EFFECTS OF BASE-PAIR SEQUENCE, NICKS AND GAPS ON DNA MINICIRCLE SHAPES: ANALYSIS AND EXPERIMENT

THÈSE N^{O} 3995 (2007)

PRÉSENTÉE LE 19 DÉCEMBRE 2007

À LA FACULTÉ DES SCIENCES DE LA VIE
CHAIRE D'ANALYSE APPLIQUÉE
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Arnaud AMZALLAG

Ingénieur des Médias et Architecture de la Communication, IMAC, Université Panthéon-Assas (Paris II), France
et de nationalité française

acceptée sur proposition du jury:

Prof. N. Stergiopulos, président du jury
Prof. J. Maddocks, directeur de thèse
Prof. V. Hatzimanikatis, rapporteur
Dr A. Stasiak, rapporteur
Prof. J. Widom, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2008

Résumé

L'ADN est un long polymère en forme de double hélice d'environ deux nanomètres (2.10^{-9} mètres) de diamètre. Il est composé d'une séquence de nucléotides qui porte le message de l'hérédité. Les quatre types de nucléotides ont des formes légèrement différentes, et l'on pense que leur séquence influence la forme et la rigidité de la double hélice à une échelle de quelques dizaines ou centaines de paires de bases, et par là même influence son activité biologique.

Les minicercles (ou miniplasmides) d'ADN sont des boucles d'une longueur de quelques dizaines ou centaines de paires de bases dans lesquelles la double hélice est fermée sur elle-même avec un certain nombre de tours d'hélice. Sa forme et son énergie de formation dépendent de la séquence, et peuvent être calculés efficacement à l'aide de modèles de polymères si les paramètres de forme et de rigidité de l'ADN sont fournis. Cette structure est donc un motif expérimental intéressant pour étudier les paramètres mécaniques associés à une séquence d'ADN.

Le sujet de cette thèse est l'étude des méthodes expérimentales pour la détermination de la forme et de l'énergie libre de formation des minicercles d'ADN. Les formes ont été déterminées à partir d'images de microscopie électronique. L'efficacité de formation a été mesurée à l'aide d'une nouvelle méthode appelée annealing-cyclization.

La cryo-microscopie électronique permet d'observer de très petites molécules (ici 17 ou 11 nm de diamètre) dans de l'eau vitrifiée, pour garder la forme tridimensionnelle des molécules aussi fidèle que possible à leur forme originelle en solution. Au chapitre 1, j'ai utilisé des paires stéréo de cryo-micrographes pour déterminer et comparer la forme de 95 minicercles dont la séquence est identique, à part dans un intervalle de 18 paires de bases qui contient soit une séquence TATA soit une séquence CAP. J'ai défini la notion de distance de forme qui a servi à estimer l'erreur de reconstruction tridimensionnelle, et grâce à laquelle j'ai pu détecter plusieurs groupes de formes similaires avec un algorithme de tri approprié. Cependant, les groupes n'avaient pas l'air spécialement associés avec la séquence variable (TATA ou CAP).

J'ai alors analysé (au chapitre 2) des formes bidimensionnelles de minicercles d'ADN plus courts (94 bp) déterminés par microscopie électronique à coloration négative. Les minicercles avaient soit deux nicks (cassure dans l'un des deux brin) soit deux gaps (nucléotides manquants formant localement un simple brin d'ADN) à des positions diamétralement opposées. J'ai observé que les minicercles avec gaps prennent une forme plus allongée que les minicercles avec nicks ; à l'aide en plus de simulations de dynamique moléculaire, j'ai conclu que la flexibilité des gaps, peut-être appuyée par la cassure de certaines paires de base aux sites gaps, est responsable de la forme allongée des minicerles avec gap.

Finalement, j'ai proposé une méthode pour mesurer l'efficacité de formation des minicercles, et qui produit un haut rendement de minicercles avec nicks ou gaps. Cette méthode évite l'utilisation de ligase et les problèmes associés aux effets de la concentration de ligase sur les résutats. J'ai déterminé une équation à partir du modèle chimique de la réaction qui décrit les données lues à partir de gels d'électrophorèse, et j'ai défini le facteur J_a pour cette réaction, qui mesure l'efficacité de cyclisation, et qui est indépendant de la concentration initiale de la réaction. La méthode confirme que les minicercles avec gaps se forment plus facilement que les autres, probablement grâce à la flexibilité des gaps.

En guise de perspective, les conclusions de cette thèse pourraient être utilisées pour le design de minicercles dont la forme serait suffisamment sensible à un changement de séquence pour que celui-ci soit détecté par microscopie électronique. Ces formes pourraient alors être comparées, grâce à l'outil de distance de forme définit ici, à des formes calculées avec des modèles d'ADN connus ou hypothétiques.

Mots clés: Minicercles d'ADN, cryo-microscopie éléctronique, reconstruction 3D, classement de formes, dynamique moléculaire, circularisation d'ADN, mécanique de l'ADN dépendente de la séquence.

Abstract

DNA is a long polymer with the form of a double helix of about two nanometers diameter (2.10^{-9} meters). It is composed of nucleotides whose sequence carry the information of heredity. The four possible nucleotides have slightly different geometries, and their sequence along the DNA molecule are thought to influence the shape and the stiffness of the double helix on a scale of few tens to a few hundred base pairs, and thereby its biological activity.

DNA minicircles (or miniplasmids) are closed loops with lengths of the order of a few tens or hundreds of base pairs in which the double helix bends around to close on its own tail with some number of twists. Its minimal energy shape and its energy of formation depend on the sequence of base pairs, and can be efficiently computed by polymer and rod models, if shape and stiffness parameters of DNA are provided as inputs. This structure is therefore an interesting experimental motif to test sequence-dependent mechanical properties of the DNA molecule.

The purpose of this thesis is to explore the experimental methods to determine the shape and free energy of formation of DNA minicircles. The shapes have been be determined from cryo-electron micrographs. Efficiency of formation has been investigated by a novel method called annealing-cyclization.

Cryo-electron microscopy allows observation of very small molecules (here 17 or 11 nm diameter) in vitrified water, in order to keep the 3D shape of the molecules as close as possible to the shape they had in solution. In Chapter 1, I used stereo cryo-electron micrographs to determine and compare the three-dimensional shape of 95 individual DNA minicircles of 158 base pairs, which were identical in sequence except within a 18 bp block which contained either a TATA box sequence or a CAP site. I defined the notion of shape-distance which I used to estimate the error of reconstruction, and I detected clusters of shapes using an appropriate sorting algorithm. However the cluster did not seem to be associated with the variable sequence (TATA or CAP).

I then analyzed (in Chapter 2) two-dimensional shapes of shorter DNA

minicircles (94 bp) determined by negative staining electron microscopy, designed with either two nicks (breaks in one of the strands) or two gaps (missing nucleotides) at diametrically opposite sites of the minicircles. I observed that the gapped minicircles have an elongated shape with respect to the nicked minicircles, and I used this result together with the results of atomic level molecular dynamics simulations to conclude that the gap flexibility, perhaps together with base unpairing at the gap site, is responsible for this elongated minicircle shape.

Finally, I proposed a ligase-free assay to measure the minicircle formation efficiency, which can give a high yield of nicked minicircles. The method avoids the use of ligase and the associated concerns about the effect of ligase concentration on the measurements. I determined an equation from a chemical model of the reaction that fits the experimental data, and I defined the J_a factor which gives a measure of the cyclization yield independent of DNA initial concentration. This method seems to confirm that minicircles with two gaps cyclize more efficiently that minicircles with two nicks, probably because of the gap flexibility.

As a perspective, the conclusions of this thesis could be used for the design of minicircle constructs whose shape would be sensitive enough to sequence mutation in order to be detected by electron microscopy. Such shapes could be then compared, thanks to the shape-distance tool defined herein, to shapes computed with known or putative DNA models.

Keywords: DNA minicircles, cryo-electron microscopy, 3D reconstruction, shape clusturing, molecular dynamics, DNA cyclization, sequence-dependent DNA mechanics.

Acknowledgments

I am grateful to Prof. John H. Maddocks for trusting me and giving me the opportunity to work on this challenging and exciting scientific project, and for his acute and critical advice throughout this process. I am also grateful to Dr. Andrzej Stasiak, who guided me with experience and kindness in the experimental part of the project and who helped a lot for the redaction of the article. I could not have done this work without the great participation of all my collaborators: Prof. Jacques Dubochet who kindly allowed me to work in his lab, always made me feel as a lab member and shared with me very instructing ideas; Prof. Jason Kahn who very kindly invited me in his lab at the University of Maryland where I could perform (with the help of his post-doctoral fellow Mein Samala) ligase-catalyzed DNA cyclization experiments, and with whom I had important discussions on experimental design; Cédric Vaillant who introduced me to the practical difficulties of this project; Jan Bednar who provided me the great images of DNA minicircles I used here, and Mathews Jacob and Michael Unser for the code and advice about DNAJ; Davide Demurtas for his great images of DNA minicircles, help at the bench, discussions and support; Leesa Heffler for the careful molecular dynamics simulations results she provided for this study, and Filip Lankas for the same reason and for his availability for scientific discussions; Daphna Tsafrir for her advice in using SPIN, and Prof. Domany and his group for stimulating comments; Prof Horst Vogel and Dr. Ruud Hovius for the fluorometer and the initiation to fluorescence in solution. I also thank for fruitful discussions, Profs Edward Trifonov, Jonathan Widom, Rob Philips, Jorg Langowski, Paulo De Los Rios and all the people who kindly discussed with me about my project. I am also indebted to my master degree advisor Damien Schoëvaërt-Brossault who first introduced me to biophysical sciences.

I thank all LCVMM members, especially our secretary Carine Tschanz who took care of all the possible administrative issues and allowed me to focus on my work; Philippe Caussignac and Mathias Carlen for keeping my computer in good shape till the last moments of the thesis writing; Prakash Tiwary, Ludovica Cotta-Ramusino, Angelo Rosa, Henryk Gerlash and Jessika Walter. I thank also all the members of LAU (UNIL) for scientific discussions and also for the good time, should it be in the lab or in the montains.

I have to give a very special thank to my wife, Sophie Shnaper-Amzallag, who supported, encouraged and adviced me all the time, and took care of our daughter Daphna and most of the house duties during the last months of the thesis redaction; I thank Daphna for being so cute. I thank my dear mother Danielle Amzallag for having raised me with such attention and trust, and for giving me the taste of learning; her memory lives in me. I thank my father Bob Amzallag and Cathy Hazouard for their constant support during all these years, and Sophie's parents Natan Shnaper and Natalie Belenkii for their encouragments and help. I thank my childhood friends who encouraged me to grow and to get interested in the world around me through discussions, games and music: Guénaël, Ouriel, Jérémie, Elie et Eva, François, and all the friends who do not find their name here, but which I think of for sure.

Contents

Chapter 0

Introduction

0.1 DNA Minicircles: A Bridge Between Theory and Experiment

The DNA molecule is a long polymer whose shape can be described by various mathematical models at differing levels of coarse-graining, for instance continuous elastic rods [1], birods [2] , or worm like chain models [3]. To be of use, these models must have mechanical parameters provided as inputs (for instance the bending stiffness of the rod) which depend on the properties of the material of interest, in our case a DNA molecule with a given sequence. If these parameters are known, the elastic energy, the free energy associated to the geometrical constraint of minicircle closure, and the minimal energy shape of the minicircle can be computed by various means [1, 3, 4, 5, 6, 7]. Therefore, direct visualization of sequence effect on DNA minicircle shapes by electron microscopy and 3D reconstruction may be a valuable tool for validation of models and the associated mechanical parameters of DNA. The free energy of DNA minicircles can be measured by an established method called ligase-catalyzed cyclization [6, 8, 9, 10]. In Chapter 1, I will discuss whether effects of sequence on DNA minicircle shapes can be observed from cryo-electron microscopy images, provided by collaborators in LAU, UNIL. Although the answer turns out to be in the negative, I will show in Chapter 2 that an appropriate analysis does reveal that the effects of nicks and gaps on DNA minicircle shapes are visible in negative staining electron microscopy images, again provided by collaborators in LAU, and interpret molecular dynamics simulations of such minicircles performed by

my colleagues in LCVMM. Finally, I will present in Chapter 3 experimental results of a variant of the cyclization assay which does not use ligase, and discuss the efficiency of formation of nicked and gapped DNA minicircles.

0.2 The DNA Molecule and its Mechanical Properties

History: Discovery of DNA structure As the support of genetic information, DNA is one of the most important molecules in life. The exact duplication of this information in any species from generation to generation assures the genetic continuity of that species. At each cell division, the total DNA (6.10^9 base-pairs for human cells) is duplicated, separated in two quasi-identical sets of chromosomes, and strongly compacted into the nucleus of each daughter cell.

It was known in the early 50's that "Nucleic acids are, with few exceptions, linear polymers of nucleotides whose phosphate bridge the 3' and the 5' positions of successive sugar residues" ([11] p. 849). A nucleotide is made of a sugar and a base. The names of the four types of bases are Adenine, Guanine, Cytosine and Thymine, commonly noted A, G, C and T. The two purines A and G are composed of two aromatic rings, whereas the pyrimidines C and T are smaller, composed of one aromatic ring. By the end of the 1940's Erwin Chargaff succeeded to chop DNA into single nucleotides by hydrolysis, and measured by paper chromatography that DNA always had equal numbers of T and A residues, and equal numbers of C and G residues [12]. This equality is known as the Chargaff rule. Then Francis Crick, who derived the equations describing the X-ray diffraction patterns of helical molecules, deduced from Rosalind Franklin's X-Ray diffraction photographs of DNA fibers that DNA is a helical molecule formed with a stack of parallel bases.

Based on the geometry of bases, Chargaff's rule and insights from X-ray diffraction patterns, James Watson and Francis Crick proposed a three dimensional model of the DNA molecule for which they received the Nobel Prize and which is widely considered as the starting point of modern molecular biology. The model they proposed is a double helix formed by two polymers of DNA –the DNA strands– wrapping around each other, with the sugar phosphate backbone on the outside, and the bases associated in

pairs inside. The base pairs are planar, and lie in parallel planes along the molecule, spaced by 0.34 nanometers (1 nm = 10^{-9} m) intervals (Figure 1b). Therefore the base-pairs are commonly said to be stacked. The most remarkable feature of the structure proposed by Watson and Crick is that the bases associate into planar base pairs of two kinds only: A with T through two hydrogen bonds, and C with G forming three hydrogen bonds. They explained that the pairs C-G and T-A are both made of one purine and one pyrimidine, and thus have very similar geometries(Figure 1a); therefore, what is now known as Watson-Crick base pairs "are interchangeable in that they can replace each other in the double helix without altering the position of the sugar-phosphate backbone's C1' atoms" [13], whereas any other combination of bases would significantly distort the double helix.

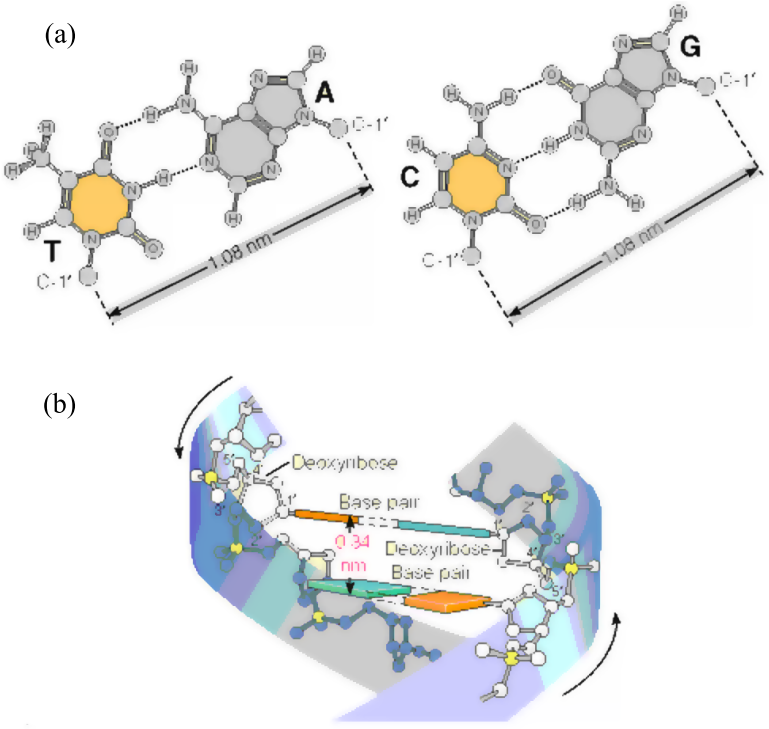

Figure 1: The Watson-Crick base pairs and the DNA double helix. Picture from [14] **(a)** The two Watson-Crick base pairs have similar geometry. **(b)** The DNA double helix. Base pairs stack, and are parallel to each other. The sugar phosphate backbones form a double helix.

This model not only explained the Chargaff rule, it provided an expla-

nation for the mechanism of heredity. Indeed, the unambiguous way of forming base pairs allows only one complementary sequence of bases to hybridize with a given DNA strand. Therefore it provides a way for cells to copy its DNA [15]. First, the two strands of the DNA molecule are dissociated. Then, on each single strand a complementary strand is rebuilt with only Watson-Crick base pairs, hence producing two identical copies of the original double stranded DNA molecule. The structure proposed by Watson and Crick is also called B-DNA, to distinguish it from two rarer structures called A-DNA and Z-DNA (Figure 2). The A form was discovered before the B form, probably because it is the dehydrated form of DNA: it was easier to produce a crystal of A-DNA than of B-DNA. Indeed, when the relative humidity is reduced to 75%, B-DNA undergoes a reversible conformational change to the A form ([11] p. 857). Its inter-base pair distance along the helical axis (or pitch) is 0.26 nm ([16], p. 57), which is slightly smaller than the B-DNA pitch (0.34 nm). Z form, in contrast to A and B form is a left-handed helix. This form was given the name Z because of the zigzag shape of its sugar-phosphate backbone. A remarkable feature of Z form is that the bases are rotated by about 180°around the bond which links the bases to their respective sugars (Figure 3). This angle of rotation is called the χ angle.

Deviation from the ideal helical shape: sequence dependence B-DNA is 2 nanometers wide, and the distance between two consecutive base pairs is 0.34 nanometers. Along the helical axis of the molecule, the base pairs are rotated around the helical axis by about 34.3°, so that one helical turn is achieved after 10.5 base pairs (Figure 4). Although the Watson-Crick base pairs are similar in geometry, the base pairs sequence can induce differences in shape along the double helix. The A and T bases associate in a base pair through two hydrogen bonds and therefore may require less energy to dissociate than the bases G and C. Also, pyrimidines are smaller than purines. For these and other reasons, the sequence can influence the geometry of the DNA double helix, and in fact the helical axis is not perfectly straight. The physical properties such as stiffness may also vary along the molecule. Models which describe the curvature of the DNA helical axis as a function of the sequence of base pairs have been proposed, based on gel retardation data [18, 19], or crystallography of DNA-protein complexes [20].

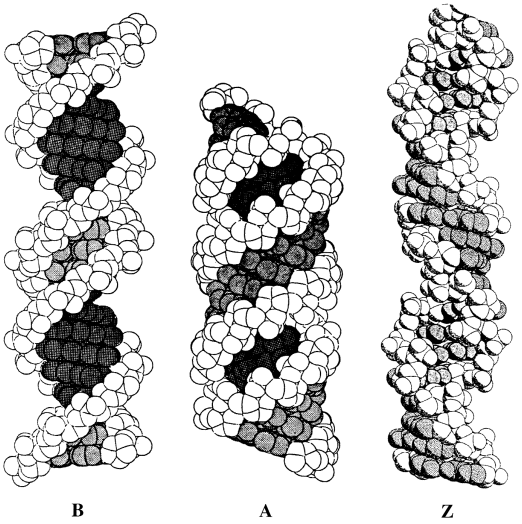

Figure 2: Three well known (but highly idealized) forms of DNA: B and A are right-handed with 10.5 and 11 base pairs per helical turn, respectively, while Z is left-handed, and has 12 base pairs per turn (Figure from [17] p. 25).

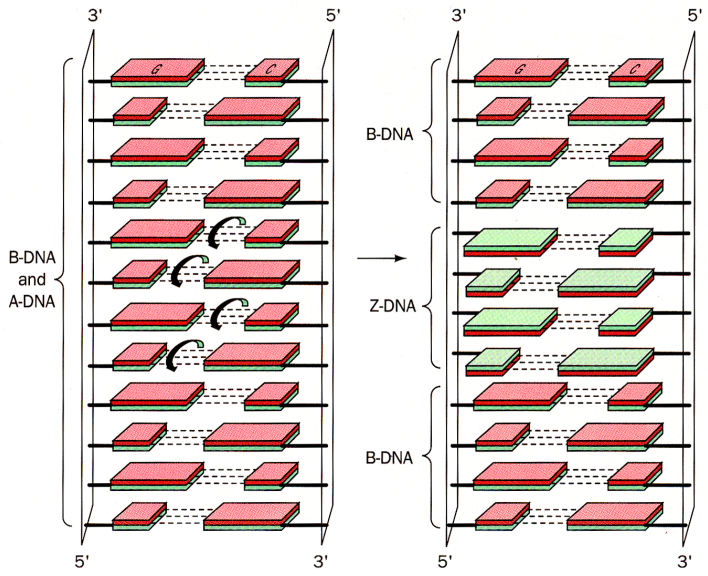

Figure 3: Transition to Z form of DNA. The different faces of the bases are colored in red and green. In Z form, the bases are rotated by 180°around the bonds which links the bases to the sugars (Figure from [11], p. 860).

A striking example of sequence influence on DNA shape is the A-tract, a sequence of 4 to 6 consecutive adenines. It is known to create a local bend of the helical axis: molecules with an A-tract repeated with a periodicity of 10.5 base pairs in average (periodicity of the double helix) have successive bends in the same plane. Indeed, it has been observed that such DNA molecules are very curved at the scale of a few hundred base pairs, by cryo-electron microscopy [21], by cyclization [22, 23], and by gel retardation [24, 25]. The sequences of the minicircles observed in Chapter 1 were designed with six A-tracts, with an average periodicity of 10.5 base pairs, in order to facilitate minicircle formation.

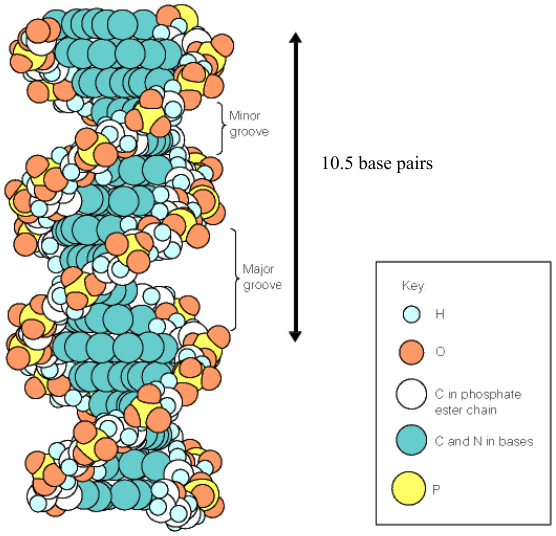

Figure 4: Side view of a B-DNA molecule. A helical turn is about 10.5 base pairs long. We can see that the distances between the two backbones along the helical axis are not equal: they are the minor and major grooves of DNA. [26]

The double helix in solution, the persistence length In order to appreciate the level of curvature of DNA in the biological examples of the next paragraph, one should first know how naked DNA behaves in solution. At room temperature, the molecules that form the solution are agitated and bumping into each other. So, DNA in solution is permanently subject to collisions of the surrounding water molecules, ions, and itself. Under the influence of these movements, called thermal motion, the polymer is randomly

bending. But the average magnitude of the bends depends on the flexibility of the polymer. "A common measure of polymer bending rigidity is the length over which the average deflection of the polymer axis caused by thermal agitation is one radian. This length is called the persistence length" [21]. Several methods were used to measure DNA persistence length, including cryo-electron microscopy, the experimental method of direct visualization of DNA, on which most of this thesis is based. Bednar and collaborators used cryo electron microscopy to measure the persistence length of double stranded DNA with random sequence [21], and confirmed the previously measured value of 50 nm [27]. This corresponds to the length of a molecule of 150 base pairs, or 15 helical turns. Thus, the helical axis is deviated from the straight line by thermal motion, and the persistence length gives a rough idea of the stiffness of DNA.

Tightly bent DNA In several biological examples DNA is significantly bent on lengths shorter than the persistence length. One reason is the packaging constraints of genomic DNA. DNA can be considered as an immensely long polymer in the sense that a single DNA molecule can be composed of hundreds of millions of base pairs. For instance, human chromosome 1 is made of 240 million base pairs and thus has a contour length of 8.2 centimeters, although it has to fit in a human cell nucleus of few microns of diameter. In fact, in human cells, bacteria and in viruses, DNA has to be tightly packed in order to fit in its container. Virus capsids diameters are on the order of tens of nanometers, whereas the viral genome length is about three orders of magnitude longer. DNA has therefore to be tightly coiled, generating pressure inside the virus; this pressure is even thought to help the DNA ejection into the infected cell [28].

Sequence-dependent DNA curvature and stiffness is also thought to play a direct role in DNA compaction. Periodic A-tracts have been detected in the genome of the bacteria *Escherichia coli*; they may ease DNA packing in the cell [30]. In eukaryotes (organisms whose cells have a nucleus) DNA is organized in chromatin, the smaller subunit of which is called a nucleosome. In the nucleosome, 147 bp of DNA wraps for almost two complete turns around protein complexes called histones (see [31, 32] for a crystal structure). The radius of curvature of the double-helix axis is between 4 and 5 nm, whereas the DNA thickness is 2 nm. In this highly bent regime,

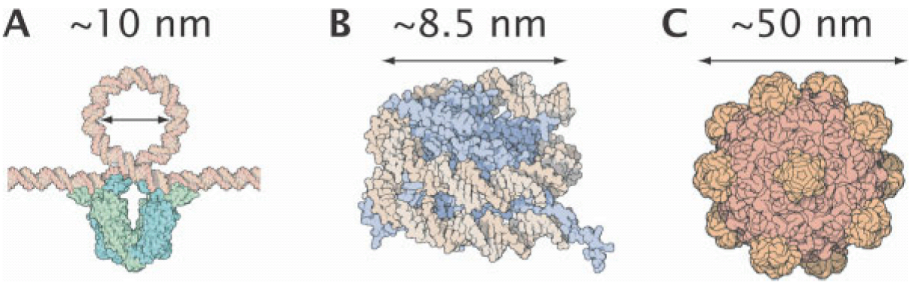

Figure 5: Biological examples of tightly bent DNA. (A) Transcription factor mediated DNA looping, (B) DNA packing in the nucleosome, (C) DNA packing in bacterial viruses. Image is taken from [29], original artwork is from David Goodsell.

the DNA sequence strongly affects the nucleosome position along the DNA molecule [33, 34, 35], and it is believed that this is due to sequence-dependent DNA deformability [36]. Bacterial gene regulation involves DNA-protein complexes forming short DNA loops [37, 38]. In DNA-protein complexes, the contribution of the DNA sequence to the binding affinity is frequently indirect, by opposition to direct recognition where nucleotide bases interact with amino-acids through hydrogen bonds. For instance, the Bovine Papilloma Virus BPV-1 E2 binds two sites spaced by four base pairs [39]. The DNA spacer does not contact the protein, but its sequence affects the stability of the DNA-protein complex [40]. The role of the spacer in the stability of the DNA-protein complex has also been shown in the case of the Cyclic AMP Receptor Protein (CRP)[41, 42], also known as Catabolite Activator Protein (CAP) [43]. Hence, in several biologically essential structures, DNA is significantly bent at a length scale shorter or equal to the persistence length, and thus subject to high mechanical constraints (for a review see [29]). Knowing the mechanical properties of DNA at the length scale of a persistence length is therefore relevant to biology. Moreover, at this length scale the base pair sequence influences the geometry of DNA, suggesting that DNA sequence may have a mechanical function in addition to the genetic code.

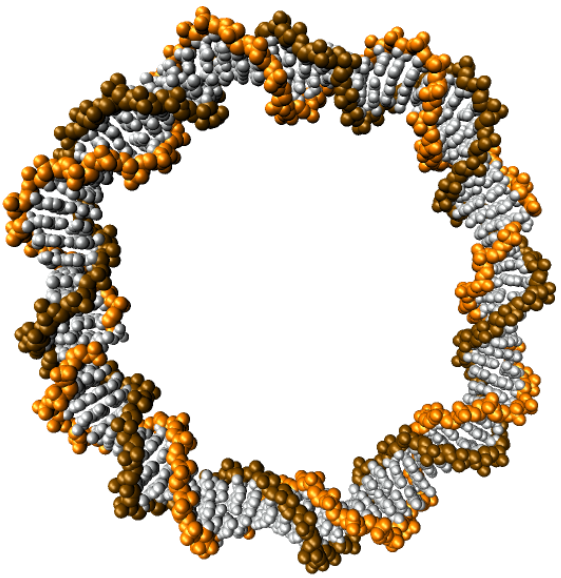

Figure 6: Atomic level representation of a 94 base-pair DNA minicircle. Data is from the molecular dynamics simulations performed at EPFL by Leesa Heefler and Filip Lankas, which I analyzed in Chapter 2. The image was produced with VMD [44].

0.3 DNA Cyclization and Minicircles

DNA minicircles are closed loops with lengths of the order of a few hundred base pairs or less in which the double helix bends around to close on its own tail with some numbers of helical turns. They can be used to probe the sequence-dependent mechanical properties of DNA, such as stiffnesses and intrinsic shape. The length of minicircles is often set to about the persistence length of DNA, around 50 nm, or 150 bp. The sequence-dependent elastic properties of DNA will have more significant effects on the physical behavior of these short DNA fragments than on longer segments of DNA, for which entropic effects, which are largely sequence independent, will gradually dominate. Such minicircles provide an attractive experimental motif within which the mechanics of DNA can be studied because the molecule is stressed, with different loadings for different topoisomers, without there being any difficulties associated with end-effects or external loading devices. The study of DNA minicircles has been introduced by an experiment which consists in measuring the rate of the reaction of formation of minicircles by joining ends of a short linear DNA molecule. This reaction is called

cyclization, or sometimes circularization.

DNA cyclization Cyclization is the reaction were the two ends of a linear polymer bind to each other to form a circular polymer. In 1981, Shore Langowski and Baldwin designed a cyclization assay for DNA in solution [8, 9]. When a DNA fragment extremities are in contact, aligned and with an appropriate twist angle between them, the DNA molecule can be cyclized by a DNA ligase. DNA ligase is an enzyme which can form a covalent bond between the phosphate group at the 5' end of one strand and the sugar ring at the 3' end of the same strand (through the C3' carbon of the sugar). To facilitate the ligation, the DNA molecules are often designed so that they have short complementary single stranded overhangs at the ends (between two and four nucleotides usually), which can hybridize and form a substrate for the DNA ligase. Complementary single stranded overhangs are therefore commonly called "sticky ends".

Cyclization of polymers in solution has be studied from the 1950's by Jacobson and Stockmayer [45]. A polymer end can hybridize with the other end of the same molecule, or it can hybridize with an end of another molecule. The first case is called cyclization, The latter case is called bimolecular association or dimerization. The product of cyclization of the initial reactant is a monomeric circle, or a circular monomer, and the product of bimolecular association is a linear dimer, which can in turn cyclize and form a circular dimer. Cyclization efficiency depends on the mechanical ability of the polymer to accept the geometrical constraint of circular closure, as opposed to bimolecular association. Jacobson and Stockmayer [45] used the principles of statistical mechanics to derive a formula, which relates the equilibrium constants of cyclization and bimolecular association to the probability of the polymer ends to be in contact [46]:

$$J = \frac{K_C}{K_D} = exp\left[\frac{-(\Delta G^\circ_C - \Delta G^\circ_D)}{RT}\right] = \frac{W(0)}{N_A}. \quad (1)$$

J is called the J factor and is by definition equal to K_C/K_D. In the context of DNA, K_C and K_D are respectively the equilibrium constants of the cyclization and bimolecular association, by hybridization of the sticky ends, without covalent ligation. ΔG°_C and ΔG°_D refer to the standard free energy changes of the two reactions; ΔG°_D accounts for the hydrogen bonding,

and ΔG_C° for both the hydrogen bonding and the free elastic energy change due to the cyclization constraint imposed on the molecule. The difference $\Delta G_C^\circ - \Delta G_D^\circ$ cancels the hydrogen bonding energy terms, which are thought to be equal because the sticky ends are identical in both reactions. Thus, J depends on the elastic energy term and temperature only. N_A is Avogadro's number, $W(r)$ is the probability density of the polymer ends to be at a distance r. Flory and co-workers [4] introduced expressions for J which include as factors the probability densities of the orientations and the twist angles of the DNA ends in addition to W, because the ends have to be aligned, and their twist angle has to be appropriate for the ends to be ligated [46]. Such probability densities depend on the mechanical properties of the polymer. Hence the J factor can be computed using mechanical polymer models, for instance with Monte Carlo methods [5, 47] or rod models [1]. The J factor can be interpreted as the effective concentration of one end of the polymer in the vicinity of the other end, with appropriate orientations and twist angle. It can be understood as a measure of the ability of a polymer to bend into a circular polymer.

Shore, Langowski and Baldwin showed that in the ligase-catalyzed assay they proposed, under certain conditions,

$$\frac{K_C}{K_D} = \frac{k_1}{k_2}, \tag{2}$$

with k_1 and k_2 being the rate constants associated to the irreversible reactions of covalent cyclization and bimolecular association, respectively [8, 9, 10]. The rate constants k_1 and k_2 are derived by measuring the products of the cyclization reaction at different time points in a gel electrophoresis image. The cyclization reaction is therefore a useful experiment to validate mathematical models of DNA and associated parameters. In the third chapter, I propose a ligase-free cyclization experiment with quasi-irreversible cyclization and bimolecular associations, and I estimate the ratio k_1/k_2 using relative concentrations of the reaction products measured in gel electrophoresis images and a reaction rate model.

0.4 Direct visualization of DNA minicircles by Cryo-Electron Microscopy

DNA minicircles have mostly been used in cyclization assays. Finding DNA parameters by this method requires extensive bench work since it should be done for molecules with different base-pair lengths for each sequence of interest. The output of a series of cyclization reactions is essentially one number, the J factor. But the shape of the minicircles, if correctly determined, contain more information about the mechanical properties of the minicircle than a single number. Since we can compute the minimal energy shape of the minicircle, the experimental determination of the minicircle shape could be extremely valuable for validation of the theory.

Cryo-Electron microscopy allows the observation of DNA molecules in nearly physiological conditions; thin aqueous layers containing suspended DNA molecules are rapidly cooled and cryo-vitrified at such speed that ice crystals do not form [48, 49, 50]. The frozen sample can be tilted, and one can obtain micrographs of individual DNA molecules visible from two different angles of view. This method has been used to reconstruct the three-dimensional path of individual DNA molecules [51], and to determine DNA persistence length [21]. In the first Chapter of the thesis, I analyzed the shapes of DNA minicircles with two different sequences, determined from tilted pairs of cryo-electron micrographs, using a program specially designed for 3D reconstruction of DNA minicircles [52]. In the second chapter, I analyze shapes of DNA minicircles obtained from electron microscopy images and I compare them with the result of molecular dynamics simulations of minicircles.

Chapter 1

Sequence Dependent Shape of DNA Minicircles Obtained by Cryo-Electron Microscopy

1.1 Introduction

The main question this first chapter aims at answering is whether or not it is possible to detect a sequence effect on the shape of DNA minicircles determined from stereo images of cryo-electron microscopy. In this chapter I used cryo-electron stereo micrographs produced by Jan Bednar at LAU, UNIL, to compare 3D shapes of 158 base-pair (bp) long DNA minicircles that differ only in the sequence within an 18 bp block containing either the regulatory sequence TATA box or a binding site for the CAP protein. I present a sorting algorithm that correlates the reconstructed shapes and groups them into distinct categories. The study described in this chapter was published in [53].

1.1.1 The TATA box

The TATA box is a short sequence in the promoter of genes that binds protein complexes and initiates transcription. Cyclization experiments showed that free DNA in solution containing a TATA box sequence exhibits greatly enhanced J factors that can be attributed to strong bends and high flexibility [54]. It was shown that prebending of DNA enhances its interaction with TATA Box Binding Protein (TBP) [55]. As there are at most a few direct

hydrogen bonds between DNA and the TBP [56, 57], it is thought that the mechanical properties of the TATA box are probably very important for its function [54]. This hypothesis is supported by a recent all-atom computation that predicts mostly indirect recognition between TBP and the TATA box [58].

Studies of DNA cyclization using DNA between 147 and 163 bp in length provided strong indications that the TATA box is highly flexible [54]. However, the exact nature of this flexibility is not known. The sequence could behave as a kink and permit high local bending. On the other hand the bending abilities of the TATA box could be approximately limited to the curvature expected for a 158 bp long DNA circle. A flexible kink should perturb the shape of the minicircle and be easily visible on cryo-EM images. However, if the bending is limited to a curvature similar to the one in the relaxed minicircle, it might not affect the shape of observed minicircles. In this study I observe and compare shapes of DNA minicircles of length 158 bp in which an 18 bp fragment contains either a TATA box or a CAP (CRP) site.

1.1.2 3D Reconstruction of DNA minicircles

To observe how DNA shape is influenced by its sequence, it is advantageous to minimize variations due to thermal fluctuations and to visualize molecules as close as possible to their minimal energy shape. DNA minicircles seem to be best suited for this purpose. Because of their short length (close to the persistence length) the closure constraint of minicircles effectively limits the range of possible fluctuations.

On the other hand, the small size of the minicircles (about 17 nm in diameter) implies that even nanometer-size errors in the 3D reconstruction procedure significantly affects the reconstructed shapes. It is therefore desirable to use specialized software that can reconstruct the filaments with sub-pixel resolution [52].

As I wish to study sequence-dependent effects, it would be advantageous to know which point in our reconstructed center lines corresponds to which base pair of the sequence, which would require the use of a molecular marker. However the intrinsic DNA shape can be altered by protein-DNA binding or by binding of specific chemicals that can be used to map specific sequences.

For this reason I did not attempt such an approach, and instead visualize totally naked DNA.

1.2 Material and methods

1.2.1 DNA constructs

The two DNA minicircles constructs were provided by my collaborator Jason Kahn (University of Maryland): t11T15 and c11T15 [54, 43]. The two minicircles are 158 bp long, and their sequences differ by 14 bp (see figure 1.1). The TATA box site in t11T15 is replaced by a CAP (CRP) binding site in c11T15 [59]. Kahn and collaborators measured the cyclization rate of t11T15 and c11T15 sequences and determined their J factors [8, 9, 43]. The J factor of the t11T15 minicircle is around 3500 nM whereas the J factor of c11T15 is 95 nM. The t11T15 and c11T15 fragments will be referred to as TATA and CAP respectively throughout the text.

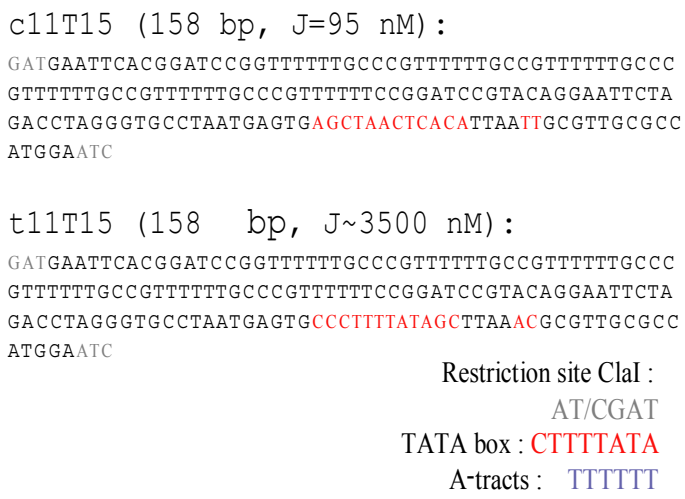

Figure 1.1: The sequences of the studied minicircles. The differences between the two sequences are highlighted in red.

1.2.2 Cryo-Electron Microscopy

The cryo-electron micrographs were produced by Jan Bednar at LAU, UNIL. The DNA minicircles were immobilized in a 50 nm layer of vitreous ice, at a temperature of $-170°C$. Images were taken at the magnification of 53000x and registered on Kodak EM negative plates. The negatives were scanned at 1800 dpi, 8 bit gray-scale. The first image is taken with the sample tilted by -15°and the second at +15°. The tilt axis is vertical in both images presented in figure 1.2.

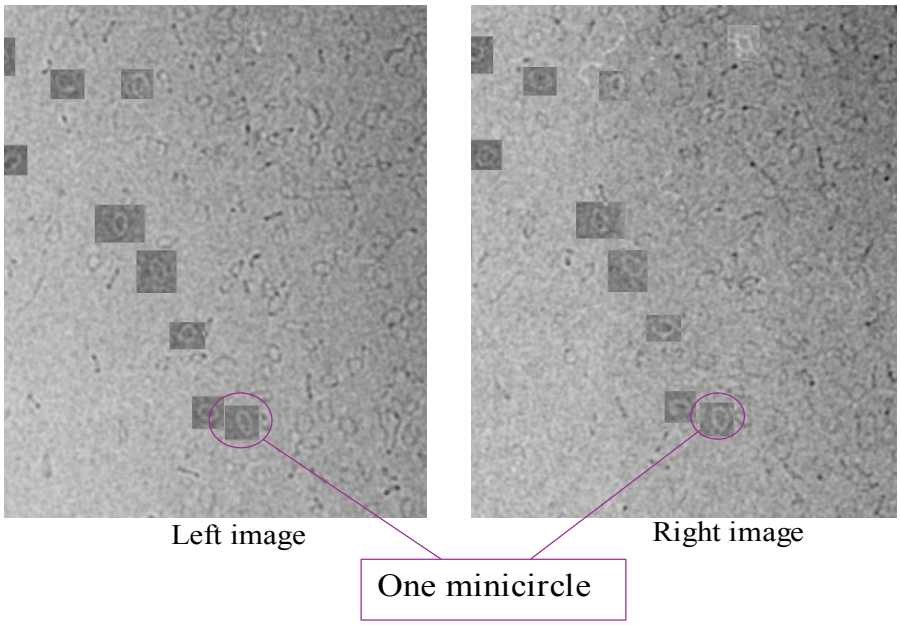

Figure 1.2: Regions of a pair of stereo micrographs. Stereo pairs of minicircles used for reconstruction are indicated with negative color frames.

1.2.3 Three-dimensional reconstruction of DNA

My colleague Cédric Vaillant and I used the software package developed by Jacob et al. [52] to reconstruct the DNA minicircles shapes from the cryo-electron micrographs. For each minicircle the user traces an initial

approximation of the visible DNA path on the two images. A smoothing filter of the images aids in this initial tracing. Our study is blind in the sense that the user does not know the sequence (TATA or CAP) of the minicircle in order to avoid bias in initial path tracing. Given this initial estimate, the program then performs the reconstructions by assuming a 3D curve model. The shape of the curve is optimized such that its 2D projections onto the micrograph planes match with the signals in the two images. The reconstructed curves are output in a list of points expressed in 3D Euclidean space. I re-sampled (using the spline function of Matlab) the output curves with a cubic spline to have 200 points per minicircle, equally spaced within one curve. I then analyzed the shape of 64 reconstructions of TATA and 31 of CAP.

1.2.4 Shape Analysis and Visualization

The main part of the code for data analysis was written in Matlab, with some Python scripts. Methods are described together with results in the next section. 3D pictures were produced with VMD [44].

1.3 Results and Discussion

1.3.1 Curvature

Given three consecutive points A, B and C on a discrete curve, the curvature at B can be approximated by the inverse of the radius of the circle that goes through A, B and C. A kink should induce high curvature in a short portion of the minicircle double-helix axis. I analyzed the distribution of such curvature values in the reconstructed minicircles. Each minicircle provided 200 entries for the curvature measured at each of the 200 indexed points. The curvature distributions of the points belonging to the TATA circles and to the CAP circles are computed separately and compared in figure 1.3. For both sequences, the curvature distribution is peaked; the maximum corresponds to the curvature of a 158 bp perfect circle (0.12 nm^{-1}). The distribution of curvature is very similar for both TATA and CAP. (Figure 1.3). Note that the shape data have no reference that indicates the location of the TATA box or CAP (CRP) site sequences.

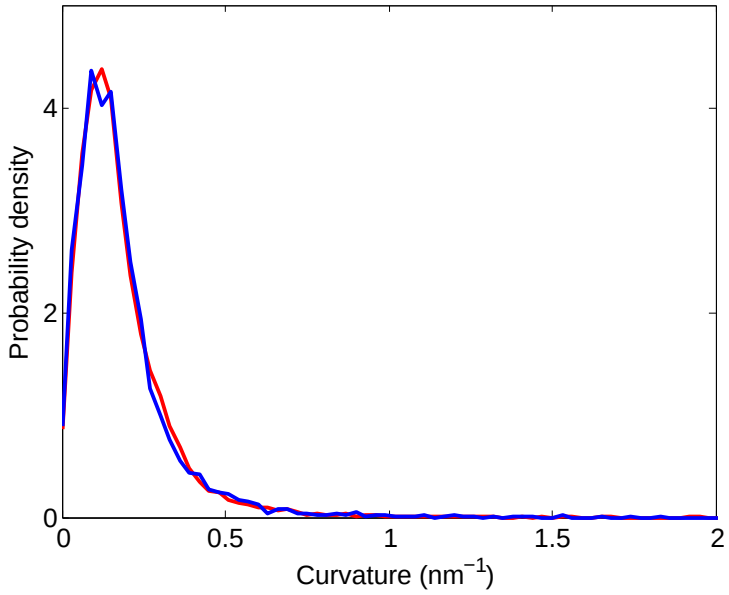

Figure 1.3: Probability density function of curvature in reconstructed TATA and CAP minicircles (the corresponding profiles are red and blue, respectively). The function is approximated by a normalized histogram counting curvature values with intervals of 0.03 nm^{-1}.

1.3.2 Superposition of DNA minicircles shapes along their principal axes of inertia.

Figure 1.4 shows axial paths of reconstructed minicircles that have been translated and rotated so that their center of mass (assuming uniform mass density), and their principal axes of inertia coincide. Such a presentation allows us to visually compare many minicircles shapes at the same time. The resulting picture does not show a clear difference between the shapes of TATA (red) and CAP (blue) minicircles.

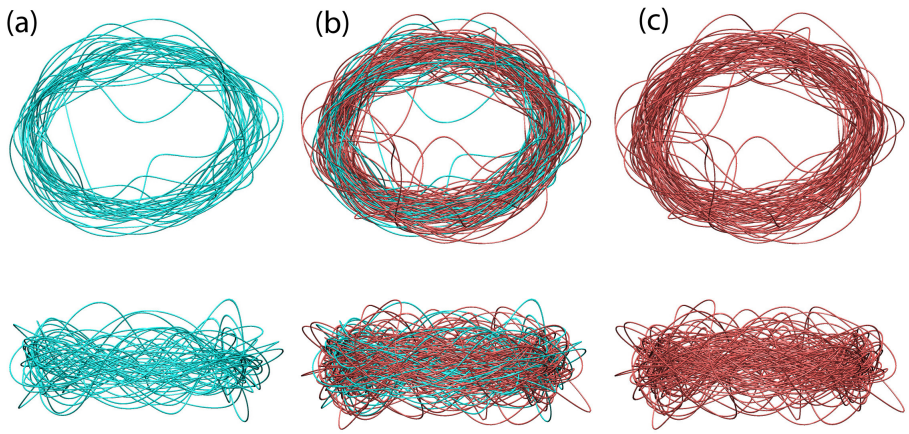

Figure 1.4: All the reconstructed shapes of DNA are aligned by superposition of their principal axes of inertia. The upper and lower views differ by a rotation of 90°around the horizontal axis: **(a)** all the 31 CAP minicircles (blue), **(b)** all the 95 minicircles and **(c)** all the 64 TATA minicircles (red).

1.3.3 Shape-Distance for curves: Minimum RMSD over all rigid-body motions, index shifts and curve orientations

Although curvature analysis and visualization did not reveal the presence of a kink in TATA in comparison to CAP minicircles, there may be a more subtle sequence-dependent shape pattern. Therefore, rather than looking for a particular shape, I designed a method to identify groups of similar shapes, and looked whether the sequence correlates with the groups or not. I first chose a distance for the determination of shape similarity, then I clustered the shapes according to their mutual similarities measured in terms of this distance.

The distance function has to be as insensitive as possible to the noise in the original image from which the reconstructions come from. For instance, curvature along arc-length is inappropriate because of the noise level in the micrographs. Indeed, because of the low contrast and the noise in the image, curvature can vary a lot between two possible reconstructions of one stereo pair, even if the fit between the reconstructions and the image pair is good. Curvature is too sensitive to the noise in the image. Therefore, I did not use curvature but rather a distance that is robust to small and local variations of the curve.

Because I do not know the correspondence between the sequence and

the curve in each image, in order to estimate the similarity between two minicircle shapes I need to adapt the standard Root Mean Square Deviation (RMSD) minimization procedure that is often used to compare the geometries of two solid objects. The standard method is as follows: For two ordered sets of N points $\mathbf{x}$ and $\mathbf{y}$, RMSD is the square root of the average over i of the squares of the Euclidean distances between two corresponding points $\mathbf{x_i}$ and $\mathbf{y_i}$. Then, to eliminate rigid-body motions, one computes a 3×3 rotation matrix $\mathscr{U}$ and a translation vector $\mathbf{r}$ which, when applied to $\mathbf{x}$, minimizes the RMSD function defined in equation (1.1), producing the best superposition of the two structures:

$$\min_{\mathscr{U},\mathbf{r}} \left[\frac{1}{N} \sum_{i=1}^{N} |\mathscr{U}\,\mathbf{x}_i + \mathbf{r} - \mathbf{y}_i|^2 \right]^{\frac{1}{2}}. \tag{1.1}$$

A fortran 95 code given in [60] was used to compute this minimum RMSD.

Our shape-distance function is then defined in (1.2) via minimization over all possible rigid-body rotations and translations in 3-D, plus further minimizations in all shifts of an index (the variable δ), and two curve orientations, clockwise or counter-clockwise (the variable α):

$$\min_{\mathscr{U},\mathbf{r},\delta,\alpha} \left[\frac{1}{N} \sum_{i=1}^{N} |\mathscr{U}\,\mathbf{x}_i + \mathbf{r} - \mathbf{y}_{(-1)^\alpha i+\delta}|^2 \right]^{\frac{1}{2}}, \quad \delta \in \{1,\ldots,N\},\ \alpha \in \{1,2\}. \tag{1.2}$$

The additional minimization over δ is necessary in our case because I do not know which point of the discretized curve $\mathbf{y}$ should correspond to the first point of the curve $\mathbf{x}$. However, if there is a common pattern between shapes of minicircles, a particular mapping of $\mathbf{x}$ onto $\mathbf{y}$ should give a minimal RMSD. The minimization over δ in equation (1.2) allows all possible phasing differences in index to compete in the fit. Minimization over α recognizes that a given curve can be discretized with two distinct orientations. Except for particular symmetrical shapes, identical curves that happen to be discretized with opposite orientations cannot be perfectly superposed by standard RMSD. As a matter of implementation the additional minimizations in (1.2) were achieved by calling the RMSD function given in [60] inside a Matlab loop for all possible shifts δ ($\delta = 1 \ldots 200$ in our data with the index of $\mathbf{y}$ to be understood *modulo* 200), and the two choices of α. The smallest RMSD value found in the loop defines the distance between the two shapes.

1.3.4 Error of Reconstruction Measurements

To measure similarity or dissimilarity between different reconstructed minicircles it is important to determine the error of reconstruction and to see how much this error could affect the comparison between different reconstructed minicircles. In order to estimate the reconstruction error, I applied the distance function to two reconstructed shapes coming from the same image pair, but obtained by two different users of the reconstruction program. I computed the user error for six image pairs (see figure 1.5). I find that the average error is 0.9 nm, with standard deviation 0.3 nm.

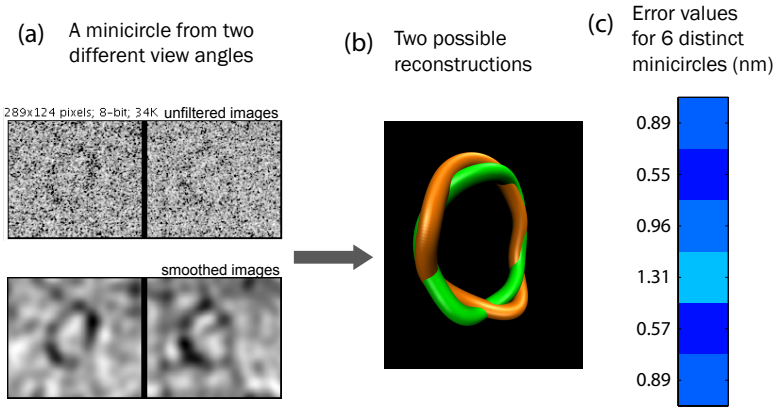

Figure 1.5: Estimation of the error of reconstruction. **(a)** The same DNA minicircle is shown from two different angles. In the right image, the sample is rotated by 30°around the vertical axis with respect to the left image. **(b)** Two reconstructions from the stereo-pair in (a) starting from two different user initializations. The shape-distance (3) between the two reconstructions is 0.89 nm. **(c)** Error values, i.e. shape-distances between two reconstructions of each of 6 stereo-pairs. To allow comparison with shape-distances between different data, the error values are expressed with the color code of figure 1.7.

I also estimated some differences induced by changing the reconstruction program (DNAJ) parameters. For instance, DNAJ can constrain the resulting curves to be of a particular length, because the length of the observed minicircles is known. I used unconstrained reconstruction for analysis because the user error does not decrease when the length of the reconstruction is constrained. In order to compute the length constraint error, I modified

the source code of DNAJ in order to automatize reconstructions for different parameters of DNAJ. I found that the constraint does not affect the reconstruction a great deal; I observe a shape-distance of 0.40 nm in average between a constrained and an unconstrained reconstruction of the same minicircle, with the same user input (length constraint error was measured on the 95 shapes).

1.3.5 Analysis of shape-distances with respect to TATA and CAP sequences

I analyzed a set of 95 distinct minicircles (64 TATA, 31 CAP) all reconstructed by the same user. I therefore had a set of 4465 (or 95*94/2) pairwise shape-distances. Figure 1.6 gives the normalized histograms, i.e. probability distributions of pairwise distances in three groups: TATA to TATA, CAP to CAP and TATA to CAP. The average shape-distances are: 2.03 nm for TATA-TATA (SD[1] 0.57 nm), 1.96 nm for CAP-CAP (SD 0.52 nm) and 1.98 nm for TATA-CAP (SD 0.55 nm). TATA-TATA and CAP-CAP shape-distances are not significantly smaller than TATA-CAP distances. Therefore I do not observe increased shape similarity between minicircles with the same sequence.

1.3.6 Shape Clustering

I cannot use classical methods for clustering our shapes, as there is not a sensible way to represent them as vectors in a multidimensional space. I also do not have reference shapes to build clusters. Accordingly I adopt the reference-free SPIN algorithm [61] that is capable of ordering elements of a set using only their pairwise distances. For an ordered list of shapes and a shape-distance function, there exists a unique shape-distance matrix defined as follows: each element (i, j) of the matrix is the shape-distance between minicircles i and j. By definition, the matrix is symmetric and the elements on the diagonal vanish; the i^{th} line (or column) is a list of the distances between minicircle i and all others. SPIN finds a permutation of an initial ordered list of shapes that minimizes the elements near the diagonal. If the resulting matrix has a block of low (dark blue) values near the diagonal, with comparatively higher values above and below (and therefore necessarily

[1] Standard Deviation

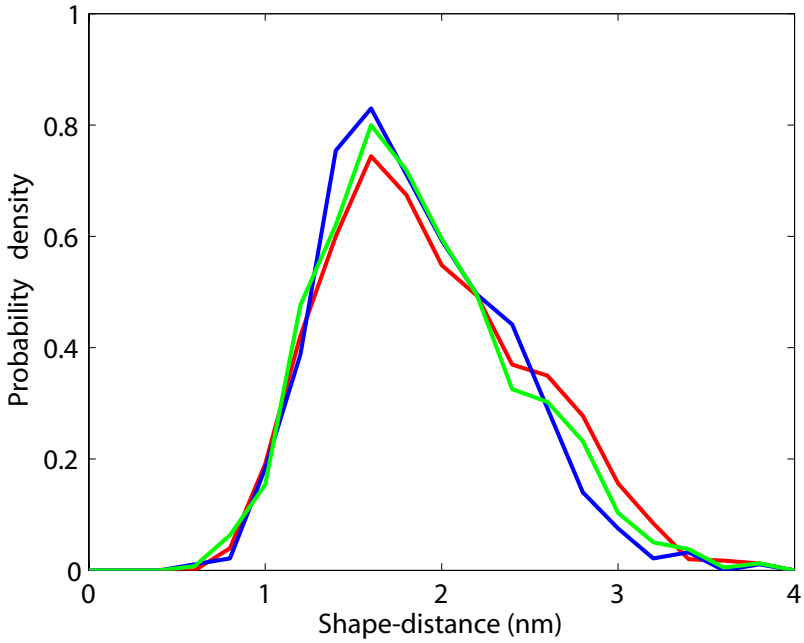

Figure 1.6: Normalized histograms, i.e. probability density, of the shape-distance values between any two TATA minicircles reconstructed shapes (red), any two CAP shapes (blue), and between one TATA shape and one CAP shape (green).

by symmetry to right and left), the shapes in the block can be considered as clusters. A SPIN sorted shape-distance matrix and the corresponding clusters are shown in figure 1.7.

Three columns were added on the left of the matrix. They show some properties of the shapes. Each line and each column of the matrix correspond to a minicircle. For each line i of the matrix, the corresponding element i of the column "Minicircle type" shows whether the corresponding minicircle i is of type TATA (gray) or CAP (white). It is clear that the TATA and CAP minicircles are spread throughout each cluster. Similarly, the i^{th} element of the column "Circle" (respectively "Ellipse") shows the distance between minicircle i and a circle (respectively an ellipse). The circle diameter is 17.1

nm (corresponding to a perimeter of 158 bp). The longer ellipse axis is also 17.1 nm while the shorter axis is 13.7 nm. These two columns and the lower part of figure 1.7 suggest that the method was able to identify clusters of circular and ellipsoid shapes, and to find another non-planar cluster. Stereo images of the cluster 7-15 are presented in figure 1.8.

Interestingly, the distance matrix apparently reveals presence of multiple clusters of shapes. It is known that DNA circles with non uniform sequence have multiple local energy minima [62]. For this reason, I believe that our clustering analysis detected sampling of at least two and possibly more energy wells in the configuration space. However, the small difference between the majority of the clusters (comparable with the error of the reconstruction method) warns against over-interpretation of the distance matrix data. Importantly, each detected cluster contains both TATA and CAP minicircles, so that the different clusters seem to be associated with the sequence-dependent features that are shared between the two sequences, e.g. the 6 phased A-tracts, rather than the differences between TATA and CAP sequences. I therefore conclude that TATA and CAP sequences produce minicircles with similar 3D shapes.

1.3.7 Conclusion

Using cryo-electron microscopy I have investigated the effect on the 3D shape of 158 bp long DNA minicircles with identical sequences except for the interchange of TATA and CAP boxes. Although, the TATA minicircles cyclize two orders of magnitude more efficiently than CAP in ligation experiments, I did not detect significant differences in the observed 3D shapes. Analysis of the reconstruction errors revealed that the average user error (0.9 nm) was two times smaller than the average shape-distance between two minicircles (2 nm). I conclude, therefore, that thermal fluctuations "blur" the possible differences in 3D shapes of DNA minicircles induced by the presence of CAP or TATA sequences.

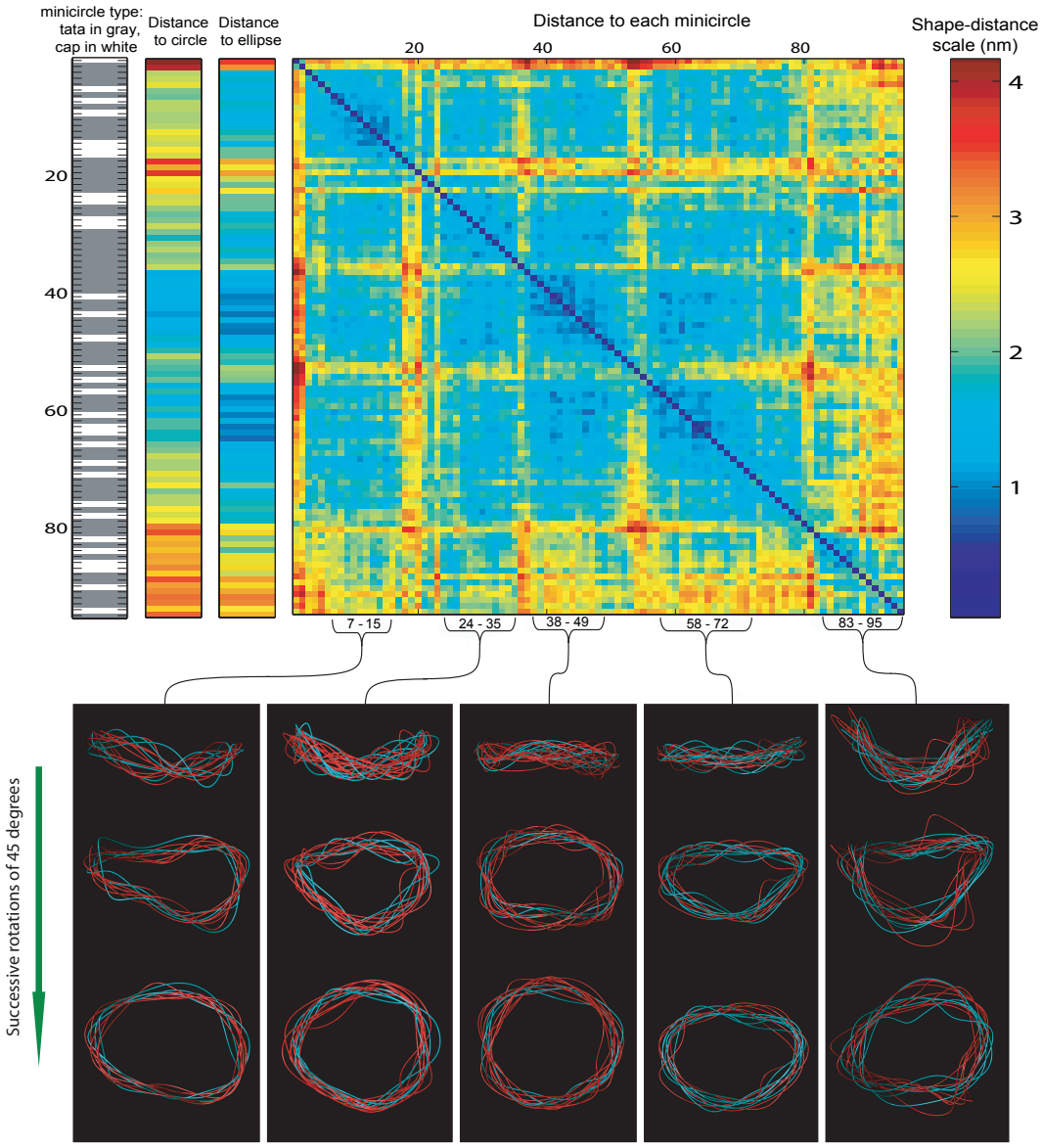

Figure 1.7: **(top)** The Shape-Distance Matrix after clustering. Each line of the figure contains information about one minicircle reconstructed shape: the sequence type (first column), the shape-distance between the shape and a perfect circle (second column), between the shape and an ellipse (third column), and between the shape and every other reconstructed shape in the set (matrix). Shape-distance values are represented by colors with scale in nanometers shown on the right. **(bottom)** Different views of the clusters are produced by successive rotations of 45°around the horizontal axis. The indices of the minicircles shown are indicated with the brackets below the matrix. TATA (resp. CAP) minicircles are colored in red (resp. blue).

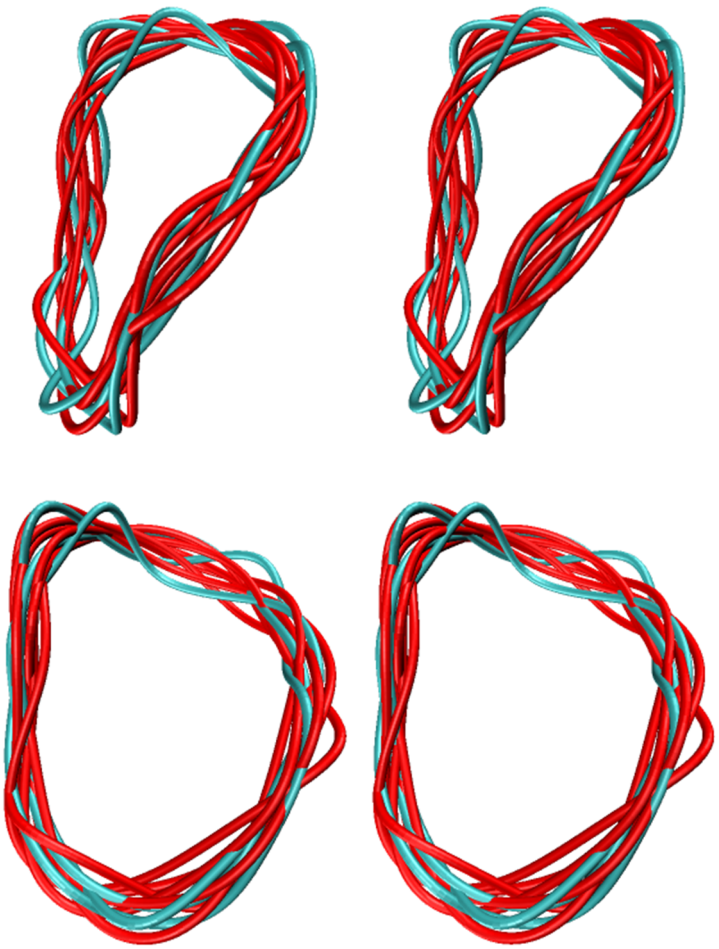

Figure 1.8: Stereo images of the cluster 7-15. Images are presented in "side by side" stereo mode.

Chapter 2

Effect of Nicks and Gaps on DNA Minicircle Shapes

2.1 Introduction

In this Chapter, I report (i) the analysis of shapes of 94 bp DNA minicircles containing nicks and gaps, obtained by cryo-electron microscopy and negative staining electron microscopy (images by Davide Demurtas at UNIL); (ii) the analysis of molecular dynamics simulations of (almost) the same minicircles (simulations performed by Leesa Heffler and Filip Lankas at EPFL). The aim of this study is to answer the same question as in Chapter 1, namely whether or not it is possible to detect a difference in shape due to the chemical composition of the minicircles, but with consideration of sequence effects replaced by comparison between two oligomers with either nicks or gaps. Now the answer is in the positive, so I also compare simulated minicircles to observed ones, and present here an atomic level description of the nick site and the gap sites in the simulated minicircles.

2.1.1 DNA Nicks and Gaps

In a single DNA strand, the sugar phosphate backbone links the bases together. In double-stranded DNA, it can happen that one of the two backbones is broken. This break is called a "nick". Sometimes there is not only a break, but also one or more nucleotides missing along one strand, this is called a "gap".

In cells, DNA strand breaking happens about 4000 times each day, resulting in nicks [63]. Nicks are detected by DNA repair proteins [64]. DNA repair enzymes can remove damaged bases on one strand of the double helix, producing gaps [65]. Also, during genetic recombination, the two DNA strands of a chromosome are nicked, and large pieces of DNA are cut and reinserted in various places in the genome; in DNA replication, the Okasaki fragments are not covalently linked until processing by DNA ligase I, and therefore genomic DNA carries nicks at the replication fork [66]. In the field of bioengineering, bacterial enzymes are used to perform DNA restriction reactions (for cloning purposes for instance); the enzymes create nicks in both DNA strands to cut the molecule in two parts. DNA oligomers are also used in nanotechnology because of the possibility to design specific interactions between DNA molecules. Self-assembled DNA structures inspired from natural DNA recombination are made of multiple short complementary DNA oligomers and therefore carry nicks or gaps. The nicking of DNA molecules is considered as "central to the generation of new motifs with potential nanotechnological utility" [67]. Recently Frank-Kamenetskii and collaborators used nicked and gapped DNA molecules in order to investigate the energetics of base pair stacking independent of base pairing [68, 69]. A description of nick dynamics at the atomic level may therefore be useful to understand the underlying mechanisms of phenomena involving DNA nicks *in vivo* or in the context of DNA nanotechnology.

2.1.2 The 94 base-pair DNA Minicircles

The sequences of the analyzed minicircles are the same as in a recent study where minicircles as short as 94 base pairs were reported to be able to cyclize with J factors between 3 and 5 orders of magnitudes higher than predicted by the current theories [70]. This result stimulated revision of the theories of cyclization [71, 72], but was also received with criticism [73] (the criticisms by Vologodskii and collaborators are explained in section 3.1.1). The sequence used here is the best cyclizer (i.e. the sequence with the highest J factor) of the study by Cloutier and Widom [70], which was obtained by an artificial selection assay (SELEX) for the best cyclizing sequences. It was therefore an interesting minicircle to observe under the electron microscope, as it might shed light on the mystery of its surprisingly high J factor. A minicircle with high J factor is attractive for observation by cryo-electron

microscopy also because this method demands high concentrations of the specimens of interest; thus sequences which allow efficient minicircle formation were of particular interest. The method that was used to produce the minicircles is called annealing-cyclization, which is the subject of the third chapter of this thesis.

The DNA minicircles observed by electron microscopy carried either two nicks at diametrically opposite sites, or two gaps of two nucleotides length at almost diametrically opposite sites (the nicked minicircle in simulations carries only one nick). Hence, contrary to the minicircles studied in Chapter 1, the observed minicircles have differences at two sites in the sequence instead of one site only, and the difference can be expected to be more significant as it involves a defect, not only a difference in sequence. The specific position of the nick or gap is determined by the length of the sticky ends in the linear DNA molecule which formed the minicircle during the annealing-cyclization reaction. When the length of the sticky ends is half of the total length (47 bp), the nicks are at diametrically opposite sites. This sticky-end length proved to increase the minicircle formation yield substantially, as demonstrated in Chapter 3. Furthermore, the positions of the nicks and the gaps were chosen such that it may increase the chances of detecting a difference between the minicircle shapes. Indeed, if one imagines a simple DNA model in which the double helix is straight when unconstrained and stiff, and in which the gaps are very flexible points, then the DNA fragments from each side of the diametrically opposed gaps, which have the same length, should straighten and move close to each other; the DNA helical axis at the gap sites would then be strongly curved. However if the gaps are not at diametrically opposed sites, one part of the minicircle will be curved because of the minicircle closure constraint. Therefore putting two gaps at diametrically opposite sites of the minicircle is likely to induce a shape which is markedly different from the ideal circular shape, namely closer to a narrow ellipse. This is indeed what we see in both observed and simulated minicircles. A quantitative analysis of the shapes is presented in this chapter.

2.2 Methods

2.2.1 Sequences, nick and gap positions

I describe here the sequence of both the observed and simulated DNA minicircles; except for the insertion of nicks and gaps, the sequences of all the minicircles are identical (Figure 2.1).

The two simulated minicircles are called Sn1 (for simulated minicircle with 1 nick) and Sg2 (for simulated minicircle with 2 gaps). The two types of minicircle observed by electron microscopy are called On2 (for observed minicircle with 2 nicks) and Og2 (for observed minicircle with 2 gaps). There are two differences between Sn1 and On2: On2 has an additional nick with respect to Sn1, and Sn1 does not have a phosphate group at the nick site. The reason for this difference is that my collaborators at LAU originally planned to seal the minicircle nicks by ligation in order to compare the shapes with existing simulations of the covalently closed minicircle [74]; therefore the oligonucleotides carried phosphate groups at the 3' end of the oligonucleotides. A minicircle with one nick only could have been produced by ligation of DNA minicircles composed of two complementary oligonucleotides, one carrying a phosphate group at the 5' end and one not. However, it was difficult to clarify whether or not the ligation completed because it was difficult to detect the presence of nicks in DNA minicircles by experimental methods; hence only the minicircles with unsealed nicks were used for imaging. On2 has two nicks at diametrically opposite sites in the sequence. They are separated by a distance of 47 base pairs in both directions along the double helix axis. Sg2 and Og2 are chemically identical. Each strand carries a gap of two nucleotides in place of the nicks of On2: specifically each strand has two nucleotides missing from the 3' side of the nick, with respect to On2. The fact that the nucleotides were removed from the 3' end in both strands means that the gap sites are not exactly at diametrically opposites sites: they are separated by a distance of 43 base pairs in one direction, and 47 base pairs in the other. Images of minicircles with two gaps at exactly diametrically opposite sites are currently being prepared, but the images are not available at the time of this thesis writing.

Sn1
strand 1:
GGCCGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGT
CTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAC
strand 2:
GTAGA **(')** GACTAGTAAGTAATCCTATTGGCGGTTAAAACGCGGTAGACAGCG
CGTACGTGCGTTTAAGCGGTGCTAGAGCTTGCTACGACCCGGCC

On2
strand 1:
GGCCGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTA**(P')**CGCGCTGT
CTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAC
strand 2:
GTAGA**(P')**GACTAGTAAGTAATCCTATTGGCGGTTAAAACGCGGTAGACAGCG
CGTACGTGCGTTTAAGCGGTGCTAGAGCTTGCTACGACCCGGCC

Sg2 and Og2
strand 1:
GGCCGGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACG **(- -)** CGCGCTGT
CTACCGCGTTTTAACCGCCAATAGGATTACTTACTAGTCTCTAC
strand 2:
GTA **(- -)** GACTAGTAAGTAATCCTATTGGCGGTTAAAACGCGGTAGACAGCG
CGTACGTGCGTTTAAGCGGTGCTAGAGCTTGCTACGACCCGGCC

Figure 2.1: Nick and gap positions in the sequences of the observed minicircles and the simulated minicircles. Sequences are written from 5' to 3'. A nick is noted by a quote sign with brackets **(')**. The phosphate groups are missing on the 5' carbon at the nick except when the nick is noted by a **(P')**. Two missing nucleotides which create a gap are noted with two hyphens in brackets **(- -)**. **Sn1** Simulated minicircle with one nick. Both strands are made of 94 nucleotides. Strand 1 is circularly closed. Strand 2 has one nick after the fifth nucleotide. **On2** Observed minicircle with two nicks. Both strands are made of 94 nucleotides. Strand 1 has a nick after the 42^{nd} nucleotide, Strand 2 has a nick after the fifth nucleotide. Both nicks have phosphates groups. **Sg2 and Og2** Simulated and observed minicircles with two gaps. Both strands are made of 92 nucleotides. Strand 1 has the 41^{st} and the 42^{nd} nucleotides missing (T and A); this gap is called Gap 1. And Strand 2 has the fourth and the fifth nucleotides missing (G and A); this gap is called Gap 2.

2.2.2 Molecular Dynamics Methods

The molecular dynamics simulations were performed by Leesa Heffler and Filip Lankas at EPFL. The simulated minicircle was simulated in explicit solvent, meaning that all the water molecules were simulated explicitly. Initial atomic structure of the 94 base-pair minicircles is exactly the same as that used in [74] which was generated with JUMNA [75]. The nick was generated by removal of a phosphate group in one of the DNA backbones. The gaps were generated by removal of two successive nucleotides with their phosphate groups. Simulations were run with the program AMBER 7, using the PARM 94 force field (see reference [74] for the details of the procedure). The helical axis of the minicircles were computed from the simulated structures with the program CURVES [76, 77].

2.2.3 Electron Microscopy Methods

The preparation, microscopy imaging and shape acquisition of the minicircles were realized by Davide Demurtas.

Minicircle preparation for imaging The minicircles were prepared by mixing two complementary synthetic oligonucleotides in STE (100 nM NaCl), heated at 95°C and slowly cooled to room temperature. The DNA oligos were designed such that we obtain circular products with nicks or gaps at the appropriate position; the yield of formation of On2 and Og2 is discussed in Chapter 3, and the design of the oligos is described in Figures 3.1a and 3.1e.

Electron Microscopy Images were taken at magnifications of 35000x or 45000x, with a 1000x1000 CDD camera, which correspond to a pixel size of 0.52 nm and 0.40 nm, respectively. For the cryo-electron microscopy procedure, the DNA minicircles were immobilized in a 20 to 30 nm layer of vitreous ice, at a temperature below -170°C. In the negative staining procedure, the minicircles were adsorbed on a carbon surface, and 2% uranil acetate was added to the sample after adsorbtion.

3D reconstruction and 2D shape acquisition 3D reconstructions (from cryo-electron microscopy images) were performed as described in the first Chapter of this thesis, with the program DNAJ [52]. For the 2D shapes

(from negative staining images), the 3D reconstruction procedure was applied to a pair of identical minicircle images with a tilt angle set to 10^{-5} degrees.

2.2.4 Analysis of the DNA minicircle shapes with the principal axes of inertia (PAI)

The 2D minicircle shapes from negative staining images and the 3D shapes from simulations were aligned on their principal axes of inertia (PAI), as was done with the 158 bp minicircles of Chapter 1. In order to compare the simulated minicircle shapes to the shapes obtained by electron microscopy, I used the DNA helical axis of the atomic structure given by the program CURVES [76, 77]. The axes ratios were computed for all the shapes. The axis ratio of a minicircle shape is defined as the long axis divided by the short axis. The long (respectively short) axis length is the length of the projection on the first (respectively second) PAI. Since the 3D helical axes that come from the simulations are not far from planar, the third PAI of the 3D shapes were not considered, and the axis ratio, which is defined with respect to the two first PAI, was used for the 3D shapes as well as for the 2D shapes.

The results of the simulations were visualized with VMD [44], at a rate of 1 frame each 100 picoseconds (ps) of simulation, except for the analysis of the fast sugar rotations at the nick site, which was visualized at 1 frame per ps (movies are available at http://lcvmwww.epfl.ch/∼lcvm/articles/T08/info.html).

2.3 Shapes of Nicked and Gapped Minicircles

2.3.1 3D Reconstructions from Cryo-Electron Microscopy

In figure 2.2, one reconstruction of Og2 (red) and one of On2 (blue) are drawn. They both have elliptic shapes, but Og2 is narrower in the sense that it has a larger axes ratio than On2. The cryo-electron microscopy data presented here is composed of only one pair of DNA minicircles because it is very difficult to realize tilted pairs of minicircle images, for the reasons mentioned below. The cryo-electron microscopy method demands large amounts of the molecule of interest because approximately nine tenths of the sample solution is absorbed on a paper, and therefore lost, in order to form a very thin layer of water (between 20 and 30 nm in the images used for this

study) in the carbon holes of the grid. Furthermore the 94 bp minicircles have low J factors because of their short length, so they have a poor formation efficiency. In addition, DNA is not visible in the "search mode" of the electron microscope, and images have to be taken without knowing if the 500x500 nm^2 frame of the camera contains minicircle images, whereas the micrographs used in Chapter 1 cover more than a micron square. Finally, if minicircles are visible in the first image, the sample has to be tilted in order to take the second photograph; during the tilt, it often happens that the photographed minicircles moved out of the visible frame and are absent from the second tilted image. For all these reasons, the technique of negative staining was used to obtain 40 two-dimensional minicircle shapes.

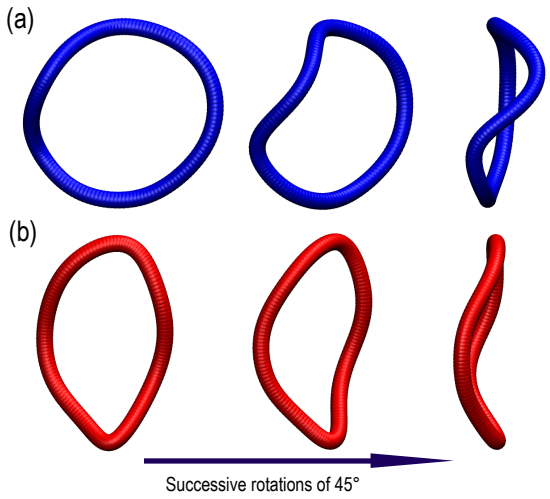

Figure 2.2: Two 3D reconstructions from cryo-electron microscopy images. Each 3D reconstruction comes from a single tilted image pair, and is represented from 3 different angles of view, separated by successive rotations of 45°around the vertical axis of the image. In **(a)** a reconstruction of On2 is drawn in blue and in **(b)** a reconstruction of Og2 is drawn in red. The reconstruction of Og2 has a higher axis ratio than the reconstruction of On2. A single pair of minicircle reconstructions is presented here because pairs of tilted micrographs of such short minicircles are very difficult to obtain (see text). However 20 two-dimensional images of each minicircle type have been obtained by negative staining electron microscopy, and the analysis of those shapes is presented in Section 2.3.2.

2.3.2 2D Shapes from Negative Staining

The negative staining technique gives a better contrast than cryo-electron microscopy. Contrary to the cryo technique, in negative staining the minicircles are adsorbed on a carbon surface, and thus constrained to lie in a plane which is parallel to the image plane. Although it does not allow observation of the minicircle 3D shapes, this technique avoids the problems related to the realization of tilted images, and my collaborator Demurtas could collect 40 shapes from his electron micrographs. The results presented in Chapter 1 using cryo-electron microscopy suggests that many of the 158 bp minicircles are mostly planar in solution; indeed, only the fifth cluster of shapes bends significantly out of the plane (Figure 1.7). The two minicircles which were reconstructed and presented in Figure 2.2 are relatively planar as well. For such minicircles, it is possible that the shape of absorbed minicircles is relatively similar to the shape in solution, and that the constraint imposed by the carbon surface is on the orientation of the minicircles only, namely that the first and second principal axes of inertia of the minicircles lie in the plane of the carbon surface (which is also the image plane). In this case, the axes ratios measured in negative staining electron microscopy may be relatively close to the axes ratios of the minicircles which are planar in solution. Therefore I measured the minicircles axes ratios using the principal axes of inertia (see methods) to compare the adsorbed minicircles between them, and to compare the 2D adsorbed minicircle shapes and the 3D shapes of simulated minicircles.

Twenty images of On2 and twenty images of Og2 were collected. The shapes were aligned on their principal axis of inertia; the result is displayed in figure 2.3. There is a clear difference between the circles which carry two nicks and the circles with two gaps. This difference can be quantified by measuring the axis ratio, which corresponds to the ratio of height over width of the shapes drawn in figure 2.3. The mean axes ratio for Og2 is 1.70 (SD 0.26) whereas it is 1.29 (SD 0.18) for On2. This is a significant difference. The result shows that the minicircles which contains two gaps at almost diametrically opposite sites are narrower (have a greater axis ratio) than the minicircles of the same length without gaps but with nicks. It seems to confirm the hypothesis that the gaps are flexible points (at least more flexible than the nicks) along a stiff DNA double helix. Following this hypothesis, the

gaps should lie on the minicircle at the extrema of the first principal axis, i.e. at the rightmost and leftmost parts of the minicircles in figure 2.3. Although the position of nicks and gaps is not visible in the electron micrographs, this hypothesis can be checked for the simulated minicircles; the principal axis of inertia analysis was performed on the simulated minicircles and is presented in the next subsection.

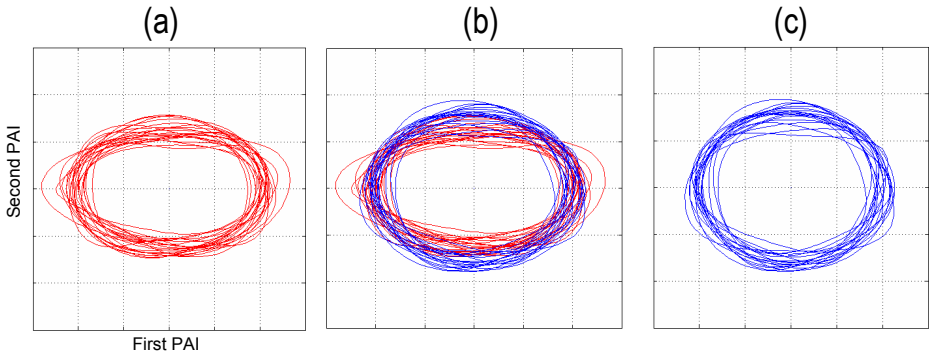

Figure 2.3: Comparison of 2D shapes of nicked and gapped DNA minicircles, from negative staining images. All the shapes are aligned on their principal axes of inertia. The gapped minicircles shapes (red) are more elliptical than the nicked minicircle shapes (blue), suggesting that the nicks or gaps lie at the most curved parts of the minicircles. **(a)** 20 Og2 shapes (two gaps) with mean axes ratio 1.70 (SD 0.26). **(b)** All 40 shapes (Og2 in red and On2 in blue). **(c)** 20 On2 shapes (two nicks) with mean axes ratio 1.29 (SD 0.18).

One signficant problem with the negative staining images is that the apparent length of the minicircle does not quite correspond to the expected length of a 94 base-pair B-DNA minicircle, which should be about 32 nm. Indeed, the mean apparent length of Og2 is 21.84 nm (SD 1.74 nm) and the mean apparent length of On2 is 22.82 nm (SD 1.28 nm). This phenomenon may be due to a transition of the minicircles DNA to the A form when adsorbed on the carbon surface: indeed, DNA on the carbon surface may have dried before or during the addition of uranyl acetate. DNA can turn to A form if the relative humidity is reduced to 75% ([11], p. 857). The rise between base pairs in A-DNA is around 0.26 nm ([16], p. 57), which predicts a length of 24.44 nm for a 94 bp fragment, which is much closer to the length observed in the images. Presumably, this transition should be uniform along the DNA minicircle, and the effects of gaps seem to be visible

even in the adsorbed minicircles. The axes ratios have the advantage to be dimensionless, and they compare well with the axes ratio of the simulated B-DNA minicircles in solution, as explained in the next subsection.

2.3.3 Helical Axis of the Double Helix in the Molecular Dynamics Simulation

Axes ratios In order to support the hypothesis and to gain insight on atomic structure, I analyzed the results of molecular dynamics simulations of the minicircles. Before analyzing the structural details of the simulated minicircles, I compared the simulation and the observation. As for the electron microscopy data presented in figure 2.3, I compared the two simulated minicircles Sn1 (which has one nick) and Sg2 (which has two gaps) by alignment on the principal axes of intertia, and I measured their axes ratio. For this purpose, I took a snapshot of each simulated minicircle every nanosecond from the 30^{th} nanosecond, so there are 70 shapes of Sn1 and 70 of Sg2. The result is presented in Figure 2.4. As in the electron microscopy data (Figure 2.3), the gapped minicircle shapes are elongated with respect to the nicked minicircle shapes. Indeed, the mean axes ratio for Sn1 is 1.35 (SD 0.14) and 1.76 (SD 0.19) for Sg2. This compares remarkably well with the experiment, since On2 axes ratio is 1.29 (0.18) and Og2 is 1.70 (SD 0.26).

Gap Positions An important difference here with electron microscopy is that the atomic level details of the simulation are available, thus the locations of the nicks and gaps are known. Therefore the nick and gap sites are plotted in addition to the minicircle shapes in Figure 2.4. We observe that the nick and gap sites indeed lie at the most curved part of the minicircles, in agreement with the simple model we described earlier, namely that double stranded DNA parts will tend to straighten and the gap sites will be strongly curved. Therefore I believe that gaps lie at the most curved parts of the minicircles observed by electron microscopy as well, and that it is the reason for the higher axes ratio observed in gapped minicircles.

PAI (see methods) and gap positions The two gaps do not lie exactly at diametrically opposite sites in the sequence (at a distance of 47 bp on one side, 43 bp on the other side). It is therefore not surprising that one of the gaps may not lie exactly on the first PAI, as is the case for Gap 1. In

addition, 17 minicircle shapes are flipped by 180°around the first PAI with respect to the 54 others minicircle shapes (one is flipped around the second PAI). Therefore Gap 1 (green circles) are more or less symmetrically distributed from each side of the first PAI. Flipped or not, the minicircle shapes are still aligned on the same PAI; thus the flips do not affect the values of the axes ratios. It is interesting to have the simulated minicircles aligned on the PAI rather than aligned by RMSD fit, because it is the procedure that was applied to the shapes obtain from electron microscopy (since the locations of nicks and gaps are not known in the observed minicircle shapes, the classic RMSD fit is not relevant to the electron microscopy data). The alignment of simulated and observed data are therefore comparable. The flip around the PAI is of note as it suggests that the same flips certainly happened in the micrograph data as well. The "faces" of the minicircles which bind to the carbon surface is not known, so we get PAI superpositions of microscopy data up to such flips, but again the axes ratios are not affected.

I have described the experimental and simulated data which can be compared, and so far they compare remarkably well: the difference between the nicked and gapped minicircles can be seen by electron microscopy, and a very similar difference is visible between the two simulated minicircles. However, the simulations contain far more detailed information, since the dynamics of the structure is available at atomic resolution and on the picosecond time scale. Such details are not comparable with the experiments. Nevertheless, in the next section I will describe the dynamics of the nick site and the gaps sites, as it may be informative as to the structures of nicks and gaps in general, and it may give insight on the molecular motions which lead to the deformation of the minicircles.

The simulations provide information about the timescale at which the minicircle changes its shape. As in [74] the starting configuration of the minicircle has a circular shape, therefore with an axes ratio of 1 (the lowest possible axes ratio). The result of the simulation suggests it takes some nanoseconds to turn from a circle to an ellipse. It is especially visible for Sg2 while looking at the axes ratio over time (figure 2.6). In comparison, Sn1 axes ratio does not change much over the 100 nanoseconds (figure 2.5), whereas Sg2 undergoes a large deformation and stabilizes after about 30

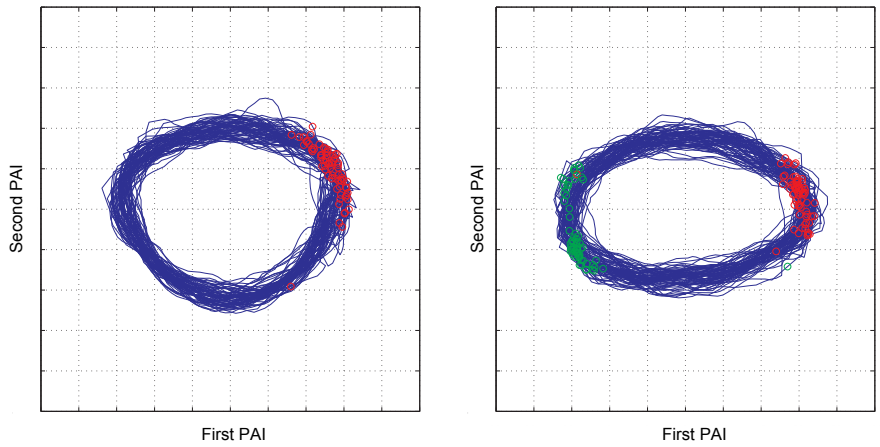

Figure 2.4: 3D Helical axes of simulated minicircles Sn1 and Sg2. Snapshots were taken from 30 ns to 100 ns, every nanosecond. 3D Helical axes are aligned on the principal axes of inertia. **(Left)** Sn1. The red circles show the nick site. **(Right)** Sg2. The green circles show the Gap 1 site and the red circles show the Gap 2 site. The gaps lie at the first PAI extrema.

nanoseconds. Therefore the beginning of the simulation certainly does not reflect an equilibrium shape of the minicircle. For this reason I considered only the snapshots after 30 nanoseconds for comparison with the microscopy data.

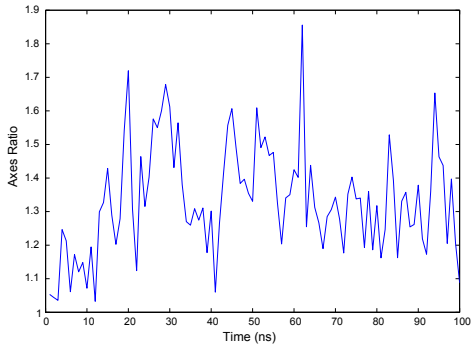

Figure 2.5: Time course of the Sn1 axes ratio. At the beginning the minicircle shape is a circle, and the axes ratio is equal to 1 (the smallest possible axes ratio, by definition). The axes ratio increases and first reaches its average value (1.35, SD 0.14) after about 15 ns.

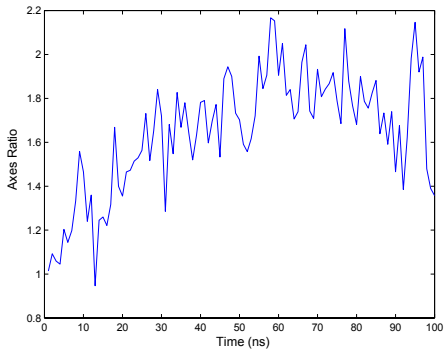

Figure 2.6: Time course of the Og2 axes ratio. The axes ratio increases and reaches its average value (1.76, SD 0.19) for the first time at about 30 ns. The snapshots used for the comparison with experiment were taken only after the 30^{th} nanosecond of simulation of Sn1 and Og2, to avoid that a bias is introduced by the circular starting structure.

2.4 Atomic Description of the Nick and Gap Dynamics

Here I describe successively the atomic details of the nick, Gap 1 and Gap 2 sites. I describe changes in stacking and pairing interactions between the bases during all the simulation, then I give some details about base flips and some interesting dihedral rotations at key time points of the processes.

2.4.1 Nick Site

The nick site dynamics can be split in three main states, described in figure 2.7. At the beginning, the nick is closed (figure 2.7a). Between 10.7 and 11.5 ns the nick opens (figure 2.7b). At the 39^{th} ns the base pair T-A3 (purple) breaks. Except for a 1.5 ns long break of the T-A3 base pair (purple) the base pairing remains stable until the 39^{th} nanosecond; then the A3-T base pair (purple) breaks and the T base stacks with the base C (orange) of the adjacent base pair, same strand, but on the other side of the nick (figure 2.7c). The T base (purple) unstacks from C (gray) and restacks to C (orange) in less than a nanosecond. This structure remains almost unchanged until the end of the simulation. Although the bases which stack together remain the same until the end of the simulation, there is still a change in the stacking interaction: at 62 ns, A3 (purple) unstacks and rotates around

the χ bond by 180°(the χ bond is between the C1' carbon of the sugar and the base) and stacks back to the G (black), thus presenting its other face to G. I will call this movement a "base flip". This type of flip occurs during transition between B-DNA and Z-DNA (see Chapter 0).

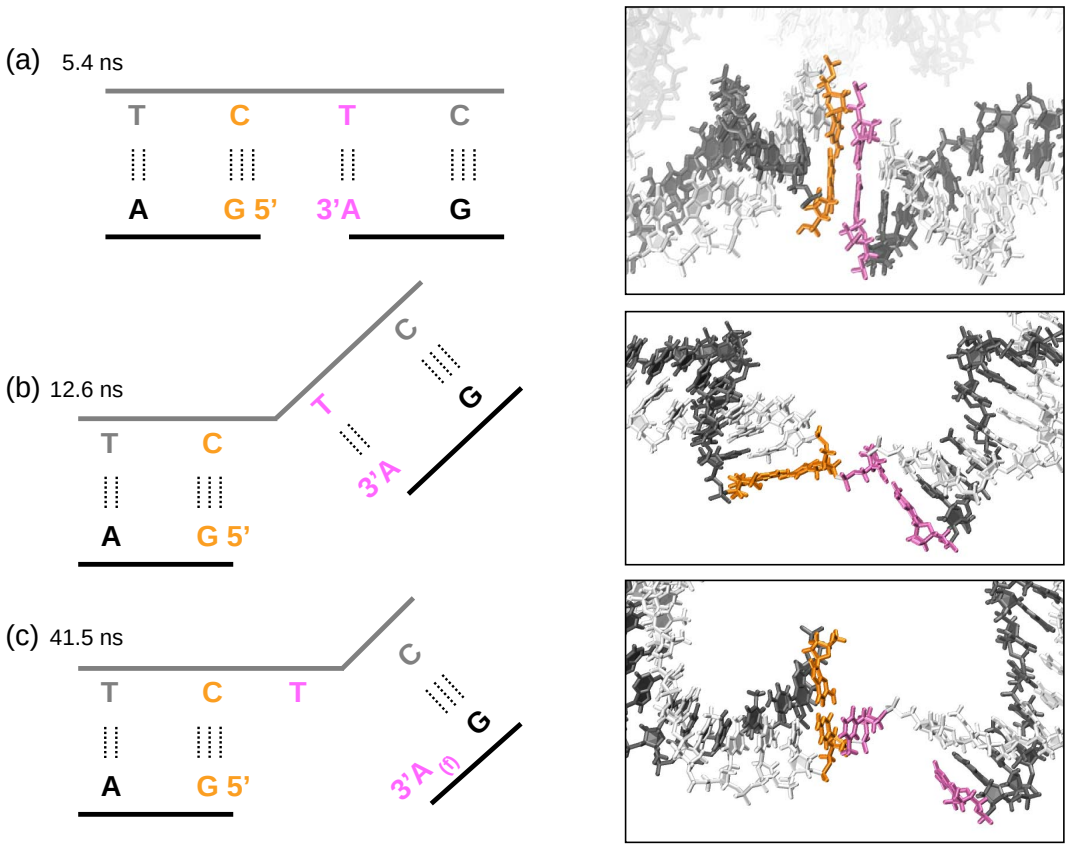

Figure 2.7: Visualization of base stacking and geometry of the nick site at 3 times emphasizing the striking differences. The base pairs at the nick site are colored in orange (C-G5) and purple (T-A3) in both the cartoon and the molecular representation. **(left)** Schematic representation of the nick site at different simulation times. Continuous lines represent DNA backbones; angles in continuous lines indicate unstacking of bases. Flipped bases (i.e. bases which underwent a χ rotation of about 180°) are indicated by the subscript f. **(right)** Corresponding visualizations of the atomic structure. **(a)** At the beginning, the nick is closed. **(b)** After 12 ns, the nick is opened. **(c)** After 40 ns the T-A3 (purple) base pair is broken, T (purple) restacks to the other side of the nick. The base pairs flanking the nick site are colored in orange (C-G5) and in purple (T-A3) in both the cartoon and molecular representations. The base A3 flips at 62 ns.

Sugar rotation and base flip It is interesting to notice that the opening of the nick is preceded by two rotations of the sugar ring of the bases flanking the nick. At 9.3 ns, the sugar of the G5 base does a 30 picosecond fast rotation around the χ bond of about 90°, to lie in the same plane as the base. At 10.6 ns the sugar of T (purple) does the same rotation as the G5 sugar did one nanosecond earlier, and the nick starts to open less than 50 picoseconds later. At 11.4 ns the planes of the two base pairs flanking the nick are orthogonal to each other. The wide opening of the nick is fast, since it took 0.8 ns for the base pair planes to move from stacked to orthogonal. Including the sugar rotations, which were not observed earlier in the simulation, the nick opening took three nanoseconds. Therefore the proximity (in time) of the two events suggests that the alignment of the sugar in the plane of G5 destabilized the base pair stack at the nick site, allowing the nick to open widely less than two nanoseconds later. It suggests that the orientation of the sugar rings may play a role in the stabilization of the base pair stacks.

2.4.2 Gap 1

At Gap 1, two bases of strand 1 are missing, leaving the two bases T and A of strand 2 unpaired (figures 2.8 and 2.9; T is drawn in green and A in orange). Although the two bases are not paired, they stay stacked together during almost all the simulation time.

At the beginning of the simulation the single bases point towards the center of the minicircle (Figure 2.9a). After 4 ns of simulation, the space of the gap is filled by the two backbone parts of strand 1 (gray), inducing a deflection angle of the double helix of about 90°(Figure 2.9b). The bases of the gap T (green) and A (orange) are unstacked. At 6 ns, the base T (green) of the gap unstacks from the double helix (precisely from G (black), Figure 2.10c). These two bases do not restack during the rest of the simulation, allowing a large deflection angle at the site of Gap 1 (see for instance figure 2.9e). T restacks with the other single base of the gap (A in orange, figure 2.9c).
At 9 ns, the base pair C-G3 (yellow) opens and C (yellow) is extruded out of the double helix (Figure 2.9d). At the same time, A (orange), which stacked to C (yellow) slides towards G3 (yellow) and stacks. There is now a base

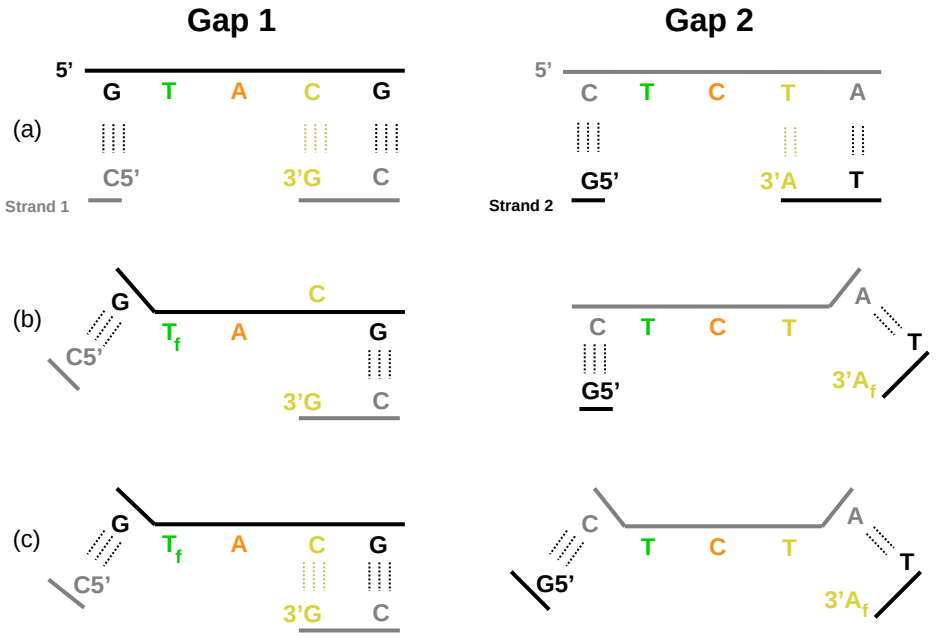

Figure 2.8: Schematic representations of the three main states of Gap 1 and Gap 2. As in Figure 2.7, continuous lines represent DNA backbones; angles in continuous lines indicate unstacking of bases. Flipped bases (i.e. bases which underwent a χ rotation of about 180°) are indicated by the subscript f. The base pairs that undergo unpairing at some point are drawn in yellow. **Gap 1 (a)** Initial state. The two single bases are T (green) and A (Orange). **(b)** T (green) flipped, stacked to A (orange) and is not stacked to G (black). The base pair C-G (yellow) is broken and C is extruded out of the double helix. **(c)** The base pair C-G is reformed. **Gap 2 (a)** Initial state. **(b)** The base pair T-A3 (yellow) is broken. A3 is flipped and T unstacks from A (gray). **(c)** T (green) is unstacked from C (gray).

of Strand 2 (A, orange) which stacks with a base of Strand 1 (G3, yellow). This conformation is stable until the 39^{th} nanosecond, when the base pair C-G3 (yellow) reforms (figure 2.9e). Then A (orange) remains parallel to the C-G3 base pair (yellow), sometimes stacking to C and sometimes to G3(figure 2.9f), until the end of the simulation. The bases T (green) and G (black) stay unstacked from 6 ns, and the deflection angle of the double helix at Gap 1 stay high during all the simulation, although lower than the Gap 2 deflection angle.

Sugar rotation and base flip The restacking of A (orange) and T (green)

at 6 ns is simultaneous with a rotation of the sugar ring of the T base (green) by about 60°. Indeed, at the beginning of the simulation the C2' carbon of T (green) (which has two visible hydrogens in Figure 2.9b) points towards A (orange), whereas at 8.5 ns (Figure 2.9c) it points towards the opposite direction. The subsequent base flip of T (green) is visible while looking at the two pictures 2.9c and 2.9d. Precisely, at 10 ns, T (green) rotates around χ by half a turn in the same direction as its associated sugar, resulting in the restacking of T (green) in the flipped conformation at 11 ns. T (green) stays stacked with this orientation until the 97^{th} second, then it flips back.

2.4.3 Gap 2

At Gap 2, two bases of strand 2 are missing, leaving the two bases T and C of strand 2 unpaired (figures 2.8 and 2.10; T is drawn in green and C in orange).

The DNA minicircle is even more curved at Gap 2, but the bending happened much later than at Gap 1. Indeed, until the 35^{th} ns, all the bases of strand 1 (gray) remain stacked, except between 19 and 26 ns, when the T base (yellow) is extruded out of the double helix. Indeed, at 19 ns the base T-A3 (yellow) breaks (Figure 2.9b) and A3 flips at 25 ns; in Figure 2.10c the large ring of the purine points away from the complementary base, whereas in the B form DNA (or in figure 2.10b) it is in the middle of the double helix as it linked to the other base of the pair by two hydrogen bonds. Although the T-A3 (yellow) base pair is broken, the bases remain stacked to their neighbors until the 37^{th} nanosecond, when the T base (yellow) unstacks from the A-T (gray-black) base pair (Figure 2.10d). This unstacking induces a large deflection angle of the double helix axis. There is then a stack of the three single bases T, C and T (yellow, orange and green) on strand 1 (gray). A3 (yellow), which flipped, slides sometimes and stacks with T (gray) instead of T (yellow), and the stack of single bases does not restack with the double helix from the T (yellow) side for the rest of the simulation. On the contrary, A3 (yellow) unstacks from the other side of the gap (Figures 2.9d and e) at 70 ns for few nanoseconds, forming a stacked single stranded linker between the two parts of the double helix, with angles of 90°at its extremities; indeed, in figure 2.9 (e) it is remarkable that, one nanosecond after figure 2.9 (d), the helical axes from each side of the gap are almost parallel, thus forming a bend of approximately 180°along only four base

pairs. Some other unstackings in the gap happens later. For instance T (green) unstacks from C (orange) and restacks to C (gray) (Figure 2.9f). The stacking of Strand 1 is not complete until the end of the simulation, and the deflection angle at the gap site stays high.

2.5 Discussion and Conclusion

In contrast with the study of Chapter 1, I observe a clear difference between the shapes of the nicked and gapped DNA minicircles (the nicked minicircles had two nicks at diametrically opposite sites, and the gapped minicircles were designed as the nicked minicircles but with two nucleotides missing from the 3' end of each single stranded oligo). The clustering analysis and the notion of shape-distance introduced in Chapter 1 were not needed, as the difference between the two groups of shapes was visible by superposition on the principal axes of inertia (Figure 2.3). It was then possible to quantify this difference by defining the notion of axes ratio, the ratio of the width and the height of the minicircle shapes drawn in Figure 2.3. The 20 nicked minicircles (On2) had an average axes ratio of 1.29 (SD 0.18), whereas the 20 gapped minicircles (Og2) had an average axes ratio of 1.70 (SD 0.26), which is a significant difference.

It is interesting to compare the nicked minicircles with the 158 bp, covalently closed DNA minicircles of Chapter 1, whose average axes ratio is equal to 1.33 (SD 0.19). Their average axes ratio is not significantly different; this could indicate that the effect of two opposite nicks on DNA minicircle shapes is not significant. However the two minicircles have different lengths so this conclusion has to be taken with caution.

Molecular dynamics simulations of minicircles that are chemically identical to Og2 (Sg2), and of minicircles similar to On2 but with one nick only (Sn1) were analyzed with the same method as the observed minicircles (Figure 2.4). The comparison of axes ratio between the simulated and observed shapes is remarkable: the average axes ratio of the nicked minicircle is 1.35 (SD 0.14), and the average axes ratio of the gapped minicircle is 1.76 (SD 0.26). The simulations, which agree with the observation to the extent of what can be compared, provide a description of the DNA minicircles

dynamics at a time and space resolution that are inaccessible to electron microscopy. I used this simulation information to gain insight in the structural details that may explain differences in the global DNA minicircle shapes.

The simulations seem to confirm that the elliptic shape of the gapped minicircles is due to increased curvature of the double helix at the gap sites, probably because of flexibility of the gaps (Figure 2.4). There are two main structural features responsible for this curvature: base-pair unstacking and unpairing. It is clear that unstacking is facilitated by the break in one of the two sugar phosphate backbone, leaving only one covalent bond on the side of each base pair flanking the nick site or the gap sites, and thus allowing large deflection of the DNA double helix (see for instance Figures 2.7c and 2.9e) . Interestingly, the less stable stacks are not found between two consecutive single bases forming gaps, but rather between single bases and base pairs (Figure 2.8). Unstacking may release bending and torsional stress induced by the circular closure constraint: therefore the minicircle nicks and gaps at equilibrium are expected to be dominated by the unstacked state, as the simulations suggest. It is then understandable that the other significant event observed at the nick and gap sites is the unpairing of some bases (yellow base pairs in the figures). Indeed, the base-pair stacking is thought to be a more significant stabilizing factor of the double helix than base pairing [68, 69]: the flanking base pairs of the opened nicks and gaps are not stabilized by stacking of neighboring base pairs, and therefore are subject to destabilization. Moreover, base unpairing is a kind of extension of the gap lengths, and thus a contribution to the release of the mechanical stress of the minicircle. If the gap length is defined as the number of unpaired bases along a DNA strand, the length Gap 2 is two nucleotides at the beginning of the simulation, and three at the end (Figure 2.8). Similarly, the nick becomes similar to a one nucleotide gap after the T-A3 base breaking.

The multiple base unpairing observed in the simulations may indicate that the equilibrium of nicks is not as simple as has been previously hypothesized, for instance by Frank-Kamenetskii and his collaborators [68, 69]. The hypothesis of their study of base-pair stacking energy is that nick equilibrium is dominated by two main states, stacked and unstacked, with intact base pairs flanking the nick site. The presented simulations show that in

some conditions the equilibrium can be more complex, with additional states involving base unpairing. However, the DNA molecules used here are different in that they are under strong bending and torsional stress, which probably shift the nick equilibrium towards the opened state and therefore facilitate base unpairing; in contrast, Frank-Kamenetskii and collaborators used linear nicked DNA which are not subject to this stress. New insight on the equilibrium of nicked DNA could be provided by simulations of nicked linear DNA molecules.

Finally, the simulation suggests that the curvature is already very high at the nick site, although it is not clear if it could be detected under the microscope. The curvature is even greater at the gap sites. At some time points, there is an almost 180°deflection angle between the double helices on either side of the gap site (see Figure 2.10e). Therefore, I conclude that the curvature of the double helix has to be very high in order to be detected by electron microscopy. Also, the fact that the gaps are at diametrically opposed sites allows a greater visibility of the deformation, that may not be seen if the flexible sites were closer to each other. Based on this observation, further experiments could be designed in order to detect sequence effect on DNA minicircle shapes. One could design a construct with a gap and a test sequence at diametrically opposite sites. In the hypothetical case where the test sequence is as curved as the gap, a shape similar to the double-gapped minicircle shape should be seen. In the opposite case, the axes ratio should be significantly smaller. Therefore this work provides a basis for the study of sequence effect on DNA minicircle shapes by electron microscopy.

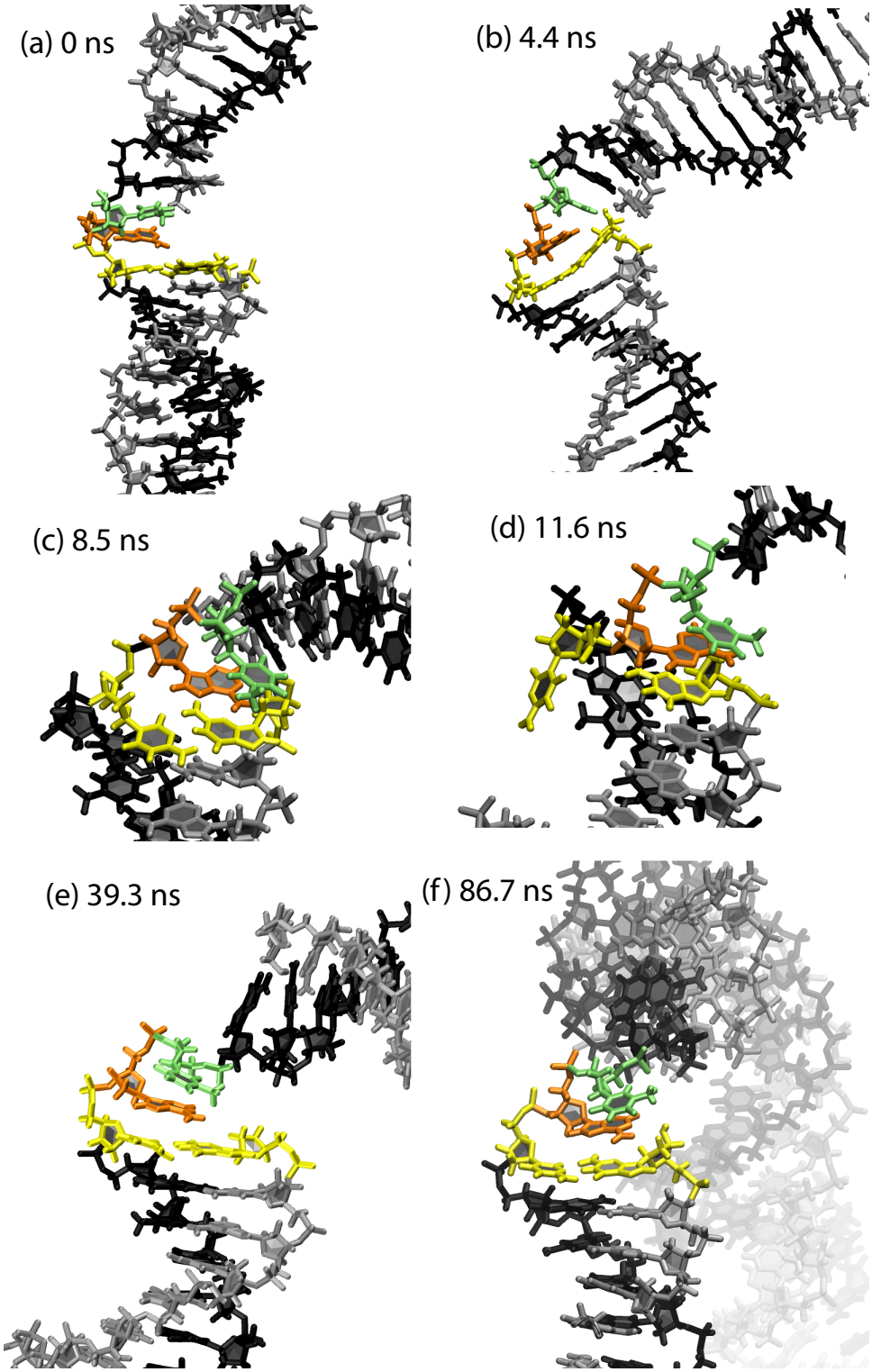

Figure 2.9: Some snapshots of Gap 1

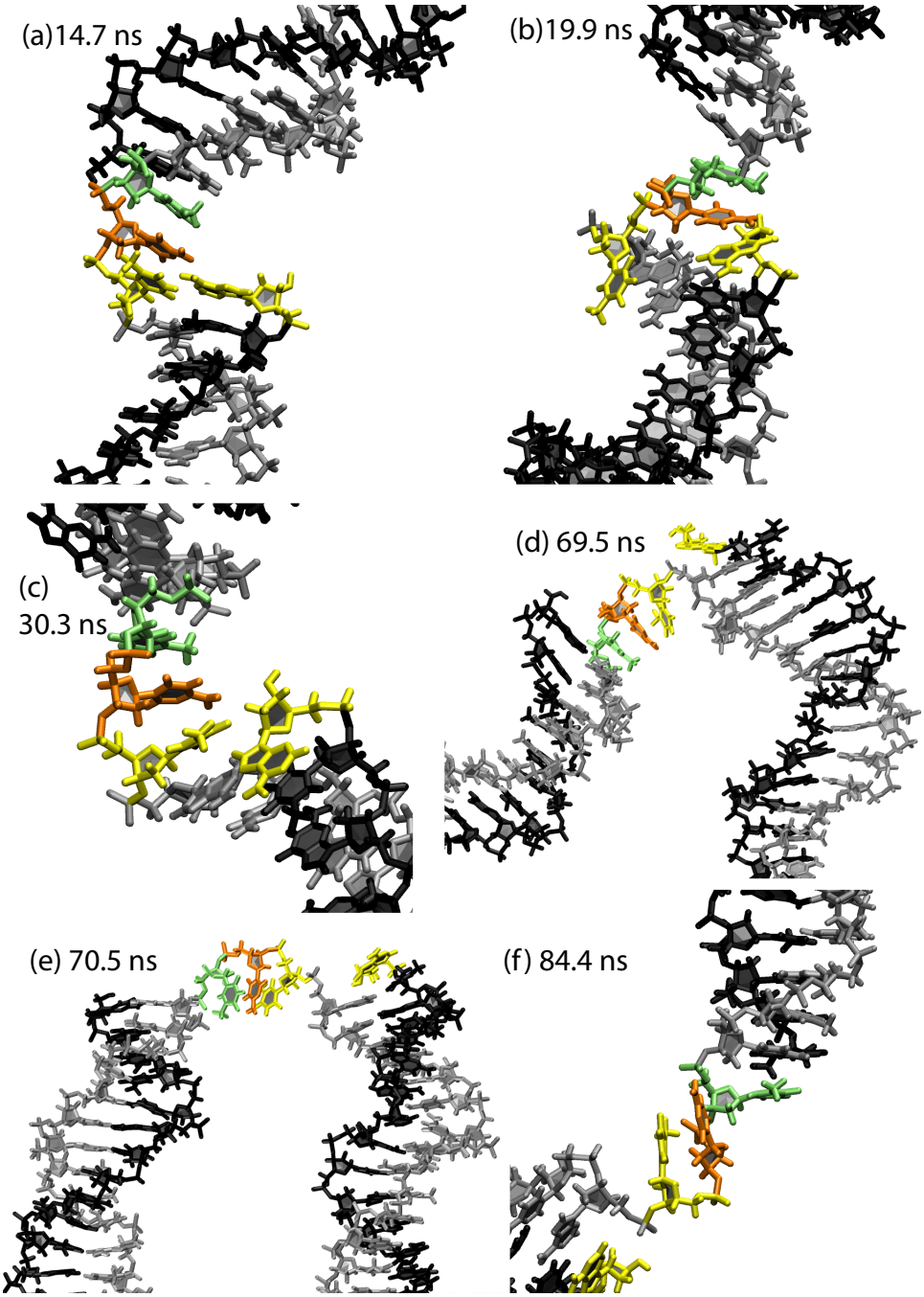

Figure 2.10: Some snapshots of Gap 2

Chapter 3

Annealing-Cyclization of 94 Base-Pair DNA Minicircles

3.1 Introduction

In the first two parts of the thesis, two studies of DNA minicircle shapes were presented. The minicircles of Chapter 1 were prepared[1] with a ligase-catalyzed cyclization reaction, whose principle is described in the introduction of the thesis (Section 0.3). The minicircles of Chapter 2 were produced by annealing-cyclization. The aim of the present chapter is to propose a novel method to measure cyclization rates, via annealing-cyclization, which allows the production of nicked and gapped DNA minicircles and does not use DNA ligase. I describe here experimental results on the circular monomer fraction of the products, and I use the analytical solution of a reaction rate model to interpret the results.

3.1.1 Annealing-Cyclization

As early as 1966, Wang and Davidson showed experimentally that for a long DNA molecule (from bacteriophage λ) the cyclization factor could be obtained by taking the ratio of the rate constant of cyclization (k_1) and the rate constant of bimolecular association (k_2), both reactions mediated by base pairing of the 12-nucleotide complementary ends. Thus, before the introduction of ligase for DNA cyclization by Shore, Langowski and Baldwin [8], Wang and Davidson [78] already used natural long sticky ends of phage

[1] and provided by Jason Kahn, University of Maryland, College Park, MD, USA

λ DNA to measure cyclization rates without using ligase. However, λ DNA is rather long, about 50'000 base pairs, much longer than the persistence length. Therefore the orientations of the ends of their DNA molecule in solution were uncorrelated, and thus not influenced by the specific DNA sequence between the ends. Therefore such long molecules are not well suited to study either sequence effects, or defects such as nicks and gaps, which is the purpose here.

A ligase-free cyclization assay is advantageous as it avoids several considerations about the effect of the ligase on DNA cyclization equilibrium. In the ligase-catalyzed cyclization assay (cf. Section 0.3), if the dissociation rate of the sticky ends is high with respect to the ligation rate, then the reactions of cyclization and bimolecular association are said to be in fast pre-equilibrium. Shore et al. [8, 9, 10] showed that in this case, k_1 and k_2 are linear functions of the ligase concentration. The linearity of k_1 and k_2 is a necessary condition for the equality $J = k_1/k_2$. Therefore linearity has to be verified while measuring J factors with the ligase-catalyzed cyclization assay. It is tedious work, as each rate constant requires several time points of the reaction to be run in a gel, and it has to be repeated for several different ligase concentrations. If the linearity condition is not met, the measures of the rate constants should not be used, since it could introduce large errors in measurements of the J factor. This consideration is at the heart of the critique by Vologodskii and collaborators [73] of the results of Cloutier and Widom who observed surprisingly high J factors of 94 base pair DNA molecules [70], i.e. molecules whose length is about two thirds of the standard persistence length of DNA. Cloutier and Widom checked the linearity of k_1 as a function of ligase concentration, but not of k_2. Du et al. [73] explained that k_2 as function of ligase may reach a plateau before k_1, and then J, which was computed as k_1/k_2, could have been over estimated if the measurements were performed with too high ligase concentration. Concerned with these issues, Widom et al. (personal communication) carefully repeated and re-analysed their cyclization study and concluded that ligase activity was perhaps too high in some of their earlier reactions and that this caused an over estimation of DNA flexibility as reported in ref. 70. But in conditions where J is provably ligase-independent they still obtain a value of J of approximately 20 pM which is still approximately 100-fold above

that expected from a standard worm-like chain model.

In order to avoid effects of ligase in DNA minicircle formation experiments, I designed DNA molecules with comparatively long sticky ends which can form stable nicked minicircles without ligase. For each experiment, I compared the amount of monomeric circles to the cumulated amount of dimers, trimers and so on. The absence of ligase in the annealing-cyclization reaction avoids the risk of errors linked to the pre-equilibrium assumption and to the excess of ligase concentration, and it avoids the tedious verification steps of the linearity of k_1 and k_2 with respect to ligase concentration. We also investigated the effect of the length of the sticky ends on cyclization, which would be problematic with the ligase-catalyzed assay. Indeed, the ligation reaction has a pre-equilibrium only if the sticky-end dissociation rate is high, thus only with short sticky ends, usually between two and four base pairs [46]. The annealing-cyclization assay does not require a pre-equilibrium assumption to be verified, so the experiments could be performed for constructs with various sticky end lengths, namely 16, 25 and 47 base pairs.

3.2 Design of the Annealing-Cyclization Experiment

The experiment consists in annealing two complementary DNA oligonucleotides with complementary single stranded overhangs (Figure 3.1) in order to obtain, among other products, DNA minicircles, but without covalently closed backbones (Figure 3.2 bottom frame). Indeed, during the annealing reaction, hydrogen bonds link the bases of each the complementary DNA oligonucleotides and form base pairs (they are then said to be hybridized) but the extremities of the sugar phosphate backbones do not undergo covalent closure, unlike a ligase-catalyzed assay. Therefore, if all the sticky ends are 100% complementary, the resulting monomeric circles will carry two nicks. If the sticky ends are only partially complementary, and if the complementary parts lie at the extremity of the sticky ends, the minicircles will carry gaps instead of nicks. In order to be observable by gel electrophoresis, the single stranded overhangs or sticky ends (SE) have to be long enough to stably hybridize together at room temperature. This

condition avoids the need of ligase for DNA cyclization.

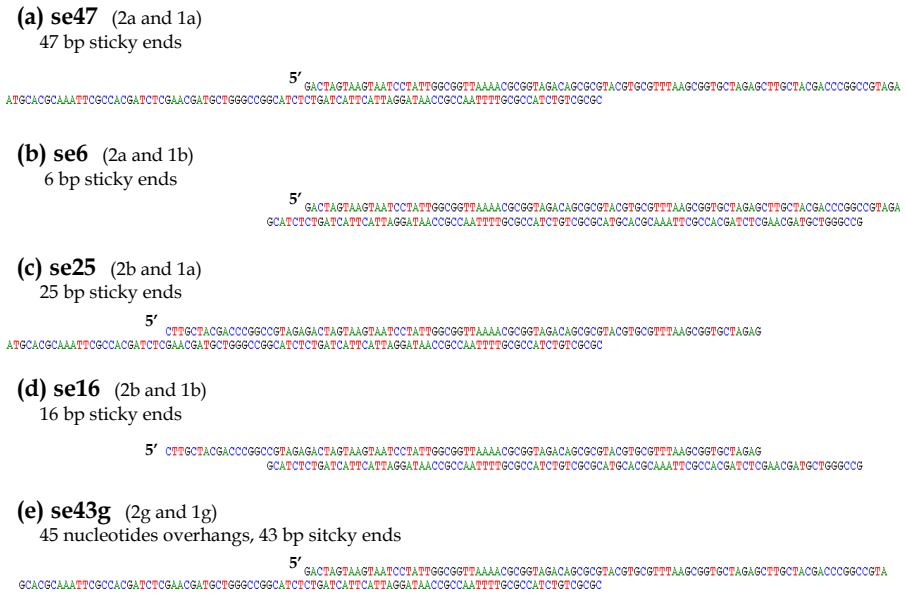

Figure 3.1: Sequences and names of the five constructs. Upon annealing, two single stranded DNA oligonucleotides can form a double stranded DNA molecule with sticky ends of various lengths. The three sequences 2a, 2b and 2g are read from 5' to 3' (top strands). 2b is a cyclic permutation of 2a, while 2g is gapped by the removal of two nucleotides at the 3' end of 2a. Similarly there are three complementary sequences 1a, 1b and 1g that are read from 3' to 5' (bottom strands). All the sequences are vertically aligned in the image. **(a b c d)** The four possible complementary pairs formed by the oligonucleotides 1a, 1b, 2a and 2b were used to assemble the four nicked constructs. The oligos are composed of 94 nucleotides each. The sticky ends are complementary. The four minicircles formed by cyclization have exactly the same sequences: only the sticky end length (and therefore the positions of the nicks in the resulting minicircles) vary. **(e)**. The single stranded overhangs of se43g are 45 nucleotides long. However they are complementary upon 43 base pairs only, leaving two gaps of two nucleotides, TpA or TpC. The minicircles formed by the constructs in (a) and (e) are described in detail in Chapter 2 of this thesis.

The reactants and possible intermediate and final products of the annealing-cyclization reactions are summarized in Figure 3.2. By design, the SE are long enough to hybridize stably at room temperature. Therefore two

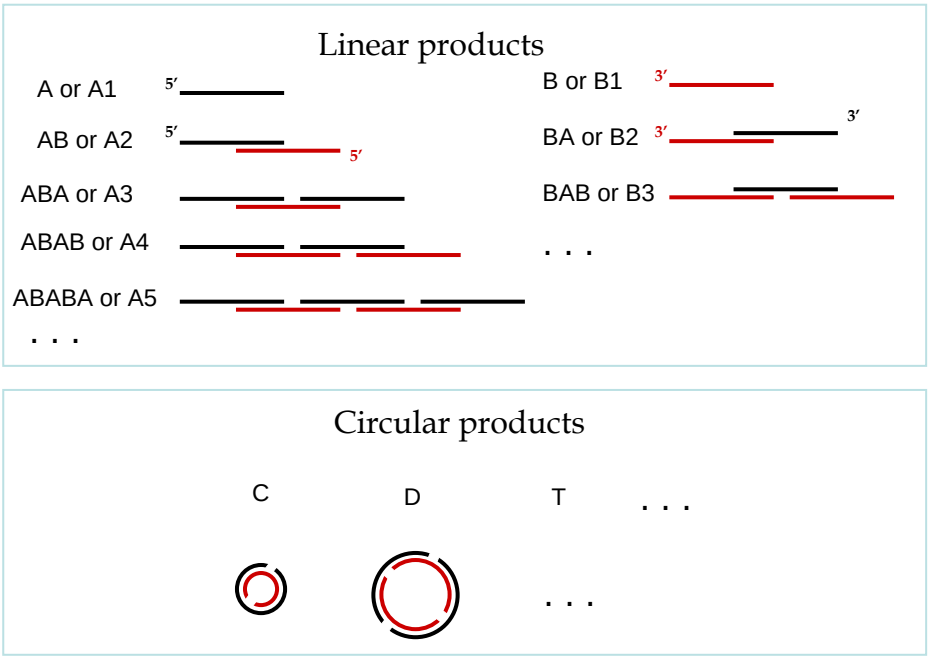

Figure 3.2: The possible products and reactants of the annealing-cyclization reaction. The two initial reactants are two complementary oligonucleotides, drawn in red and black. When the reactants undergo hybridization, they can form linear molecules of various lengths (upper frame). The molecules with complementary sticky ends can have their two extremities hybridized together, and form a circular molecule (lower frame). Only the linear molecules composed of an even number of oligos can cyclize. In any circular product the number of nicks (or gaps) is equal to the number of single stranded oligos.

oligonucleotides can hybridize either their SE, or the rest of sequence (which I will call the "body"), leading to two different molecules, A2 or B2 (Figure 3.2). By definition, the SE are shorter than the body, and therefore hybridization of the body is more energetically favorable than hybridization of the SE, so that one of the two species A2 or B2 will be rare. A2 and B2 have complementary ends and can both form monomeric circles (C in Figure 3.2). They can also hybridize with an additional oligonucleotide, either A or B, or with a longer product with suitable SE. Not any pair of product has mutually complementary SE; for instance A2 cannot hybridize with B2, or B4. All the products with an even number of oligos can cyclize, meaning that they can form circular double stranded DNA molecules. None

the molecules with an odd number of oligos (the "odd products", A3, B3, A5 ...) can cyclize, because their SE are not complementary. Therefore the annealing cyclization can be viewed as a polymerization process, where DNA polymers grow until they cyclize. As in the classic annealing protocols, the solution is heated at 95°C, and cooled slowly to room temperature. At the beginning of the annealing, the DNA molecules cannot hybridize because the temperature is higher than the melting temperature of the oligos. As the temperature decreases, the DNA strands start to hybridize in a reversible way, and the hybridization becomes more and more stable as the temperature decreases. All the SE of our constructs (except for se6) have melting temperatures of at least 25°C above room temperature, so the double stranded products are stably hybridized at the end of the annealing. Therefore the annealing-cyclization reaction is an irreversible reaction. As the reactions last for several hours, we assume that the reactions go to completion, and all the reactants should hybridize if there is some available complementary DNA in the solution. Furthermore, if the two oligos are in equal amounts, then all the linear products should either associate with another linear molecule or cyclize. Eventually, all the linear DNA cyclize, and there should remain in principle only circular products in the solution. If there is a significant imbalance between the reactant oligos, a distribution of odd products will be found at the end of the reaction.

As in classic ligase induced cyclization experiments, I measured the ratio of minicircles to other products of the reaction. The products were separated by gel electrophoresis and their ratios were measured from the width and light intensities of bands in the gel image. As the reaction nearly goes to completion before the gel electrophoresis is performed, the kinetics of the reactions are not observable with this method. Rather, I observed what I believe to be the final products of the reaction. However, the amounts of the final products depends on the kinetics of the competing reactions of cyclization and bimolecular association, and on the initial concentration of the reactants. A mathematical model of the annealing-cyclization can be used to interpret the measurements and to deduce some kinetic values of the reaction, namely the ratio of the rate constants k_1 and k_2. I varied independently three parameters and observed their effect on the monomer circle fraction (defined in section 3.4.3): (i) the initial concentration of DNA oligos, (ii) the length of the sticky ends, and (iii) the replacement of nicks

by gaps (cf. Section 3.7).

3.3 Experimental protocol

3.3.1 Oligonucleotide Sequences

The four possible complementary pairs formed by the oligonucleotides 1a, 1b, 2a and 2b were used to assemble the four nicked constructs se6, se16, se25 and s47 with sticky end lengths of 6, 16, 25 and 47 base pairs respectively, and two supplementary oligonucleotides (1g and 2g) were used to construct the gapped construct se43g (Figure 3.1). The sequence is the best cyclizer obtain in an *in vitro* selection assay (SELEX) by Cloutier and Widom [70]; se47 was used to produce the minicircles On2 of Chapter 2, and se43g was used to produce the minicircles Og2. This sequence is somehow famous because its J factor was observed to be orders of magnitude larger than predicted by the then current theories. It was therefore an interesting minicircle to observe under the electron microscope.

3.3.2 Annealing

The oligonucleotides were ordered from Microsynth^{TM}and Sigma-Aldrich^{TM}. For each annealing, two oligos were mixed in sodium Tris-EDTA (STE), 100 mM NaCl. To obtain series of different initial DNA concentration reactions, the mix was divided in aliquots which were successively diluted with STE to the desired concentration. Then, the solutions were completely immersed in a 94°Celsius water bath. The water bath heater was turned off, and samples slowly cooled to room temperature. Then the specimens were analyzed by gel electrophoresis.

3.3.3 Gel preparation and imaging

The 2.5% to 4% GTG^{TM}agarose gels were prepared in TBE buffer. The gels were either poured directly with either 1X SYBR^{TM}safe colorant or ethidium bromide, or stained with SYBR^{TM}safe in 1X TBE after migration. The gels were run between three and four hours at constant voltage, between 40 and 60 Volts. The DNA samples were mixed with Cambrex^{TM}5X loading buffer. The gels were illuminated with UV light in a BioRad 2000 gel doc station and photographed with an 8-bit gray scale digital camera.

3.4 Gel Image Processing and Analysis

3.4.1 Dust signal removal by image processing

The gel images that were obtained contain some white spots, that seem to be associated to dust and have nothing to do with the DNA products of the annealing-cyclization reactions (Figure 3.3, left image). The profile of the light intensities for each lane may be significantly changed by these dust spots since they appear with maximum intensity levels in the image. However they are easily distinguishable from signal associated with DNA because DNA bands are always (i) orthogonal to the migration direction (horizontal in the image), (ii) their length is always equal to the width of a lane, and (iii) their intensity should be approximately constant along the horizontal band. On the contrary, signal from dust creates random shapes, close to circular, and several times smaller than the DNA bands. Therefore I was able to design and to program a semi-automatic method to filter the dust spots from the gel image. It works as follows. The user selects a lane with the rectangular selection tool in ImageJ [79]. For each horizontal pixel line of a vertical gel lane, there is no reason for the pixels to have large intensity differences between them, except if there are dust spots; therefore, if some pixels are above a threshold defined by 10% higher than the median of the line of pixels, it is changed to the median value. I did several trials of this program (a macro for ImageJ [79]), I did not notice modification of pixels inside bands, as the pixels usually do not overwhelm the threshold; all the dust spots inside the lanes disappeared (Figure 3.3). Note that it is crucial that the user selects correctly a lane before applying the function, otherwise the assumptions ii) and iii) will not hold, and DNA bands could be erased if they are short with respect to the width of the selection. In addition, dust spots that occupy the full width of a selection will not be erased, and some dust spots remain between lanes because of this (Figure 3.3 right). This procedure was applied to gels 1 and 3 (Figures 3.11 and 3.13) before measurements.

3.4.2 Intensity measurements in the gel image

The agarose gels 1 and 3 were stained with SYBR safe, Gel 2 with Ethidium Bromide. These molecules intercalate non-specifically between the DNA base pairs, and are fluorescent to UV light. For band intensities measure-

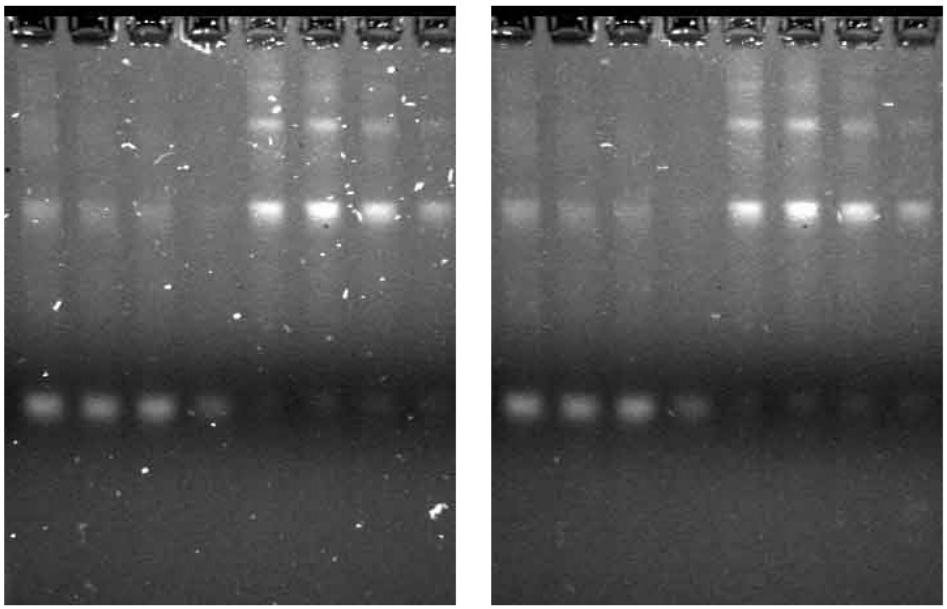

Figure 3.3: An agarose gel image before (left) and after (right) the dust removal procedure by image processing. For each line of a user selected lane, the procedure replaces pixels that are at least 10% greater than the median by the median value. The selections were also applied between lanes; the stains which occupied the full width of a selection remain visible.

ments, radioactive labels are also often used; it was interesting here to investigate whether UV fluorescence could replace radioactive labels for this purpose, because it simplifies further the cyclization assay. A photograph of the gel under UV light reveals (i) where the DNA is in the gel and (ii) the amount of DNA in each band of the gel. Band intensity should be linear with the amount of DNA in the gel because of the non-specific binding of the stain along the DNA molecule. The intensity profiles of each lane was plotted with ImageJ in the following way: after rectangular selection of a lane, the average intensity of each pixel line along the lane is computed (see figure 3.4). Then I computed the area of the peak. The area of the peak is limited by the curve above, and below by a straight line between the rightmost and leftmost points of the peak on the curve. This way, the intensity at the left and the right of the peak becomes the base intensity level, or the background.

The background light is produced by reflections of fluorescent light in the gel, staining molecules unassociated to DNA in the gel and by rare DNA

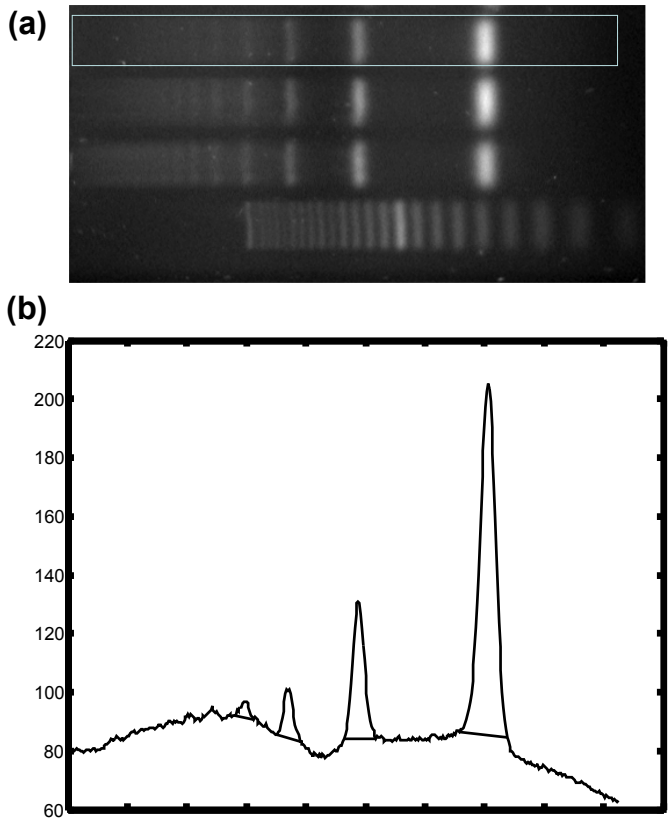

Figure 3.4: Computation of DNA band intensities in the gel images. **(a)** A lane of Gel 2 (also shown in Section 3.7.6) is selected (light blue rectangle) in the gel image. The intensity profile along the selected lane is plotted exactly below the gel, so that bands in the gel lie precisely above the corresponding peaks of the curve in (b) . **(b)** Intensity profile of the lane selected in (a) (see section 3.4.2). Band intensities are computed as areas of the peaks. The lower limit of a peak surface is drawn as a straight line below the peak. The extremities of the lower limit lie on the curve, one extremity at each side of a peak. The right and left limits of a peak were assigned by eye, so that the peak surface is just above the surrounding inter-peak regions of the curve. The four main bands of this lane are, from right to left, monomeric circles (94 bp), dimers, trimers, and tetramers (see next section).

species coming from the reaction tube, perhaps DNA knots, or structures where the oligos are partially hybridized, which produce a smear in the lanes. I tried to measure the intensity of such species by subtraction of the light intensity profile of a lane with one oligo only. The oligo lane should not have a smear of DNA species, and can serve as a background reference. However

the background signal was sometimes higher than rare DNA species signals, resulting in negative measurements of DNA amounts. Therefore this method was not satisfactory to measure band intensities. Hence, both signals from rare DNA species and from free staining molecules were considered as background, i.e. the light intensity level that lies between clearly distinct bands was considered as background.

The right and left limits of the peak were delimited by eye for each curve. Indeed the shape of the peaks vary, and there is no truly objective and automatic way to assign the limits of a band peak. However, if the bounds of the peaks are set in a consistent way for every lane in a gel, it is possible to measure ratios of band intensities within lanes, and compare these ratios. A visual example and the criteria to set the limits of the peaks are illustrated in figure 3.4.

3.4.3 The Monomer Fraction

The quantitative analysis presented in this chapter is based on measurements of the circular monomer fractions obtained by the analysis of the gel images presented in Figures 3.11, 3.12 and 3.13 (Section 3.7.6). The circular monomer fraction, or monomer fraction, is the amount (the total mass) of monomer circles divided by the total DNA in the reaction tube. It is approximated by the ratio of the monomer circles band intensities over the sum of all the clearly measurable bands in the same lane. Since the band intensity should be linear to the DNA amount, this measure should give a good approximation to the monomer fraction.

Under a specific assumption, the monomer fraction is also the fraction of double stranded DNA linear monomers (or simply linear monomers) which turned into monomeric circles. This definition of the monomer fraction is important because the mathematical model proposed in the next section uses it. The assumption is that after a given time, all the single stranded oligos hybridize and form linear monomers (A2 in Figure 3.2), and they are the only DNA species in the solution at this time point. The two complementary oligos are mixed in equal amounts so that they can all hybridize. If this assumption is true, the linear monomers (L) can be considered as the only reactants; at this particular time, the concentration of L is equal to the initial

concentration of each oligonucleotide (L_0). Then, the monomer fraction is the fraction of double stranded DNA linear monomers which turned into monomeric circles.

This assumption is believed to be realistic since there exists a temperature at which the body can hybridize and the sticky ends (SE) cannot; if the reaction stays a long time at this temperature, each oligo will have its body hybridized with the body of one complementary oligo and form a linear monomer, which will not undergo any further reaction until temperature allows its SE to hybridize. The assumption is certainly true for the se25 and se16 constructs, which have very different lengths for the body and the SE, and may also be true for se47, whose body and SE may have different melting temperatures because of their sequence of bases. The validity of the assumption depends on the time during which the reaction temperature stays between the melting temperature of the body and the melting temperature of the SE. The annealing-cyclization experiments described here took several hours to reach room temperature from an initial temperature of 95°C, and we assume that the reaction of annealing of single stranded oligos into double stranded monomers goes to completion before the SE are able to hybridize.

3.4.4 Monomer Fraction Correction

I measured the monomer fractions of the constructs in three gels presented in the Gel Images Section (3.7.6). It happened that, because of differences in gel imaging conditions (for instance exposure time), and differences in gel preparation (agarose concentration or the staining molecule used), the apparent monomer fraction for the same DNA solution was different between two gels. For instance the monomer fraction of se47 at 0.2 μM differs between Gel 2 and Gel 3: $M_f^{Gel2} = 0.72$ and $M_f^{Gel3} = 0.59$. It is therefore interesting to verify how the measurement of a given DNA species ratio from a given DNA solution vary when the solution is loaded in two different gels. In order to do so, I measured the band intensities of the DNA ladder, which is the same in all the gels. I measured the intensity of the 120 bp ladder band (B_{120}, four band below 200 bp bright band in the gel images) and of the 260 bp band (B_{260}, three bands above the 200 bp bright band). I chose the 120 and 260 base pair bands because they are the closest ladder bands to the monomeric circles and dimeric circles bands. The intensity profiles of

the ladders are visible in Figure 3.5.

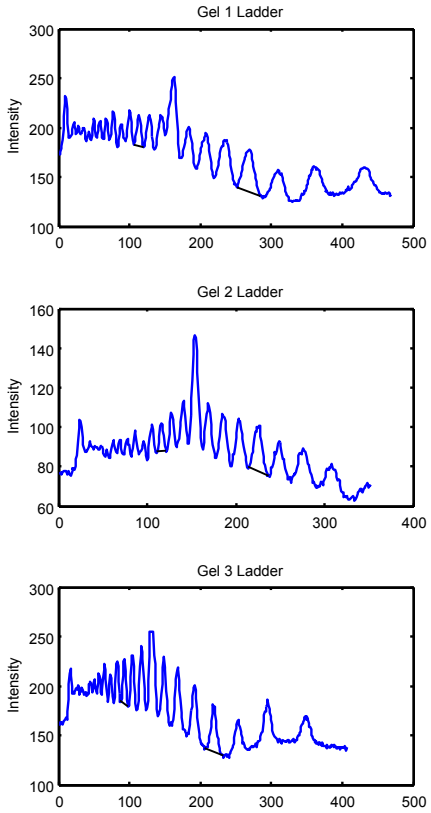

Figure 3.5: Ladder intensity profiles for gels 1, 2 and 3. The peak areas of the 120 bp band and the 260 bp band are delimited by black straight lines below the peaks. their ratios were computed for each gel. The ratio is equal to 3.76 for Gel 1, 3.33 for Gel 2 and 1.82 for Gel 3. This ratios were used to compute a correction of the monomer fractions as described in equation (3.1). It seems that background is higher toward the left part of the profile curves (upper part in gel images) in Gel 1 and 3. This effect is less visible in gel 2. For this reason, Gel 2 was taken as reference for the monomer fraction correction.

If the band intensities were perfectly linear to the amount of DNA in the bands, the ratio $b = B_{120}/B_{260}$ should be identical in all the gels because the same ladder solution was used in all gels. However the observed band ratios vary between gels: $b_1 = 3.76$, $b_2 = 3.33$ and $b_3 = 1.82$ (the index of b

indicates the gel number). This variation of the ladder band intensities indicates that there can be variations of a factor of two between measurements of C/D ratios, from the same annealing-cyclization reaction but in different gels.

In order to reduce the error in measuring the DNA band intensity ratios, I applied a correction to the band intensities based on the values of b_i. I first made the assumption that the error in one given gel is uniform along the horizontal axis of the gel (perpendicular to migration). Then, I applied the same linear correction to all the bands which are at the height of B_{120}, namely the monomer circles bands of gels 1 and 3:

$$X_i^{corrected} = \frac{b_2}{b_i} X_i, \tag{3.1}$$

b_i being the ladder ratio of Gel i, and X_i being either B_{120} or a monomeric circle band intensity in Gel i. By definition, the correction applied to the B_{120} value of Gel i makes the ladder band ratio $b_i^{corrected}$ equal to b_2, which should be the case if the band intensities were indeed linear to DNA amounts. In that sense, the error along the migration axis was corrected, and if the error is uniform along the horizontal axis, then it has been corrected for all the lanes. The ladder ratio of Gel 2 was taken as a reference, because its background looks homogeneous along the migration axis of the gel, and is smaller (Figure 3.5 and Gel Images Section (3.7.6)). All the monomer fraction measurements used in this chapter, before and after correction, are reported in Figure 3.6. For instance, there is a good agreement between the corrected monomer fractions of se47 at $0.2\mu M$ in Gel 2 and Gel 3: $M_f^{gel2} = 0.72$ and $M_f^{gel3} = 0.71$. The correction improves the quality of the fit by the mathematical model presented later and the data. The correction improves slightly the agreement between the monomer fractions of se25 and se16 in different gels, but the discrepancy is larger than in se47. The main reason is that their monomer fraction are much smaller than those of se47: their monomeric circles band are faint, and their signals are difficult to measure accurately. Nevertheless, the corrected monomer fraction were used for the quantitative analysis presented in this chapter.

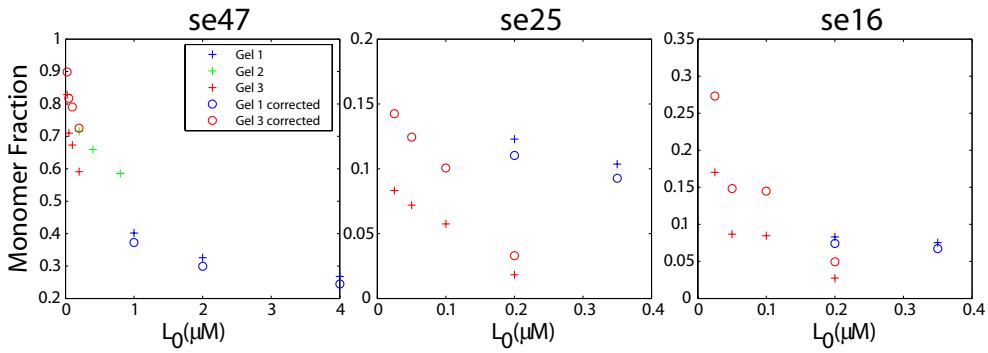

Figure 3.6: Corrected (circles) and uncorrected (crosses) monomer fractions for the three constructs se47, se25 and se16. Legend box for all three figures is inserted in the left graph (se47). **(se47)** A linear correction was applied to the monomer circles band intensities of Gels 1 and 3 (blue and red crosses) according to the ladder ratio of Gel 2 (see Equation 3.1). After correction, Gel 3 monomer fractions (red circles) fit better data of Gel 2 (green crosses). **(se25 and se16)** The correction is less efficient than for se47. The values for $L_0 = 0.2\mu M$ are closer than before correction, but the error stays large, especially for se25. The main reason appears to be that the signals from se25 and se16 monomer bands are particularly weak: their intensities are therefore difficult to measure accurately, as are the corresponding monomer fractions.

3.5 Interpretation of the Results

3.5.1 Identification of the Products

The first part of the results section consists in observing the obtained products in gels and comparison with the model described in Figure 3.2. The gel images and the uncorrected monomer fractions are presented in the Gel Images section (3.7.6).

As expected in Gel 2 (Figure 3.12a), there are bands of various sizes in the lanes with long sticky ends constructs, whereas there is only one band in the lane of the construct se6, whose sticky ends (6 bp) are too short to form stable unligated circles or longer products (see Figure 3.12a). This indicates that the sticky ends of length 47 bp, 25 bp and 16 bp are long enough to hybridize and to stay stable in the agarose gels (Figures 3.11, 3.12 and 3.13). The speed of migration of se6 is slightly greater than those of the fastest products of se47 and se43g (corresponding to the lowest bands in the gel image), which suggests that the se47 and se43g lowest bands correspond to

monomeric circles, as opposed to the linear monomers formed by se6. We also observe that se43g monomeric circles, which contain two gaps in place of se47 nicks, migrate slightly faster than se47 monomeric circles, probably because of their elongated shape (see Chapter 2, where se43g is called Og2 and se47 is called On2).

The hypothesis that the lower bright bands are monomeric circles is supported by electron micrographs of the annealing products of se47 and se43g, where we see minicircles, but no linear DNA molecules of the length of the shortest minicircles (Figure 3.7). It seems that all the monomeric linear molecules underwent cyclization or bimolecular association by the end of the reaction. However longer molecules are found in both linear and circular forms. Therefore the hypothesis that the fastest bands are only composed of circular products seems to be verified, and it is possible to use this band to compute the circular monomer fraction (a precise definition of which is given in Section 3.4.3).

In the lanes of se43g (Figure 3.12), we can see a faint band between the monomeric circles and the dimeric circles. This band probably corresponds to what I will refer to as an odd product. An odd product is a linear product composed of an odd number of oligonucleotides. In Figure 3.2, these products are named A3, A5 or B3: they are all the products with a name which finishes by an odd number greater than one (with A1 and B1 not being included because they are not products). The faint band we observe in Figure 3.12 is probably of type A3 or B3, because it lies between the monomeric and dimeric circles. In order to confirm this hypothesis, I made an annealing of oligonucleotides 2a and 1a with a large excess of oligonucleotide 2a (four times more 2a than 1a). Because of the large excess of 2a, odd products with two 2a and one 1a are very likely to form. Furthermore, odd products cannot cyclize because their sticky ends are not complementary. Therefore, when oligonucleotides 1a are no longer available in solution, odd products cannot undergo further polymerization reaction, and are visible in the gel. The result is visible in the lanes 4 and 5 of Figure 3.8. All the other annealings products shown in this gel were done with equal amounts of oligonucleotides. In lanes 4 and 5 we see bright bands which do not appear in other lanes. The band of the odd product lies in the middle between the monomeric and the dimeric circles, as in Gel 2 (Figure 3.12). It is interesting to notice that this

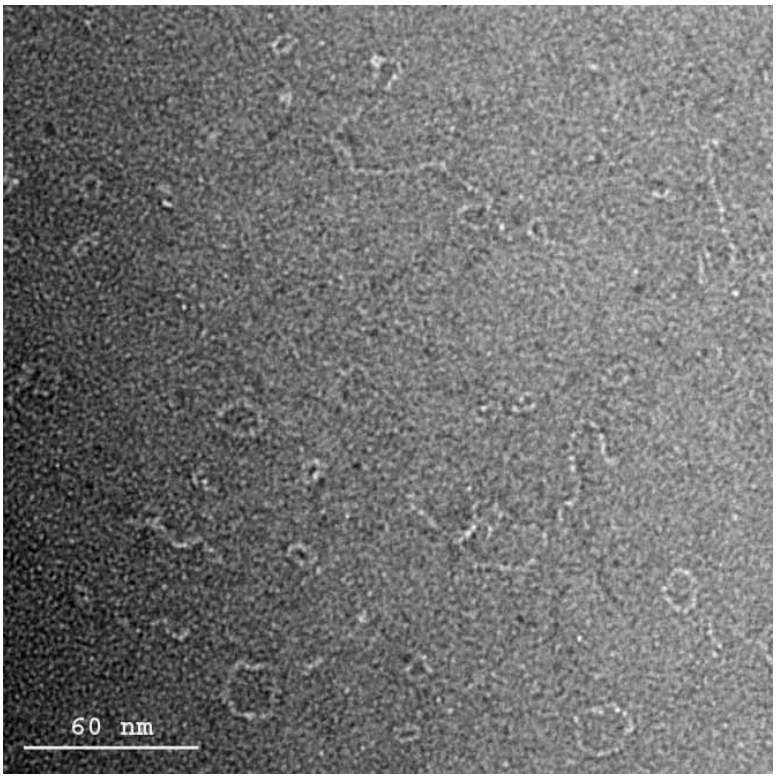

Figure 3.7: Electron micrograph of the products of the annealing-cyclization reaction of the construct se43g. Several short circular products are visible, but no short linear molecules, suggesting that all the linear monomers where consumed by hybridization reaction. Therefore the fastest gel band seems to be made mainly of circular products, and can be used to compute the circular monomer fraction. Image courtesy of Davide Demurtas.

excess of 2a inhibited the formation of longer species: there is no smear at the top of lanes 4 and 5, in contrast to the other lanes. The lack of oligonucleotide 1a did not allow the linear DNA molecules to grow by hybridization.

3.5.2 Initial Concentration of the DNA oligos

Gels 1, 2 and 3 (Figures 3.11, 3.12 and 3.13) show the products of annealing-cyclization reactions at various concentrations and the associated monomer fractions. For every construct, the monomer fraction increases with decreasing concentration. Combining the three experiments, I measured the monomer fraction of se47 (47 bp SE) from 25 nM initial oligonucleotides

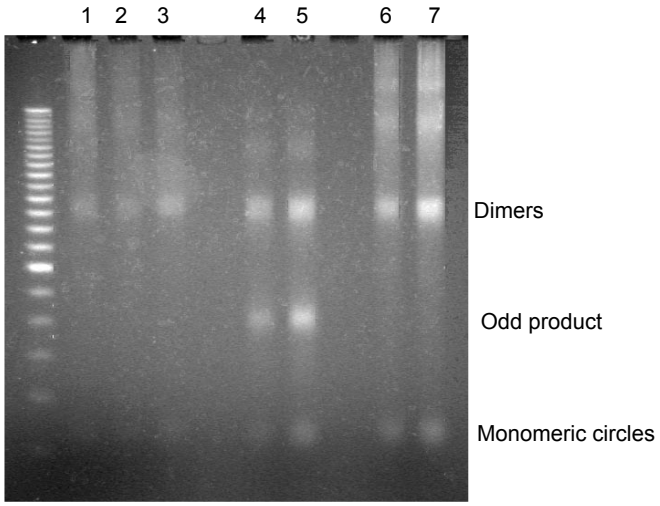

Figure 3.8: Excess of oligonucleotide 2a (lanes 4 and 5). A thick band appeared between the monomeric and dimeric circles. It is probably the odd product formed by three oligonucleotides (A3 or B3 in Figure 3.2).

concentration (L_0) up to 4000 nM, which represents a 160 fold difference. Although the monomer fractions come from three different gels with very different L_0 values, after correction, the monomer fraction monotonically decreases with concentration (Figure 3.6, se47). This result can be understood intuitively for the point of view of a linear monomer: the lower the concentration, the less probable it is that the monomer will find another molecule to form a dimer, and the monomer fraction increases.

In order to interpret further the monomer fractions with respect to initial concentration, a reaction rate model for the annealing-cyclization reaction is described in the next section. The fit between the model and the corrected data is presented in Section 3.7.

3.6 The Mathematical Model

A simplified model of the annealing-cyclization equation can be described by the system chemical equations:

$$L \xrightarrow{k_1} C, \tag{3.2}$$

$$2L \xrightarrow{k_2} D. \tag{3.3}$$

This is the system used to model the ligase-catalyzed cyclization reaction [46]. L represents the linear monomers of double stranded DNA, C the circular monomers and D is the linear dimers. The single stranded oligos are not considered in the chemical system: as discussed in the definition of the monomer fraction (section 3.4.3), it is assumed here that all the single stranded oligos form double stranded linear monomers (L) before the system reaches a temperature at which the sticky ends can hybridize.

In the ligase-catalyzed assay, k_1 and k_2 are the reaction rate constants of the irreversible DNA ligation reactions. As discussed in the experimental design (section 3.2), the annealing cyclization goes to completion so it can be viewed as irreversible. The backward reactions are therefore neglected.

The reaction of association between linear monomers and longer linear products is neglected here, as in previous works using ligase (for instance in [6]). Indeed, if the quantity of trimers and longer DNAs produced by the reaction is low, the chemical reaction $L + D \rightarrow T$ may be neglected. In these conditions, the cyclization and association of longer linear products between them will not affect the monomer fraction, therefore they are not considered in the model.

The evolution of the reactants and product concentrations as a function of time is described by a system of ordinary differential equations (ODE) called reaction rate equations, which follow from chemical equations (3.2) and (3.3):

$$\frac{d[L]}{dt} = -k_1[L] - 2k_2[L]^2, \tag{3.4}$$

$$\frac{d[C]}{dt} = k_1[L], \tag{3.5}$$

$$\frac{d[D]}{dt} = k_2[L]^2. \tag{3.6}$$

This system satisfies mass conservation, which states that the total mass of compounds is constant in time:

$$\frac{d[L]}{dt} + \frac{d[C]}{dt} + \frac{2d[D]}{dt} = 0. \tag{3.7}$$

Equation (3.4) has only one variable and can be solved by integration: when it is combined with the initial condition $[L]_0 = L_0$,

$$[L]_t = \frac{1}{(\frac{1}{L_0} + \frac{2k_2}{k_1})e^{k_1 t} - \frac{2k_2}{k_1}}. \tag{3.8}$$

Equation (3.8) is equivalent to equation (16) given by Crothers and collaborators in [46], the only difference being that k_2 has a factor of 2 and not of 4. Crothers et al. use a slightly different definition of k_2 which accounts for the palindromic sticky ends, which allow four ways of forming linear dimers whereas in our case the sticky ends are not palindromic, so there are only two ways of forming linear dimers from two linear monomers. We next use the expression for $[C]_t$,

$$[C]_t = \frac{k_1}{2k_2} ln\Big(1 + \frac{2k_2}{k_1} L_0(1 - e^{-k_1 t})\Big), \tag{3.9}$$

which can be derived by substitution of (3.8) in (3.5) and integration, using the initial conditions $[C]_0 = 0$ and $[D]_0 = 0$. I defined earlier the monomer fraction as the proportion of linear monomers which turned into monomeric circles at the end of the reaction. The monomer fraction M_f is given by

$$\lim_{t\to\infty} \frac{[C]_t}{L_0} = \frac{k_1}{2k_2} \cdot \frac{ln(1 + \frac{2k_2}{k_1} L_0)}{L_0}. \tag{3.10}$$

Not surprisingly, equation (3.10) satisfies

$$\lim_{L_0\to 0} M_f = 1, \qquad \lim_{L_0\to\infty} M_f = 0.$$

These limits can be understood intuitively in the following manner: at very low initial concentration L_0 the DNA molecules are very distant from one another and have a very low probability to undergo a bimolecular association. Therefore the monomer fraction goes to one. When on the contrary L_0 is very high, the probability of bimolecular association becomes much higher than the probability of cyclization and the monomer fraction goes to zero.

I denote $J_a = k_1/k_2$, as in reference [46] where this ratio is referred to as the apparent J factor. Indeed, in the ligase-catalyzed cyclization assay the J factor is determined by measuring k_1 and k_2 under conditions in which $J = k_1/k_2$ (cf. Section 0.3). It can be shown that the function M_f defined in

equation (3.10) is an increasing function of J_a. Thus, the higher the apparent J factor, the higher the monomer fraction is, which is also coherent with the definition of the real J factor.

3.7 Interpretation of the experimental results using the mathematical model

Equation (3.10) provides a mathematical model for the monomer fractions that were measured from the gel images. The monomer fraction of equation (3.10) can be fit to the data by a least squares minimization over the single parameter k_1/k_2 ($= J_a$), and thus J_a can be computed.

3.7.1 Initial Concentration of the DNA oligos

The monomer fraction of se47 and the fitted model are plotted for various initial DNA concentrations in Figure 3.9. The apparent J factor given by the least square fitting procedure is $0.56\mu M$. It seems like there is still discrepancy between the data of different gels. In order to estimate the error in the measurement of J_a, I measured J_a from Gel 2 data only. The result is $0.71\mu M$, which is 27% higher than the previous value: this error is not very high in regard to J factor values, which can often vary by orders of magnitude.

J_a is extremely high even compared to the J factors reported by Cloutier and Widom with the same sequence, but with the ligase-catalyzed cyclization method. The method is different in several points and the values cannot be compared straightforwardly. Nevertheless, a striking difference between the two experiments is the length of the sticky ends used: 4 base pairs in the Cloutier and Widom experiment, in contrast with the 47 base pairs of the construct se47. In fact J_a depends on the SE length. Indeed, the constructs with 47 bp SE have a much higher J_a than the constructs with 25 or 16 bp SE.

3.7.2 Length of the Sticky Ends

Monomer fractions of the constructs se25 and se16 were measured in gels 1 and 3. In Figures 3.6, 3.11b and 3.13b it is visible that in each gel, se25

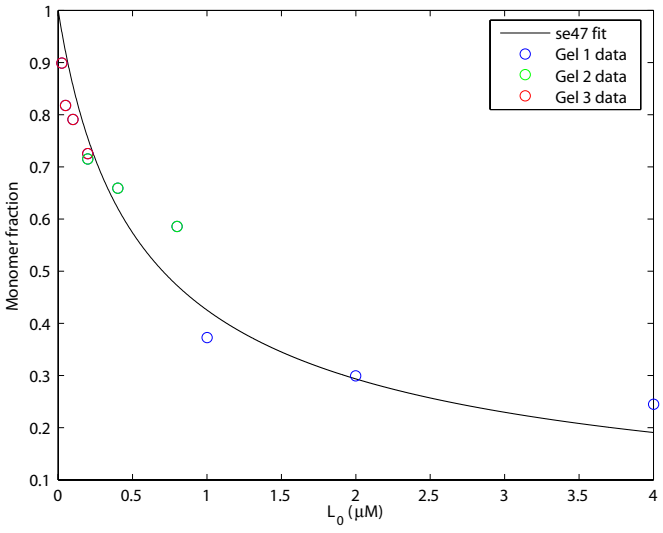

Figure 3.9: Fit of the model to the corrected monomer fractions for construct se47. The model is

$$M_f(L_0) = \frac{k_1}{2k_2} \cdot \frac{ln(1 + \frac{2k_2}{k_1} L_0)}{L_0},$$

where M_f is the monomer fraction, L_0 is the initial concentration of oligonucleotides, and k_1/k_2 is the ratio of the cyclization and the bimolecular association rate constants. The fitting procedure gives $k_1/k_2 = 0.56\mu M$. Data from Gels 1, 2 and 3 are represented by colored circles, as indicated in the inserted box.

and se16 monomer fractions are increasing while L_0 is decreasing. As for se47, I compared the monomer fractions of the same construct at $L_0 = 200$ nM. After corrections, I obtained the following values: $Mr_{se16}^{Gel1} = 0.075$, $Mr_{se16}^{Gel3} = 0.049$, $Mr_{se25}^{Gel1} = 0.110$ and $Mr_{se25}^{Gel3} = 0.033$. Even after correction, the values found do not agree well between the gels: the values in Gel 3 are smaller than the ones of Gel 1. This might be explained by the fact that the monomer fractions of se25 and se16 are much smaller than the ones of se47. When in addition the initial concentration of the reaction L_0 is low, the amount of monomer circles in solution is particularly low and the corresponding bands in the gels are very faint. Thus, the error in measuring the monomer ratio is big. There was less DNA loaded in Gel 3 because the annealing-cyclizations were performed at lower L_0 (between 25 and 200

nM) than in Gel 1, so the monomer circles are difficult to see in Gel 3. For these reasons, I did not consider data from Gel 3 in the comparison of se25 and se16, and rather used the data from Gel 1.

The monomer fractions and their corresponding J_a values for the three constructs are reported in the following table:

	se47	se25 (Gel 1)	se16 (Gel 1)
$M_f(0.2\mu M)$	0.71	0.11	0.07
J_a (nM)	560	14.2	8.6

There is a large difference between se47 and the two other constructs. It seems that the shorter the SE, the lower J_a is. Hence, I conclude that the SE length has a strong effect on the monomer circle formation efficiency, with the highest efficiency when the SE are as long as half of the monomer length (the longest possible SE): this design increases J_a by between about one or two orders of magnitude, and should be used if large amounts of minicircles are needed, for instance in minicircle preparation for visualization by electron microscopy.

3.7.3 Nicked and Gapped Minicircles

I present here the comparison of the annealing-cyclization which leads to the formation of the minicircles investigated in Chapter 2, On2 and Og2. Figure 3.10 shows that the monomer fractions are slightly higher for se43g (Og2) than for se47 (On2), for the three values of L_0 tested (see also Figure 3.12b). The J_a values were computed by the least square minimization procedure (Figure 3.10). The value for se43g was fit from the monomer fractions measured in Gel 2, and is equal to $1.06\mu M$. The value of J_a of se47 measured only from Gel 2 is $0.71\mu M$, whereas it is $0.56\mu M$ when fitted to the full dataset (from gels 1, 2 and 3). Therefore the J_a of se43g is greater than the one of se47 by a factor of 1.5 to 2.

This difference is not striking compared to the effect of SE length on monomer fraction. The two constructs also have a difference in SE length because four nucleotides were removed from the constructs to create the gaps: the SE of Og2 hybridize on 43 base pairs, whereas the sticky ends

of two nicks hybridize on 47 base pairs. The J_a values of se47, se25 and se16 constructs suggest that the shorter the sticky ends, the smaller J_a is. Thus, if we consider only SE length, se43g should in principle have a smaller J_a value than se47. However, se43g cyclizes better than se47; its monomer fraction is higher. At the end of the cyclization reaction, most of the single stranded oligos are hybridized, whether they form circular monomers or dimers. Thus, se43g and se47 different monomer fractions may not be explained by hydrogen binding energy differences, but rather by the elastic energy of the products, suggesting that se43g minicircles have a lower elastic energy than se47 minicircles. It supports the hypothesis that the higher J_a value of se43g in comparison to se47 is due to the special properties of gaps compared to nicks, presumably their increased flexibility.

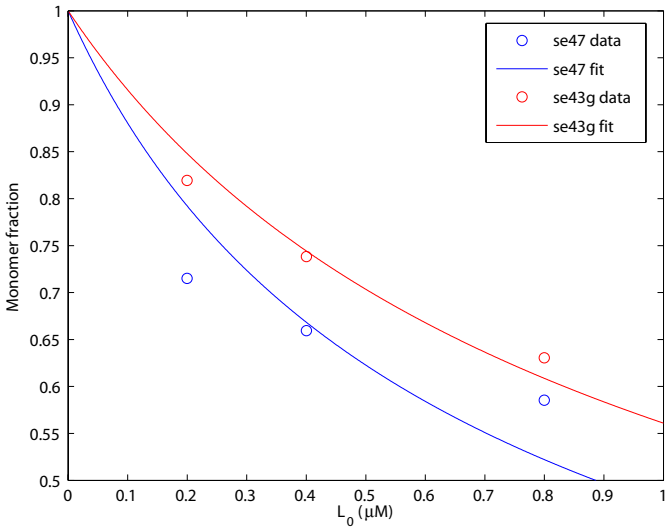

Figure 3.10: Comparison of se47 (blue) and se43g (red) (in Chapter 2 notation: On2 blue and Og2 red) with annealing cyclization. For On2, the fitting procedure gives $k1/k2 = 0.71\mu M$, and for Og2 $k1/k2 = 1.06\mu M$.

3.7.4 k_1 and k_2 in the annealing-cyclization model

The meaning of k_1 and k_2 here are not exactly the same as in the ligase-catalyzed assay. The ligation reaction is chemically irreversible, whereas the annealing-cyclization reaction is irreversible only at the end of the re-

action, when temperature is low enough to consider that the sticky ends do not dissociate. Nevertheless the problem can be simplified. It can be reasonably considered that the double strand fraction of DNA is always increasing during the reaction; hence annealing reaction rates are always positive. Additionally if one assumes the reaction rates to depend only on the reactant concentration, the model can be written with irreversible rate constants k_1 and k_2 (equations (3.5) and (3.6)), k_1 or k_2 accounting for both the on and off rates of cyclization. There remains the question whether J_a depends on reaction cooling speed. Temperature variation will mainly affect the strength of the sticky-end binding, so k_1 and k_2 are in fact not constant. But the two reactions are occurring in the same test tube (thus at the same temperature), and through the same sticky ends, so that k_1 and k_2 temperature dependences are probably similar. Therefore it is not unreasonable to suppose that the ratio $J_a = k_1/k_2$ does not depend on cooling speed or on time. This assumption could easily be tested in future work with a temperature controlled water bath. If this assumption is true, the J_a value defined here depends on the DNA construct tested and not on the reaction conditions (cooling speed or initial concentration).

Furthermore, although the model is very simple and has only one parameter (J_a), it yields an acceptable fit with the data. Because J_a is the only parameter of the fit, in principle as few as one annealing-cyclization reaction is enough to determine J_a (if it does not depend on cooling speed), whereas ligase-catalyzed cyclization assays require time courses of the reaction to fit k_1 and k_2 rate constants, and it has to be repeated for several different ligase concentrations.

3.7.5 Conclusion

In this chapter results of the annealing-cyclization reaction were described.

Data The data observed in the experiments is the monomer fraction: it is the amount (the total mass) of monomer circles divided by the total amount of DNA in the reaction tube. Equivalently, it is the fraction of double stranded DNA linear monomers which turned into monomeric circles. The monomer fraction was measured from gel images, by computing the

ratio of the monomer circles band intensities over all the clearly measurable bands in the same lane. There exists error in measurement of the monomer fractions. In order to compare values measured in different gels, I computed a correction using the DNA ladder bands.

Model I proposed to model the annealing-cyclization reaction using the reaction rate model of the ligase-catalyzed reaction given by Crothers and co-workers in [46] (coming from Shore, Langowski and Baldwin [8, 9, 10] originally). I derived an equation of the monomer fraction with respect to L_0 and the ratio $J_a = k_1/k_2$ (L_0: initial concentration of linear monomers, k_1: cyclization rate constant and k_2: bimolecular association rate constant). In the ligase-catalyzed method, J is defined as K_C/K_D (the ratio of the equilibrium constants at the constant temperature of the experiment), which is equal to k_1/k_2 under some conditions. But our experiment is not at constant temperature. To discover how the J_a of the annealing-cyclization relates to the J factor (K_C/K_D), and more generally how J_a relates to the mechanical parameters of DNA is certainly a long term goal initiated by this study.

Results I compared the monomer fractions obtained from the annealing-cyclization experiments to the model equation. I found the values of J_a which give the best fit between the model and data. (i) Our model provides a coherent interpretation of the annealing-cyclization reaction, even though it is difficult to measure the monomer fraction from the gel. Indeed the monomer fraction needed to be corrected for the sole purpose of consistency between gels; measurement may be improved by using specific ladders for band intensity measurements, or by using radioactive labels and a phospho-imager to get more precise results, without background. Nevertheless the mathematical model of the reaction fits the data over a large range of initial oligonucleotide concentrations. (ii) It seems that the shorter the sticky ends, the lower the monomer fraction, and thus the lower J_a. It may be because of the flexibility of the sticky ends compared to the stiffness of the double stranded part of the molecule. We find that for constructs whose sticky-end length is half the total length (in fact the longest possible sticky-end length) the monomer circle formation is greatly enhanced. (iii) J_a is higher for the gapped construct than for the nicked construct with approximately the same sticky-end length; it is probably because the gaps are more flexible

than nicks.

Conclusion Annealing-cyclization is a very simple experimental procedure compared to the ligase-catalyzed reaction which involves many ligation reactions and quenching in order to deduce the cyclization rate constants. Nevertheless, it allows the study of cyclization of very short DNA fragments (94 bp in this study) with a large yield of monomeric circles (the monomer fractions are often greater than 0.5 when the sticky ends are very long, see Figure 3.12). Finally, it successfully detected differences in the cyclization ability of gapped and nicked minicircles, as for the three different initial concentrations the monomer fraction of the gapped construct was always higher than in the nicked construct. Annealing-cyclization is therefore a simple method that can probe the ability of DNA molecules to cyclize, and that is sensitive to the difference between nicks and gaps. With further improvements in the model and quality of the gel measurements, it may be a valuable tool to investigate the mechanical parameters of DNA.

3.7.6 Gel Images

The gel images and the associated monomer fractions (uncorrected) are presented in this section. The monomer fractions are not corrected in the next figures to allow the reader to compare the images and the measured values. The data of the quantitative analysis were measured from the three gel images presented here.

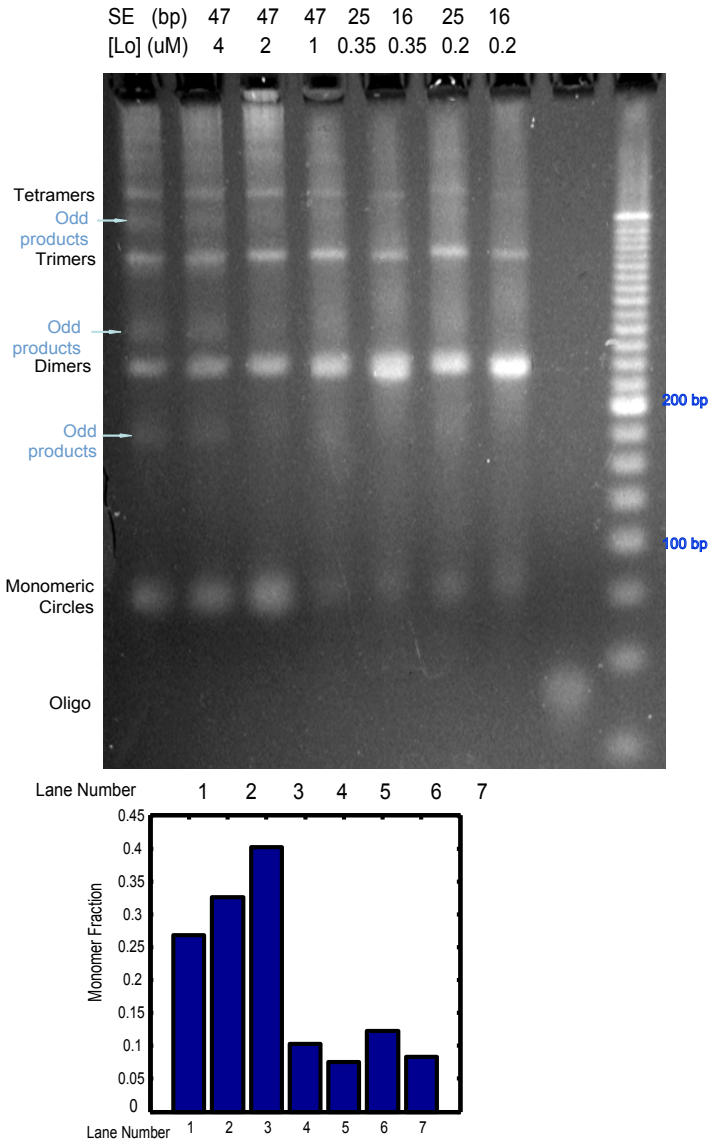

Figure 3.11: Gel 1. Products of annealing-cyclization reactions of several constructs, with different initial oligonucleotide concentrations. **(a)** Each numbered lane (numbers are written below the gel) contains the products of one annealing-cyclization reaction. Sticky ends length and initial oligo concentration are indicated above each lane. The rightmost lane is the DNA ladder. The species names are written near the bands on the left of the gel image. **(b)** Uncorrected monomer fractions of the numbered lanes of gel shown in (a). It is visible that the monomer fraction increases with (i) initial oligonucleotide concentration L_0 and (ii) the sticky-end length.

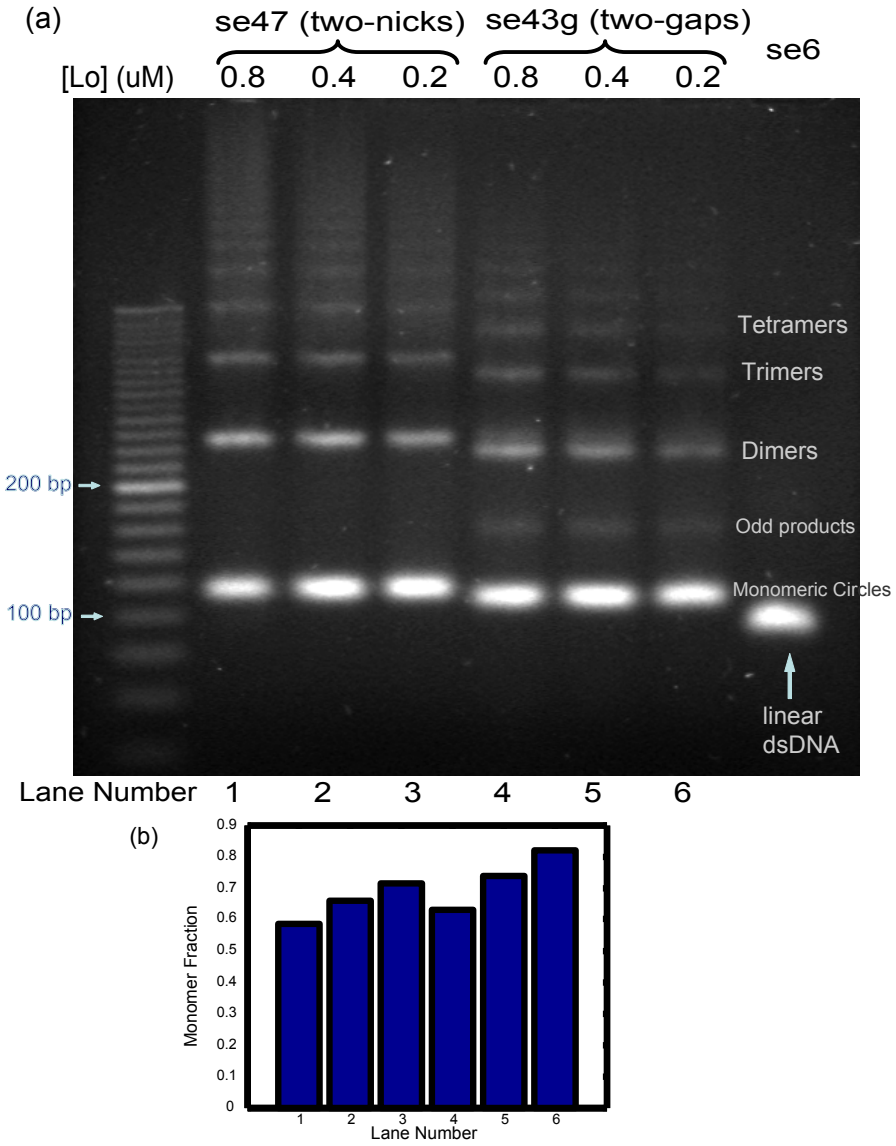

Figure 3.12: Gel 2. **(a)** Gel image of the annealing-cyclization products of the three constructs se6, se47 and se43g. The construct se6 whose sticky ends are too short to hybridize produced only linear monomers of double stranded DNA. In contrast the constructs se47 and se43g produce DNA polymers of various lengths. The fastest products migrate slower that se6 linear monomer of the same length, and are circular products. The circular nature of the shortest products of se47 and se43g has been confirmed by electron microscopy (Figure 3.7). **(b)** The circular monomer fractions have been measured for the se47 and se43g constructs at various initial oligo concentrations. It is visible that (i) the monomer fraction increases with decreasing initial concentration, and that (ii) for each given initial concentration, the monomer fraction of se43g is higher than the one of se47. This gel was prepared in collaboration with Davide Demurtas.

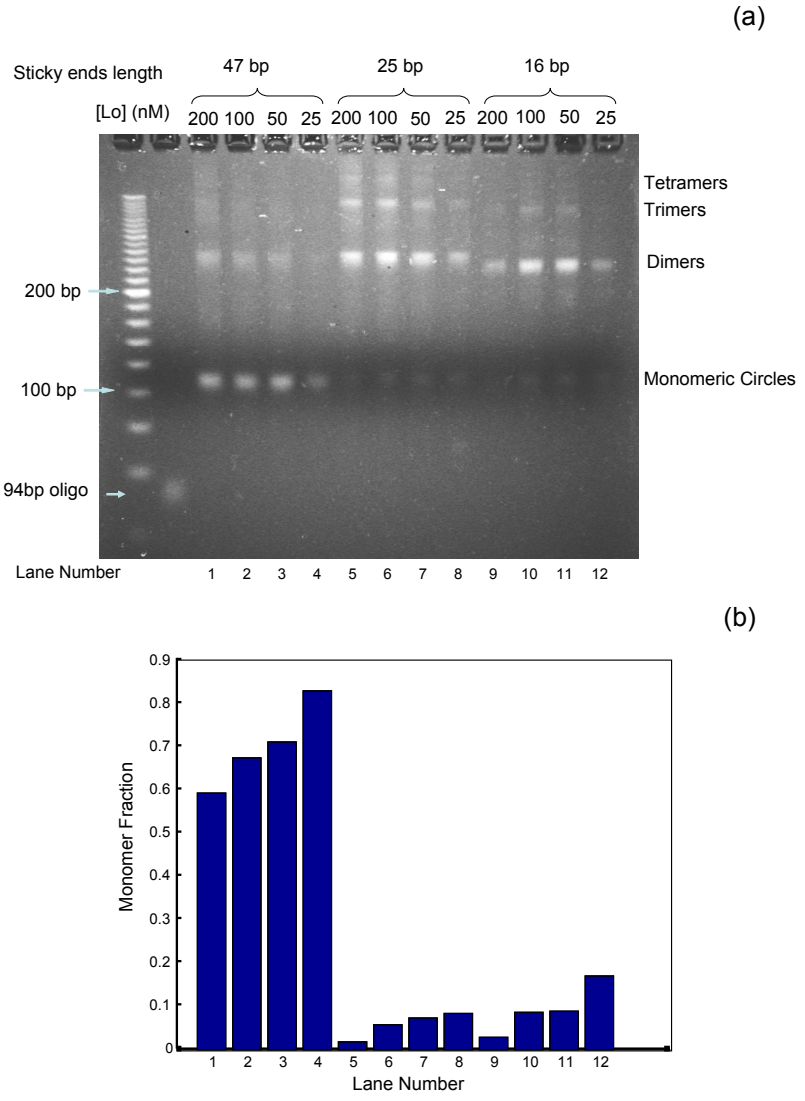

Figure 3.13: Gel 3. **(a)** Gel image of the annealing-cyclization products of the constructs se47, se25 and se16, at different initial oligonucleotide concentrations (L_0). The dimers of se16 migrate faster than se25 or se47 dimers. The position of the four nicks along the DNA molecule differs between the three circular dimers, which may affect the shape and therefore the migration speed of the molecules. This difference in migration speed is not seen in gel Gel 1, which was stained with SYBR$^{\text{TM}}$Safe only after migration This gel was cast with SYBR$^{\text{TM}}$safe. **(b)** Uncorrected monomer fractions of the numbered lanes of gel shown in (a). As in Gel 1, it is visible that the monomer fraction increases with (i) initial oligonucleotide concentration L_0 and (ii) the stick-end length.

Chapter 4

Conclusion

In this thesis I investigated effects of nucleotide sequence mutation and of nick and gap insertions on the shape and formation of DNA minicircles.

The results of Chapter 1 suggest that a large difference in J factors between two different minicircle sequences is not enough to imply a visible difference in the minicircle shapes using stereo cryo-electron microscopy. Even though the two tested 158 bp sequences differed only in 14 bp that contained either a CAP site or a TATA box, they had J factors of 75 and 3500 nM, respectively. The two sequences shared six periodic phased A-tracts, which are believed to be curved and stiff. I designed a method that successfully detected clusters of shapes in the analyzed data, suggesting that the conformational energy landscape of the studied minicircles have probably at least two minima, one composed of planar shapes and one composed of circular shapes bent out of plane (Figure 1.7). However both the TATA and CAP minicircles populated all the detected clusters, suggesting that the set of minimal energy shapes was dominantly determined by the mechanical properties of the A-tracts and not by the TATA or CAP sequences. It is possible that effects induced by the presumed greater flexibility of the TATA box compared to the CAP site were not detected because the large common portion of both minicircle sequences containing six phased A-tracts (about 70 base pairs) was already curved so much as to dominate the observed shapes.

Thanks to these observations, new minicircles were designed and produced in order to allow easier detection of minicircle shape differences by

electron microscopy. Such minicircles were analyzed in Chapter 2. First, the new sequences had more drastic differences, namely the removal of one phosphate group (a nick), or the removal of two succesive nucleotides along one strand of the DNA molecule (a gap). Second, there were two test sequences (nicks or gaps) per minicircle instead of one. Third, the test sequences were placed at diametrically opposite sites of the minicircle. Fourth, the sequences were shorter (94 bp instead of 158 bp), which also raised the possibility of performing molecular dynamics simulations of the minicircles in explicit solvent. Thanks to this design, I could observe a clear signature of the gap effect in both micrographs and simulations, namely that the axes ratios (or aspect ratios) were larger in gapped minicircles than in nicked minicircles (Figure 2.3). The interpretation is that the DNA molecule tends to be straight, except at the gap sites, where the gap flexibility allows high curvature of the DNA double helix. This interpretation was supported by the molecular dynamics simulations, which in turn suggested that base unpairing events can happen at the nick and gap sites and facilitate DNA double helix bending. However, I did not observe a significant difference in the axes ratios of the 94 bp nicked minicircles and the longer minicircles of Chapter 1 (158 bp). Because of the exceptionally small size of the minicircles, it was difficult to obtain many stereo cryo-electron micrographs. Therefore most of the shapes presented in Chapter 2 come from minicircles that were adsorbed on a carbon surface. Hence they were perhaps less subject to thermal motion than free DNA in solution, and the observed shapes may then have had a smaller variation than if the minicircles had been observed by stereo cryo-electron microscopy. This issue will be probably clarified with new images of such minicircles by cryo-electron microscopy, an ongoing work by Davide Demurtas at UNIL.

The successful shape detection of Chapter 2 hints that it may be possible to design DNA minicircles for the purpose of sequence-effect detection by electron microscopy. A new sequence could be designed with two presumably flexible test sequences, for instance TATA boxes, at diametrically opposite sites of a minicircle. Such minicircle shapes could be compared with shapes of minicircles which carry gaps, or nicks, instead of the TATA boxes. Alternatively, one minicircle could have a gap and a TATA box at a diametrically opposite site. If the TATA box is indeed particularly flexible,

its effect should resemble that of a gap.

In Chapter 3, I presented the analysis of short minicircle formation efficiency using a novel method, annealing-cyclization, a ligase-free method used to produce the nicked and gapped minicircles of Chapter 2. I showed that with a maximum sticky-end length design, a very high yield of (nicked) minicircles is achieved: for instance, annealing-cyclization of the construct se47 at $0.2\mu M$ initial DNA concentration leads to a proportion of about 70% of circular monomers at the end of the reaction (Figure 3.9). This high yield is useful when high concentrations of exceptionally short DNA minicircles are needed, for instance for cryo-electron microscopy. I proposed an equation that describes the circular monomer fraction as function of the initial DNA concentration. Thanks to this model I defined a measure (J_a) associated to the tested DNA construct which does not depend on DNA initial concentration, and which reflects the yield of formation of DNA minicircles. I showed that J_a depends sensitively on sticky-end length, or equivalently on the nick positions in the resulting minicircle sequences. Also, the experimental analysis suggests that the gapped minicircle Og2 has a higher formation efficiency than the nicked minicircle On2. I believe the reason is the increased flexibility of DNA at the gap sites, as observed in Chapter 2. As a perspective, the relation –which is not yet clear– between J_a and the previously defined J factor, could be further investigated in order to compute and predict J_a with sequence-dependent polymer models of DNA. Also, this novel method allows us to measure cyclization ability of DNA constructs in a much simpler way than ligase catalyzed cyclization assays. Additionally, the precision of the band intensity measurements by gel electrophoresis could be improved by using radioactive labels and a phosphoimager. Monomer circles with short sticky ends may then be more easily detectable in the gel, and different sequences could be tested, and indicate whether or not one can detect sequence effect by measurement of J_a.

Finally, I showed that with an appropriate design of DNA constructs, electron microscopy and annealing-cyclization allow detection of differences induced by insertion of gaps. Although in the current state of the art it is not as sensitive as ligase-catalyzed cyclization, direct visualization could provide important information on the nature of DNA as a material, and the

relevance of rod models for DNA. As a perspective, further improvements in resolution may allow detection of more subtle sequence-dependent effects on DNA minicircle shapes determined by electron microscopy; improvements in resolution could be measured thanks to the notion of shape-distance defined in Chapter 1. Appropriate polymer models accounting for gap flexibilities as well as sequence effects may then be used for computation of minimal energy shapes of the proposed DNA minicircles (with a gap and a TATA box at diametrically opposite sites), and comparison with 3D reconstructions from stereo cryo-electron microscopy using the shape-distance tool. This way, direct visualization of DNA minicircle could be used in the aim to quantify stiffness of tested sequences with respect to gaps. I also proposed a ligase-free experiment to measure efficiency of formation of minicircles. If polymer models can be adapted to compute the J_a factor, annealing-cyclization –a reaction which involves only naked DNA– would allow investigation of the mechanical properties of DNA.

Bibliography

[1] Manning, R. S., Maddocks, J. H., and Kahn, J. D. (1996) A continuum rod model of sequence-dependent DNA structure. *J. Chem. Phys.,* **105**(13), 5626–5646.

[2] Moakher, M. and Maddocks, J. (2005) A Double-Strand Elastic Rod Theory. *Archive for Rational Mechanics and Analysis,* **177**(1), 53–91.

[3] Shimada, J. and Yamakawa, H. (1984) Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules,* **17**(4), 689–698.

[4] Flory, P., Suter, U., and Mutter, M. (1976) Macrocyclization equilibriums. 1. Theory. *Journal of the American Chemical Society,* **98**(19), 5733–5739.

[5] Levene, S. D. and Crothers, D. M. (1986) Ring closure probabilities for DNA fragments by Monte Carlo simulation. *J. Mol. Biol.,* **189**(1), 61–72.

[6] Taylor, W. H. and Hagerman, P. J. (1990) Application of the method of phage t4 dna ligase-catalyzed ring-closure to the study of dna structure. *J. Mol. Biol.,* **212**, 363–376.

[7] Zhang, Y. and Crothers, D. M. (2003) High-throughput approach for detection of DNA bending and flexibility based on cyclization. *Proc. Natl. Acad. Sci. U S A,* **100**(6), 3161–3166.

[8] Shore, D., Langowski, J., and Baldwin, R. L. (1981) DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U S A,* **78**(8), 4833–4837.

[9] Shore, D. and Baldwin, R. L. (1983) Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J. Mol. Biol.,* **170**(4), 957–981.

[10] Shore, D. and Baldwin, R. L. (1983) Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol.,* **170**(4), 983–1007.

[11] Voet, D. and Voet, J. (1994) Biochemistry, second edition, John Wiley & Sons, New York, .

[12] Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia,* **6**(6), 201–209.

[13] Watson, J. D. and Crick, F. H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* **171**(4356), 737–738.

[14] Mathews, C. and Van Holde, K. (1990) Biochemistry, Benjamin/Cummings Redwood City, Calif, .

[15] Watson, J. D. and Crick, F. H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature,* **171**(4361), 964–967.

[16] Van Holde, K., Johnson, W., and Ho, P. (2000) Principles of physical biochemistry, second edition, Pearson Prentice Hall Upper Saddle River, NJ, .

[17] Calladine, C. and Drew, H. (1997) Understanding DNA, Academic Press San Diego, .

[18] Bolshoy, A., McNamara, P., Harrington, R. E., and Trifonov, E. N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U S A,* **88**(6), 2312–2316.

[19] De Santis, P., Palleschi, A., Savino, M., and Scipioni, A. (1990) Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry,* **29**(39), 9269–9273.

[20] Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U S A,* **95**(19), 11163–11168.

[21] Bednar, J., Furrer, P., Katritch, V., Stasiak, A. Z., Dubochet, J., and Stasiak, A. (1995) Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.,* **254**(4), 579–594.

[22] Ulanovsky, L., Bodner, M., Trifonov, E., and Choder, M. (1986) Curved DNA: Design, Synthesis, and Circularization. *Proc. Nat. Acad. Sci. U S A,* **83**(4), 862–866.

[23] Koo, H., Drak, J., Rice, J., and Crothers, D. (1990) Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry,* **29**(17), 4227–4234.

[24] Wu, H. and Crothers, D. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature,* **308**(5959), 509–513.

[25] Hagerman, P. (1984) Evidence for the Existence of Stable Curvature of DNA in Solution. *Proc. Nat. Acad. Sci. U S A,* **81**(15), 4632–4636.

[26] Watson, J., Hopkins, N., Roberts, J., Steitz, J., and Weiner, A. (1987) Molecular biology of the gene., Benjamin/Cummings Publishing. Menlo Park, CA, .

[27] Hagerman, P. J. (1988) Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.,* **17**, 265–268.

[28] Grayson, P., Evilevitch, A., Inamdar, M., Purohit, P., Gelbart, W., Knobler, C., and Phillips, R. (2006) The effect of genome length on ejection forces in bacteriophage lambda. *Virology,* **348**(2), 430–436.

[29] Garcia, H., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P., Phillips, R., Widom, J., and Wiggins, P. (2006) Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers,* **85**(2), 115–130.

[30] Tolstorukov, M. Y., Virnik, K. M., Adhya, S., and Zhurkin, V. B. (2005) A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Res.,* **33**(12), 3907–3918.

[31] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* **389**(6648), 251–260.

[32] Richmond, T. J. and Davey, C. A. (2003) The structure of DNA in the nucleosome core. *Nature,* **423**(6936), 145–150.

[33] Widlund, H. R., Cao, H., Simonsson, S., Magnusson, E., Simonsson, T., Nielsen, P. E., Kahn, J. D., Crothers, D. M., and Kubista, M. (1997) Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.,* **267**(4), 807–817.

[34] Anderson, J. D. and Widom, J. (2000) Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.,* **296**(4), 979–987.

[35] Virstedt, J., Berge, T., Henderson, R. M., Waring, M. J., and Travers, A. A. (2004) The influence of DNA stiffness upon nucleosome formation. *J. Struct. Biol.,* **148**(1), 66–85.

[36] Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I., Wang, J., and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature,* **442**, 772–778.

[37] Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem.,* **61**, 199–223.

[38] Vilar, J. M. G. and Leibler, S. (2003) DNA looping and physical constraints on transcription regulation. *J. Mol. Biol.,* **331**(5), 981–989.

[39] Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature,* **359**(6395), 505–512.

[40] Zhang, Y., Xi, Z., Hegde, R. S., Shakked, Z., and Crothers, D. M. (2004) Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. U S A,* **101**(22), 8337–8341.

[41] Ivanov, V. I., Minchenkova, L. E., Chernov, B. K., McPhie, P., Ryu, S., Garges, S., Barber, A. M., Zhurkin, V. B., and Adhya, S. (1995) CRP-DNA complexes: inducing the A-like form in the binding sites with an extended central spacer. *J. Mol. Biol.,* **245**(3), 228–240.

[42] Emmer, M., deCrombrugghe, B., Pastan, I., and Perlman, R. (1970) Cyclic AMP receptor protein of E. coli: its role in the synthesis of inducible enzymes. *Proc. Natl. Acad. Sci. U S A,* **66**(2), 480–487.

[43] Kahn, J. D. and Crothers, D. M. (1992) Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. U S A,* **89**(14), 6343–6347.

[44] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics,* **14**, 33–38.

[45] Jacobson, H. and Stockmayer, W. H. (1950) Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *J. Chem. Phys.,* **18**, 1600–1606.

[46] Crothers, D. M., Drak, J., Kahn, J. D., and Levene, S. D. (1992) DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.,* **212**, 3–29.

[47] Levene, S. D. and Crothers, D. M. (1986) Topological distributions and the torsional rigidity of DNA. A Monte Carlo study of DNA circles. *J. Mol. Biol.,* **189**(1), 73–83.

[48] Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W., and Schultz, P. (1988) Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.,* **21**(2), 129–228.

[49] Dubochet, J., Adrian, M., Dustin, I., Furrer, P., and Stasiak, A. (1992) Cryoelectron microscopy of DNA molecules in solution. *Methods Enzymol.,* **211**, 507–518.

[50] Adrian, M., tenHeggeler-Bordier, B., Wahli, W., Stasiak, A. Z., Stasiak, A., and Dubochet, J. (1990) Direct visualization of supercoiled DNA molecules in solution. *EMBO J.,* **9**(13), 4551–4554.

[51] Dustin, I., Furrer, P., Stasiak, A., Dubochet, J., Langowski, J., and Egelman, E. (1991) Spatial visualization of DNA in solution. *J. Struct. Biol.,* **107**(1), 15–21.

[52] Jacob, M., Blu, T., Vaillant, C., Maddocks, J., and Unser, M. (January, 2006) 3-D shape estimation of DNA molecules from stereo cryo-electron micro-graphs using a projection-steerable snake. *IEEE Transactions on Image Processing,* **15**(1), 214–227.

[53] Amzallag, A., Vaillant, C., Jacob, M., Unser, M., Bednar, J., Kahn, J., Dubochet, J., Stasiak, A., and Maddocks, J. (2006) 3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy. *Nucleic Acids Res.,* **34**(18), e125.

[54] Davis, N. A., Majee, S. S., and Kahn, J. D. (1999) TATA box DNA deformation with and without the TATA box-binding protein. *J. Mol. Biol.,* **291**(2), 249–265.

[55] Parvin, J. D., McCormick, R. J., Sharp, P. A., and Fisher, D. E. (1995) Pre-bending of a promoter sequence enhances affinity for the TATA-binding factor. *Nature,* **373**(6516), 724–727.

[56] Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., and Dickerson, R. E. (1996) How proteins recognize the TATA box. *J. Mol. Biol.,* **261**(2), 239–254.

[57] Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal structure of a yeast TBP/TATA-box complex. *Nature,* **365**(6446), 512–520.

[58] Paillard, G. and Lavery, R. (2004) Analyzing protein-DNA recognition mechanisms. *Structure (Camb),* **12**(1), 113–122.

[59] Gartenberg, M. R. and Crothers, D. M. (1988) DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nature,* **333**(6176), 824–829.

[60] Coutsias, E. A., Seok, C., and Dill, K. A. (2004) Using quaternions to calculate RMSD. *J. Comput. Chem.,* **25**(15), 1849–1857.

[61] Tsafrir, D., Tsafrir, I., Ein-Dor, L., Zuk, O., Notterman, D. A., and Domany, E. (2005) Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinformatics,* **21**(10), 2301–2308.

[62] Furrer, P. B., Manning, R. S., and Maddocks, J. H. (2000) DNA rings with multiple energy minima. *Biophys J,* **79**(1), 116–136.

[63] Billen, D. (1990) Spontaneous DNA damage and its significance for the "negligible dose" controversy in radiation protection. *Radiat. Res.,* **124**(2), 242–245.

[64] deMurcia, G. and Menissier deMurcia, J. (1994) Poly(ADP-ribose) polymerase: a molecular nick-sensor. *Trends Biochem Sci,* **19**(4), 172–176.

[65] Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature,* **362**(6422), 709–715.

[66] Timson, D. J., Singleton, M. R., and Wigley, D. B. (2000) DNA ligases in the repair and replication of DNA. *Mutat. Res.,* **460**(3-4), 301–318.

[67] Seeman, N. C. (2001) DNA nicks and nodes and nanotechnology. *Nano letters,* **1**(1), 23.

[68] Protozanova, E., Yakovchuk, P., and Frank-Kamenetskii, M. D. (2004) Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.,* **342**(3), 775–785.

[69] Yakovchuk, P., Protozanova, E., and Frank-Kamenetskii, M. D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.,* **34**(2), 564–574.

[70] Cloutier, T. E. and Widom, J. (2004) Spontaneous sharp bending of double-stranded DNA. *Mol. Cell,* **14**(3), 355–362.

[71] Yan, J. and Marko, J. F. (2004) Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.,* **93**(10), 108108.

[72] Wiggins, P. A., Phillips, R., and Nelson, P. C. (2005) Exact theory of kinkable elastic polymers. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.,* **71**(2 Pt 1), 021909.

[73] Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., and Vologodskii, A. (2005) Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U S A,* **102**(15), 5397–5402.

[74] Lankas, F., Lavery, R., and Maddocks, J. H. (2006) Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure,* **14**(10), 1527–1534.

[75] Lavery, R., Zakrzewska, K., and Sklenar, H. (1995) Jumna (junction minimization of nucleic acids). *Comput. Phys. Commun.,* **91**, 135–158.

[76] Lavery, R. and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn,* **6**(4), 655–667.

[77] Lavery, R. and Sklenar, H. (1988) The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dyn,* **6**(1), 63–91.

[78] Wang, J. C. and Davidson, N. (1966) On the probability of ring closure of lambda DNA. *J. Mol. Biol.,* **19**(2), 469–482.

[79] Rasband, W. (1997) ImageJ. *National Institutes of Health, Bethesda, Maryland, USA: http://rsb.info.nih.gov/ij,* **2005**.

AVENUE DE LA VALLOMBREUSE, 81 • 1008 PRILLY
PHONE 021 646 39 52 • E-MAIL ARNAUD.AMZALLAG@EPFL.CH

ARNAUD AMZALLAG

EDUCATION

2003-2007 Swiss Federal Institute of Technology Lausanne (EPFL)

***PhD**, Bioengineering and Biotechnology Doctoral Program*

- Theory and experimental observations of the DNA double-helix shape and flexibility, under the supervision of Prof. J.H. Maddocks (Institute of Mathematics B, EPFL) and Dr A. Stasiak (LAU, UNIL)

2002 Institut d'Informatique d'Entreprise, CNAM, Evry, France

***Mastère** (postgrade) in Bioinformatics*

2001 IMAC, Université Panthéon-Assas PARIS II, Paris, France

***Master Degree** (fr: diplôme d'ingénieur) in Multimédia and Architecture of Communication*

1998 Université PARIS XII, Créteil, France

***DEUG** in Mathematics and Computer science Applied to Sciences*

PROFESSIONAL EXPERIENCE

From 2003 (in progress) LCVMM, IMB, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

Junior researcher and teaching assistant

- See Teaching and Publication
- Webmaster of the LCVMM website and events (see for ex. the Workshop on Nanomechanics of Biomolecules web page)

2002 Naama Barkai Lab, Weizmann Institute of Sciences, Rehovot, Israel

Internship on systems biology

- Work on signal transduction in yeast, using a GFP reporter gene to monitor transcription factor concentration in vivo. Simulation and evolution of MAP-kinase network in silico.

2001 AIPC, Institut Universitaire d'Hématologie, Saint-Louis Hospital, Paris, France

Master degree project

- Simulation of cytoskeleton intermediate filaments network with L-systems.

2000

Developments and maintenance on web applications,

- www.tigersushi.com (2 days a week)
- www.ivision.fr, wapcorporate, summer 2000.

PUBLICATION

3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy , Arnaud Amzallag, Cédric Vaillant, Mathews Jacob, Michael Unser, Jan Bednar, Jason D. Kahn, Jacques Dubochet, Andrzej Stasiak and John H. Maddocks, *Nucleic Acids Research*, Oct 2006;**34**(18):e125

IMAGES AND ILLUSTRATIONS

- Cover page of the journal Nucleic Acids Research, Oxford University Press, vol. 34 issue 18
- Home page of the EPFL Basic Sciences Faculty, and an article summary.
- A stereo-pair of DNA minicircle cluster in the EPFL virtual tour

TEACHING

Teaching assistant at the chair of Applied Analysis, section of Mathematics, EPFL

- **2006—2007**
  - Linear Algebra, by Prof. JH Maddocks, for 1st semester Computer Science and Communication Systems sections
  - Analysis II (differential equations, vector fields), by Dr. P Chabloz, 2nd semester, Microtechnics section
- **2005—2006**
  - Linear Algebra, by Prof. M Cibils, for 1st semester Computer Science and Electric Engineering sections
  - Analysis II, by Dr. P Chabloz, for 2nd semester Microtechnics section
- **2004—2005**
  - Linear Algebra, by Prof. JH Maddocks, for 1st semester Computer Science and Electric Engineering sections
- **2003—2004**
  - Mathematical Modeling of DNA I, by Prof. JH Maddocks, master degree, Mathematics section
  - Mathematical Modeling of DNA II, by Prof. JH Maddocks, master degree, Mathematics section
  - Linear Algebra, by Prof. JH Maddocks, 1st semester Computer Science and Electric engineering sections

Lectures

- In Mathematical Modeling of DNA II course, on DNA experiments (2004)
- In Analysis II course, on differential equations applied to physics (2007)

POSTERS AND SEMINARS

- **2007**
  - Invitation to give a seminar at the 3D-EM Gordon Research Conference 2007
  - Group seminar Jason Kahn's lab, University of Maryland, College Park, MD, USA
- **2006**
  - Group seminar at Eytan Domany's lab, Weizmann Institute of Science, Rehovot, Israel

- Poster presentation at the Workshop on Nanomechanics of Biomolecules, Ascona, Switzerland

- **2005**
  - Poster at the NSF/EPFL research day, EPFL, Switzerland
  - Poster at the Bio-Image Summer School Paris 2005, Ecole Normale Supérieure, Paris
  - IMA Workshop, Experiments in Physical Biology at the University of Minnesota, USA
- **2004**
  - Seminar at the workgroup on bioinformatics (in french), ENS, Paris

LANGUAGES

French: fluent. English: excellent. Hebrew: very good (spoken). German: notions.

HOBBIES AND INTERESTS

Musical composition and playing (guitar and keyboard), home studio, Rollerskating.