = new Vertex[ih][iw];
    for (int i=0; i<ih; i++)
        for (int j=0; j<iw; j++)
            vertices[i][j] = new Vertex(j,i);
    final Vertex vbegin = vertices[begin.getIntY()][begin.getIntX()];
    final Vertex vend = vertices[end.getIntY()][end.getIntX()];
    vbegin.cost = 0;
    queue.add(vbegin);
    final double shift = 2 * map.getAverage();
    final Message m = MessageManager.publishProgressMessage("Solving the
graph...", 2);
    while (true) {
        if (queue.isEmpty())
            return null; //No possible connection!
        final Vertex v = queue.poll(); //Take the vertex with the lowes
cost
        if (v == vend)
            break; //Solved - return
        for (int i=-1; i<=1; i++)
            for (int j=-1; j<=1; j++) {
                if (i == 0 && j == 0)
                    continue;
                final int x = j + v.x;
                final int y = i + v.y;
                if (x < 0 || x >= map.getWidth() || y < 0 || y >=

```

Appendix A. Listings of the code

```
        map.getHeight())
            continue;
        final Vertex v2 = vertices[y][x];
        if (v2 == v.prev || v2.cost <= v.cost)
            continue;
        final double tcost = Map2D.getEdgeWeight(shift -
            map.getPoint(v.x, v.y) -
            map.getPoint(v2.x, v2.y)) * Math.sqrt(i*i+j*j);
        final double newCost = v.cost + tcost;
        if (newCost < v2.cost) {
            queue.remove(v2);
            v2.cost = newCost;
            v2.prev = v;
            queue.add(v2);
        }
    }
    if (Thread.currentThread().isInterrupted())
        throw new InterruptedException(); //interrupted by user
}
MessageManager.remove(m);
final List<ImagePoint> returnList = Lists.newArrayList();
Vertex prev = vend;
while (prev != null) {
    returnList.add(new ImagePoint(prev.x+0.5, prev.y+0.5));
    prev = prev.prev;
}
Collections.reverse(returnList);
System.out.println("Solved in:" + (System.currentTimeMillis() -
time));
return returnList;
}
}
```

Bibliography

- [1] J. Adamcik, J.M. Jung, J. Flakowski, P. De Los Rios, G. Dietler, and R. Mezzenga. Understanding amyloid aggregation by statistical analysis of atomic force microscopy images. *Nature nanotechnology*, 5(6):423–428, 2010.
- [2] J. Adamcik, D.V. Klinov, G. Witz, S.K. Sekatskii, and G. Dietler. Observation of single-stranded DNA on mica and highly oriented pyrolytic graphite by atomic force microscopy. *FEBS letters*, 580(24):5671–5675, 2006.
- [3] Stephanie Allen, Xinyong Chen, John Davies, Martyn C Davies, Adrian C Dawkes, John C Edwards, Clive J Roberts, Joanna Sefton, Saul JB Tendler, and Philip M Williams. Detection of antigen-antibody binding events with the atomic force microscope. *Biochemistry*, 36(24):7457–7463, 1997.
- [4] Paulo FF Almeida, Winchil LC Vaz, and TE Thompson. Lateral diffusion and percolation in two-phase, two-component lipid bilayers. topology of the solid-phase domains in-plane and across the lipid bilayer. *Biochemistry*, 31(31):7198–7210, 1992.
- [5] Livan Alonso-Sarduy, Giovanni Longo, Giovanni Dietler, and Sandor Kasas. Time-lapse AFM imaging of DNA conformational changes induced by daunorubicin. *Nano letters*, 13(11):5679–5684, 2013.
- [6] Chunli Bai. *Scanning tunneling microscopy and its application*, volume 32. Springer, 2000.
- [7] Harry Bateman, Arthur Erdélyi, Hendrik van Haeringen, and LP Kok. *Tables of integral transforms*, volume 1. McGraw-Hill New York, 1954.
- [8] J. Bednar, P. Furrer, V. Katritch, A. Stasiak, J. Dubochet, and A. Stasiak. Determination of DNA persistence length by cryo-electron microscopy. separation of the static and dynamic contributions to the apparent persistence length of DNA. *Journal of molecular biology*, 254(4):579–594, 1995.
- [9] George I Bell. Models for the specific adhesion of cells to cells. *Science*, 200(4342):618–627, 1978.

Bibliography

- [10] Fabrizio Benedetti, Cristian Micheletti, Giovanni Bussi, SK Sekatskii, and G Dietler. Nonkinetic modeling of the mechanical unfolding of multimodular proteins: Theory and experiments. *Biophysical journal*, 101(6):1504–1512, 2011.
- [11] G.K. Binnig. Atomic force microscope and method for imaging surfaces with atomic resolution, February 9 1988. US Patent 4,724,318.
- [12] GH Bergold and J Weibel. Demonstration of yellow fever virus with the electron microscope. *Virology*, 17(4):554–562, 1962.
- [13] M. Bezanilla, S. Manne, D.E. Laney, Y.L. Lyubchenko, and H.G. Hansma. Adsorption of DNA to mica, silylated mica, and minerals: characterization by atomic force microscopy. *Langmuir*, 11(2):655–659, 1995.
- [14] Joshua Bloch. *Effective java (the java series)*. Prentice Hall PTR, 2008.
- [15] Maria V Bogachek, Elena V Protopopova, Valery B Loktev, Boris N Zaitsev, Melanie Favre, Sergey K Sekatskii, and Giovanni Dietler. Immunochemical and single molecule force spectroscopy studies of specific interaction between the laminin binding protein and the West Nile virus surface glycoprotein E domain II. *Journal of Molecular Recognition*, 21(1):55–62, 2008.
- [16] MV Bogachek, BN Zaitsev, SK Sekatskii, EV Protopopova, VA Ternovoi, AV Ivanova, AV Kachko, VA Ivanisenko, G Dietler, and VB Loktev. Characterization of glycoprotein E c-end of West Nile virus and evaluation of its interaction force with $\alpha\beta 3$ integrin as putative cellular receptor. *Biochemistry (Moscow)*, 75(4):472–480, 2010.
- [17] C Bouchiat, MD Wang, J-F Allemand, T Strick, SM Block, and V Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical journal*, 76(1):409–413, 1999.
- [18] Carlos Bustamante, James Vesenka, Chun Lin Tang, William Rees, Martin Guthold, and Rebecca Keller. Circular DNA molecules imaged in air by scanning force microscopy. *Biochemistry*, 31(1):22–26, 1992.
- [19] Annalisa Calo, Pablo Stoliar, Eva Bystrenova, Francesco Valle, and Fabio Biscarini. Measurement of DNA Morphological Parameters at Highly Entangled Regime on Surfaces. *Journal of Physical Chemistry B*, 113(15):4987–4990, 2009.
- [20] Mariano Carrion-Vazquez, Andres F Oberhauser, Susan B Fowler, Piotr E Marszalek, Sheldon E Broedel, Jane Clarke, and Julio M Fernandez. Mechanical and chemical unfolding of a single protein: a comparison. *Proceedings of the National Academy of Sciences*, 96(7):3694–3699, 1999.
- [21] Andrea Cerreta, Dusan Vobornik, Giovanni Di Santo, Susana Tobenas, Livan Alonso-Sarduy, Jozef Adamcik, and Giovanni Dietler. FM-AFM constant height imaging and force curves: high resolution study of DNA-tip interactions. *Journal of Molecular Recognition*, 25(9):486–493, 2012.

- [22] Andrea Cerreta, Dusan Vobornik, and Giovanni Dietler. Fine DNA structure revealed by constant height frequency modulation AFM imaging. *European Polymer Journal*, 49(8):1916–1922, 2013.
- [23] Leonid V Chernomordik and Michael M Kozlov. Mechanics of membrane fusion. *Nature structural & molecular biology*, 15(7):675–683, 2008.
- [24] Lilia A Chtcheglova, George T Shubeita, Sergey K Sekatskii, and Giovanni Dietler. Force spectroscopy with a small dithering of AFM tip: a method of direct and continuous measurement of the spring constant of single molecules and molecular complexes. *Biophysical journal*, 86(2):1177–1184, 2004.
- [25] Pietro Cicuta, Sarah L Keller, and Sarah L Veatch. Diffusion of liquid domains in lipid bilayer membranes. *The Journal of Physical Chemistry B*, 111(13):3328–3331, 2007.
- [26] Hauke Clausen-Schaumann, Markus Seitz, Rupert Krautbauer, and Hermann E Gaub. Force spectroscopy with single bio-molecules. *Current opinion in chemical biology*, 4(5):524–530, 2000.
- [27] IR Cooke and DRM Williams. Stretching polymers in poor and bad solvents: Pullout peaks and an unraveling transition. *EPL (Europhysics Letters)*, 64(2):267, 2003.
- [28] P.G. De Gennes. *Scaling concepts in polymer physics*. Cornell university press, 1979.
- [29] PG De Gennes. Kinetics of collapse for a flexible coil. *Journal de Physique Lettres*, 46(14):639–642, 1985.
- [30] Marzena de Odroważ Piramowicz, Paweł Czuba, Marta Targosz, Květoslava Burda, and Marek Szymoński. Dynamic force measurements of avidin–biotin and streptavidin–biotin interactions using AFM. *Acta Biochimica Polonica*, 53(1):93–100, 2006.
- [31] FL Degertekin, AG Onaran, M Balantekin, W Lee, NA Hall, and CF Quate. Sensor for direct measurement of interaction forces in probe microscopy. *Applied Physics Letters*, 87(21):213109, 2005.
- [32] Edsger W Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.
- [33] A.V. Dobrynin. Electrostatic persistence length of semiflexible and flexible polyelectrolytes. *Macromolecules*, 38(22):9304–9314, 2005.
- [34] Feng Dong, M Vijayakumar, and Huan-Xiang Zhou. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophysical journal*, 85(1):49–60, 2003.
- [35] M. Dong, M.B. Hovgaard, S. Xu, D.E. Otzen, and F Besenbacher. AFM study of glucagon fibrillation via oligomeric structures resulting in interwoven fibrils. *Nanotechnology*, 17(16):4003, 2006.

Bibliography

- [36] S Doniach, T Garel, and H Orland. Phase diagram of a semiflexible polymer chain in a θ solvent: Application to protein folding. *The Journal of chemical physics*, 105(4):1601–1608, 1996.
- [37] William A Ducker, Robert F Cook, and David R Clarke. Force measurement using an AC atomic force microscope. *Journal of applied physics*, 67(9):4045–4052, 1990.
- [38] Andreas Ebner, Linda Wildling, Rong Zhu, Christian Rankl, Thomas Haselgrübler, Peter Hinterdorfer, and Hermann J Gruber. Functionalization of probe tips and supports for single-molecule recognition force microscopy. In *STM and AFM Studies on (Bio) Molecular Systems: Unravelling the Nanoworld*, pages 29–76. Springer, 2008.
- [39] E. Ercolini, F. Valle, J. Adamcik, G. Witz, R. Metzler, P. De Los Rios, J. Roca, and G. Dietler. Fractal dimension and localization of DNA knots. *Physical review letters*, 98(5):58102, 2007.
- [40] Evan Evans and Ken Ritchie. Dynamic strength of molecular adhesion bonds. *Biophysical journal*, 72(4):1541–1555, 1997.
- [41] M Favre, LA Chtcheglova, DA Lapshin, SK Sekatskii, F Valle, and G Dietler. Force-clamp spectroscopy with a small dithering of AFM tip, and its application to explore the energy landscape of single avidin–biotin complex. *Ultramicroscopy*, 107(10):882–886, 2007.
- [42] E. Ficarra, L. Benini, E. Macii, and G. Zuccheri. Automated DNA fragments recognition and sizing through AFM image processing. *Information Technology in Biomedicine, IEEE Transactions on*, 9(4):508–517, 2005.
- [43] Darel Rex Finley. Efficient polygon fill algorithm with c code sample, 2007.
- [44] E-L Florin, Matthias Rief, H Lehmann, Markus Ludwig, C Dornmair, Vincent T Moy, and Hermann E Gaub. Sensing specific molecular interactions with the atomic force microscope. *Biosensors and Bioelectronics*, 10(9):895–901, 1995.
- [45] Ernst-Ludwig Florin, Vincent T Moy, and Hermann E Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264(5157):415–417, 1994.
- [46] P.J. Flory. *Statistical Mechanics of Chain Molecules*. Hanser Gardner Publications, 1989.
- [47] Michael L Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM (JACM)*, 34(3):596–615, 1987.
- [48] Claudia Friedsam, Angelika K Wehle, Ferdinand Kühner, and Hermann E Gaub. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *Journal of Physics: Condensed Matter*, 15(18):S1709, 2003.

- [49] Richard Fritz, Janja Blazevic, Christian Taucher, Karen Pangerl, Franz X Heinz, and Karin Stiasny. The unique transmembrane hairpin of flavivirus fusion protein E is essential for membrane fusion. *Journal of virology*, 85(9):4377–4385, 2011.
- [50] Richard Fritz, Karin Stiasny, and Franz X Heinz. Identification of specific histidines as pH sensors in flavivirus membrane fusion. *The Journal of cell biology*, 183(2):353–361, 2008.
- [51] Sergi Garcia-Manyes, Jasna Brujić, Carmen L Badilla, and Julio M Fernández. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophysical journal*, 93(7):2436–2446, 2007.
- [52] A Garcia-Valenzuela. Limits of different detection schemes used in the optical beam deflection method. *Journal of applied physics*, 82(3):985–988, 1997.
- [53] C. Goldsbury, J. Kistler, U. Aebi, T. Arvinte, and G.J.S. Cooper. Watching amyloid fibrils grow by time-lapse atomic force microscopy. *Journal of molecular biology*, 285(1):33–39, 1999.
- [54] SW Gollins and JS Porterfield. Flavivirus infection enhancement in macrophages: an electron microscopic study of viral cellular entry. *Journal of General Virology*, 66(9):1969–1982, 1985.
- [55] Rudolf Gorenflo and Francesco Mainardi. Random walk models for space-fractional diffusion processes. *Fractional Calculus and Applied Analysis*, 1(2):167–191, 1998.
- [56] IS Gradstein and IM Ryshik. Tables of series, products, and integrals. *Verlag MIR, Moskau*, 1981.
- [57] N Michael Green. Avidin and streptavidin. *Methods in enzymology*, 184:51, 1990.
- [58] Leo Gross, Fabian Mohn, Nikolaj Moll, Peter Liljeroth, and Gerhard Meyer. The chemical structure of a molecule resolved by atomic force microscopy. *Science*, 325(5944):1110–1114, 2009.
- [59] Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.
- [60] Nikhil Gunari, Anna C Balazs, and Gilbert C Walker. Force-induced globule-coil transition in single polystyrene chains in water. *Journal of the American Chemical Society*, 129(33):10046–10047, 2007.
- [61] Petra Gutjahr, Reinhard Lipowsky, and Jan Kierfeld. Persistence length of semiflexible polymers and bending rigidity renormalization. *EPL (Europhysics Letters)*, 76(6):994, 2006.

Bibliography

- [62] Neal A Hall and F Levent Degertekin. Integrated optical interferometric detection method for micromachined capacitive acoustic transducers. *Applied Physics Letters*, 80(20):3859–3861, 2002.
- [63] A Halperin and EB Zhulina. On the deformation behaviour of collapsed polymers. *EPL (Europhysics Letters)*, 15(4):417, 1991.
- [64] A Halperin and EB Zhulina. Stretching polymer brushes in poor solvents. *Macromolecules*, 24(19):5393–5397, 1991.
- [65] Stephen C Harrison. Viral membrane fusion. *Nature structural & molecular biology*, 15(7):690–698, 2008.
- [66] BJ Haupt, TJ Senden, and EM Sevick. AFM evidence of Rayleigh instability in single polymer chains. *Langmuir*, 18(6):2174–2182, 2002.
- [67] Julian Havil and J Gamma. Exploring euler’s constant, 2003.
- [68] Peter Hinterdorfer, Werner Baumgartner, Hermann J Gruber, Kurt Schilcher, and Hansgeorg Schindler. Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proceedings of the National Academy of Sciences*, 93(8):3477–3481, 1996.
- [69] Osawa Hisao. Cantilever and scanning force microscope, 10 1994.
- [70] Sergei Izrailev, Sergey Stepaniants, Manel Balsera, Yoshi Oono, and Klaus Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical journal*, 72(4):1568–1581, 1997.
- [71] Kelly V Jamieson, Jinhua Wu, Stevan R Hubbard, and Daniel Meruelo. Crystal structure of the human laminin receptor precursor. *Journal of Biological Chemistry*, 283(6):3002–3005, 2008.
- [72] Sandor Kasas, Beat M Riederer, Stefan Catsicas, Brunero Cappella, and Giovanni Dietler. Fuzzy logic algorithm to extract specific interaction forces from atomic force microscopy data. *Review of scientific instruments*, 71(5):2082–2086, 2000.
- [73] F. Kienberger, VP Pastushenko, G. Kada, T. Puntheeranurak, L. Chtcheglova, C. Riethmueller, C. Rankl, A. Ebner, and P. Hinterdorfer. Improving the contrast of topographical AFM images by a simple averaging filter. *Ultramicroscopy*, 106(8):822–828, 2006.
- [74] Stefan Kreitmeier, Markus Wittkop, and Dietmar Göritz. Computer simulations on the cyclic deformation of a single polymer chain above and below the θ temperature. *Physical Review E*, 59(2):1982, 1999.
- [75] A. Kumar, HL Chan, JJ Weimer, and L. Sanderson. Structural characterization of pulsed laser-deposited AlN thin films on semiconductor substrates. *Thin Solid Films*, 308:406–409, 1997.

- [76] Chih-Kung Lee, Yu-Ming Wang, Long-Sun Huang, and Shiming Lin. Atomic force microscopy: determination of unbinding force, off rate and energy barrier for protein–ligand interaction. *Micron*, 38(5):446–461, 2007.
- [77] Gil U Lee, David A Kidwell, and Richard J Colton. Sensing discrete streptavidin-biotin interactions with atomic force microscopy. *Langmuir*, 10(2):354–357, 1994.
- [78] Lee-Peng Lee and Bruce Tidor. Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. *Protein Science*, 10(2):362–377, 2001.
- [79] Maofu Liao, Claudia Sánchez-San Martín, Aihua Zheng, and Margaret Kielian. In vitro reconstitution reveals key intermediate states of trimer formation by the dengue virus membrane fusion protein. *Journal of virology*, 84(11):5730–5740, 2010.
- [80] Yu-Shiu Lo, Ying-Jie Zhu, and Thomas P Beebe. Loading-rate dependence of individual ligand-receptor bond-rupture forces studied by atomic force microscopy. *Langmuir*, 17(12):3741–3748, 2001.
- [81] Jannis Lübbe, Matthias Temmen, Philipp Rahe, Angelika Kühnle, and Michael Reichling. Determining cantilever stiffness from thermal noise. *Beilstein journal of nanotechnology*, 4(1):227–233, 2013.
- [82] Yuri Lyubchenko, Lyuda Shlyakhtenko, Rodney Harrington, Patrick Oden, and Stuart Lindsay. Atomic force microscopy of long DNA: imaging in air and under water. *Proceedings of the National Academy of Sciences*, 90(6):2137–2140, 1993.
- [83] Yuri L Lyubchenko and Luda S Shlyakhtenko. AFM for analysis of structure and dynamics of DNA and protein–DNA complexes. *Methods*, 47(3):206–213, 2009.
- [84] Jason M Mackenzie, Malcolm K Jones, and Paul R Young. Improved membrane preservation of flavivirus-infected cells with cryosectioning. *Journal of virological methods*, 56(1):67–75, 1996.
- [85] AC Maggs, DA Huse, and S Leibler. Unbinding transitions of semi-flexible polymers. *EPL (Europhysics Letters)*, 8(7):615, 1989.
- [86] Berenike Maier and Joachim O Rädler. Conformation and self-diffusion of single DNA molecules confined to two dimensions. *Physical review letters*, 82(9):1911, 1999.
- [87] A. Majumdar, J. Schäfer, P. Mishra, D. Ghose, J. Meichsner, and R. Hippler. Chemical composition and bond structure of carbon-nitride films deposited by CH_4/N_2 dielectric barrier discharge. *Surface and Coatings Technology*, 201(14):6437–6444, 2007.
- [88] AA Malygin, EI Bondarenko, VA Ivanisenko, EV Protopopova, GG Karpova, and VB Loktev. C-terminal fragment of human laminin-binding protein contains a receptor domain for venezuelan equine encephalitis and tick-borne encephalitis viruses. *Biochemistry (Moscow)*, 74(12):1328–1336, 2009.

Bibliography

- [89] J. Marek, E. Demjénová, Z. Tomori, J. Janáček, I. Zolotová, F. Valle, M. Favre, and G. Dietler. Interactive measurement and characterization of DNA molecules by analysis of AFM images. *Cytometry Part A*, 63(2):87–93, 2005.
- [90] Celestina Mariani, Veronique Gossele, Marc De Beuckeleer, Marc De Block, Robert B Goldberg, Willy De Greef, and Jan Leemans. A chimaeric ribonuclease-inhibitor gene restores fertility to male sterile plants. *Nature*, 357:384–387, 1992.
- [91] Sascha Martens and Harvey T McMahon. Mechanisms of membrane fusion: disparate players and common principles. *Nature Reviews Molecular Cell Biology*, 9(7):543–556, 2008.
- [92] Andrey Mikhaylov, Sergey K Sekatskii, and Giovanni Dietler. DNA trace: A comprehensive software for polymer image processing. *Journal of Advanced Microscopy Research*, 8(4):241–245, 2013.
- [93] J. Mou, D.M. Czajkowsky, Y. Zhang, and Z. Shao. High-resolution atomic-force microscopy of DNA: the pitch of the double helix. *FEBS letters*, 371(3):279–282, 1995.
- [94] Vincent T Moy, Ernst-Ludwig Florin, and Hermann E Gaub. Intermolecular forces and energies between ligands and receptors. *Science*, 266(5183):257–259, 1994.
- [95] Henning Mueller, Hans-Jürgen Butt, and Ernst Bamberg. Force measurements on myelin basic protein adsorbed to mica and lipid bilayer surfaces done with the atomic force microscope. *Biophysical journal*, 76(2):1072–1079, 1999.
- [96] Fuji Nagami, Giampaolo Zuccheri, Bruno Samorí, and Reiko Kuroda. Time-lapse imaging of conformational changes in supercoiled DNA by scanning force microscopy. *Analytical biochemistry*, 300(2):170–176, 2002.
- [97] ML Ng, J Howe, V Sreenivasan, and JJJ Mulders. Flavivirus West Nile (Sarafend) egress at the plasma membrane. *Archives of virology*, 137(3-4):303–313, 1994.
- [98] World Health Organization. Dengue and severe dengue fact sheet, March 2014.
- [99] World Health Organization. Yellow fever fact sheet, March 2014.
- [100] Alexey A Polotsky, Tatiana M Birshstein, Mohamed Daoud, and Oleg V Borisov. Conformations of amphiphilic polyelectrolyte stars with diblock copolymer arms. *Macromolecules*, 44(22):8999–9012, 2011.
- [101] Andrei D Polyanin. *Handbook of linear partial differential equations for engineers and scientists*. CRC press, 2010.
- [102] Elena V Protopopova, Aleksandr V Sorokin, Svetlana N Konovalova, Alla V Kachko, Sergey V Netesov, and Valery B Loktev. Human laminin binding protein as a cell receptor for the tick-borne encephalitis virus. *Zentralblatt für Bakteriologie*, 289(5):632–638, 1999.

- [103] EV Protopopova, SN Konavalova, and VB Loktev. [isolation of a cellular receptor for tick-borne encephalitis virus using anti-idiotypic antibodies]. *Voprosy virusologii*, 42(6):264–268, 1996.
- [104] Lord Rayleigh. On the stability, or instability, of certain fluid motions. *Proceedings of the London Mathematical Society*, 1(1):57–72, 1879.
- [105] Claudio Rivetti, Martin Guthold, and Carlos Bustamante. Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *Journal of molecular biology*, 264(5):919–932, 1996.
- [106] C. Roduit, B. Saha, L. Alonso-Sarduy, A. Volterra, G. Dietler, and S. Kasas. OpenFovea: open-source AFM data processing software. *Nature Methods*, 9(8):774–775, 2012.
- [107] Prince E Rouse Jr. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *The Journal of Chemical Physics*, 21(7):1272–1280, 1953.
- [108] A. Sanchez-Sevilla, J. Thimonier, M. Marilley, J. Rocca-Serra, and J. Barbet. Accuracy of AFM measurements of the contour length of DNA fragments adsorbed on mica in air and in aqueous buffer. *Ultramicroscopy*, 92(3):151–158, 2002.
- [109] Massimo Sandal, Fabrizio Benedetti, Marco Brucale, Alberto Gomez-Casado, and Bruno Samorì. Hooke: an open software platform for force spectroscopy. *Bioinformatics*, 25(11):1428–1430, 2009.
- [110] Dror Sarid. *Scanning force microscopy*. Oxford University Press, 1991.
- [111] Gideon Schreiber and Alan R Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry*, 32(19):5145–5150, 1993.
- [112] Falk Schwesinger, Robert Ros, Torsten Strunz, Dario Anselmetti, Hans-Joachim Güntherodt, Annemarie Honegger, Lutz Jermutus, Louis Tiefenauer, and Andreas Plückthun. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proceedings of the National Academy of Sciences*, 97(18):9972–9977, 2000.
- [113] Sergey K Sekatskii, Mounir Mensi, Andrey G Mikhaylov, Giovanni Dietler, et al. The use of light diffracted from grating etched onto the backside surface of an atomic force microscope cantilever increases the force sensitivity. *Journal of Surface Engineered Materials and Advanced Technology*, 3(04):29, 2013.
- [114] SK Sekatskii, F Benedetti, and G Dietler. Dependence of the most probable and average bond rupture force on the force loading rate: First order correction to the bell–evans model. *Journal of Applied Physics*, 114(3):034701, 2013.
- [115] SK Sekatskii, M Favre, G Dietler, AG Mikhailov, DV Klinov, SV Lukash, and SM Deyev. Force spectroscopy of barnase–barstar single molecule interaction. *Journal of Molecular Recognition*, 23(6):583–588, 2010.

Bibliography

- [116] Felix B Sheinerman and Barry Honig. On the role of electrostatic interactions in the design of protein–protein interfaces. *Journal of molecular biology*, 318(1):161–177, 2002.
- [117] Pei-Yong Shi. *Molecular Virology and Control of Flaviviruses*. Horizon Scientific Press, 2012.
- [118] Alexander Spaar, Christian Dammer, Razif R Gabdoulhine, Rebecca C Wade, and Volkhard Helms. Diffusional encounter of barnase and barstar. *Biophysical journal*, 90(6):1913–1924, 2006.
- [119] Frank Spitzer, Frank Spitzer, Frank Spitzer, and Austria Mathematician. *Principles of random walk*. Springer, 1964.
- [120] Karin Stiasny, Richard Fritz, Karen Pangerl, and Franz X Heinz. Molecular mechanisms of flavivirus membrane fusion. *Amino Acids*, 41(5):1159–1163, 2011.
- [121] Karin Stiasny and Franz X Heinz. Flavivirus membrane fusion. *Journal of general virology*, 87(10):2755–2766, 2006.
- [122] Torsten Strunz, Krisztina Oroszlan, Rolf Schäfer, and Hans-Joachim Güntherodt. Dynamic force spectroscopy of single DNA molecules. *Proceedings of the National Academy of Sciences*, 96(20):11277–11282, 1999.
- [123] Joan K Stuart and Vladimir Hlady. Effects of discrete protein-surface interactions in scanning force microscopy adhesion force measurements. *Langmuir*, 11(4):1368–1374, 1995.
- [124] T Sulchek, RJ Grow, GG Yaralioglu, SC Minne, CF Quate, SR Manalis, A Kiraz, A Aydine, and A Atalar. Parallel atomic force microscopy with optical interferometric detection. *Applied Physics Letters*, 78(12):1787–1789, 2001.
- [125] J Suss. Tick-borne encephalitis in Europe and beyond—the epidemiological situation as of 2007. *Euro surveillance: bulletin europeen sur les maladies transmissibles= European communicable disease bulletin*, 13(26):717–727, 2008.
- [126] Pieter Rein ten Wolde and David Chandler. Drying-induced hydrophobic polymer collapse. *Proceedings of the National Academy of Sciences*, 99(10):6539–6543, 2002.
- [127] F Valle, M Favre, P De Los Rios, A Rosa, and G. Dietler. Scaling exponents and probability distributions of DNA end-to-end distance. *Physical review letters*, 95(15):158105, 2005.
- [128] J Vesenka, M Guthold, CL Tang, D Keller, E Delaine, and C Bustamante. Substrate preparation for reliable imaging of DNA molecules with the scanning force microscope. *Ultramicroscopy*, 42:1243–1249, 1992.

- [129] Kirstin A Walther, Jasna Brujić, Hongbin Li, and Julio M Fernández. Sub-angstrom conformational changes of a single molecule captured by AFM variance analysis. *Bio-physical journal*, 90(10):3806–3812, 2006.
- [130] Patricia C Weber, DH Ohlendorf, JJ Wendoloski, and FR Salemme. Structural origins of high-affinity biotin binding to streptavidin. *Science*, 243(4887):85–88, 1989.
- [131] John W Weisel, Henry Shuman, and Rustem I Litvinov. Protein–protein unbinding induced by force: single-molecule studies. *Current opinion in structural biology*, 13(2):227–235, 2003.
- [132] Paul A. Wiggins, Thijn Van der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C. Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, 1(2):137–141, 2006.
- [133] Paul A Wiggins, Thijn Van Der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C Nelson. High flexibility of dna on short length scales probed by atomic force microscopy. *Nature nanotechnology*, 1(2):137–141, 2006.
- [134] Mark C Williams and Ioulia Rouzina. Force spectroscopy of single DNA and RNA molecules. *Current opinion in structural biology*, 12(3):330–336, 2002.
- [135] Guillaume Witz, Kristian Rechendorff, Jozef Adamcik, and Giovanni Dietler. Conformation of Circular DNA in Two Dimensions. *Physical Review Letters*, 101(14), oct 2008.
- [136] Guillaume Witz, Kristian Rechendorff, Jozef Adamcik, and Giovanni Dietler. Conformation of circular DNA in two dimensions. *Physical review letters*, 101(14):148103, 2008.
- [137] Atsushi Yarai, Yoichiro Fukunaga, Kazuo Sakamoto, and Takuji Nakanishi. High-frequency and high-gain amplification of photothermal beam deflection angle using cylindrical reflection mirror. *Japanese journal of applied physics*, 33(5S):3251, 1994.
- [138] GG Yaralioglu, A Atalar, SR Manalis, and CF Quate. Analysis and design of an interdigital cantilever as a displacement sensor. *Journal of Applied Physics*, 83(12):7405–7415, 1998.
- [139] Christopher M Yip, Cecil C Yip, and Michael D Ward. Direct force measurements of insulin monomer-monomer interactions. *Biochemistry*, 37(16):5439–5449, 1998.
- [140] Chunbo Yuan, Aileen Chen, Pamela Kolb, and Vincent T Moy. Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy. *Biochemistry*, 39(33):10219–10223, 2000.

Bibliography

- [141] Boris N. Zaitsev, Fabrizio Benedetti, Andrey G. Mikhaylov, Denis V. Korneev, Sergey K. Sekatskii, Tatyana Karakouz, Pavel A. Belavin, Nina A. Netesova, Giovanni Dietler Protopopova, Elena V. Svetlana N. Konovalova, and Valery B. Loktev. Force-induced globule–coil transition in laminin binding protein and its role for viral–cell membrane fusion. *Journal of Molecular Recognition*, 2014.
- [142] Wei Zhang, Paul R Chipman, Jeroen Corver, Peter R Johnson, Ying Zhang, Suchetana Mukhopadhyay, Timothy S Baker, James H Strauss, Michael G Rossmann, and Richard J Kuhn. Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nature Structural & Molecular Biology*, 10(11):907–912, 2003.
- [143] Jordanka Zlatanova, Stuart M Lindsay, and Sanford H Leuba. Single molecule force spectroscopy in biology using the atomic force microscope. *Progress in biophysics and molecular biology*, 74(1):37–61, 2000.

Andrey MIKHAYLOV
Riva Paradiso 36
CH-6900, Paradiso
Switzerland
Tel : +41 76 344 14 25
Mail : andrey.mikhaylov.86@gmail.com

Born in Moscow on 01.12.1986
Russian
Married, one child



Experience

- 2014**
(up to now) BLOOMBERG L.P., LUGANO. FINANCIAL SOFTWARE DEVELOPER.
Development of financial software platform.
▷ *C++ software development.* ▷ *Financial models.* ▷ *Financial predictions.*
- 2010**
(6 months) INSTOCKTECH, MOSCOW. JAVA DEVELOPER.
Worked on the development of client-server software managing geositional data of telecommunication companies.
▷ *Java software development.* ▷ *Client-server architecture.* ▷ *Oracle database.*
- 2008**
(2 months) ALIGN TECHNOLOGIES, MOSCOW. C++ DEVELOPER.
Internship in a team developing automated software aimed to construct "aligners", unique polymer parts used to align teeth of the patients.
▷ *C++ software development.* ▷ *Subversion control.*
- 2005-2006**
(1 year) KURCHATOV INSTITUTE, MOSCOW. INTERNSHIP ON FUEL COMBUSTION SIMULATION.
Internship in a laboratory simulating the dynamics of hydrogen combustion.
▷ *C++ software development.* ▷ *Numerical simulations.*
- 2004-2005**
(1 year) AMPLITUDA, MOSCOW. VBA DEVELOPER.
Internship in a company developing equipment for gamma spectroscopy. The project aimed to develop a user interface and data treatment modules of the software.
▷ *Data processing automation.* ▷ *UI development.*

Education

- 2010-2014**
(expected) ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE. PH.D. IN PHYSICS.
▷ *Software development.* ▷ *Hardware development.* ▷ *Data processing.* ▷ *LabVIEW.* ▷ *MatLab.*
▷ *Mathematica.* ▷ *OpenCL.*
- 2007-2009**
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY. MS IN PHYSICS.
▷ *GPA : 5/5.* ▷ *Work experience in a lab.*
- 2003-2007**
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY. BS IN PHYSICS.
▷ *GPA : 4.85/5.* ▷ *Mathematics.* ▷ *Physics.* ▷ *Algorithms.* ▷ *Statistics.* ▷ *Software development.*

Languages

Russian	Native
English	Fluent
French	Fluent
German	Upper-intermediate

Publications

- 2014 B.N. Zaitsev et al. *Journal of Molecular Recognition.*
- 2013 A.G. Mikhaylov et al. *Journal of Advanced Microscopy Research.*
- 2013 S.K. Sekatskii et al. *Journal of Surface Engineered Materials and Advanced Technology.*
- 2010 S.K. Sekatskii et al. *Journal of Molecular Recognition.*

Other information

Organization Management	Organized International Physicists' Tournament in EPFL in 2013 and 2014. Project manager in ShARE, international student organization. We have delivered a consulting-based project supervised by Deloitte.
Awards	2 nd place in Helvetic Coding Contest.
Teaching	Significant experience of work as a tutor of mathematics, physics and programming in Moscow. During this time I prepared schoolboys for university entrance exams (mostly MIPT). I was teaching physics in EPFL (1st - 3rd year students, in French).
Communication	Oral presentations in different conferences.
Hobbies	Traveling, photography, sports.