THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

NANOFLUIDICS FOR STATIC AND DYNAMIC DNA-PROTEIN INTERACTION STUDIES

- REPAIR OF DOUBLE-STRAND BREAKS FROM A SINGLE-MOLECULE PERSPECTIVE

ROBIN ÖZ

Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2020

NANOFLUIDICS FOR STATIC AND DYNAMIC DNA-PROTEIN INTERACTION STUDIES - REPAIR OF DOUBLE-STRAND BREAKS FROM A SINGLE-MOLECULE PERSPECTIVE
ROBIN ÖZ
ISBN 978-91-7905-343-7

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4810
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
Illustration of a nanoconfined, broken DNA segment, with proteins threaded on the free ends. The intact DNA is wrapped around histone octamers to form nucleosomes, which are folded into a chromosome. Created with BioRender.com

Chalmers digitaltryck
Gothenburg, Sweden 2020

NANOFLUIDICS FOR STATIC AND DYNAMIC
DNA-PROTEIN INTERACTION STUDIES

- REPAIR OF DOUBLE-STRAND BREAKS FROM A SINGLE-MOLECULE PERSPECTIVE

ROBIN ÖZ

Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

Double-strand breaks (DSBs) is one of the most lethal forms of DNA damage. A single DSB may result in stalling of vital cellular machineries, and thus, requires immediate measures by the cell. Although the main hallmarks of DSB repair mechanisms are known for most prokaryotes and eukaryotes, details on crucial intermediate steps are still to be explained, such as how the free DNA ends are kept in close proximity during the repair process.

In the original work, upon which this Thesis is based, the two main DSB repair mechanisms, homologous recombination (HR) and non-homologous end-joining (NHEJ), have been studied from a molecular perspective. A fluorescence-based single-molecule nanofluidics assay has been developed and employed to characterize and visualize static biomolecular interactions between DNA and key DSB repairing proteins. Furthermore, a novel dynamic nanofluidic device has been developed to enable dynamic interaction studies in real time, allowing analytes to be introduced on-demand to stretched DNA molecules.

Single-molecule characterization of the NHEJ mechanism in *Bacillus subtilis*, comprising the Ku and Ligase D proteins, revealed that the end-joining activity is mediated by C-terminal protrusions on the homodimeric Ku complex. Using the novel dynamic nanofluidic device, the Ku homodimer was further demonstrated to stay bound to the DNA ends and junctions after completed repair, similar to the human Ku70/80 heterodimer. In addition, the traditional static nanofluidic device was used to identify a previously unknown potential DNA-bridging role of CtIP, a key protein in the human HR process. This could possibly explain how broken DNA ends are kept in close proximity during the initial steps of DSB repair through HR in humans. A similar method was employed to show that the Xrs2 component is indispensable for the end-joining activity of the Mre11-Rad50-Xrs2 complex of *Saccharomyces cerevisiae*.

Keywords: Single-molecule, DNA repair, NHEJ, HR, protein, nanofluidics biomolecule interactions, dynamic, fluorescence, microscopy.

List of Publications

This Thesis is based on the work contained in the following research papers:

I. **A nanofluidic device for real-time visualization of DNA-protein interactions on the single DNA molecule level**
Robin Öz, Sriram KK, Fredrik Westerlund
Nanoscale, **11(4)**, (2019), 2071-2078.

II. **Dynamics of Ku and bacterial non-homologous end-joining characterized using single DNA molecule analysis**
Robin Öz[1], Jing L. Wang[1], Raphael Guerois, Sriram KK, Virginie Ropars, Rajhans Sharma, Firat Koca, Jean-Baptiste Charbonnier, Mauro Modesti, Terence R. Strick, Fredrik Westerlund
(Manuscript submitted, 2020).

III. **Phosphorylated CtIP bridges DNA to promote annealing of broken ends**
Robin Öz, Sean M. Howard, Rajhans Sharma, Hanna Törnkvist, Ilaria Ceppi, Sriram KK, Erik Kristiansson, Petr Cejka, Fredrik Westerlund
Proceedings of the National Academy of Sciences, **117(35)**, (2020), 21403-21412.

IV. **MRX promotes annealing of DNA ends in an Xrs2-dependent manner**
Rajhans Sharma, Robin Öz, Carl Möller, Giordano Reginato, Elda Cannavo, Sriram KK, Petr Cejka, Fredrik Westerlund
(Manuscript)

[1] Authors contributed equally to the work.

Additional research papers not included in this Thesis:

V. Interspecies plasmid transfer appears rare in sequential infections with extended-spectrum β-lactamase (ESBL)-producing Enterobacteriaceae
Anna Lindblom, Sriram KK, Vilhelm Müller, Robin Öz, Hilda Sandström, Christina Åhrén, Fredrik Westerlund, Nahid Karami.
Diagnostic Microbiology and Infectious Disease, **93(4)**, (2019), 380-385.

VI. Quantifying DNA damage induced by ionizing radiation and hyperthermia using single DNA molecule imaging
Vandana Singh, Pegah Johansson, Dmitry Torchinsky, Yii-Lih Lin, Robin Öz, Yuval Ebenstein, Ola Hammarsten, Fredrik Westerlund
Translational Oncology, **13(10)**, (2020), 100822.

Contribution Report

Below follows a description of my contributions to the papers appended in this Thesis.

I. Developed the nanofluidic device together with S.K.K. and F.W. Planned and performed all experiments and analyzed the data. Wrote the paper together with F.W.

II. Expressed and purified the bacterial proteins, used for all experiments in the paper. Labelled the Ku protein. Planned and performed nanofluidics experiments and SPR, EMSA and ligation assays, including data analysis. Wrote the paper together with T.S. and F.W.

III. Developed the nanofluidic assay. Planned and performed nanofluidics experiments and AFM imaging, including all data analysis. Wrote the paper together with P.C. and F.W.

IV. Performed the nanofluidics experiments together with R.S. and C.M. and analyzed the data. Wrote the paper together with R.S. and F.W.

Preface

This dissertation was submitted for the partial fulfilment of the degree of Doctor of Philosophy. The original work presented in this dissertation was carried out between April 2016 and November 2020 at the Department of Biology and Biological Engineering, Chalmers University of Technology, under the supervision of Professor Fredrik Westerlund. The research was funded by the Swedish Research Council, the European Research Council and the Olle Engkvist Byggmästare foundation.

Robin Öz
November 2020

Table of content

1 Introduction – *The Context*

"All models are wrong, but some are useful"

- George E. P. Box

Deoxyribonucleic acid (DNA), usually referred to as the genetic code of life, functions as a central information storage system for most living organisms[1, 2]. The genetic information is stored within the order of the nucleobases, which encodes for proteins that enable and control most of the elemental functions in the cell. DNA is constantly subjected to internal and external stress, which may induce damage[3]. Factors such as UV-radiation and exposure to genotoxic foreign compounds and metabolites can give rise to deleterious changes in the structure of the DNA molecule. It is of immense importance that such damages are quickly repaired in order to avoid genetic alterations that may pose short and long-term lethal effects to the cell. Although estimated to occur only ten times per day per cell, double-strand breaks (DSBs) are

considered one of the most lethal types of DNA damage[4]. These typically occur as a result of multiple nicks in the sugar-phosphate backbone of the intertwined single DNA strands. Two proximal nicks on the opposite strands will cause rupturing of the DNA molecule, resulting in stalling of vital cellular machineries, such as replication and transcription, which is lethal for the cell[5]. Fortunately, evolution has generated a plethora of repair mechanisms to ensure efficient repair of such DNA lesions. Homologous recombination (HR) and non-homologous end-joining (NHEJ) are examples of the most common mechanisms for repair of DSBs *in vivo*[4, 6]. There are numerous existing models explaining the main hallmarks of DSB repair by HR and NHEJ, however, details on crucial intermediate steps are still yet to be explored in order to build on the present models, thereby making them more accurate.

The original work, upon which this Thesis is based, demonstrate how nanofluidics tools can be used to better understand DNA-protein interactions on a molecular level. Single-molecule analysis provides the opportunity to study properties that are typically omitted in traditional bulk assays due to the ensemble averaging effect that is inherent to such methods[7]. However, most used techniques for single DNA molecule analysis require tethering of one or both of the DNA ends to stretch the molecule and thereby visualize biomolecular interactions *in vitro*. Obstructing the free DNA ends provides a significant disadvantage when studying DSB repair, where most reactions take place at those ends. This typical limitation can be overcome by employing nanofluidics, making use of the fact that DNA tends to spontaneously stretch when confined in nanofluidic channels in order to accommodate[8-10]. Detailed characterization of DNA-protein interactions along the entire contour and at the very ends of the molecule is thus made possible.[11]

In this Thesis, DSB repair by NHEJ and HR is discussed in the light of the four appended papers, referred to as **Paper I-IV**. The following sections 2-5 provide a theoretical background on fundamental topics and methods used in the appended papers. In section 6, the original work is summarized, which constitutes the developments and findings presented in **Paper I-IV**. Finally, in section 7, future perspectives are discussed based on the reported original results. The work presented in this Thesis demonstrates how DNA-protein interactions can be studied by nanofluidic methods, which have been used in combination with more traditional assays, to answers fundamental questions on DSB repair. In **Paper I**, a novel device is presented, which enables dynamic DNA-protein interaction studies in nanofluidic channels. The traditional nanofluidic chip design does not allow for the *in situ* addition of analytes to confined and stretched DNA molecules. There has thus been a need for a device, which enables active manipulation of the local environment within the

nanofluidic channels, while observing stretched DNA molecules. The functionality of the novel device was demonstrated by a series of proof-of-concept experiments, showing that it can very well be used to add up to two analytes on-demand to a single nanoconfined DNA molecule, while observing the response under a fluorescence microscope.

In the second project, the *Bacillus subtilis* NHEJ pathway was studied, as reported in **Paper II**. Bulk experiments and single-molecule techniques were employed to better understand DSB repair in bacteria. The two NHEJ components, Ku and LigD, were expressed, purified and subsequently characterize by their abilities to interact with DNA. It was shown, among other things, that the protruding C-terminal arms of the Ku homodimer are crucial for the recruitment of LigD to the point of repair, while the core structure of the homodimer is important to hold DNA ends together.

In **Paper III**, the tetrameric CtIP protein, which is involved in human HR, was characterized on the single DNA-molecule level, by developing and employing an unconventional and novel nanofluidics assay. The assay allows for high-throughput conformational analysis of DNA, where circular molecules are easily distinguished from linear ones. A library of CtIP derivatives were studied to better understand how the tertiary structure of CtIP affects the DNA-protein interaction. It was demonstrated that CtIP possesses a structure dependent DNA bridging ability, potentially important in the alignment of free DNA ends during HR *in vivo*. A similar approach was further employed in **Paper IV**, in which the yeast HR protein complex Mre11-Rad50-Xrs2 (MRX) was studied. It was here shown that the Xrs2 component is crucial for the DNA end-joining activity of MRX.

To conclude, solid understanding on DNA-protein interactions *in vitro* is essential to develop models, which may be useful for describing fundamental and highly complex biological systems *in vivo*. This Thesis highlights the usefulness of single-molecule nanofluidics tools as a complement to traditional biochemical assays for DNA-protein interaction studies. Furthermore, the Thesis provides details on intermediate steps in the HR and NHEJ mechanisms, potentially contributing to the overall understanding of DSB repair from a single-molecule perspective.

2 DNA – *"The Code of Life"*

The DNA molecule has fascinated the human mind ever since it was first isolated in 1869 by the Swiss physician Friedrich Miescher[12]. In 1944 the hereditary role of DNA was postulated by Oswald Avery and co-workers[13], eventually leading to the common definition of DNA being the *genetic material* of most living organisms. By that time the DNA molecule was well known to play an important role, not only in the genetic transmission, but also in the evolution of organisms and the progression of life. An important advancement was made in 1953, when James Watson and Francis Crick first described the molecular structure of DNA[14], marking a paradigm shift in the history of biology. Besides explaining the three-dimensional architecture of the molecule, the deciphering of the DNA structure also resulted in a detailed understanding of the fundamental properties of life.

2.1 The Chemical Structure of DNA

Watson and Crick proposed, based on an X-ray diffraction pattern, a model for the structure of DNA, where two individual strands are intertwined to form a right-handed double-helix (Figure 1)[14]. The molecule is built up by nucleotides, which constitute the structural components of DNA. The nucleotides are composed of the five-carbon sugar deoxyribose, three phosphate groups and a variable aromatic nucleobase, being either adenine (A), thymine (T), guanine (G) or cytosine (C). Linked by a phosphodiester bond between the 3′-hydroxyl group on one deoxyribose and the phosphate group on the 5′-carbon on the neighboring one, the nucleotides are connected to form a long polynucleotide chain, referred to as single-stranded DNA (ssDNA). The release of an inorganic pyrophosphate through a condensation reaction provides the energy for a covalent bond to form as new nucleotides are added to the 3′-hydroxyl group of the terminal nucleotide in the ssDNA. The repeated addition of nucleotides in this specific manner gives the DNA an orientation, where the polymerization proceeds in the 5′to 3′ direction. Thus, the nucleobases, linked to the

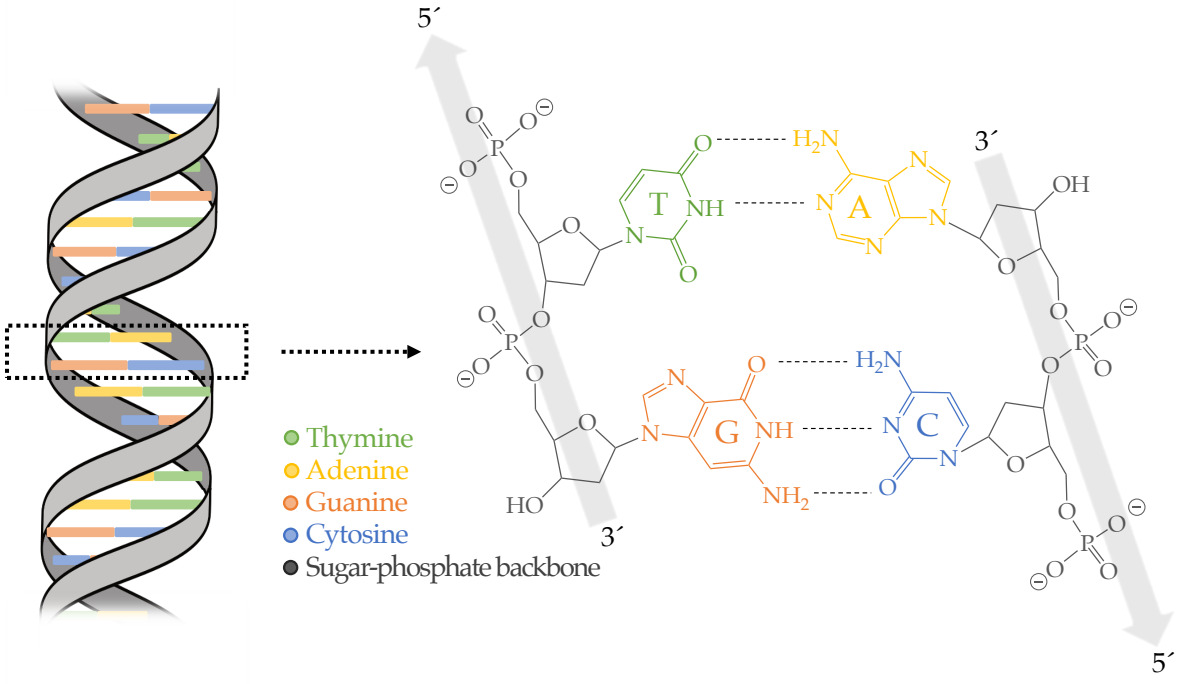

Figure 1: *The double-helical structure of DNA is composed of two intertwined polymeric strands of nucleotides connected through a phosphodiester bond between a phosphate group and a neighboring ribose sugar, which forms the DNA sugar-phosphate backbone. The nucleobases on separate strands hybridize through hydrogen bonding (dashed lines), resulting in a double-stranded DNA molecule. The arrows indicate the anti-parallel direction of the two polynucleotide strands.*

1′-carbon of each deoxyribose, will align to generate an ATGC-code, referred to as the DNA sequence. The helical structure of double-stranded DNA (dsDNA) arises from the combination of two ssDNA molecules, which are held together by non-covalent hydrogen bonds between the complementary nucleobases. Two hydrogen bonds allow base-pairing between A and T, while G and C are held together by three hydrogen bonds. This consequently means that the two strands in dsDNA are complementary and anti-parallel. The genetic information of most living organisms is stored in both of the two strands.

2.2 Physical Properties of DNA

The dsDNA molecule can exist in different conformations, of which A-, B-, and Z-DNA are the most common[14-16]. At physiological conditions, dsDNA will attain the B-form, which has the characteristic right-handed helical structure with a typical width of 2 nm, corresponding to the lateral distance between the intertwined sugar-phosphate backbones[17]. The stacked nucleobase-pairs are separated by 0.34 nm with a twist angle of approximately 36 degrees. This gives a periodicity of 10 base-pairs (bp), corresponding to 3.4 nm at physiological conditions[17]. As a consequence of the twisting nature of dsDNA, major and minor grooves are formed along the helical axis adjacent to the base-pairs. These grooves can serve as binding sites for proteins and small molecules. The phosphate groups of the backbone give the DNA molecule a net negative charge at physiological pH. Therefore, changes in the salt composition of the surrounding environment may have an immediate impact on the physical properties of the molecule. As an example, the B-to-Z transition is favored by high ionic strength, while the transition from the B-form to A-form is driven by lower hydration levels[18].

2.2.1 DNA Free in Solution

When suspended in solution, dsDNA will attain a coiled structure to minimize its free energy. Being a large polymer, the DNA molecule has highly asymmetric dimensions, making rigid models, such as cylinders, rods and spheres, inadequate to describe the macromolecular structure. Instead, a more flexible representation has been proposed, usually referred to as the *Worm-like chain* model[19] (Figure 2a), where the dsDNA is described as a continuous and semi-flexible polymer, taking into account the local rigidity of the DNA. The worm-like chain is defined by two physical quantities: the contour length (L) and the persistence length (P). L corresponds to the end-to-end distance of the polymer when fully stretched, while P can be described as the

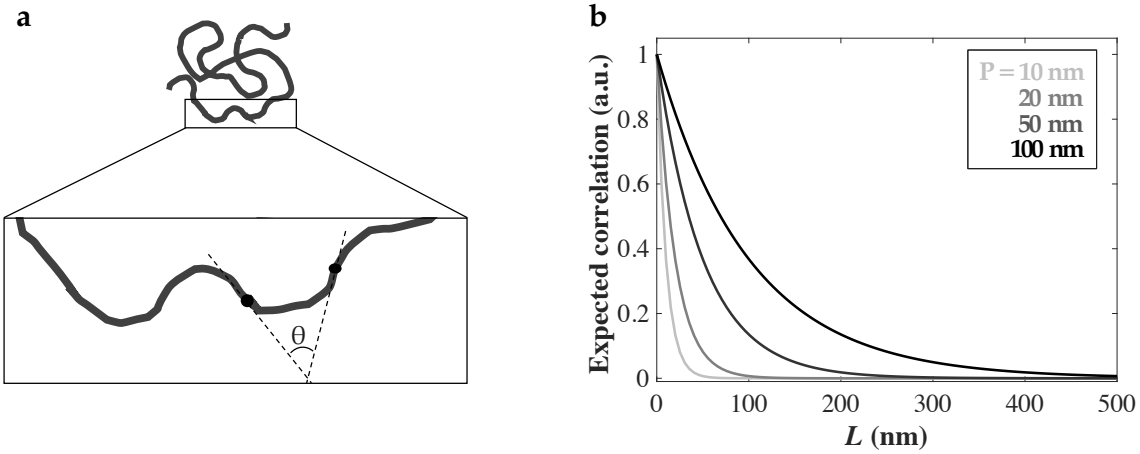

Figure 2: ***(a)*** *A schematic illustration of a coiled polymer in accordance with the worm-like chain model, where the angular difference (θ) between tangents along the contour indicates the flexibility.* ***(b)*** *The expected relative correlation between tangents, given by the average cosine of θ at given values for the persistence length (P), calculated from Eq.1. The expected correlation decreases exponentially with increasing distance (L).*

---

minimum length over which the polymer cannot be bent. The bending stiffness of a polymer can also be described as the distance over which correlations in the direction of the tangents along the molecule extension are lost[17, 20]. The correlation can be estimated from the average cosine of the angle θ between the tangents of the chain from a fixed position, given by Eq. 1.

$$< \cos(\theta) >_L = e^{-L/P} \qquad \text{(Eq. 1)}$$

The expected average cosine of the angle will decrease exponentially with L, suggesting that the correlation between the direction of tangents is lost over long distances (Figure 2b). At short distances ($L \ll P$) the tangents will be highly correlated with an angle θ close to zero degrees, giving an average cosine of one. Thus, the polymer will behave as a rigid rod, which is also supported by Eq. 2.

$$< L^2 >= 2P(L - P + Pe^{-L/P}) \qquad \text{(Eq. 2)}$$

If $L \ll P$, expansion of Eq. 2 gives $< L^2 >= L^2$, which is the behaviour expected for a rigid rod. The bending stiffness of a polymer is directly proportional to P, which is about 50 nm (~150 bp) for dsDNA at physiological conditions[17]. Long DNA molecules, satisfying the condition $L \gg P$ will coil up in solution and attain an energetically more favorable conformation with highest possible entropy. The size of a coiled DNA molecule is described by the radius of gyration (R_G), which is defined as the root-mean-square distance of segments from the center of mass of the coil (Figure 3). The inherent

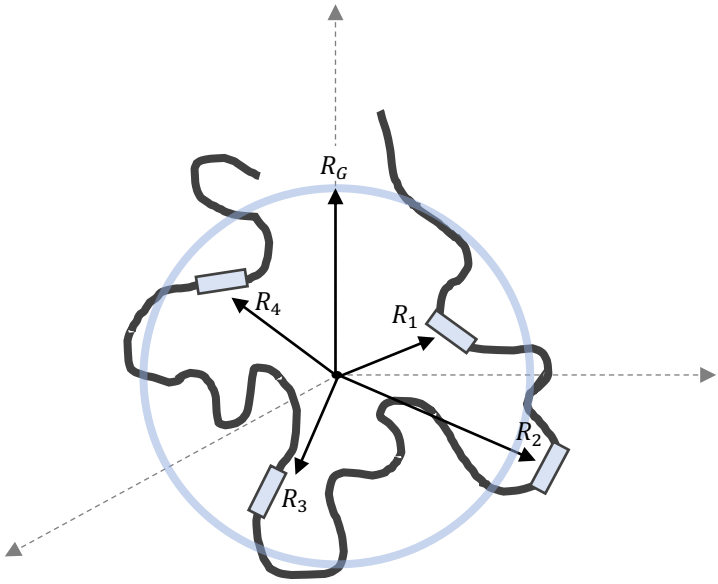

Figure 3: *Schematic illustration of a polymer with its center of mass at the origin in a three-dimensional space. The radius of gyration (R_G) is calculated from root-mean-square of the radial distance (R_n) from the center of mass to each segment along the polymer contour.*

---

net negative charge of the sugar-phosphate backbone has a direct impact on the R_G as the electrostatic repulsive forces will restrict the DNA to collapse into a compact torus. At physiological conditions, the DNA has a Debye length of approximately 1 nm[17, 21], corresponding to the distance from the sugar-phosphate backbone, at which no electrostatic forces are exerted on surrounding objects. The Debye length will increase at low ionic strength, and hence lead to a larger R_G.

2.2.2 DNA in Confinement

The flexible nature of DNA allows it to minimize its free energy through dynamic macromolecular rearrangements to mitigate changes in the local environment. However, spatial restrictions to the polymer will limit its movements and thereby affect the physical properties of the molecule. Enclosing DNA in narrow channels will cause the molecule to stretch along the channel extension, which in turn disrupts its coiled structure. The level of confinement and its effect on the macromolecular structure depend on the channel dimensions, given by the average cross-section area (D_{av}), defined by Eq. 3.

$$D_{av} = \sqrt{D_w \cdot D_h} \qquad \text{(Eq. 3)}$$

D_w and D_h correspond to the width and the height of a rectangular channel, respectively. Confined in a channel with a $D_{av} > R_G$, the DNA will keep its coiled structure, behaving as if freely suspended in solution (Figure 4). To describe the effect on polymers at higher degrees of confinement de Gennes proposed a model[22], which predicts the organization of the molecule in channels with dimensions satisfying $P \ll D_{av} < R_G$. The principle of this model is that a confined polymer can be divided into a series of non-interacting isometric (spherical) units, which have a diameter equal to D_{av} (Figure 4)[22, 23]. The DNA in each unit is considered as freely suspended in solution according to de Gennes. The self-avoiding property of DNA causes the molecule to uniformly distribute along the channel extension. This means that the extension of DNA scales linearly with L, which in turn implies that observed features along the contour of the molecule are directly related to the corresponding position in the nucleotide sequence. As the channel dimensions are reduced to $D_{av} \leq P$ the bending energy of the molecule will increase further, causing the entropy to drop. The energy required to form a loop is now greater than the thermal energy, which causes the DNA to attain an elongated conformation, where the double-helix is undulating between the channel walls (Figure 4). This level of confinement is known as the Odijk regime[24]. There is an intermediate level of confinement, where the average cross-section area is close to the persistence length of the polymer ($D_{av} > P$). This is usually referred to as the *extended de Gennes regime*, where the previously described non-interacting isometric units instead become anisometric (cylindrical) due to compression by the channel walls[23, 25-27]. Despite the deformation of the DNA-containing units, the linear scaling between the sequence and L will remain unaffected.

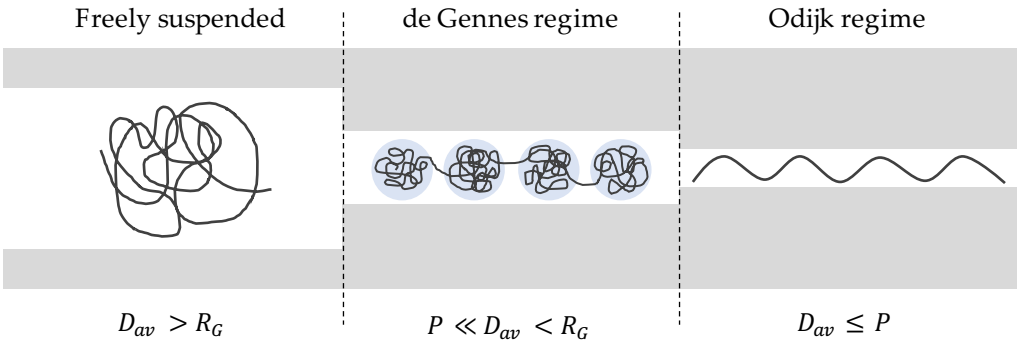

Figure 4: *The behavior of a polymer under different degrees of confinement in a channel. The coiled structure will be retained at channel dimensions (D_{av}) larger than the radius of gyration (R_G). At smaller dimensions, the extended polymer can be described as a series of non-interacting units, according to the model proposed by de Gennes. At dimensions smaller than the persistence length (P), the polymer will enter into the Odijk regime.*

2.3 The Central Dogma of Molecular Biology

Detailed characterization of the DNA architecture in the late 1950s quickly induced further advancements in the field and resulted in important explorations, which have shaped our fundamental understanding of life. In 1958, Francis Crick presented the *central dogma of molecular biology*, stating the transfer of genetic information, hidden within the DNA sequence, to other biomolecules in the cell, such as ribonucleic acids (RNAs) and proteins[1]. The deciphering of the double-stranded nature of DNA and its ability to separate into single strands, where the nucleobase sequence is exposed, allowed Crick to postulate the possible flow-path of the genetic information[2]. The general central dogma (Figure 5) states that the sequential order of the nucleobases can be copied to generate an identical DNA molecule, a process referred to as replication, which is crucial for the genetic inheritance among organisms. The DNA sequence can also be transcribed into messenger RNA (mRNA), which is further translated into a protein. In the translation process, the mRNA sequence is decoded to produce an amino acid chain. Three consecutive nucleobases are recognized by a specific transfer RNA (tRNA) molecule carrying an amino acid. The amino acid corresponding to each consecutive nucleobase triplet in the mRNA sequence will be linked to form a protein. The translation process is irreversible and once completed, the amino acid sequence of a protein cannot be translated back to produce RNA or DNA. The central dogma further emphasizes the presence of atypical flow-paths of the genetic information, such as reverse transcription of RNA to DNA, replication of RNA and direct translation of DNA to produce a protein[2]. Although not considered to be main routes for information flow, these processes have been proven to exist in nature[28-30].

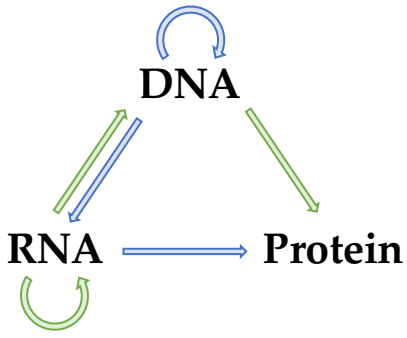

Figure 5: *The central dogma of molecular biology. Blue arrows indicate the general flow-paths in biological systems. Green arrows show atypical routes for transfer of genetic information.*

2.4 Interactions with DNA – Forces and Binding Modes

The versatile structural and physical properties of the DNA molecule enable molecular interactions with a plethora of compounds through different binding modes. In biological systems, the interactions of molecules with DNA are essential for sustaining life. As an example, the association of histone octamers to DNA is crucial for accommodating the molecule in the human cell nucleus. The protein acts as a spool, around which the DNA can be wound, causing the molecule to attain a compact structure, up to five orders of magnitude shorter than naked DNA[31]. Transient modifications to the histone octamer will alter the binding properties of the protein, which is vital for downstream processes, such as DNA transcription and replication, manifesting the importance of dynamic interactions between DNA and other compounds. The binding of histones to DNA is a typical external electrostatic interaction, where the negatively charged sugar-phosphate backbone interacts with the histone octamer, causing the DNA to be folded with approximately 80 bp covering the circumference of the core[32]. The positively charged residues on the histone octamer will exert an electrostatic shielding effect on the DNA, sufficient to overcome the energetic penalty attributable to bending of the molecule at a length-scale shorter than the persistence length.

The minor and major grooves in the DNA structure, formed as a consequence of the periodic helicity, can serve as binding pockets for small molecules, such as transcription factors. The TATA-binding protein (TBP) is an example of a minor-groove binder, which helps position the RNA-polymerase on the DNA to initiate transcription of specific genes[33]. Groove-binding molecules are usually associated with DNA through hydrogen bonding with functional groups on the nucleobases, inferring the sequence specificity of this biding mode[34]. TBP, in particular, is known to recognize and bind to a specific DNA sequence called the TATA-box, which is usually found in conjunction with the transcriptional starting sites of eukaryotic genes[33].

Ligands can also interact with DNA through intercalation, where the molecule is partially or entirely inserted into the double-helix. Intercalators are usually characterized by their planar aromatic structure, which can incorporate into the space between the stacked nucleobase-pairs, perpendicular to the helical axis. This can cause changes to the DNA structure, such as decrease in the helical twist, elongation of the DNA molecule and stiffening of the helix[35, 36]. These effects are typically reversible upon dissociation of the intercalator.

2.4.1 The Effect of Intercalators on the DNA Structure

The homodimeric cyanine dye YOYO-1[37] is an intercalator, typically used for the fluorescent visualization of DNA (Figure 6a). YOYO-1 is a so-called bis-intercalator, meaning that the two aromatic structures are inserted into the DNA simultaneously, altering the photophysical properties of the intercalator, making it highly fluorescent (further discussed in section 4.2). The poly-cationic nature of YOYO-1 gives it a high binding affinity to dsDNA, with a K_a-value on the order of $10^{10}\ M^{-1}$ [38, 39]. Upon binding, the intercalated aromatic structures will cover a total of four bp along the helical axis (Figure 6b), causing the DNA to elongate 0.51 nm and unwind approximately 24 degrees per bound intercalator[38]. The potential effect of YOYO-1 on the persistence length of DNA still remains unclear due to conflicting reports on the matter[35, 38, 40]. Besides intercalation, electrostatic interactions between the oppositely charged dye and the DNA backbone are known to exist[41]. The intercalated molecule may cause local stiffness to the DNA, while non-specific interactions will exert electrostatic shielding of the sugar-phosphate backbone, hence neutralizing the potentially stiffening effect attributable to the intercalation.

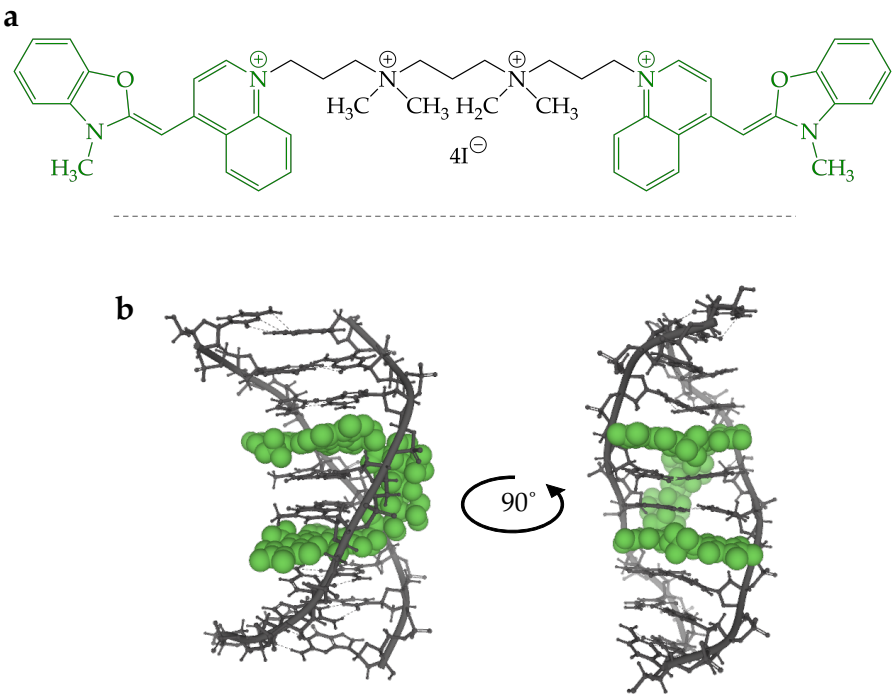

Figure 6: ***(a)*** *Molecular structure of YOYO-1, where intercalating moieties are highlighted in green.* ***(b)*** *Structural model of DNA (grey) with an associated bis-intercalator (green), visualized by JMol. The images were derived from data submitted to Protein Data Bank (PDB: 108D)*[42].

3 Proteins – *Maintaining the Genome Integrity*

The DNA molecule is constantly subjected to stress induced by endogenous cellular metabolites and external factors, such as exposure to UV-light and genotoxic compounds[3]. DNA lesions have been estimated to occur up to 10^5 times per day in a single human cell, of which the majority affects only the single strands of the duplex molecule[43]. Despite being detected and repaired efficiently by specialized cellular mechanisms, DNA damage and mutations are known to accumulate over time and are correlated with the development of various cancer diseases, as well as the natural aging process[44]. Evolution has generated an elaborate apparatus that regulates the genomic maintenance in living organisms[3]. A plethora of highly dedicated DNA repair proteins serve to ensure faithful transmission of genetic information among generations and preserve the functional integrity of cells[3, 45].

3.1 The Cellular Building Blocks

Proteins are the primary constituents of the cell biomass, with an abundance corresponding to up to 60% of the total cellular dry weight[46, 47]. Proteins are involved in almost all cellular functions, orchestrating biochemical reactions and processes in a highly structure-dependent and specific manner. They serve as structural scaffolds in the cell architecture, catalyze reactions, transport molecules, translate the genetic code into new proteins and drive the nucleic acid replication, among many other functions. In 1838 the Dutch chemist Gerardus Johannes Mulder was the first person to describe proteins in terms of their composition[48]. Mulder performed elemental analysis and found that the chemical composition was highly similar for a number of common plant and animal proteins. More than a century later, the first complete protein sequence was reported by the British biochemist Frederick Sanger in 1949, which conclusively demonstrated that proteins are linear polymeric chains of amino acids[49].

3.1.1 Protein Chemistry and Structure

A total of 20 natural amino acids exist, which are the structural components of proteins, each with different chemical properties. Amino acids are organic compounds comprising an amine ($-NH_2$) and a carboxyl (-COOH) group, along with a variable side chain, all attached to the α-carbon atom (Figure 7a). The characteristics of an amino acid, such as the polarity, charge and size are primarily determined by the side chain. At physiological conditions, the α-amine and α-carboxyl groups will be protonated and deprotonated, respectively, making the amino acids zwitterionic if disregarding possible charges on the side chain. Out of the 20 amino acids involved in protein synthesis, there are two cations (Asp, Glu) and three anions (Arg, Lys, His). The remaining amino acids are uncharged and either polar or hydrophobic (Figure 7b). The net charge of a protein is generally estimated from its isoelectric point (pI). This physical quantity reports the pH, at which the protein carries no net electrical charge. Proteins will thus be positively charged at pH below pI and negatively charged at pH above pI.

During translation of the transcribed mRNA sequence, amino acids corresponding to specific nucleobase triplets will be linked through a peptide bond between the α-carboxyl group and the α-amine group on two adjacent amino acids (Figure 8a). The order of the amino acids in the resulting polypeptide chain corresponds to the primary structure of the protein. The ends of a polypeptide chain are denoted as the N- and C-terminals, respectively, where the first refers to the free α-amine group on the first amino acid, and the latter to the α-carboxyl group on the

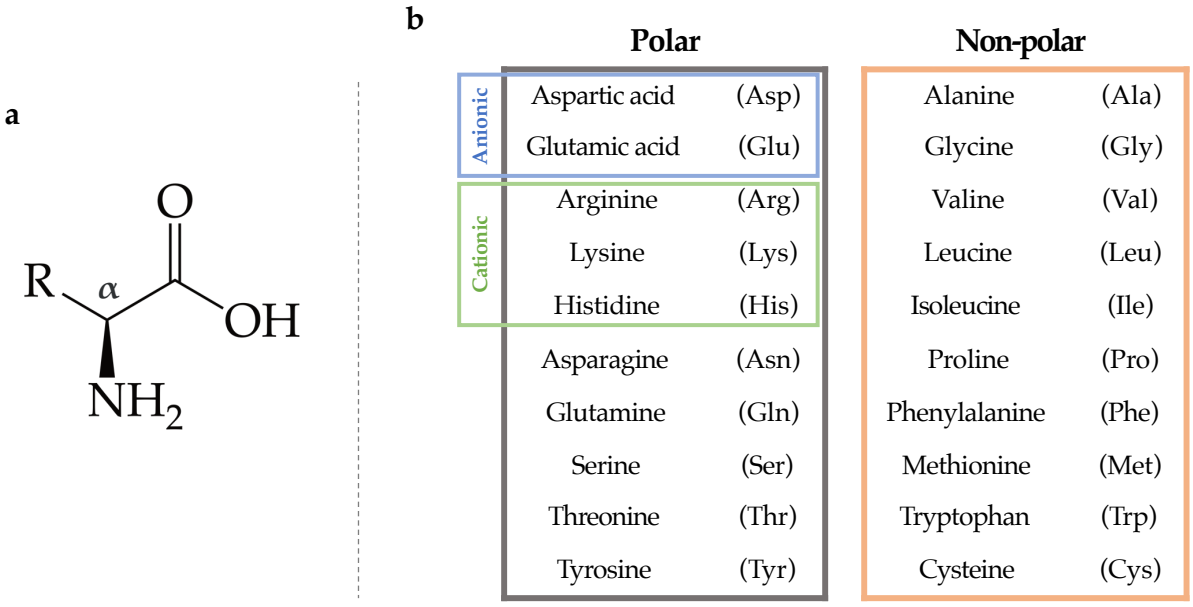

Figure 7: ***(a)*** *Molecular structure of a generic amino acid. The α-amine and α-carboxyl groups along with the variable side chain, denoted as R, are connected to the α-carbon atom.* ***(b)*** *Summary of amino acids, used in the protein synthesis, with the corresponding three-letter abbreviations. The overall polarity is determined by the variable side chain.*

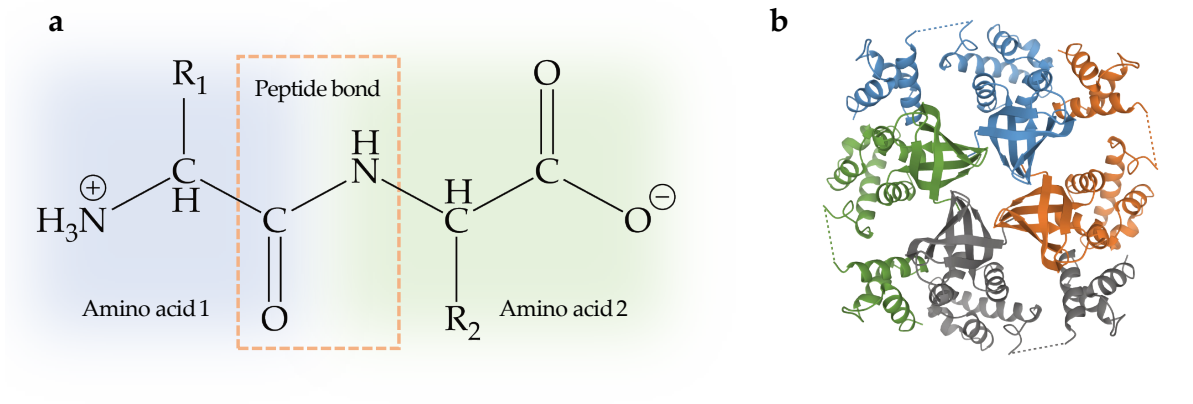

Figure 8: ***(a)*** *Two amino acids linked through a peptide bond.* ***(b)*** *Quaternary structure of the Escherichia coli homotetrameric RuvA helicase, where each subunit is highlighted in a different color. The α-helical structural motifs are mainly covering the outer surface, while β-sheets are hidden in the core of the protein complex. The image was derived from data submitted to Protein Data Bank (PDB: 1CUK)*[51].

last amino acid in the sequence. Intramolecular interactions between side chains will cause the protein to attain a thermodynamically stable conformation, where the free energy is minimized. In 1951 the American scientist Linus Pauling proposed two structural motifs that are formed upon folding of polypeptides, referred to as α-helices and β-sheets[50]. These motifs were found to be regularly occurring in various protein structures and are stabilized by intramolecular hydrogen bonding between amino acid

structural motifs along the polypeptide chain and corresponds to an intermediate folded state, referred to as the secondary structure. The protein will attain a three-dimensional tertiary structure upon non-covalent interactions between the side chains. Hydrophobic residues will typically be hidden in the core of the protein structure, while hydrophilic side chains are exposed to the aqueous solution. Some proteins are composed of multiple subunits, which are expressed and produced individually, before being combined into a so-called quaternary structure (Figure 8b).

3.2 Repair of DNA Double-Strand Breaks *in Vivo*

Double-strand breaks (DSBs) are considered one of the most lethal types of DNA damage, where both strands in the double helix have been severed, resulting in an interruption of the nucleobase sequence. Just a single event is enough to trigger cell-cycle arrest and in turn cause apoptosis[5]. Compared to ssDNA lesions, DSBs occur at a much lower frequency, with an estimated number of approximately ten events per day per cell[4]. However, the high lethality of such breaks requires immediate measures in order to quickly resume important cellular processes. Except for specialized mechanisms in lymphocytes of the vertebrate immune system, such as V(D)J recombination[52], most DSBs are accidental and occur throughout the entire cell cycle[4]. Reactive oxygen species (ROS), typically created during normal oxidative respiration, are a major cause of DSBs [53]. Upon reaction with DNA, ROS will generate nicks on the individual strands of the double-helix, which if closely spaced on the opposite strands, may cause a DSB[4, 53]. Environmental ionizing radiation also poses a risk, mainly via the generation of nicks by intracellular water-originating radicals. X- and γ-rays are primarily responsible for the radiation-caused ssDNA lesions, where roughly 65% appears to be generated by the indirect effects of radicals and about 35% by direct ionization[54]. Additional causes include inadvertent action of nuclear enzymes, replication of DNA across nicks and mechanical stress induced during the mitotic chromosome segregation[4]. The severity of DSBs stresses the importance of efficient and robust repair mechanisms in the cell to maintain the genome integrity. In order to ensure genomic stability, evolution has generated two major pathways for DSB repair: homologous recombination (HR) and non-homologous end-joining (NHEJ). The two processes are activated at different levels during the cell cycle to rejoin broken DNA ends in a faithful and efficient manner.

3.2.1 Homologous Recombination

Homologous recombination is evolutionarily conserved across all three domains of life[55]. The mechanism relies on recombining DNA ends based on distinct sequence homology, thereby allowing for complete restoring of the DNA sequence upon rupture of the molecule[56]. In human cells, the sister chromatid is typically required for DNA repair by HR. The corresponding intact sequence of the sister chromatid, homologous to that of the broken DNA, serves as a template during DNA synthesis to generate an identical copy. Thus, HR is mostly active during the late S to G_2 phases of the cell cycle, when the sister chromatids are readily available[6, 57]. HR may also occur during the G_1 phase in diploid organisms, although this is rather rare[57, 58].

HR mediated DSB repair can be divided into three conceptual steps, referred to as pre-synapsis, synapsis and post-synapsis (Figure 9). In the pre-synaptic stage the broken DNA is resected to generate a single-strand sequence that can be matched to the complementary sequence of the template[56, 59]. In humans the MRE11-Rad50-Nbs1 (MRN) complex, together with CtIP, is mainly responsible for resecting DNA in the 5′to 3′direction, resulting in 3′-overhangs of varying sizes[60]. Their counterparts in yeast are Mre11-Rad50-Xrs2 (MRX) and Sae2, respectively[61]. Following the end resection, the single-strand overhangs are initially bound by replication protein A (RPA) to neutralize secondary structures and protect from nonspecific nucleolytic activity. The RPA is subsequently replaced by Rad51, which forms a filament along the contour of the ssDNA. In the synaptic stage of HR, this nucleoprotein filament performs a so-called strand invasion, where the Rad51-bound ssDNA is inserted into the template molecule. A homology search is executed to ultimately find the site, which corresponds to the complementary sequence of the ssDNA. This is followed by D-loop formation, where the two strands in the template DNA are separated to allow the invading ssDNA sequence to anneal. In the final post-synaptic stage, the DNA is synthesized using the intact homologous sequence as template to restore the broken DNA molecule[56, 59, 62, 63]. Apart from being one of the main repair mechanisms of DSBs, HR is also essential for the exchange of genetic information between allelic sequences, thereby giving rise to genetic diversity, as well as controlling the segregation of chromosomes during meiosis[62].

In the original work presented in this Thesis, emphasis has been put on deciphering the various roles of the individual proteins involved in the pre-synaptic stage of HR. The human CtIP protein and the yeast MRX complex have been characterized on the single-molecule level based on their DNA-binding and bridging abilities to better understand their potential roles in bringing DNA ends close together during DNA repair (further discussed in section 6.4 and 6.5).

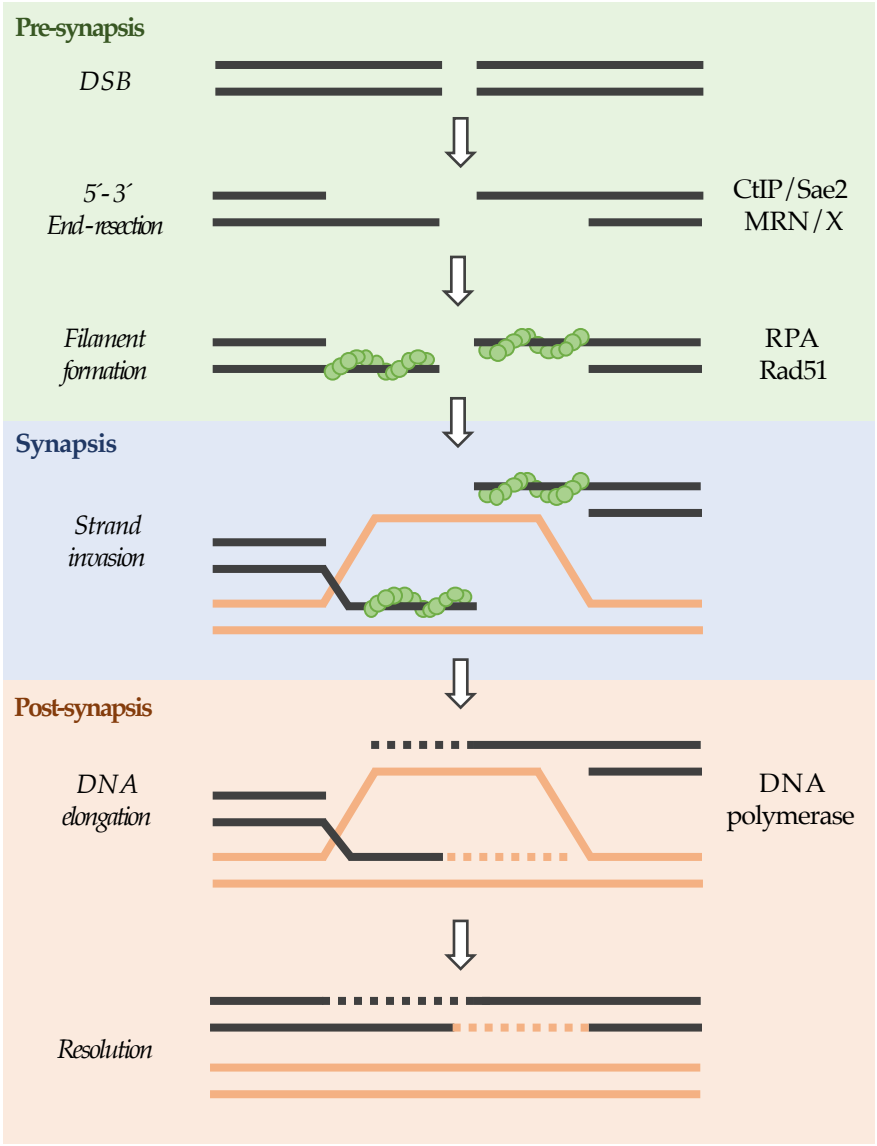

Figure 9: *DNA repair by homologous recombination. Hallmarks in the mechanism are highlighted together with the main responsible proteins in the human/yeast systems, divided into three synaptic stages. In the pre-synaptic stage, the DNA is resected to generate 3´-overhangs followed by binding of RPA, which is subsequently replaced by a Rad51 filament (green spheres). Strand invasion during the synaptic stage will allow for complementary sequences to anneal and further be elongated in the post-synaptic stage of the process. Crossover products may form as a result of interchanged strands during the resolution.*

The Human CtIP Protein

The pre-synaptic end-resection is performed by MRN in the presence of CtIP, which interacts physically with the MRN complex[60]. CtIP was initially identified as a binding partner for the C-terminal binding protein 1 (CtBP)[64] and the breast cancer type 1 susceptibility protein (BRCA1)[65]. CtIP was later established as a key factor in DNA repair by HR, where it plays an influential role in the selection of the pathway for DSB repair[60, 66]. As an example, it has been shown on the single DNA-molecule level that

CtIP and MRN in conjunction are able to remove the DNA-dependent protein kinase (DNA-PK), a key factor in NHEJ, from DNA ends through endonucleolytic incisions, thereby channel repair from NHEJ to HR[67]. Apart from promoting nucleolytic activity, recent studies suggest that the multi-phosphorylated CtIP[68] may possess additional functions, crucial for repairing DSBs[69-71]. The tetrameric structure[72] of CtIP has proven to play an important role in the bridging of DNA molecules, potentially critical for bringing DNA ends close together during the repair process. The shape of the homotetramer has been described to resemble a dumbbell, with two globular domains attached through a linker, with an end-to-end size on the order of 30 nm[70]. The oligomerization domain was proven to reside close to the N-terminal of the monomeric protein[72]. It is anticipated that the globular domains are responsible for binding distant DNA segments and thereby holding DNA ends in close proximity during the pre-synaptic stage of HR, to allow for re-joining of the ends during the subsequent steps of HR (further discussed in section 6.4 and **Paper III**). Similar functional observations have been made for the homologous Ctp1 protein complex in *Schizosaccharomyces pombe*[69] and Sae2 in *Saccharomyces cerevisiae*[71].

The MRN/X complexes

The human dimeric MRN complex comprises of three subunits, the meiotic recombination 11 homolog (MRE11), Rad50 and Nijmegen breakage syndrome 1 protein (Nbs1)[73-75]. In *S. cerevisiae* the functions of MRE11 and Rad50 are conserved[76], while Nbs1 is replaced by Xrs2[77], forming the MRX complex. Crystal structures of the complete MRN/X complexes have not been resolved up to date. However, structural studies of the subunits have reported two distinctive architectural features of the complex, a globular head and an elongated coiled coil (Figure 10a)[78, 79]. It has been proposed that MRE11 homodimerizes and associates with Nbs1/Xrs2 to form the globular head domain, to which two Rad50 entities are attached as elongated coiled coils. A crystal structure of the Rad50 coiled coil region reveals a terminal interlocking hook that binds one Zn^{2+} ion[79] (Figure 10b). This hook can potentially provide an interface, where two MRN/X complexes interact to form a bridge between the two DNA-binding domains through the association of MRE11 and Nbs1/Xrs2 with the DNA. Although both Nbs1 and Xrs2 are known to interact with CtIP and Sae2, respectively[80, 81], it has recently been shown that Xrs2 is dispensable for the Sae2 dependent end-resection by MR[82]. This suggests that Xrs2, and potentially also Nbs1 in *Homo sapiens*, may possess alternative functions in DSB repair through HR *in vivo.* Xrs2 itself has an intrinsic DNA-binding activity and has proven to specifically recognize the annealed single-strands in a DNA hairpin structure[77]. This emphasizes

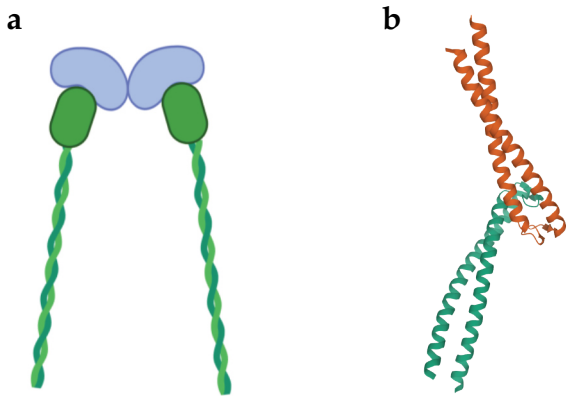

Figure 10: ***(a)*** *Schematic illustration of the MR structure. Two elongated Rad50 coiled coils (green) are attached to the globular MRE11 complex (blue), with which Nbs1/Xrs2 interacts.* ***(b)*** *Crystal structure of the Zn^{2+} hook domain, formed in the apex of two Rad50 coiled coils. The image was derived from data submitted to Protein Data Bank (PDB: 1L8D)*[79].

---

the potential role of Xrs2 (and Nbs1) in processes, other than end-resection, where it may also act as a mediator of the DNA end-joining (further discussed in section 6.5 and **Paper IV**).

3.2.2 Non-Homologous End-Joining

NHEJ is an alternative pathway for DSB repair in both eukaryotes and prokaryotes[83]. As opposed to HR, the NHEJ process does not depend on a homologous template, and thus allows for joining of any free DNA ends[4]. It is therefore typically more error-prone than HR, although this is compensated by the high efficiency of NHEJ proteins to detect and repair lethal DSBs *in vivo*[84, 85]. Upon DNA damage, the NHEJ repair mechanism will be activated to rejoin the free DNA ends within a few minutes, while the sequence homology mediated repair would require several hours for restoring the DNA molecule[85]. NHEJ is active throughout the entire cell cycle and is the markedly preferred option outside of the S/G_2 phases, when DSB repair is dominated by HR[6, 57]. Thus, NHEJ is important for maintaining the genome integrity, in particular when cells are exposed to stress-factors that may induce DSBs during any phase of the cell cycle. The highly promiscuous nature of NHEJ further allows for a wide diversity of ends to be joined, independent of overhang length, sequence or abnormal DNA end structures[4, 86].

DSB repair by NHEJ in mammals is typically initiated by the binding of the Ku70/80 heterodimer, potentially protecting the broken DNA ends from unspecific nucleolytic degradation[4, 87]. Once bound, Ku70/80 acts as a hub to recruit downstream components, necessary for the complete rejoining of the DNA ends. The DNA-

dependent protein kinase catalytic subunit (DNA-PKcs) is the first component to form a complex with the DNA-bound Ku70/80, forming the DNA-PK complex[88, 89]. The nucleolytic protein Artemis is subsequently recruited and phosphorylated at the site of repair to eliminate potential secondary DNA structures and single-stranded overhangs[90]. A complex of the X-ray cross-complementing protein 4 (XRCC4) and DNA Ligase IV (LigIV) is recruited by Ku70/80 to the DNA, followed by binding of the XRCC4-like factor (XLF)[91]. XRCC4 itself does not exhibit any enzymatic activity and only seems to serve as a scaffolding protein, to which other NHEJ factors can bind[91]. The polynucleotide kinase-phosphatase (PNPK)[92] is such a factor, which is involved in the DNA end-processing along with Artemis and other proteins[91]. XRCC4 in complex with XLF has also been shown to form helical filaments[93], which are able to interact with and bridge DNA[94]. The stabilized DNA ends are covalently attached by LigIV, which does not discriminate ends based on their compatibility, and hence allows for error prone repair of DSBs.

In contrast to the intricate eukaryotic NHEJ machinery, the corresponding bacterial process comprises only two key factors, the Ku homodimer (Ku) and Ligase D (LigD)[95]. The Ku gene of most bacterial species are generally found to be organized in operons that also encode genes for the corresponding ligases[96]. The bacterial LigD is typically stimulated by the operon-associated Ku and exhibits limited ligase activity when combined with eukaryotic Ku70/80[95]. This suggests that both proteins form a species-specific NHEJ complex. In the initial step of the bacterial NHEJ (Figure 11), Ku binds to the free DNA ends in order to protect them from unspecific nucleolytic activity, similar to the eukaryotic system. This is followed by the recruitment of LigD, which typically comprises an ATP-dependent ligase domain, a polymerase domain and a phosphoesterase domain[97]. LigD is mainly responsible for the DNA end-processing, while Ku is anticipated to modulate the extent of DNA resection by staying bound to the free ends[95]. The *Bacillus subtilis* LigD, which has been studied in the original work presented in this Thesis (further discussed in section 6.3 and **Paper II**), lacks the phosphoesterase domain[98, 99]. The minimalistic nature of the prokaryotic NHEJ machinery may suggest that evolution has generated a more efficient system to repair DSBs in bacteria, potentially due to their higher proliferation rate compared to eukaryotic cells.

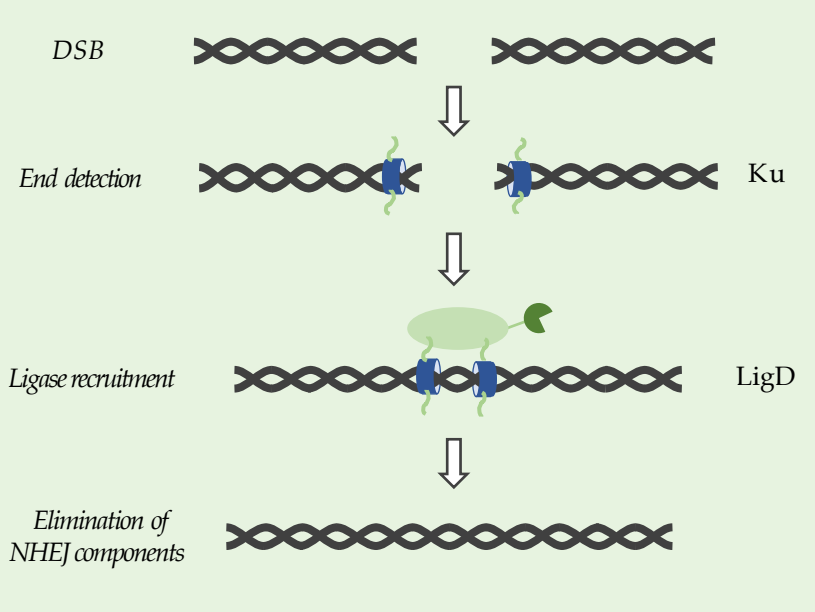

Figure 11: *The minimalistic B. subtilis NHEJ mechanism comprises only two key components, the Ku homodimer and the multifunctional LigD. The ring-shaped Ku homodimer binds to the broken DNA ends and recruits LigD to the site of repair for covalent rejoining of the ends.*

The Ring-Shaped Ku Dimer

The Ku dimer is a conserved component of the NHEJ machinery across all domains of life and is implicated in several cellular processes other than DSB repair, such as telomere maintenance, transcription and apoptosis[100]. The complex was first discovered in the serum from a patient suffering from a connective tissue disorder, and was further identified as an autoantigen[101]. The protein complex is primarily localized in the cell nucleus[102], however there is evidence supporting that Ku may also be found in the cytosol of some eukaryotic species[103, 104]. It is anticipated that there are roughly around 4-5 x 10^5 Ku70/80 heterodimers in a single human cell[4, 105]. This rather high protein abundance, combined with an equilibrium dissociation constant in the nanomolar regime for dsDNA ends, strongly supports the fact that Ku70/80 is the first protein to bind to the broken DNA ends resulting from a DSB. Crystallographic data on the Ku70/80 heterodimer (Figure 12a) reveals an asymmetric ring-shaped structure with a central hole, through which a dsDNA molecule can be threaded[106]. Upon binding, the complex can translocate along the DNA contour to allow for multiple Ku70/80 heterodimers to enter by the ends[100, 107]. The binding is believed to be sequence independent, as no direct contact was observed between the protein subunits and the nucleobases in the crystallographic data[106]. Owing to its ring-shaped structure, Ku70/80 likely becomes topologically trapped on the DNA molecule upon completion of DSB repair through NHEJ. However, the fate of the heterodimer post DNA repair

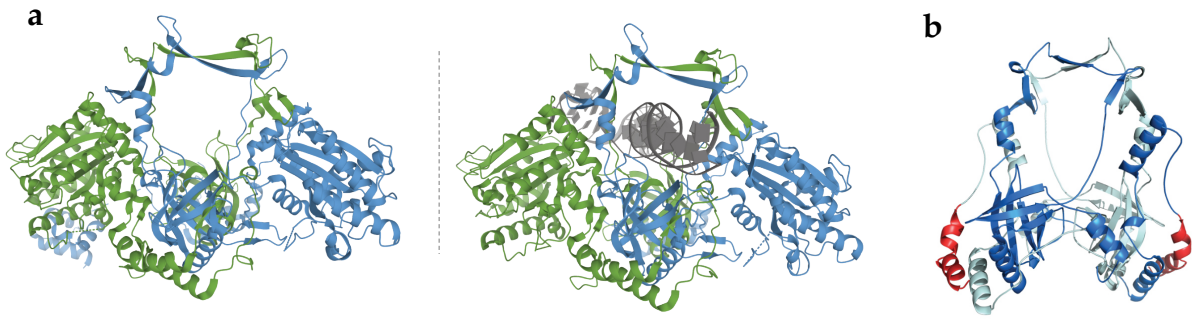

Figure 12: ***(a)*** *Structure of the human Ku70/80 heterodimer in the absence (left) and presence (right) of DNA (grey). The Ku70 and Ku80 subunits are visualized in blue and green, respectively. The images were derived from data submitted to Protein Data Bank (PDB: 1JEQ and 1JEY)*[106]. ***(b)*** *Structural prediction of the bacterial Ku homodimer, where the two identical subunits are visualized in different shades of blue. The protruding C-terminal regions are highlighted in red. The structure was generated by comparative modeling, using the eukaryotic Ku70/80 heterodimer in* ***(a)*** *as template.*

---

is still not fully understood. Several reports have demonstrated that the Ku70/80 complex can be released from the DNA through different mechanisms, predominantly based on active modifications to the Ku80 subunit[108-110].

In contrast to its eukaryotic counterpart, the prokaryotic Ku complex is composed of two identical 30-40 kDa subunits, forming a homodimer[96]. Although it is almost half the size of Ku70/80, the bacterial homodimer also forms a ring-shaped structure with a central hole, through which the DNA can be threaded (Figure 12b)[95]. The *B. subtilis* Ku homodimer has previously been demonstrated to harbor a core domain with two lysine-rich C-terminal protrusions, originating from each of the two subunits. These 58 amino acid long protrusions appear to be crucial for the recruitment of LigD to the site of repair[111]. In this Thesis, the dynamics of *B. subtilis* Ku has been characterized on the single-molecule level, with regards to joining DNA ends, alone and in the presence of LigD (further discussed in section 6.3 and **Paper II**).

4 Fluorescence – *Visualizing DNA and Proteins Using Light*

Fluorescence is the emission of light by a molecule upon relaxation from higher energy states. An early observation of a fluorescent liquid solution was reported in 1565 by Nicolás Mondares, who described the blue color of the infusion of *Lignium nephriticum*, known as the "kidney wood"[112]. In 1845 a similar observation was made by Sir John Herschel, who realized that the exposure of a quinine solution to sunrays will make the otherwise colorless solution emit blue light[113]. Since then, the phenomenon of fluorescence has evolved into a primary research tool in biological and chemical sciences. The sensitive nature of fluorescence makes it powerful in many different applications, such as molecule detection, characterization and visualization.

4.1 Light and Matter

Light can be described as propagating waves, composed of oscillating magnetic and electric fields at a right angle to each other and to the direction of propagation. The energy of light is quantized, its smallest unit being the *photon*, which is a discrete elementary particle with zero mass and electric charge. Photons propagate at the speed of light and translocate energy in space. The energy of photons is related to the oscillation frequency of the wave. The human eye is sensitive to electromagnetic radiation in the visible spectrum, typically corresponding to wavelengths in the range of 400–700 nm[114]. The relation between energy (E) and wavelength (λ) of a photon is given by Eq. 4,

$$E = h\nu = \frac{hc}{\lambda} \qquad \text{(Eq. 4)}$$

where h is Plank´s constant, ν is the oscillating frequency and c is the speed of light. The interaction of light with matter is dependent on the electron configuration of the atoms in the molecule. To describe the atomic structure, Niels Bohr proposed a model, where a positively charged nucleus is surrounded by electrons revolving in orbits with discrete energy levels[115]. This model was later extended to describe the relative position of an electron by a probability distribution function, which gives an estimate of the electron density around the nucleus. Ludwig Boltzmann stated that the distribution of an electron population can be described by an exponential function, taking into consideration the temperature and the energy difference between two levels. An immediate consequence of this rationale is that more energy levels will be occupied at higher temperature. The energy gap between two levels (ΔE) corresponds to the energy required for an electron to transfer to the higher level, given by Eq. 5.

$$\Delta E = E_{n+1} - E_n = \frac{hc}{\lambda} \qquad \text{(Eq. 5)}$$

When a molecule is excited by light, energy will be added to the system, causing the molecule to transition to an excited state. As the excitation is terminated, excess energy can be dissipated through various processes, allowing the excited molecule to relax back to its ground state. Upon absorbing energy of a certain wavelength, electrons will transition from the ground electronic level, usually the singlet ground state (S_0), to an excited electronic singlet state (S_n). The electronic states and transitions between them can be illustrated in a *Jablonski diagram* (Figure 13). The energy difference between the ground state and excited electronic levels is typically on the order of 10^0 eV, whereas the corresponding number for vibrational and rotational levels are 10^{-2} eV and 10^{-4}

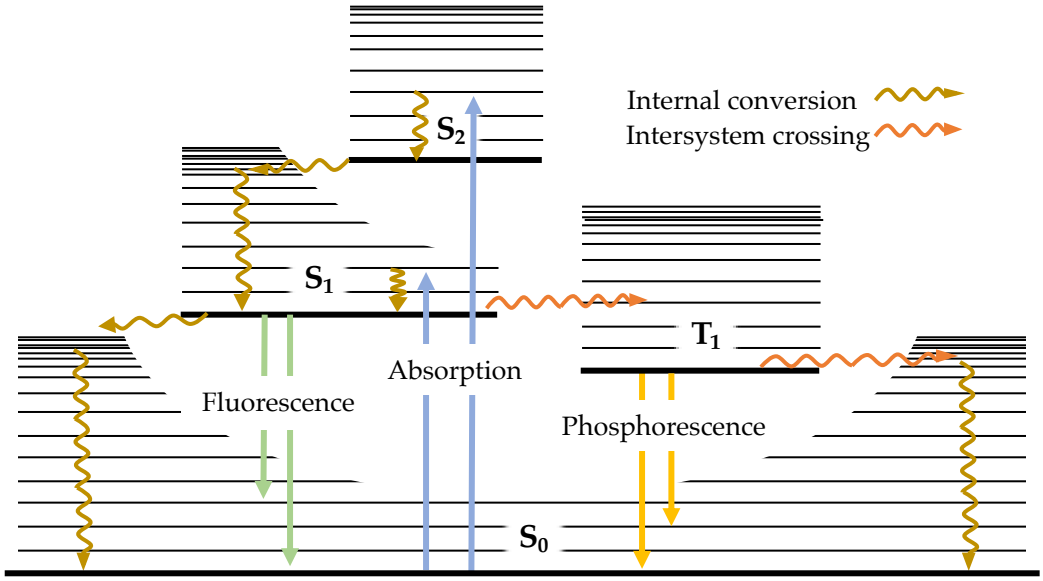

Figure 13: *Electronic transitions visualized in a Jablonski diagram. Upon excitation (straight blue arrows) from the ground electronic level (S_0) electrons will occupy different vibrational levels in the excited state. The molecule will lower its energy mainly through non-radiative processes, such as internal conversion and intersystem crossing (undulating arrows), to finally relax back to S_0. Relaxation can also occur through fluorescence and phosphorescence, which are radiative processes (straight yellow and green arrows, respectively), where photons are emitted upon deactivation of the electrons from the excited singlet (S_1) or triplet (T_1) levels to S_0, respectively.*

---

eV[116], respectively. Following absorption of energy, the molecule will typically enter a higher vibrational state (V_n) of either S_1 or S_2. Most molecules will relax to the lowest vibrational state of S_n through a process known as *internal conversions* (IC), where excess energy is emitted to the surroundings as heat. Subsequently, the molecule can relax to S_0, either through IC or through a radiative process called *fluorescence*, where photons are emitted with a wavelength corresponding to ΔE. The fluorescence lifetime, *i.e.* the average time a molecule spends in its excited state, is usually on the order of nanoseconds[117].

The energy of the emitted photons is generally less than what was absorbed by the molecule when excited to S_n. This is partially a consequence of IC, where the electrons decay to the lowest vibrational level of S_1 before light emission, and also due to electrons decaying to higher vibrational levels of S_0 post photon emission. Solvent molecules can also influence the energy of the emitted photons. Upon energy absorption, the dipole moment of the excited molecule will change, causing rearrangement of surrounding solvent molecules. The alignment of the solvent, known as *solvent relaxation*, will reduce the energy level of the excited state, resulting in a lower energy of the emitted photon. The energy difference between the immediate post absorption and pre emission states of an electron is referred to as *Stokes shift*, and

is usually reported as the shift between the wavelength, at which maximum absorption and emission is observed, respectively. Following the excitation, electrons can also undergo *intersystem crossing* (ISC), which is a transition to the excited triplet state (T_n) before relaxing back to S_0. Electrons exist in pairs, each one spinning in the opposite direction of the other, which makes the net angular momentum zero. Under certain conditions, one of the electrons can undergo a spin transition, changing its spin direction and consequently the net angular momentum of the electron pair, causing the electrons to end up in T_n. This transition is kinetically unfavorable and progresses at significantly slower timescales than singlet-singlet transitions, due to the forbidden nature of this process. The relaxation from T_n back to S_0 can go either through non-radiative ISC and IC or a radiative process called phosphorescence. This typically happens when other routes of relaxation are energetically unfavorable. In contrast to fluorescence, the phosphorescence lifetime is typically in the order of milliseconds to seconds[117].

4.2 Fluorescent Dyes

Fluorophores are molecules with an inherent ability to emit photons upon relaxation from the excited singlet state. They are generally divided into two classes – intrinsic and extrinsic. Intrinsic fluorophores are naturally occurring and can be found for instance in green plants, which have a pigment called chlorophyll. The metabolic co-factor NADH is another example of an intrinsic fluorophore, emitting in the range of the spectrum corresponding to blue light. However, most molecules do not display fluorescent properties and must therefore be modified to emit light in the visible spectrum. To do so, extrinsic fluorophores can be used, such as for instance YOYO-1 and ethidium bromide (EtBr), which are both DNA-binding dyes.

The brightness of fluorophores is characterized by two parameters, the molar absorptivity and the quantum yield. The molar absorptivity reflects the ability of a molecule to absorb light of a certain wavelength, while the quantum yield reports the fraction of absorbed photons that are emitted as fluorescence. Furthermore, the spectral separation between the excitation and emission wavelengths plays an important role when considering a fluorophore for imaging purposes. Typically, a large Stokes shift is preferred, where the excitation and emission spectra are well separated. Major overlaps in the spectra can result in lower quantum yields due to self-quenching (further discussed in section 4.3), thereby complicating the distinction between the incident light and the emitted photons, for instance during fluorescence

imaging. Labelling of a substrate with external fluorophores may change the overall physical properties of the target molecule. To minimize perturbations on the substrate, fluorophores with appropriate size and shape should be used for external labelling.

4.2.1 Staining DNA – The Intercalating Dyes YOYO-1 and EtBr

Due to the very weak intrinsic emission of the nucleobases, binding of external fluorophores is crucial to enable fluorescence studies of DNA. YOYO-1 (introduced in section 2.4) is a common fluorescent nucleic acid intercalator, frequently used in quantitative and qualitative DNA studies. When intercalated, YOYO-1 displays a molar absorptivity at about 10^5 $M^{-1}cm^{-1}$ at 491 nm, which is where maximum absorbance is observed, and an emission maximum at 509 nm (Figure 14a)[37]. The emissive properties of YOYO-1 are greatly increased upon intercalation, displaying a 3200-fold increase in brightness, compared to when it is freely suspended in solution[39]. Due to steric hindrance of the oxazole yellow (YO) units, decreased rotational mobility around the methine bridge, which links the two aromatic ring systems, will restrict relaxation from the excited state through IC and hence promote deactivation by fluorescence[118]. The virtually non-existing background emission contribution from unbound dye molecules, and the strong affinity to DNA, makes YOYO-1 an ideal probe in applications where DNA is to be visualized. Another frequently used DNA dye is EtBr. The planar aromatic structure of EtBr allows it to intercalate between the DNA nucleobases, as for YOYO-1, and thereby exhibits a 20-fold increase in brightness[119]. EtBr has an absorption and emission maximum at 520 nm and 610 nm,

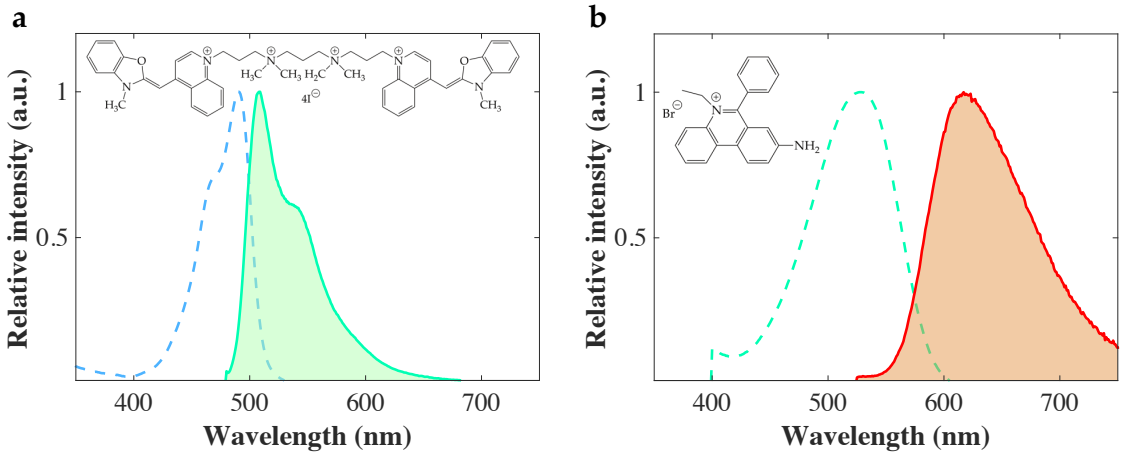

Figure 14: *Absorption (dashed line) and emission (solid line) spectra for* ***(a)*** *YOYO-1 and* ***(b)*** *EtBr, together with their respective molecular structures.*

respectively (Figure 14b)[120]. The large Stokes shift allows for efficient distinction between the incident light and the emission, making EtBr a preferred DNA dye. However, the severe health risks associated with EtBr (it is a potent mutagen), has resulted in the development of new and more safe DNA dyes, covering the entire spectral range of visible light.

4.2.2 Protein Labelling

Biopolymers generally exhibit poor emissive properties. Some proteins, on the other hand, are unique in this respect, as they can display intrinsic fluorescence. The aromatic amino acids phenylalanine, tyrosine and tryptophan are responsible for this attribute, and it is the latter that is most abundant in proteins, usually present at about 1 mol%[117]. With emission maxima in the range of 280–350 nm, the fluorescence of tryptophan is predominantly found to be in the ultraviolet spectral range. Extrinsic fluorophores can be added to provide emissive properties in the visible spectrum, enabling detailed characterization and visualization of proteins by fluorescence methods. There are various *in vitro* labelling methods available, which essentially differ by their reactivity towards specific entities on the protein. Primary amines, carboxyls and sulfhydryls are typically used for functionalization of proteins, where an external molecule, such as a dye, is conjugated to the labelling site of interest. However, these strategies may result in inconsistent labelling, where multiple dye molecules are attached, ultimately perturbing the structure and function of the protein. To overcome this, proteins can be fused with an extra domain on the N- or C-terminal, which may serve as a single reactive site for *in vitro* labelling. A commonly used fusion substrate is the SNAP-tag, which reacts with O_6-benzylguanine derivatives to covalently attach a label to the protein of interest[121, 122]. This particular approach was used to fluorescently label the bacterial Ku protein (further discussed in section 6.1 and **Paper II**). There are also naturally occurring photoproteins available, which can be used as fusion substrates. The green fluorescent protein (GFP) is such an example with an inherent fluorescence emission maximum at 509 nm[123]. Ever since first isolated from the jellyfish *Aequorea victoria* in 1962, the GFP has been used extensively in various protein studies[124]. The excellent photophysical characteristics of GFP has resulted in the development of derivatives emitting light in the whole visible spectral range[125].

4.3 Resonance Energy Transfer Between Molecules

Fluorescence is highly responsive to the immediate surrounding of the fluorophore, where the presence of quenchers in the close vicinity can cause substantial reduction of the emitted light. Resonance energy transfer (RET) is a non-radiative mechanism, by which energy can be transferred between a fluorophore in the excited state and an acceptor molecule. This process can occur whenever the emission spectrum of the fluorophore overlaps with the absorption spectrum of the acceptor molecule, giving rise to energy transfer through dipole-dipole interactions. The extent of the interaction is determined by the distance, where RET can occur between molecules up to 10 nm apart, and also the alignment of the transition dipole moments[126].

Based on their spectral overlap (Figure 14), the combination of YOYO-1 and EtBr makes a suitable RET-pair, the first being a donor and the latter being an acceptor. When intercalated in DNA, both fluorophores will be stacked in perpendicular orientation with respect to the helical axis, at a distance in the sub-nanometer scale. If the conditions for RET are fulfilled, the excitation quanta can be transferred to the neighboring EtBr upon excitation of YOYO-1. The EtBr in turn becomes excited and can emit light (Figure 15). Thus, by combining two fluorophores with different spectral properties, excitation light in the blue spectral region can be used to induce emission from a fluorophore with an absorption maximum in the green range of the electromagnetic spectrum. Self-quenching also occurs through RET and is usually observed for fluorophores with a small Stokes shift. The emitted energy is then reabsorbed by neighboring molecules of the same kind, which reduces the overall fluorescence emission from the sample.

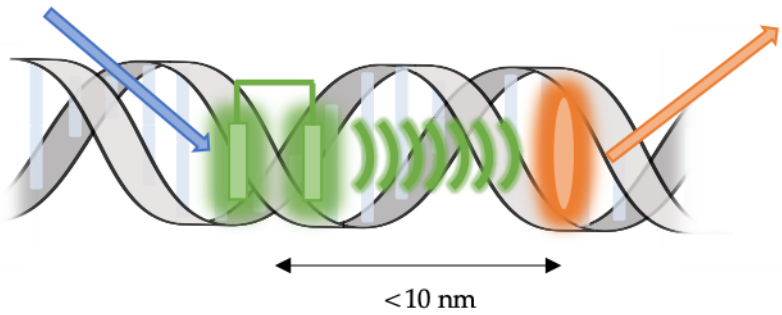

Figure 15: *Schematic illustration of energy transfer between YOYO-1 (green) and EtBr (orange), when intercalated in DNA. Upon excitation with blue light, the excitation energy from YOYO-1 is transferred to EtBr. The subsequent relaxation of EtBr back to ground electronic state can result in orange fluorescence emission.*

4.4 Principles of Fluorescence Microscopy

The human eye has typically a resolution of around 200 μm[127]. Optical microscopes use the interaction of light with matter to enable visualization of objects smaller than the resolving power of the human eye. Using a set of lenses, contained in an objective, magnified images can be created by illuminating the specimen and recording the resulting transmitted, reflected, or emitted light. The spatial resolution (d) reflects the smallest discernible details in an image and measures the closest distance between two objects that are observed as separate entities, defined by the Rayleigh criterion, given by Eq.5,

$$d = 0.61\frac{\lambda}{NA} \qquad \text{(Eq. 5)}$$

where λ is the wavelength of the light and NA is the *numerical aperture* of the objective. The numerical aperture is a dimensionless quantity describing the range of angles, over which an objective can collect light in various immersion media. It is determined from Eq. 6,

$$NA = n\sin\theta \qquad \text{(Eq. 6)}$$

taking into account the half-angle (θ), which definiens the size of the cone of light that can enter the objective, and furthermore the refractive index (n) between the objective and the specimen. Given λ = 509 nm, corresponding to the emission maximum of YOYO-1, the theoretical spatial resolution of a YOYO-1-stained DNA molecule is ~220 nm, using an oil-immersion objective with NA = 1.4.

In fluorescence microscopy the specimen to be observed is illuminated with light of certain wavelengths to excite the fluorophores in the sample. The concomitant fluorescence emission is recorded to generate an image. The main components of a fluorescence microscope comprise a monochromatic light-source, such as lasers or light-emitting diodes (LEDs), a dichroic mirror, an emission filter, an objective and a detector. Alternatively, broad-spectrum light-sources, such as an arch-discharge lamp, which emits most wavelengths in the visible spectrum, can be used to excite the sample. To obtain monochromatic excitation light when using a broad-spectrum light-source, an excitation filter must be included in the light path. In an *epi*-configured fluorescence microscope (Figure 16), the sample is placed on top of the objective and is thus excited from below. The light passes through an excitation filter with appropriate bandwidth, before being reflected by a dichroic mirror through the objective onto the specimen. The emitted photons travel in the opposite direction

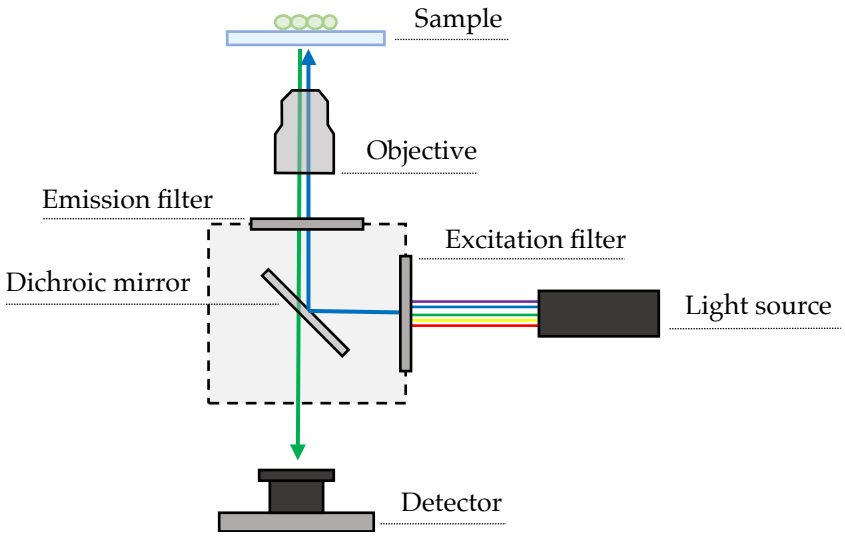

Figure 16: *Schematic illustration of the light-path through the main components of an epi-configured fluorescence microscope. Light is passed through an excitation filter before being reflected by a dichroic mirror through an objective on to the sample, which will be excited. The emitted photons from the fluorescent sample will pass through the objective and subsequently be transmitted through the dichroic mirror before hitting the detector.*

---

through the objective and subsequently hit the dichroic mirror again. Since a dichroic mirror is color selective, wavelengths below a certain threshold (typically set to match the excitation range of a fluorophore) will be reflected, whereas higher wavelengths (the fluorophore emission) will be transmitted. The transmitted light will pass through an emission filter to isolate the back-scattered light from the sample, before it is directed to the detector. As light travels through the objective, diffraction will occur as a result of the light wave encountering obstacles in its path, such as slits and apertures. The fluorescence emission from a point source will then exhibit a three-dimensional intensity distribution, with a central peak surrounded by rings at increasing distance from the center, corresponding to the diffracted photons. This is referred to as the *point spread function* (PSF). The *Abbe diffraction limit* defines the radius (r) of the PSF, given by Eq. 7.

$$r = \frac{\lambda}{2NA} \qquad \text{(Eq. 7)}$$

Considering again the emission from YOYO-1, the diffraction limit of a fluorescence microscope, equipped with an immersion objective with NA = 1.4 is ~180 nm, suggesting the perceived size of an imaged single fluorophore. This is not to be confused with the spatial resolution given by the Rayleigh criterion (Eq. 5), which defines the minimum distance at which two PSFs in close proximity are distinguishable.

5 Studying Biomolecules – *Ensemble and Single-Molecule Techniques*

The immense complexity of biology requires the utilization of innovative experimental techniques to gain a deep understanding of biological systems on the molecular level. Biomolecular interaction studies have long been a matter of interest to scientists, which has consequently pushed the development of new and more advanced methods. Well-established *in vitro* ensemble techniques, such as gel electrophoresis and surface plasmon resonance, are being complemented with techniques allowing thorough characterization of single molecules. Important advancements were made when techniques were developed for conjugating fluorophores to biomolecules, allowing analysis on the single-molecule level, to directly visualize biomolecular interactions.

5.1 Studying DNA and Proteins in Bulk

Ensemble is usually referred to as the joint contribution from individual entities within a population. When studying a sample of DNA or proteins in an ensemble experiment, the combined properties of several millions of individual molecules are observed, resulting in an average signal being recorded for the sample. Analytical techniques, such as gel electrophoresis and surface plasmon resonance, record the joint signal of a population of molecules and are hence categorized as ensemble techniques.

5.1.1 Gel Electrophoresis

Separation based on electrophoretic mobility has long been used to characterize DNA and proteins *in vitro.* In gel electrophoresis, molecules are separated on agarose or polyacrylamide gels by applying an electric field, causing migration of residues in the gel matrix. For DNA, the inherent net negative charge is directly proportional to the length of the sugar-phosphate backbone and in turn, the number of nucleotides. Thus, the mass-to-charge ratio will remain constant, allowing DNA molecules of different sizes to migrate at the same rate in a friction-free environment under the influence of a given electric field. Electrophoretic separation of DNA fragments in a gel matrix, which causes friction, will mainly depend on the gel composition. The pore-size distribution in the gel can be tuned to favor the migration of molecules within a particular size range. A high agarose or polyacrylamide concentration will produce pores with a small size distribution, favoring the migration of short polymers, which can allow for separation of DNA fragments differing in length by as little as a single nucleotide[46]. By staining the gel with for instance YOYO-1 or EtBr, the migrated DNA molecules can be fluorescently visualized and quantified. Molecular weight size markers can be used to estimate the size of the DNA fragments in a given sample. Large polymers will be hampered by the gel matrix and may thus get stuck in the loading well or migrate at a much lower rate. This is usually referred to as the *sieving effect*. To separate large DNA molecules in the kilo- to mega-base-pair regime, pulsed-field gel electrophoresis (PFGE) can be employed. This is a type of agarose gel-electrophoresis, where the direction of the electric field is periodically switched to reorient the molecules and thereby obtain a more efficient size-based separation[128].

In contrast to DNA, proteins are non-uniformly charged molecules, which are often folded into different tertiary structures. Electrophoretic separation of native proteins will thus depend not only on the size, but also on the mass-to-charge ratio and the shape of the molecules. Native polyacrylamide gel electrophoresis (PAGE) is typically employed to separate proteins in a mixture without, disrupting the tertiary

structures. However, to obtain a size-based electrophoretic separation, denaturing gels can be used, such as SDS-PAGE. The gel is cast in the presence of the anionic detergent sodium dodecyl sulphate (SDS), which is also included in the running buffer during the electrophoresis and in the sample. By heating the protein sample prior to loading on the gel, the secondary and tertiary protein structures are disrupted. The addition of reducing agents will further break any disulfide bridges, ultimately resulting in a linear protein chain, wrapped in detergent. The intrinsic charge of the amino acids is masked by the anionic SDS, giving an almost uniform mass-to-charge ratio, similar to that of DNA, which allows for size-dependent separation in the gel matrix.

Electrophoretic Mobility Shift Assay

Gel electrophoresis is commonly employed in DNA-protein interactions studies. Electrophoretic mobility shift assay (EMSA) is a sensitive method based on electrophoretic mobility of molecules in a gel matrix, typically made of agarose or polyacrylamide, to detect complex formation between nucleic acids and proteins. The assay is performed under native conditions to preserve the tertiary structure of the protein, and thereby obtaining information on its specific interactions with DNA. Although mostly used for qualitative analysis, EMSA can also provide quantitative information on binding stoichiometry, affinity and kinetics. When performing EMSA, the substrate (usually a small DNA molecule) is titrated with an analyte (a protein of interest) before separation on the gel. By observing the migration of the substrate with increasing concentration of the analyte, complex formation can be detected as the DNA is retarded upon structural and physical changes due to protein binding. Multiple proteins can be analyzed simultaneously, which may provide information on combinatory effects. By labelling both the substrate and the analytes, all species and their relative abundance in each band can be identified, giving additional insights on the biomolecular interactions. This may be crucial, especially when analyzing proteins that can form different higher-order complex structures with DNA.

5.1.2 Surface Plasmon Resonance

Gel-based assays typically provide information on static interactions once the DNA and protein have formed a complex. However, to monitor the dynamics of the binding events surface plasmon resonance (SPR) can be employed. SPR is a highly sensitive method, used to detect real-time biomolecular interactions. It is a quantitative analytical tool, which can provide information on binding specificity, affinity and kinetics. The method is based on detecting changes on a surface as a result of

biomolecular interactions. A typical SPR setup consists of a light-source, a prism coated with an electrically conductive metal (typically gold) on one side, a reaction flow-cell and a detector (Figure 17a). A beam of light is directed towards the gold-coated side of the prism at a certain resonance angle, where some photons will excite plasmons, which can be described as electrons propagating through the material. The remaining photons will be reflected towards the detector. The oscillating plasmons will create an evanescent electric field, which extends beyond the gold film into the reaction flow-cell. Local changes in the refractive index will affect the resonance angle, which can be detected through the reflected light. A substrate molecule, such as a DNA fragment, can be immobilized on the gold surface facing the reaction flow-cell. By controlled injection of an analyte at certain concentrations, the binding event can be monitored. The number of bound molecules is recorded with time through the variable resonance angle, which is coupled to the local change of refractive index upon biomolecular interactions[129]. The response is presented as a sensogram (Figure 17b), which typically displays the association, steady-state and dissociation phases. Quantitative information, such as rate constants and binding affinity, can be obtained by fitting the sensograms to mathematical models.

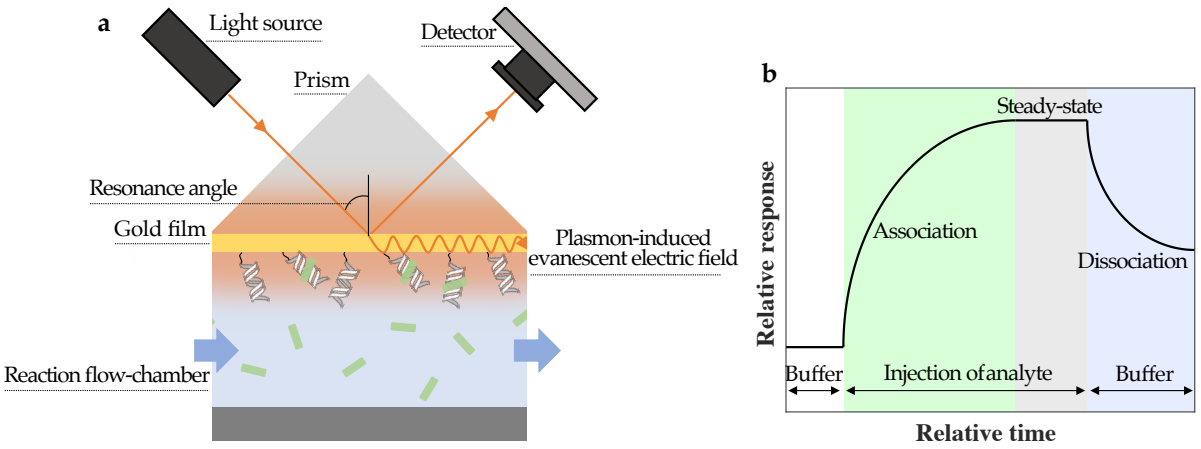

Figure 17: ***(a)*** *Schematic illustration of a typical SPR setup. Upon binding of analyte (green) to the substrate (immobilized DNA) the local refractive index will change, which is monitored by measuring the resonance angle with time.* ***(b)*** *A characteristic SPR sensogram is generated, from which the binding kinetics can be obtained. The relative response reflects variations in the resonance angle and in turn the interaction between substrate and analyte with time. The association event is followed by a steady-state phase, where equilibrium is reached. Post-injection buffer flow will allow the analyte to dissociate from the substrate, resulting in a declining relative response.*

5.2 Single-Molecule Analysis

Most biomolecules behave as individual entities in a stochastic and heterogeneous manner. Unique and transient features, potentially crucial for the complete understanding of a biochemical reaction, are likely to be lost in bulk assays due to the averaging effect inherent to such techniques. Hence, single-molecule analysis provides important insight into the function and behavior of individual molecules in a population by avoiding the ensemble averaging effect[7]. Studies on interactions between proteins and single DNA-molecules have attracted major interest, which has resulted in the development of many different techniques to stretch DNA and characterize long DNA molecules[11, 130]. Extending the coiled DNA molecule into linear conformation is crucial in order to visualize static and dynamic interactions with other molecules. This can be achieved by, for instance surface immobilization of DNA stretched by hydrodynamic flow[131, 132], exerting active stretching force on the molecule using magnetic[133] and optical tweezers[134], or through confinement in nanofluidic channels, allowing the DNA to spontaneously extend[10].

5.2.1 Immobilized DNA-Protein Complexes – Atomic Force Microscopy

One of the most common analytical techniques for studying surface immobilized and stretched DNA is atomic force microscopy (AFM). First described by Gerd Binnig and co-workers in 1986 as an alternative application of scanning tunneling microscopy, AFM was proposed to enable studies of surfaces on the atomic scale[135, 136]. Ever since its introduction, AFM has been used frequently to study various species on modified surfaces. Owing to its negative charges along the sugar-phosphate backbone, the DNA molecule is an suitable substrate for immobilization on positively charged surfaces and thereby an excellent subject for AFM studies on the single-molecule level[137]. The principle by which AFM operates is based on the detection of forces generated upon interaction between the specimen and a probe, typically made of gold, which is positioned on a cantilever (Figure 18). When approaching the face of an object, interactions between the outermost atoms of the sharp-tipped probe and the specimen will give rise to forces, resulting in deflection of the cantilever. These deflections are monitored through the reflection of a laser light onto a point-sensitive photodetector, which allows for accurate positioning of the probe and thereby visualization of the surface topography[135]. There are different modes of AFM operation, of which the intermittent contact mode, usually known as tapping mode, is most commonly used for mapping static topographical structures on a surface, including immobilized DNA and proteins[138]. In the tapping mode, the cantilever is oscillating up and down with a

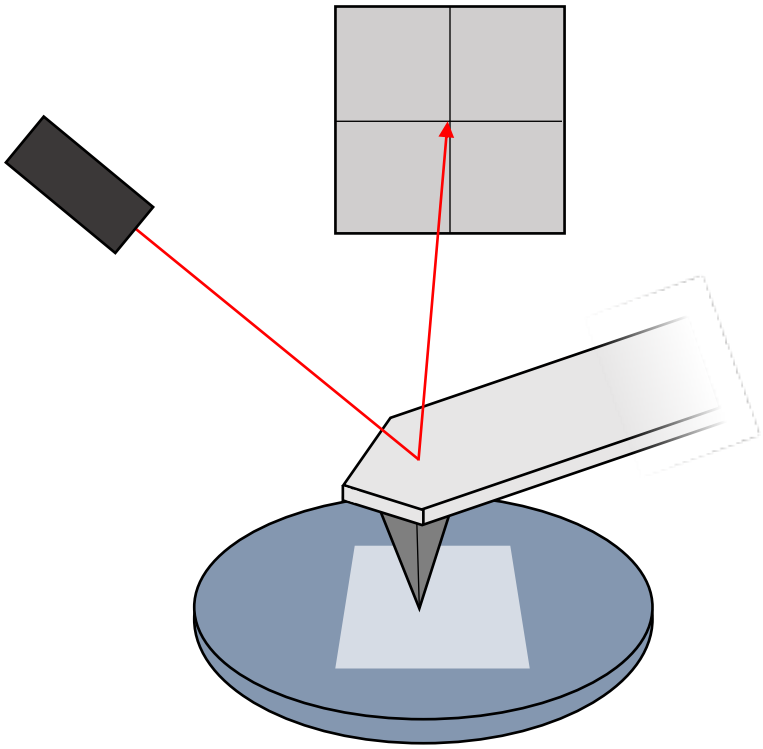

Figure 18: *Working principle of an atomic force microscope. A sharp-tipped probe is attached to a cantilever and scanned over the surface to be imaged. Interactions between the probe and the surface-immobilized specimen will result in deflection of the cantilever, which is detected by the reflection of a laser light onto a point-sensitive photodetector.*

---

given amplitude as it approaches the surface. At a sufficiently small distance to the object of interest, interacting forces between the specimen and the tip will give rise to changes in the oscillating phase and the amplitude of the cantilever vibrations. Such changes can reflect topographical variations on the surface with a lateral resolution around 1 nm and a vertical resolution on the sub-nanometer scale[139]. Apart from generating high-resolution images, an important feature of AFM is the ability to study various samples in different media, including in air and in aqueous solutions. This is of particular advantage in biological studies, where the hydration level may influence the observed structures[140].

5.2.2 Stretching DNA by Handles – Magnetic and Optical Tweezers

In 1992 Carlos Bustamante and co-workers revolutionized the field of single DNA-molecule analysis by reporting a novel experimental setup to probe the elasticity of DNA, using magnetic beads as handles. A method nowadays known as magnetic tweezers[133]. By trapping the beads in a magnetic field, the tethered DNA ends can be pulled apart while monitoring the force exerted on the molecule. Owing to the

property of magnets being able to impose rotation, magnetic tweezers are particularly suited for studying the DNA topology and torsional effects on the molecule upon interaction with proteins, such as topoisomerases[141].

A particularly intriguing application of magnetic tweezers was reported in 2018, where Wang *et al.* presented a novel DNA substrate for single-molecule nanomanipulation, allowing for mechanical probing of DNA bridging in real time (Figure 19)[142]. The DNA substrate, referred to as single-molecule forceps, was constructed as such that two dsDNA molecules of 1510 bp each were connected through a bridging dsDNA fragment of 690 bp. Anchoring each end of the bridging fragment to one 1510 bp DNA at a distance of 58 bp from the end of the molecule being tethered, resulted in a substrate constructed to have two freely fluctuating ends facing each other. By attaching the outer end of each 1510 bp construct to a surface and a magnetic bead, respectively, controlled extension of the substrate was made possible. Thus, using this substrate in combination with magnetic tweezers, protein-induced bridging between the freely fluctuating ends can be detected. Upon extending the construct, a reduction will then be observed in the total extension of the substrate at a

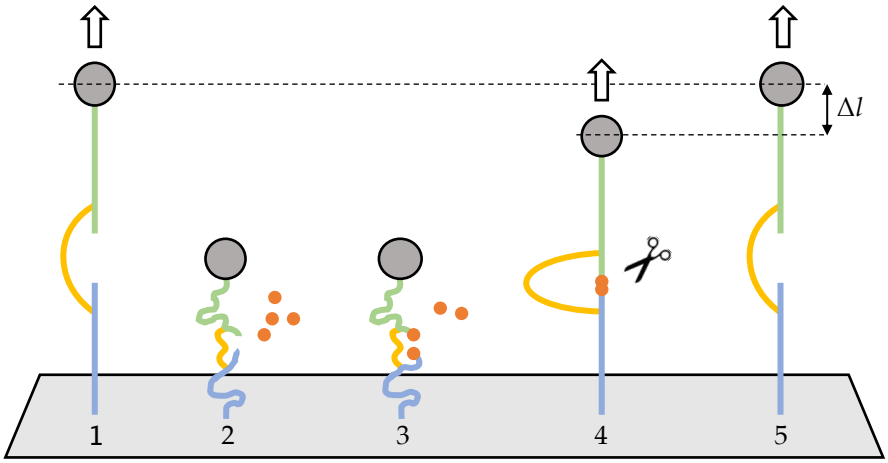

Figure 19: *Nanomanipulation of DNA with single-molecule forceps. Two 1510 bp dsDNA molecules (blue and green) are connected through a bridging 690 bp molecule (yellow). By attaching the outer end of the blue and green molecules to a surface and a magnetic bead, respectively, the entire construct can be stretched to its maximum extension by applying a magnetic field to pull the bead in the vertical direction (1). When no pulling force is applied the DNA construct will collapse, allowing interaction between the free DNA ends facing each other (2). Proteins in the surrounding buffer will bind to the free ends (3) and a potential synapsis can be detected by applying a magnetic field to pull the construct to its maximum extension (4). The reduced extension between stage 1 and 4, corresponding to Δl, suggests formation of a synapsis. Interruption of the synapsis will allow the DNA construct to attain its original extension (5).*

given pulling force compared to that of the non-bridged construct. This technique has proven to be highly suitable for studying proteins involved in DNA repair processes, such as the bacterial NHEJ (further discussed in section 6.3 and in **Paper II**).

Short after the demonstration of magnetic tweezers, a similar DNA stretching technique emerged, using optical traps instead of magnets. The technique is based on the pioneering work by Arthur Ashkin reported in 1970, where he first demonstrated that micron-sized particles can be accelerated upon irradiation with visible laser light[143]. This was later developed into a method for trapping and manipulating atoms[144] and dielectric particles[145], using electromagnetic radiation. Current optical tweezers use dielectric beads, typically made from polystyrene, to which DNA can be attached by the free ends[134]. By trapping the beads in laser beams, the DNA can be extended in a controlled manner, while recording the stretching force exerted on the molecule. The implementation of optical tweezers in a fluorescence microscopy setting allows for visualization of DNA-protein interactions in real time, alongside quantitative measurements of the force and the extension[146]. Using fluorescence coupled optical tweezers, numerous studies have reported the dynamic activities of different proteins on DNA, for instance, the translocation of polymerases and restriction enzymes, protein-induced bridging and formation of condensed DNA structures[146, 147].

5.2.3 Stretching DNA by Confinement – Nanofluidics

The stretching force exerted on DNA molecules in a magnetic or optical tweezer setup is typically in the piconewton (pN) range[148]. Interactions between DNA and some proteins, such as the prokaryotic RNA polymerase[149] and the human histone octamers[150], correspond to forces within this range. However, *in vivo*, DNA is often subjected to conformational restrictions and spatial confinement, resulting in tension on the molecule, orders of magnitude weaker and typically in the femtonewton (fN) regime[151]. Hence, the excessive force applied to stretch DNA using magnetic and optical tweezers will underscore the inherent flexibility of the molecule and may thereby potentially affect biomolecular interactions[151, 152]. An alternative approach for stretching DNA without obstructing the free ends is to use nanofluidics, where the molecule is confined in narrow channels. This will allow DNA to spontaneously attain an extended conformation due to spatial restrictions[10]. In 2004 it was demonstrated that single genomic-length DNA molecules can be uniformly stretched in nanofluidic channels with dimensions in the 100 nm regime[9]. Employing confinement-mediated DNA stretching, using such narrow channels, will eliminate the need for external

pulling forces and hence limit the tension on the molecule, to resemble that of *in vivo* conditions[152]. Another major benefit of nanofluidics is that DNA can be stretched without obstructing the free ends, which is of particular importance when studying processes such as HR and NHEJ. A simple nanofluidic device typically comprises an array of nanochannels spanning between two feeding microchannels, which are connected on each end to a loading reservoir (Figure 20). The nanofluidic chips used in **Paper I-IV** were fabricated on silicon wafers using techniques for high-precision imprinting of the nanostructures, such as photolithography, electron-beam lithography and dry reactive ion-etching. By sealing the chip with a transparent Pyrex glass lid, confined objects in the nano- and micro-channels can be observed using optical microscopy. For detailed descriptions of the fabrication process, the reader is referred to the following reviews[8, 23, 153].

As a consequence of the spatial constrictions in nano-confinement, DNA molecules will encounter an entropic barrier, which restricts them from spontaneously entering into a space with dimensions smaller than the R_G. In order to compensate for the increasing entropy upon confinement, active force has to be applied to drive the molecules into the channels. In the nanofluidic setup described above, pressurized nitrogen gas is applied to the reservoirs to create a flow of liquid through the channels. Hence, by manipulating the flow, DNA molecules can overcome the entropic barrier and thereby enter the nanochannels. The use of nitrogen gas also eliminates dissolved oxygen in the medium, reducing the light-induced DNA damage due to the generation of reactive oxygen species during optical imaging. Translocation of DNA in nanofluidic systems can also be achieved through electrophoresis. The molecules are

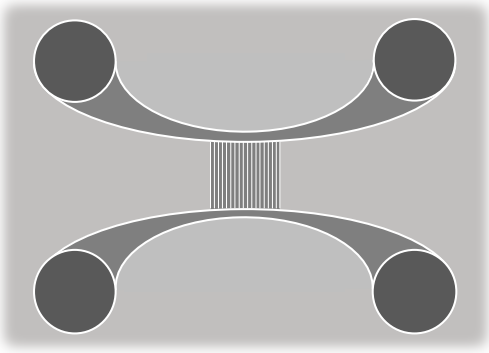

Figure 20: *In the traditional nanofluidic chip design the two upper and lower loading reservoirs are each connected through a feeding microchannel. The nanofluidic channels are spanned between the microchannels in the vertical direction.*

then driven by an electric field. Once confined, no active forces are required to keep the molecule extended in the nanochannels.

In this Thesis, the traditional nanofluidic chip (Figure 20) has primarily been used for static biomolecular interaction studies, where preformed DNA-protein complexes were being stretched and analyzed[154, 155]. Alternative devices have previously been reported, enabling *in situ* addition of analytes by passive diffusion of reagents to nano-confined DNA, allowing for semi-dynamic biomolecular interaction studies[156, 157]. By complementing the traditional nanofluidic design with additional functionalities, a novel device has been developed, which enables active buffer exchange while observing a nanoconfined DNA molecule in real time (further discussed in section 6.2 and **Paper I**). Using this device, multiple analytes can be added simultaneously to DNA *in situ*, allowing for the mapping of complex dynamic biochemical processes[158].

6 Original Work – *Results and Discussion*

DNA damage signaling and repair pathways are essential to prevent from developing various cancer diseases in humans[159]. Many components of the HR and NHEJ pathways have been linked to tumorigenesis, which emphasizes the importance of studying DSB repair, both *in vitro* and *in vivo*. This section summarizes the original work, based on the four appended papers (**Paper I-IV**), in which DSB repair has been considered from the single-molecule perspective. In section 6.1, a general description is provided on the single-molecule nanofluidics techniques and assays, which have been used in **Paper I-IV**. The following sections 6.2-6.5 summarize the main advancements and findings in the respective papers, which includes the development of a nanofluidic toolbox. The novel nanofluidics tools have been used in the analysis of key proteins involved in DSB repair in humans, yeast and bacteria. The main aim is to better understand critical features of proteins involved in the intricate DSB repair processes, which are essential to maintain the genome integrity.

6.1 Nanofluidics for Studying Static DNA-Protein Interactions

In this Thesis, nanofluidics has been used as the main tool to stretch DNA-protein complexes by confinement. Thereby, the need of attaching foreign entities to the DNA ends has been overcome, leaving the ends free and unobstructed. Biomolecular interactions have been characterized on the single DNA-molecule level, using fluorescence microscopy to visualize the binding of proteins along the DNA contour.

The Traditional Nanofluidic Device

In order to stretch DNA to an extension close to the contour length, the nanofluidic channel dimensions must be considered in accordance to Eq. 3, where dimensions satisfying $P \ll D_{av} < R_G$ will result in an extended DNA molecule (Figure 4). The channels in the traditional nanofluidic chip (Figure 20) that was primarily used for experiments in **Paper II-IV**, have a width of 100 nm and a depth of 150 nm, yielding a D_{av} = 122 nm. Given that the R_G of lambda-phage DNA (λ-DNA) is ~500 nm at 0.1 M salt concentration[17], the DNA will extended upon confinement in the channels. To visualize static interactions between DNA and different proteins, the components were mixed in appropriate buffer conditions and subsequently stained with YOYO-1, before being loaded in one of the reservoirs on the chip. Pressurized nitrogen gas was applied to the reservoir to drive the solution of DNA-protein complexes into the micro- and nanochannels.

Single-Channel Fluorescence Imaging of DNA-Protein Complexes

Information on potential biomolecular interactions can be obtained through variations in the fluorescence emission intensity along the DNA contour. In single-channel fluorescence imaging, the emission from the DNA dye is typically used as a reporter of potential interactions. The imaged DNA-protein complexes are usually visualized as *kymographs*, time-trace images. To generate a kymograph from an image stack, the individual frames are collapsed into one-dimensional arrays, each one reflecting the momentary mean fluorescence signal intensity along the stretching axis of the molecule. By stacking the one-dimensional arrays on top of each other, a kymograph is generated, where the vertical and horizontal axes correspond to time and molecular extension, respectively. The characteristic appearance of a fluorescently stained λ-DNA molecule (Figure 21a) typically shows an even emission intensity along the entire DNA contour. However, the addition of proteins, such as the bacterial (further discussed in section 3.2 and 6.3) may result in the formation of higher-order structures (Figure 21b). These could potentially report on biomolecular interactions on the single

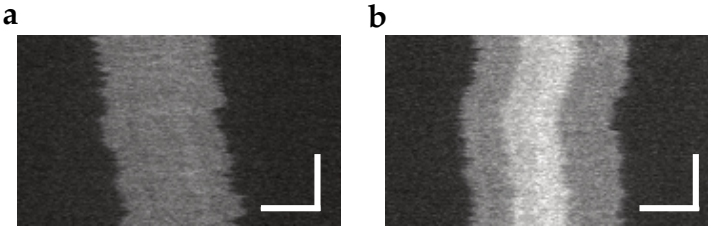

Figure 21: ***(a)*** *Kymograph of a λ-DNA molecule, labelled with YOYO-1. Upon confinement in a nanofluidic channel, the molecule will stretch to accommodate. For this particular molecule the total extension corresponds to approximately 59% of the contour length of bare λ-DNA.* ***(b)*** *Kymograph of a higher-order DNA-protein complex, formed in the presence of the bacterial Ku. The central bright patch reflects the local increase of YOYO-1 as a result of two overlapping λ-DNA molecules. The vertical and horizontal scale bars correspond to 3 s and 5 μm, respectively.*

DNA-molecule level. Furthermore, the extension of the confined DNA-protein complex may also indicate potential interactions *in vitro*. For example, DNA can be joined to form so-called concatemers, where multiple molecules have been connected by the ends to generate larger complexes[160]. Thus, by characterizing the size distribution of molecules in the sample, information on potential protein-mediated end-joining activity can be obtained.

Dual-Channel Fluorescence Imaging of DNA-Protein Complexes

Although DNA-protein complexes may be detected from the fluorescence emission of YOYO-1 along the DNA contour, this only allows for indirect observations of potential interactions. In order to localize the DNA-bound proteins, dual-channel fluorescence imaging must typically be employed, where the different biomolecules are visualized separately. This further requires that the protein of interest exhibits fluorescent properties, either intrinsically or through the addition of an extrinsic dye.

As an example, dual-channel fluorescence imaging was used to visualize the binding of the bacterial Ku on DNA in **Paper II**. To do so, a fusion protein of Ku and a SNAP-tag was constructed, which was expressed, purified and subsequently fluorescently labelled *in vitro*. The SNAP-tag allows for the addition of a wide variety of different external labels to the protein post purification. Ku was made fluorescent by reacting the SNAP-tagged protein with an Alexa-647-O_6-benzylguanine derivative to covalently attach the dye to the tag. The ability of the fluorescently labelled Ku (Ku_{AF647}) to stimulate the LigD activity was confirmed by a ligation assay, where covalent joining of a 1000 bp DNA substrate was performed, resulting in permanent concatemers of increasing size (Figure 22a). The protein was confirmed to be active and appears not affected to a notable extent by the labelling procedure. To visualize the biomolecular interactions on the single-molecule level, Ku_{AF647} was incubated with

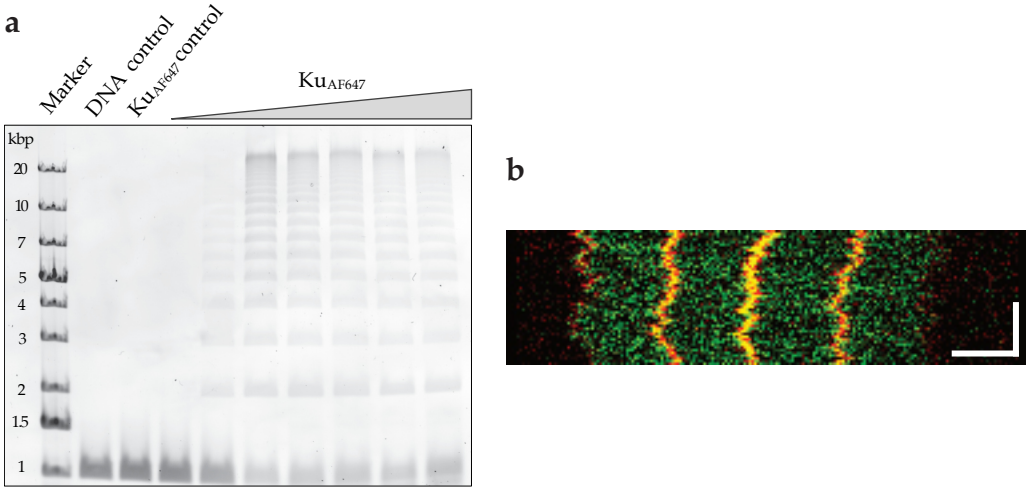

Figure 22: ***(a)*** *Agarose gel showing that Ku_{AF647} stimulates the activity of 100 nM LigD to generate concatemers from 100 ng of a 1000 bp blunt-ended DNA substrate in the presence of 0-1000 nM Ku_{AF647} (increments of 200 nM). The concentration of protein in the LigD-free Ku_{AF647} control corresponds to 1000 nM.* ***(b)*** *Kymograph of a Ku_{AF647}-λ-DNA-complex. The protein (red) appears to connect the ends of λ-DNA (green), given from the co-localization of emission signals. The vertical and horizontal scale bars correspond to 8 s and 4 μm, respectively.*

λ-DNA in the absence of LigD. The resulting kymograph (Figure 22b) shows co-localization of DNA and protein, confirming that the protein has successfully been fluorescently labelled. The activity of LigD and Ku is further discussed in section 6.3.

Thermal Fluctuations Reveal Information on the Conformation of Nanoconfined DNA

The versatility of fluorescence-coupled nanofluidics distinguishes this technique from other single-molecule approaches. It provides both qualitative and quantitative information on DNA-protein complexes, freely suspended in solution. In **Paper III**, a novel nanofluidic-based approach is presented, which allows for the detection of different structural populations within a sample of DNA-protein complexes. The extension of DNA in nanofluidic channels will mainly be determined by the channel dimensions and the salt concentration of the surrounding buffer. Thermal energy will give rise to fluctuations of the molecule extension, which can be quantified, by for instance the standard deviation of the extension. The double-folded nature of circular molecules will cause smaller axial flexibility of the polymer, resulting in lower thermal fluctuations, compared to unfolded linear molecules with free ends (Figure 23a)[161]. Thus, circular molecules display a lower standard deviation (STD) of the extension compared to linear molecules of similar apparent extension. By plotting the STD against molecule extension for each entity, a so-called scatterplot can be generated. The 12-nucleotide (nt) single-strand overhangs of λ-DNA allows for generating stable circular molecules as well as long concatemers, which can be readily distinguished in

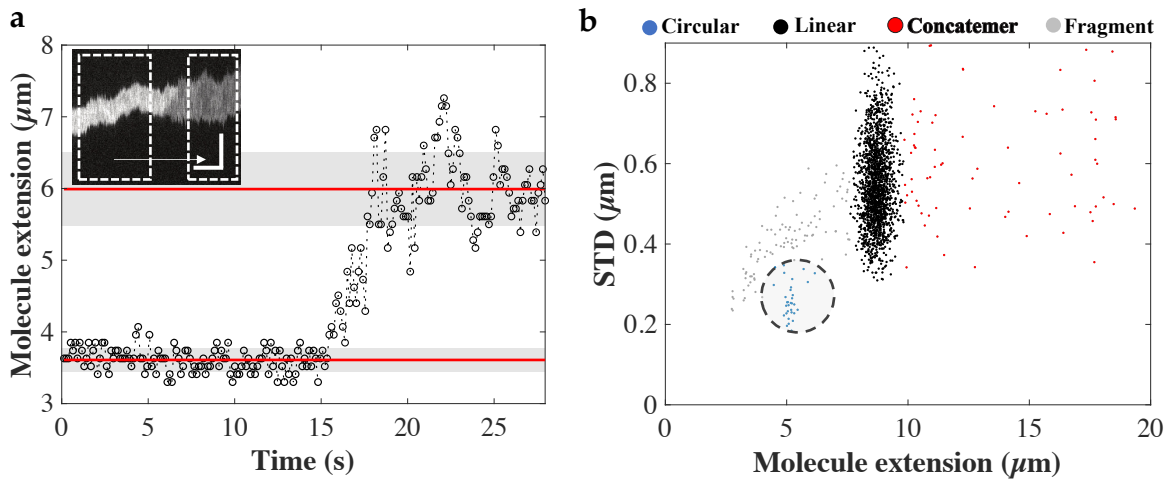

Figure 23: ***(a)*** *Changes in the thermal fluctuations of a circular molecule undergoing linearization upon photoinduced DSB, visualized by the framewise molecule extension obtained from the kymograph (inset). The time elapses in the direction of the arrow. The red lines represent the mean extension for the circular and linear conformations, respectively, and the grey patches reflect the corresponding standard deviations. As the circular molecule is linearized, variations in the framewise molecule extension increase significantly. The vertical and horizontal scale bars correspond to 5 μm and 5 s, respectively.* ***(b)*** *A characteristic scatterplot for λ-DNA, where the standard deviation (STD) has been plotted against the extension for each molecule (N = 1876). The encircled cluster (blue) corresponds to λ-DNA molecules that have circularized by intramolecular annealing of the 12-nt overhangs. All molecules larger than the full-length λ-DNA cluster (black) are considered to be concatemers (red) as a result of intermolecular annealing of the 12-nt overhangs. Fragmented linear molecules (grey) of different extensions give rise to a pattern, reflecting the extension-STD relation for linear molecules.*

---

the characteristic scatterplot of a λ-DNA sample (Figure 23b). Bare λ-DNA typically generates few circles and concatemers due to spontaneous annealing of the complementary ends. The circular fraction of molecules stands out from the linear ones by forming a distinct cluster, deviating from the main trend for the extension-STD relation given by the linear fragments of λ-DNA.

In **Paper III-IV**, different clusters of the scatterplot were quantified to better understand the structural influence of the human CtIP protein and the yeast MRX complex on λ-DNA (further discussed in section 6.4 and 6.5). The general workflow for analyzing a scatterplot starts by hierarchical clustering of the dataset based on the Euclidian distance, to distinguish the different populations of DNA-protein complexes. The circular fraction of molecules is directly deduced from the output of the clustering. However, in contrast to the circular molecules, the more scattered linear fraction of full-length λ-DNA typically results in multiple sub-clusters (Figure 24). In order to account for all linear full-length molecules in the dataset, sub-clusters corresponding to the linear fraction of full-length λ-DNA molecules, are visually picked as representatives for the entire population. Those, that are not specific to only the linear full-length molecules, including either shorter fragments or concatemers, are

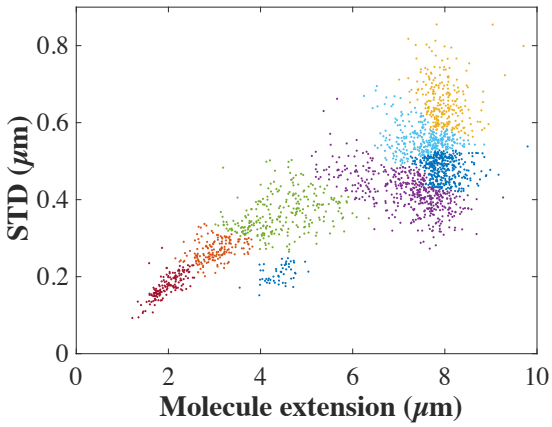

Figure 24: *Clustering of a λ-DNA-CtIP sample. The circular fraction of molecules (blue) at around 4 μm is directly deduced from the output of the clustering. The yellow and navy sub-clusters, belonging to the more scattered linear full-length fraction of molecules, are pooled and fitted to a normal distribution function, from which μ and σ is determined. The purple and light-blue sub-clusters include both linear full-length molecules as well as fragments and are hence considered not representative for the linear full-length population of molecules.*

---

not considered representative sub-clusters for the linear full-length population of molecules. A normal distribution function is fitted to the visually picked representative datapoints, from which the representative mean molecule extension (μ) is determined, together with the corresponding standard deviation (σ). Only molecules satisfying $\mu \pm 2\sigma$ (3σ for **Paper IV**) are considered to belong to the linear fraction of full-length λ-DNA molecules. The mean molecule extension and the corresponding standard deviation is finally calculated for the entire population and used for further analysis.

The development of a method to quantify the relative abundance of sub-populations has enabled the detection of conformational variations of DNA in large datasets. Although circularization of DNA and the formation of concatemers might not always be directly related to any specific biological phenomenon *in vivo*, it can be used to better understand the DNA bridging and end-joining properties of various proteins.

6.2 A Novel Nanofluidic Device for Dynamic DNA-Protein Interactions

The traditional nanofluidic device (Figure 20) serves as an excellent tool for studying qualitative and quantitative properties of pre-formed DNA-protein complexes. However, active flow of liquid during image recording is not possible, as this will cause the confined molecule to move out of the field of view. This implies that the

buffer conditions cannot be manipulated *in situ*, once the DNA-protein complexes are confined in the nanochannels. In previous studies, analytes have been introduced to stretched DNA molecules by diffusion within the nanofluidic system[156, 157]. Although diffusion-based methods allow for dynamic studies of DNA-analyte interactions, repeated alteration of the local environment in the nanofluidic channels cannot be obtained within a reasonable timeframe. In order to overcome this, a novel nanofluidic device, reported in **Paper I**, was designed and fabricated, allowing for active addition of analytes to single stretched DNA molecules. In the new design (Figure 25a-b), 100 x 100 nm^2 (width x depth) channels have been complemented with a so-called reaction

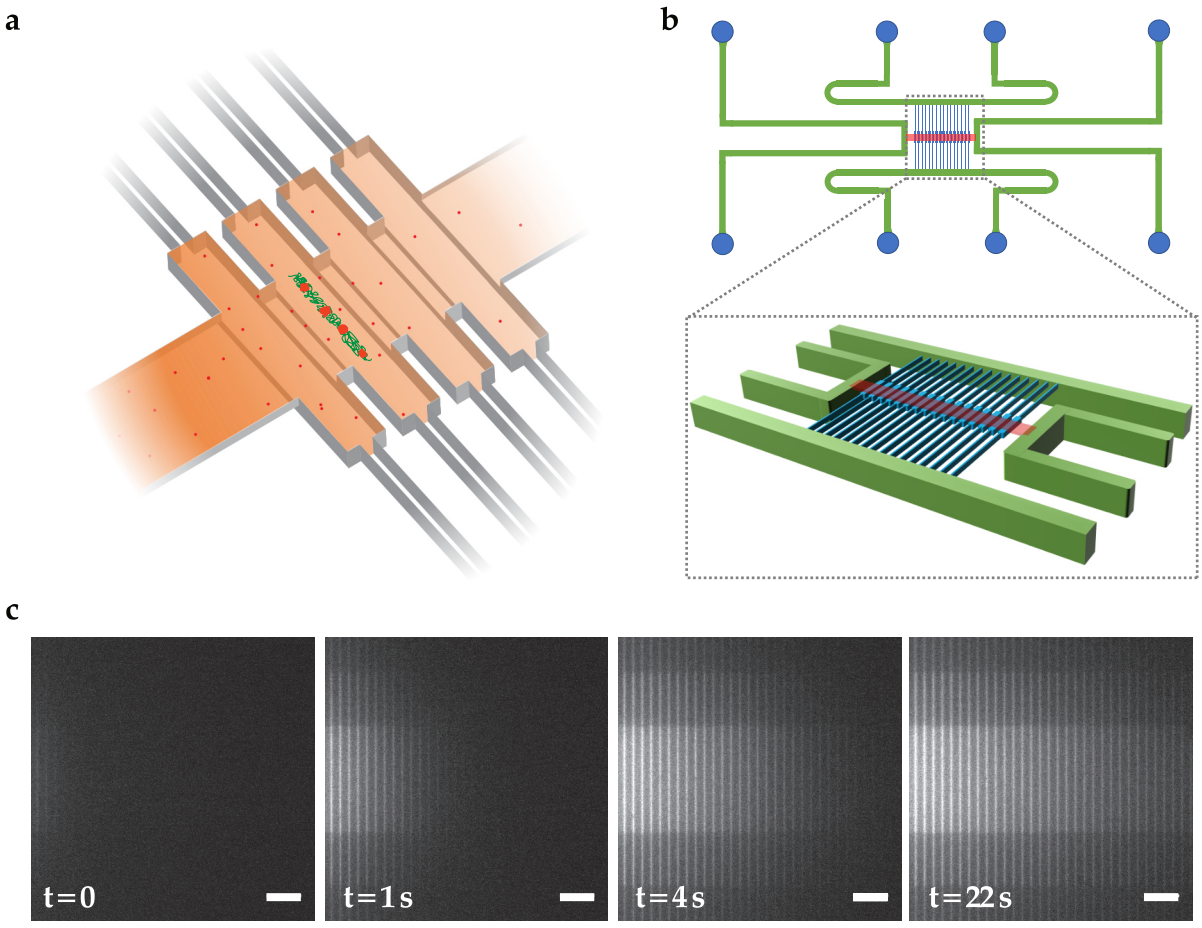

Figure 25: ***(a)*** *Schematic overview of the reaction chamber, wherein a DNA is confined. Analytes are flushed into the stretched DNA molecule through the nanoslit and interactions can be monitored in real time.* ***(b)*** *The dynamic nanofluidic device has eight reservoirs (blue circles), four microchannels (green), a nanoslit (red) and 120 nanochannels (blue).* ***(c)*** *Snapshots at different timepoints during addition of a fluorescein solution through the nanoslit, giving rise to a gradual increase in fluorescence emission with time. The nanochannels, reaction chambers and the nanoslit are clearly distinguishable when filled with fluorescein solution. Scale bars correspond to 10 µm.*

chamber, a 60 μm region in the center of the 500 μm long nanochannel, where the width is increased to 300 nm and the depth to 130 nm. This locally larger chamber serves as an entropic barrier, hindering the stretched DNA molecule from escaping out of the chamber. In addition, a 30 μm x 30 nm (width x depth) nanoslit has been aligned orthogonally across all 120 reaction chambers, allowing for the addition of analytes directly to the stretched DNA molecule. Each end of the nanoslit is connected to two reservoirs, in which analytes of interest can be loaded and flushed into the system (Figure 25b). By applying pressure to both reservoirs on one end of the nanoslit, a fluid of interest can be flushed into the nanochannels in an active manner, changing the local environment within a few seconds (Figure 25c). The shallow nanoslit does not allow the confined DNA molecules to be ejected when flushing across the nanochannels. Thus, the novel device enables real-time, on-demand addition of analytes to single DNA molecules, thereby allowing for biomolecular interaction studies in situ.

A series of proof-of-concept experiments were performed in order to highlight the functionality of the novel device. As an example, DNase I was introduced to the reaction chamber to visualize digestion of stretched DNA in real time. The resulting kymograph (Figure 26a) shows that the λ-DNA molecule is disintegrated within seconds after the endonuclease has generated the first DSB. Since the DNA is stretched solely by confinement, individual cuts can be distinguished as long as the generated fragments are larger than the resolving power of the microscope. Moreover, the cationic (4+) polyamine spermidine was added to a stretched λ-DNA molecule *in situ*

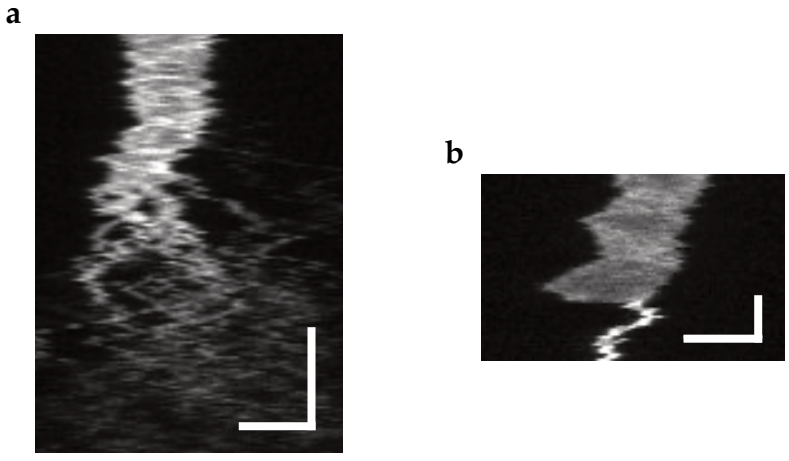

Figure 26: ***(a)*** *Kymograph showing the endonycleolytic digestion of λ-DNA upon introduction of DNase I.* ***(b)*** *Kymograph showing the compaction of a stretched λ-DNA molecule upon addition of spermidine. The compaction appears to start from one of the DNA ends. Vertical and horizontal scale bars correspond to 5 s and 5 μm, respectively.*

(Figure 26b). The corresponding kymograph shows that spermidine immediately compacts the λ-DNA molecule into a dense coil. A closer inspection of the kymograph reveals that the compaction appears to initiate at one of the ends, further highlighting the advantage of nanofluidics in biomolecular interaction studies, where the DNA ends are free and unobstructed. The novel device has subsequently been used to visualize the compaction of nanoconfined DNA by introducing the hepatitis C virus core protein (HCVcp) to extended molecules[162]. The compaction was proven to be reversible as the molecule extension was recovered upon *in situ* digestion of the protein by the addition of Proteinase K.

Sequential Addition of Analytes to a Single DNA Molecule

The nanoslit is connected to a pair of loading reservoirs, one on each side of the 120 nanochannels. This means that two different analytes can be introduced to the same confined DNA molecule in a sequential manner, thereby enabling characterization of multi-component biomolecular interactions. To demonstrate this functionality of the device, an assay was used based on RET between YOYO-1 and EtBr. λ-DNA was pre-stained with YOYO-1 and confined in a reaction chamber, followed by the addition of EtBr through the nanoslit. By simultaneously recording the fluorescence emission from both dyes, RET was observed in real time between YOYO-1 and EtBr (Figure 27). As EtBr intercalated into the sparsely stained λ-DNA (one YOYO-1 molecule per 50 base-pairs), the resulting emission from YOYO-1, upon excitation with blue light of 470/24 nm, was quenched by EtBr, which became excited through RET and subsequently emitted light in the red spectrum. In the next step, SDS was added to the same molecule, which resulted in the dissociation of the less charged EtBr from the DNA. The YOYO-1 emission was almost fully recovered, suggesting that only EtBr has been removed by the SDS. Longer exposures or higher concentration of SDS resulted in the complete dissociation of YOYO-1 from the DNA. The sequential addition of EtBr and SDS was repeated several times and the same response was observed in each cycle. This clearly demonstrates that the novel nanofluidic device can indeed be used for *in situ* dynamic interaction studies including multiple components.

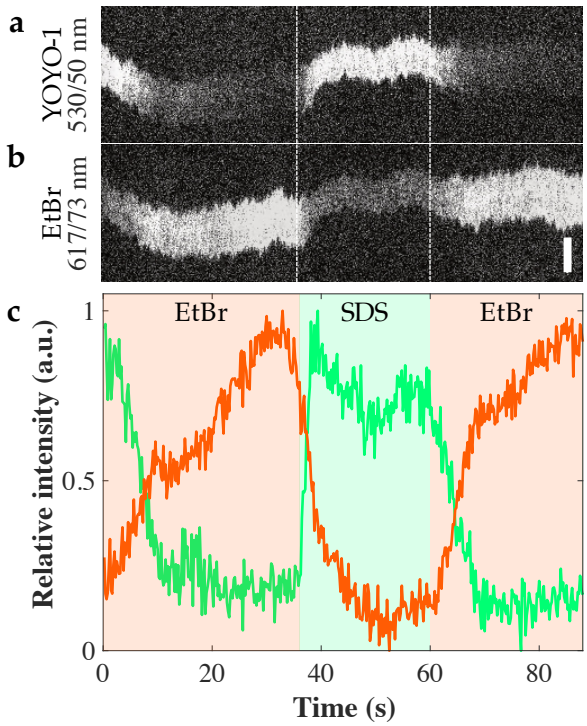

Figure 27: *Sequential addition of EtBr and SDS to a YOYO-1 stained λ-DNA molecule. The kymographs show the fluorescence emission of* ***(a)*** *YOYO-1 and* ***(b)*** *EtBr over time, using only blue light of 470/24 nm for excitation.* ***(c)*** *Relative fluorescence intensities of YOYO-1 (green) and EtBr (orange) with time, respectively. The background color shades represent the timepoints, at which transitions are detected in the emission pattern upon change of analyte between EtBr (orange) and SDS (green). Scale bar corresponds to 5 μm.*

6.3 Dynamics of Ku in Bacterial NHEJ

To better understand the dynamics of the highly minimalistic NHEJ mechanism in bacteria, the *B. subtilis* Ku and LigD components were studied in **Paper II**, using both traditional bulk assays and single-molecule techniques. The main aim was to characterize the DNA-protein and protein-protein interactions that may pose important evolutionarily conserved roles in DSB repair across the different domains of life.

Ku and LigD were expressed in *Escherichia coli* and purified for *in vitro* characterization. Nickel affinity chromatography was employed for both proteins in the initial stage of the purification. The eluted homodimeric Ku was further purified by size-exclusion chromatography at 1 M salt concentration to reduce potential interactions between the protein and DNA-fragments in the solution. The final eluate was stored at -20°C for use in all subsequent experiments. The same procedure was repeated for the purification of Ku_{core}, which is a truncated version of the wild-type Ku

(Ku_{wt}), where 58 amino acids have been deleted from the C-terminal. Following the nickel affinity purification of LigD, the protein was loaded on a cation exchange chromatography column and eluted in a gradient of increasing salt concentration. LigD was finally passed through a size-exclusion column and stored at -20°C for further use in subsequent experiments.

Ku Binds to DNA Ends in a Co-Operative Manner

The Ku homodimer is the first component to identify and bind to the free DNA ends post a DSB. To further characterize this DNA-protein interaction, EMSA was performed, where Ku was titrated to a 50 bp DNA substrate. A characteristic ladder was observed in the resulting gel (Figure 28a), indicating that the protein units interact with DNA in a highly stable manner. The presence of distinct and unsmeared bands suggests that the dissociation rate constant (k_{off}) is significantly lower than that of the association (k_{on}). It is also evident from the gel that the 50 bp DNA substrate can harbor up to three Ku_{wt} homodimers. At elevated concentrations, the number of DNA molecules with one Ku_{wt} homodimer bound appears to reduce while complexes with two and three homodimers become more prominent despite the presence of free substrate DNA. This suggests that the DNA-protein interaction is co-operative, and the binding of Ku_{wt} is favored by proteins already located on the DNA. For Ku_{core}, up to four homodimers were observed to co-operatively interact with the 50 bp DNA substrate. No interaction with DNA was observed for LigD alone under the same experimental conditions.

SPR was further performed to obtain DNA-binding parameters for the Ku derivatives. Biotinylated 50 bp DNA, with the same sequence as for the substrate used in the EMSA, was deposited on a streptavidin coated chip followed by injection of Ku_{wt} and Ku_{core} at various concentrations. The association and dissociation of protein was recorded for 180 seconds and 500 seconds, respectively. The resulting sensograms (Figure 28b) were fitted to a standard kinetic model for interactions between an analyte (Ku) and a heterogeneous ligand (DNA). In this particular model, it was assumed that a total of two Ku homodimers bind to the substrate DNA, and hence, two sets of rate and equilibrium constants were obtained. It appeared that the equilibrium dissociation constant for the interaction between the first protein unit and the substrate DNA ($K_{D,1}$) is similar for both Ku derivatives, while $K_{D,2}$ is 30-fold stronger for Ku_{wt} (Table 1). This significant difference derives mainly from the $k_{off,2}$ values, which suggests an almost 20-fold lower dissociation rate for Ku_{wt} compared to Ku_{core}. The deconvolved curves, representing the binding of each protein unit to the DNA substrate (Figure 28c),

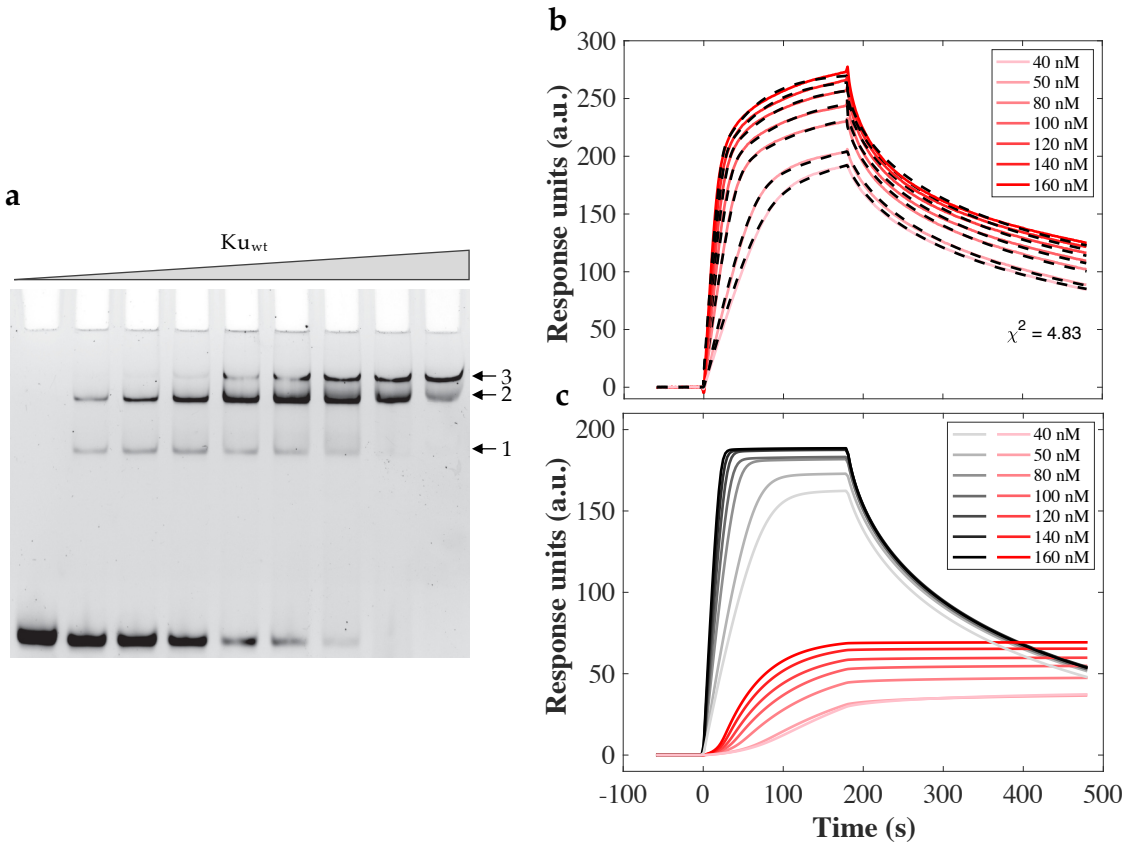

Figure 28: ***(a)*** *EMSA of a 50 bp blunt-ended DNA substrate at increasing concentration of Ku$_{wt}$ (0-0.4 μM, 50 nM increments). Arrows indicate the number of Ku$_{wt}$ homodimers bound to the DNA substrate. The leftmost lane corresponds to the DNA control.* ***(b)*** *SPR sensograms (solid red line) showing the binding of Ku$_{wt}$ to a 50 bp blunt-ended DNA substrate at various protein concentrations, together with the corresponding mathematical fit for interaction between an analyte and a heterogeneous ligand (dashed black line). The χ^2 of the fit was calculated to 4.83.* ***(c)*** *Component analysis of the SPR sensograms, where the binding pattern of the first (black) and second (red) homodimer is revealed at the different protein concentrations.*

Table 1: *Kinetic parameters obtained from the SPR analysis of Ku$_{wt}$ and Ku$_{core}$ using an immobilized 50 bp DNA substrate.*

	Kinetic parameters					
	$k_{on,1}$ (10^6 /Ms)	$k_{off,1}$ (10^{-2} /s)	$K_{D,1}$ (nM)	$k_{on,2}$ (10^5 /Ms)	$k_{off,2}$ (10^{-5} /s)	$K_{D,2}$ (pM)
Ku$_{wt}$	4.07	2.7	6.66	1.38	0.13	9.41
Ku$_{core}$	3.17	2.9	9.27	0.74	2.42	330

display a typical 1:1 binding behavior for the first binding Ku homodimer and a sigmoidal binding pattern for the second homodimer. This suggests co-operative binding, in agreement with the EMSA results. The results further indicate that the C-terminal protrusions are dispensable for stabilization of the first Ku_{wt} homodimer onto the DNA. They however seem to be involved in the binding of the second protein unit to the DNA substrate.

Ku Stabilizes Annealing of DNA with Single-Strand Overhangs

Using the fluorescently labelled Ku_{AF647} in the single-molecule nanofluidics setup (further discussed in section 6.1), the protein was found to be localized at the ends and junctions of annealed λ-DNA with 12-nt complementary ends (Figure 22b). This proves that Ku has an end-specific DNA-binding property, where it is most probably threaded on the ends, promoting or stabilizing the annealing of the complementary single strands. Noteworthy is that the 12-nt overhangs can form a stable duplex in solution, without the need of any stabilizing entity (Figure 23b). Therefore, to better understand the effect of Ku on the synaptic stability, a 4-nt sticky-ended DNA substrate was combined with Ku_{wt}. The resulting size distribution of the nanoconfined and stretched DNA-protein complexes verifies that Ku can also stabilize annealed short complementary sticky ends, which are otherwise unable to stay hybridized in solution (Figure 29). Complementing magnetic tweezers-based single-molecule

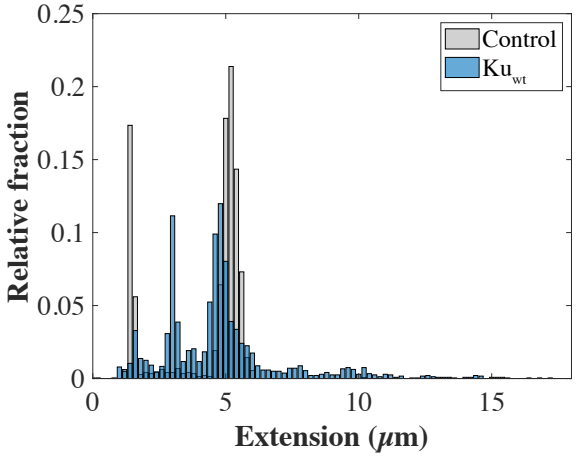

Figure 29: *Size histogram of PciI-digested λ-DNA with 4-nt sticky overhangs in the absence of protein (grey) and in the presence of Ku_{wt} (blue). The peak at ~5 μm corresponds to full-length substrate DNA molecules with 4-nt complementary overhangs. Residual DNA fragments from the digestion reaction appear as a peak at ~1 μm. The peak at ~2.5 μm corresponds to circularized full-length substrate DNA molecules. The presence of circles and concatemers shows that Ku_{wt} stabilizes the annealing of 4-nt complementary overhangs.*

forceps experiments (further discussed in section 5.2) were performed by the group of Terence Strick´s at IBENS in Paris, France, for quantitative assessment of the synaptic stability. In the absence of protein, no synapses were generated for 4-nt sticky ends, nor for blunt DNA ends. However, when adding Ku_{wt} to the sticky-ended DNA construct, stable annealing was observed for the complementary overhangs. The generated complex could not be torn apart at 1.4 pN stretching force. A similar behavior was observed for the truncated Ku_{core}, suggesting that the core structure stabilizes the annealing of ssDNA, which is in agreement with results from the nanofluidics experiments. For blunt-ended DNA, short-lived synapses with a lifetime of 1.6 ± 0.3s for Ku_{wt} were observed. The corresponding synaptic lifetime for Ku_{core} was found to be 2.1 ± 0.2s.

LigD Stabilizes DNA End-Joining in the Presence of Ku

Although Ku alone can generate short- and long-lived DNA synapses, LigD is required to covalently join the free ends, and thereby complete the NHEJ process. A previous study has demonstrated that the C-terminal protrusions of Ku (Figure 12b) are important for stimulating the LigD activity[111]. EMSA was performed to further characterize how Ku and LigD interact at the site of repair (Figure 30). Ku_{wt} or Ku_{core} were combined with LigD and a 50 bp DNA substrate, before being separated on

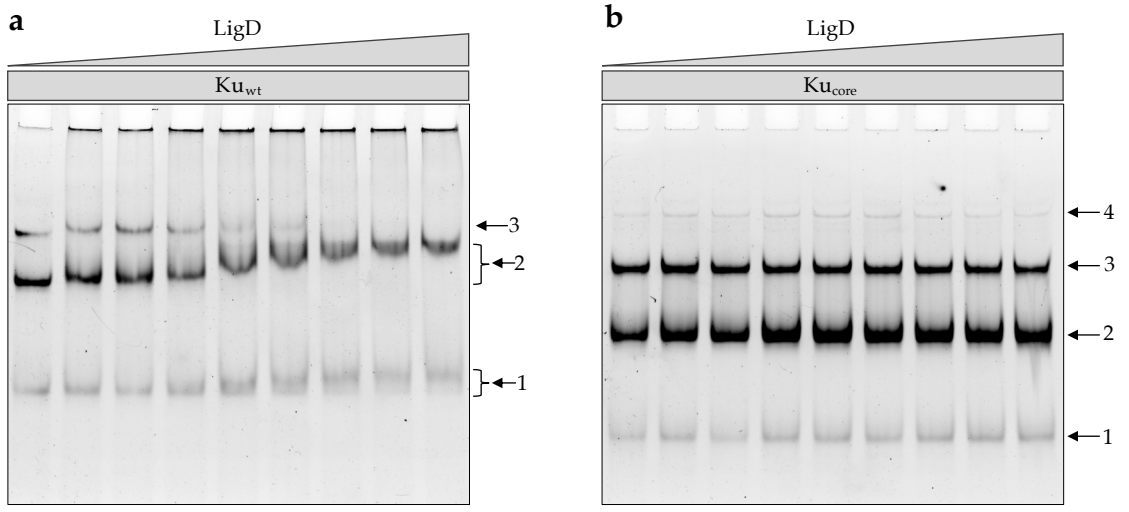

Figure 30: *EMSA gel showing the mobility of a 50 bp blunt-ended DNA substrate in the presence of* ***(a)*** *Ku_{wt} at a constant concentration of 0.2 μM and LigD at increasing concentration (0-0.8 μM, 0.1 μM increments). Arrows indicate the number of Ku_{wt} homodimers loaded on the DNA substrate. A substantial super-shift is observed for the DNA, which has two Ku_{wt} homodimers attached.* ***(b)*** *As for* ***(a)*** *but for Ku_{core}. No super-shift is observed for the DNA-Ku_{core} complexes in the presence of LigD.*

native polyacrylamide gels. Ku_{wt} and LigD display a prominent super-shift when bound to the DNA substrate (Figure 30a). Noteworthy is that this unsmeared super-shift is more obvious for the DNA that has two Ku_{wt} homodimers bound, while the effect is smaller for the DNA with only one homodimer. No super-shift was observed for the DNA-Ku_{core} complexes (Figure 30b). The prominent super-shift of the two Ku_{wt}-DNA complex (Figure 30a) suggests that LigD may need two C-terminal protrusions to stably "dock" onto the DNA. A following covariation analysis between the Ku_{wt} and LigD sequences, performed by Raphael Guerois at I2BC in Paris, France, further showed that there is a high probability that the Ku_{wt} C-terminal protrusion can interact with two particular sites on the polymerase domain of LigD. This could possibly explain why the two Ku_{wt}-DNA complex displays a more prominent super-shift in the EMSA analysis. Single-molecule forceps experiments revealed consistent data, where the synaptic stability of blunt-ended DNA was significantly increased, displaying a lifetime of 18.5 ± 1.2s in the presence of Ku_{wt} and LigD. The corresponding lifetime for Ku_{core} was 5.1 ± 0.4s.

Ku Remains Stably Bound to DNA Post Ligation

The bacterial Ku homodimer is suggested to be threaded onto DNA ends, similar to the human Ku70/80 complex. *In vitro* nanofluidic experiments with fluorescently labelled Ku_{AF647} demonstrated how the homodimer stays attached to the ends and junctions post annealing of λ-DNA (Figure 22b), suggesting that the protein is topologically trapped. This further raises questions on the fate of Ku, post ligation *in vivo*. Previous studies have reported that there may be specific mechanisms for the removal of human Ku70/80 from DNA upon completion of NHEJ[108-110]. To better understand the fate of *B. subtilis* Ku post ligation, the fluorescently labelled Ku_{AF647} was combined with LigD to ligate 10 000 bp blunt-ended DNA fragments, followed by studying the generated complexes using nanofluidics. Dual-color imaging revealed that the Ku_{AF647} homodimers is topologically attached to the concatemerized DNA-protein complexes (Figure 31a). The distance between each dot is close to constant, with a mean value of 1.56 ± 0.20 μm, corresponding to an extension of 46% with respect to the contour length of a 10 000 bp DNA fragment. To confirm that the DNA had been covalently attached, the dynamic nanofluidic device was employed. Proteinase K was introduced to DNA-protein complexes confined in the reaction chamber and the dual-color response was recorded in real time. Upon arrival of Proteinase K, the fluorescence emission from Ku_{AF647} was depleted within seconds of exposure, whereas the YOYO-1 emission remained constant (Figure 31b). 36 ligated DNA-protein complexes were exposed to Proteinase K, out of which most complexes were fully

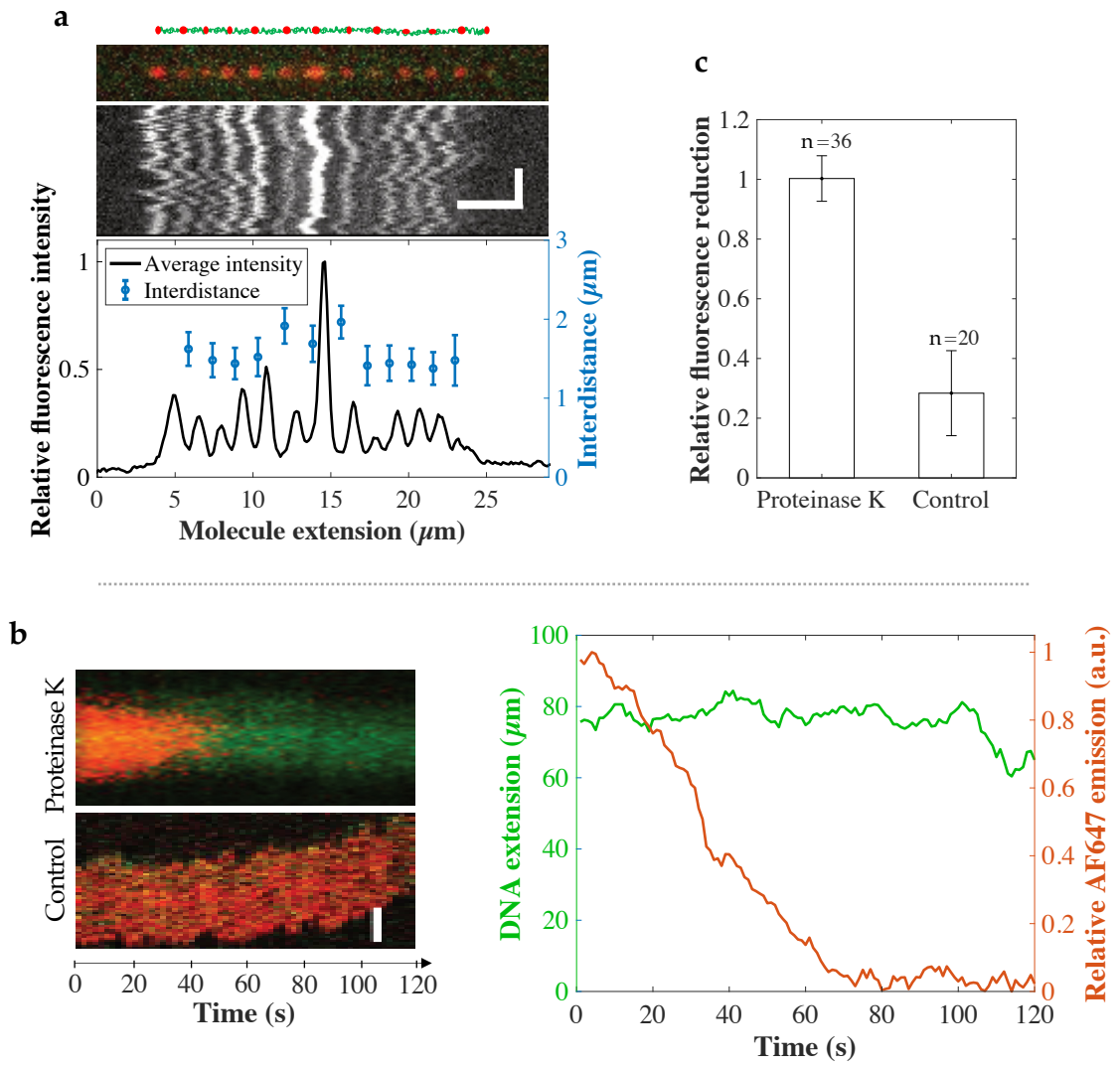

Figure 31: ***(a)*** *Dual-channel fluorescence image of a concatemer of 10 000 bp blunt-ended DNA fragments (green), ligated in the presence of* Ku_{AF647} *(red) and LigD. The corresponding kymograph shows the immobilized* Ku_{AF647} *clusters along the DNA extension with time. The relative fluorescence intensity of* Ku_{AF647} *reveals a periodic appearance along the DNA extension, with a constant distance between the clusters. Vertical and horizontal scale bars correspond to 2 s and 5 μm, respectively.* ***(b)*** *Addition of Proteinase K through the nanoslit to a confined ligation product (top kymograph). Proteinase K digests* Ku_{AF647} *(red), resulting in fluorescence emission depletion, leaving the ligated DNA (green) intact. The quantified extension of the DNA appears constant with time, while the AF647 emission is depleted within seconds after Proteinase K is introduced. The minimal level of photobleaching is displayed by the control complex (bottom kymograph). Scale bar corresponds to 5 μm.* ***(c)*** *The AF647 emission of most complexes, exposed to Proteinase K, is depleted within two minutes, while photobleaching of control complexes reveal a ~28% emission depletion within the same timeframe.*

depleted in AF647 emission within two minutes. The emission depletion for 20 control complexes was around 28% (Figure 31c). This emission decrease was attributed to photobleaching of the AF647 dye. The results show that the bacterial Ku homodimer stays topologically trapped on the DNA molecule post ligation, suggesting that there might be additional mechanisms activated *in vivo* for removal of Ku from the DNA post ligation, similar to the proposed human systems.

6.4 CtIP Mediated DNA Bridging in the Initial Stages of Human HR

The homotetrameric CtIP protein complex has been established as a key component in the initial stages of HR, where the DNA ends are resected to generate 3′ single-strand overhangs[60]. In **Paper III,** a previously unknown function of CtIP was reported, demonstrating that the protein can act as a linker to bridge distant segments of single DNA molecules. A novel nanofluidic approach was developed (further discussed in section 6.1), using λ-DNA as a substrate for probing intra- and intermolecular annealing in the presence of different CtIP derivatives.

The Tetrameric Structure of CtIP is Crucial for Intramolecular Annealing of λ-DNA

The 12-nt single-strand overhangs of λ-DNA allow for limited inter- and intra-molecular annealing in the absence of proteins (Figure 23b). Upon mixing wild type CtIP ($CtIP_{wt}$) with λ-DNA, a major difference was observed in the resulting scatterplot (Figure 32a). The relative number of circular DNA-protein complexes increased significantly in the presence of the protein, compared to bare λ-DNA, suggesting that $CtIP_{wt}$ has promoted intramolecular annealing of the complementary DNA ends. The fraction of circular complexes was found to be ~71% (Figure 32b) using the established clustering approach (further discussed in section 6.1). The relative number of concatemers did not increase significantly, suggesting that $CtIP_{wt}$ exhibits a specific

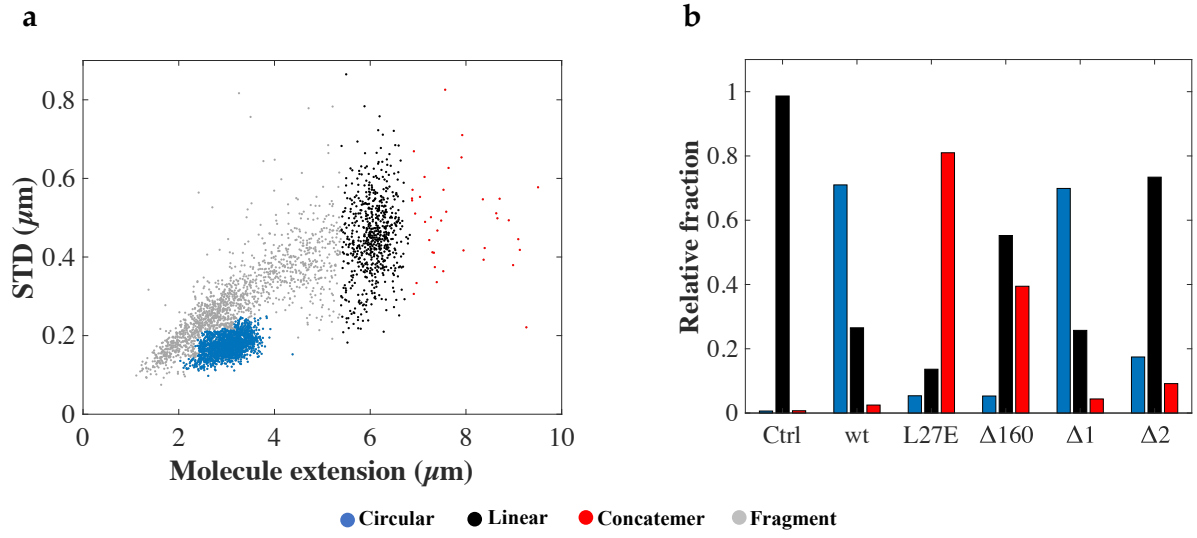

Figure 32: ***(a)*** *Scatterplot of λ-DNA in the presence of $CtIP_{wt}$ at a concentration ratio equivalent to 1000 tetramers per DNA molecule (N = 3886).* ***(b)*** *Relative fractions of circular, full-length linear and concatemerized DNA-protein complexes at a concentration equivalent to 1000 tetramers / 2000 dimers / 4000 monomers per λ-DNA molecule for different derivatives of CtIP.*

function, which promotes intramolecular annealing of the λ-DNA ends. To further investigate how this outcome is affected by variations on the protein quaternary structure, different CtIP derivatives were characterized by their abilities to promote circularization of λ-DNA and formation of concatemers. The tetramerization of $CtIP_{wt}$ is mediated by an N-terminal domain, which enables dimer-dimer interactions to generate a dumbbell-shaped three-dimensional protein structure with two globular parts, held together by a linker[70]. Introducing the L27E point-mutation will block this interaction, generating a dimeric CtIP derivative ($CtIP_{L27E}$). By completely removing 160 residues from the N-terminal, the quaternary interactions are further disrupted, resulting in a monomeric protein ($CtIP_{\Delta160}$). The circularization was heavily reduced in the presence of $CtIP_{L27E}$ and $CtIP_{\Delta160}$, suggesting that the dumbbell shape is important for this particular effect (Figure 32b). Two additional derivatives of CtIP were also characterized, $CtIP_{\Delta1}$ and $CtIP_{\Delta2}$. The first lacks residues 350-600 while the latter lacks residues 165-790. However, both derivatives retain their tetrameric structures. $CtIP_{\Delta1}$ was found to be equally efficient as $CtIP_{wt}$ in generating circularized λ-DNA molecules, while $CtIP_{\Delta2}$ generated much less circles (~17%). Although not as effective in circularizing λ-DNA as other tetrameric derivatives, $CtIP_{\Delta2}$ was more efficient in promoting intramolecular annealing of DNA compared to the dimeric and monomeric derivatives (Figure 32b). It is hypothesized that the two globular DNA-binding units of the CtIP dumbbell can interact with distant regions on DNA. By creating a DNA loop the ssDNA ends are allowed to come close, and intramolecular annealing is thereby promoted. The decrease in circularization efficiency of $CtIP_{\Delta2}$ could potentially be explained by the reduced size of the protein tetramer. Formation of a DNA loop by $CtIP_{\Delta2}$ will theoretically require the negatively charged backbone of the DNA molecule to come closer than for $CtIP_{wt}$.

CtIP Possesses a DNA Bridging Ability

Under the current experimental conditions, circular λ-DNA molecules displayed a mean extension of 4.7 μm in the absence of protein, while the corresponding extension of linear full-length molecules was 8.8 μm. However, when adding $CtIP_{wt}$ the mean extension reduced by approximately 36% and 31% for the circular and linear molecules, respectively (Figure 33a). A 97 kbp circular plasmid was used to further see how the absence of free DNA ends affects the protein-mediated compaction. The extension of the plasmid reduced by 17% as a result of global protein-DNA backbone interactions. This suggests that the compaction of λ-DNA is partially attributable to this particular mode of interaction. The significantly higher reduction in extension of linear and circularized λ-DNA upon addition of $CtIP_{wt}$ indicate that there are

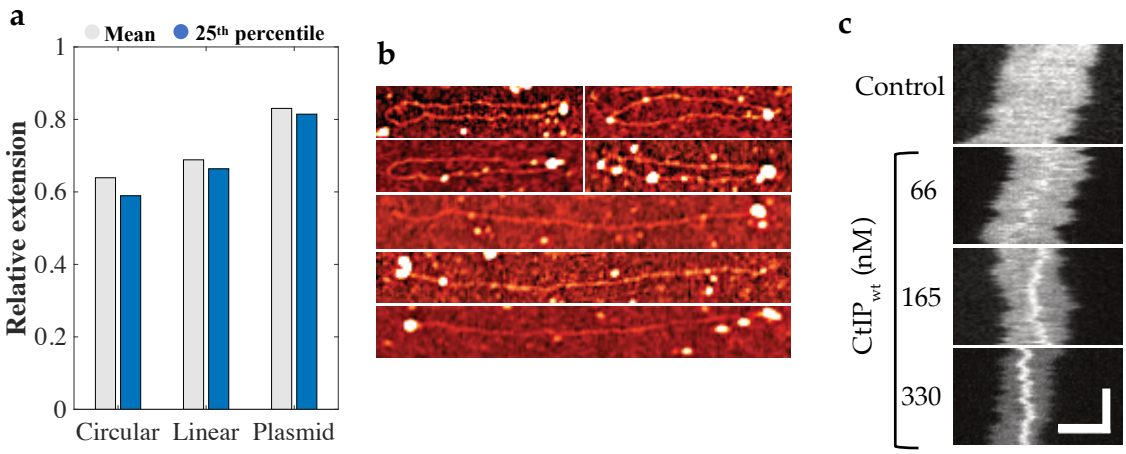

Figure 33: ***(a)*** *Mean extension of all detected DNA-CtIP$_{wt}$ complexes (grey) and complexes in the 25th percentile (blue), relative to their corresponding controls (no protein added) at a protein concentration equivalent to 1000 tetramers per λ-DNA molecule.* ***(b)*** *AFM images showing circular and linear DNA-CtIP$_{wt}$ complexes, where features are reflected through the height difference. CtIP$_{wt}$ forms clusters along the contour of both circular and linear DNA molecules. Protein clusters are also observed at ends of the linear molecules.* ***(c)*** *Kymographs of circular DNA-CtIP$_{wt}$ complexes at varying protein concentrations, equivalent to 200, 500 and 1000 CtIP$_{wt}$ tetramers per DNA molecule. The size of the condensed region appears to increase with higher protein concentration. Vertical and horizontal scale bars correspond to 3 s and 3 μm, respectively.*

---

potentially also DNA end-specific interactions in play. This was further confirmed by AFM imaging of DNA-CtIP$_{wt}$ complexes, which showed both end-specific interactions and general binding of the protein along the DNA contour (Figure 33b). From fluorescence based nanofluidics experiments, kymographs reveal that the compaction of circularized λ-DNA molecules can partially be attributed to local condensates along the extension of the complexes (Figure 33c). The bright dot along the stretching axis indicates that the DNA has locally condensed as a result of protein binding. Given that the tetrameric CtIP derivatives are efficient in bridging DNA, observed from the circularization of λ-DNA, the generation of this condensate indicates probable protein-mediated DNA looping. The size of the local condensate appears to be dose dependent, suggesting that more protein is accumulating on the DNA at higher protein concentrations. Higher-order complexes were also detected, such as for instance circular DNA with two mobile dots (Figure 34a), as well as complexes of two λ-DNA molecules, concatemerized and subsequently circularized, harboring up to two dots along the extension (Figure 34b). Complexes were also detected, where two circularized λ-DNA molecules are joined by a central static dot (Figure 34c), suggesting that proteins that have accumulated on DNA can mediate interactions with other DNA molecules. The results demonstrate that CtIP$_{wt}$ possesses a DNA bridging ability, allowing the broken DNA ends to be kept in close proximity post a DSB, which may be important in the initial stages of HR.

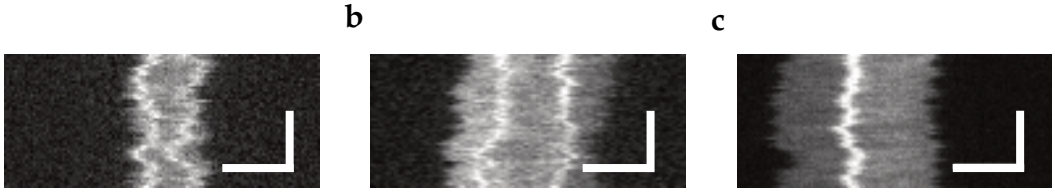

Figure 34: ***(a)*** *Kymographs of a single circularized λ-DNA-CtIP$_{wt}$ complex with two mobile local condensates along the extension.* ***(b)*** *A higher-order complex of two concatemerized λ-DNA molecules, subsequently circularized in the presence of CtIP$_{wt}$, harboring two mobile local condensates.* ***(c)*** *Two circularized λ-DNA-CtIP$_{wt}$ complexes joined through a central, static condensate. Vertical and horizontal scale bars correspond to 3 s and 3 μm, respectively.*

Unfolding Circular Complexes Reveal Dynamic Features of the DNA-CtIP Interaction

Stretched circular DNA molecules will unfold as a result of a single DSB, in order to attain the thermodynamically more favorable linear conformation (Figure 23a)[161, 163]. To further study the characteristics of the DNA-CtIP interaction, locally condensed circularized DNA-protein complexes were subjected to extensive illumination to generate a single photoinduced DSB *in situ*, while recording the subsequent unfolding event. The resulting kymographs reveal that the local condensate does not prevent the DNA molecule from unfolding (Figure 35a-b). Furthermore, the bright feature seems to disappear as the unfolding end reaches the locally condensed region and the molecule is relaxed to its linear conformation, displaying a homogeneous fluorescence emission along the extension. This phenomenon can be explained as such that the protein is dynamically associated with the DNA and potentially falls off as the unfolding end reaches the condensed region. The same procedure was repeated for higher-order complexes of two single circular λ-DNA molecules, joined through a static central dot (Figure 35c-d). Upon introduction of a DSB, only the broken molecule appears to unfold into the linear conformation. As the unfolding proceeds, the linearized molecule slides out of the junction and detaches from the intact circular molecule, suggesting that the bridging is reversible. In conjunction to the detachment, the accumulated proteins seem to fall off the DNA since the bright feature disappears. These results substantiate the DNA bridging function of CtIP, further demonstrating the dynamics behind the DNA-protein interactions. Given that CtIP bridges DNA during HR, the ability of the proteins to slide along the DNA could potentially pose a critical function during DSB repair. By providing flexibility, sliding of the proteins along DNA will allow for the ends to be dynamic and mobile, while firmly held in position at the site of repair.

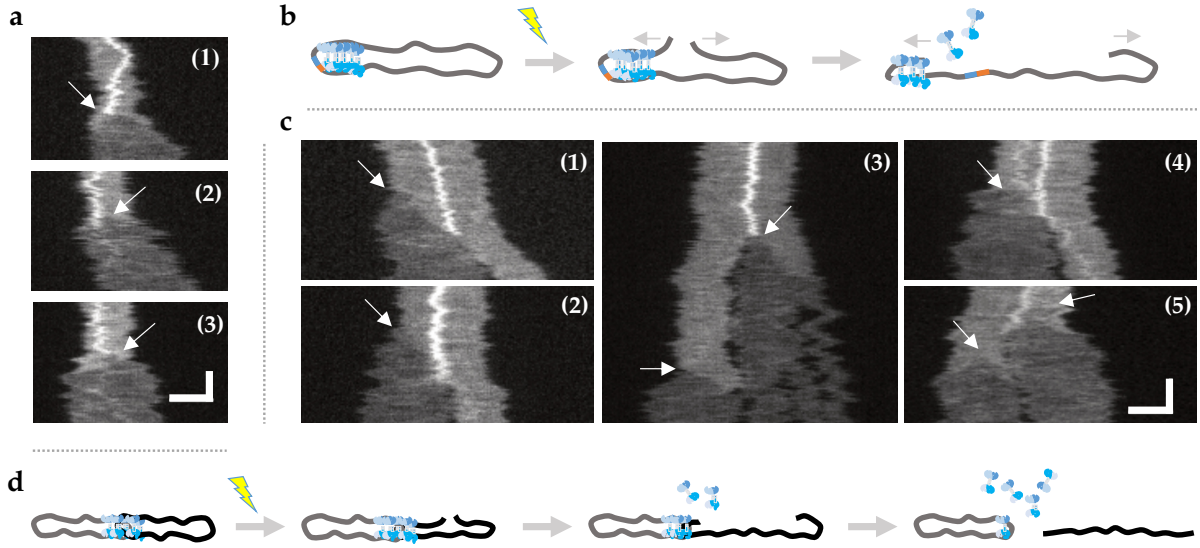

Figure 35: ***(a)*** *Kymographs displaying the unfolding of locally condensed and circularized λ-DNA-CtIP$_{wt}$ complexes upon photoinduced DSB. Arrows indicate initiation of the unfolding events.* ***(b)*** *Schematical illustration of the unfolding event in* ***(a)*** *demonstrating how the protein tetramers potentially fall of the DNA upon breaking of the molecule.* ***(c)*** *Kymographs displaying the unfolding of higher-order complexes, where two circularized λ-DNA molecules are joined through a central static CtIP$_{wt}$ cluster. As the broken DNA end unfolds, the local condensate disappears once the end has reached the center and the linear molecule is subsequently separated from the intact circular molecule.* ***(d)*** *Schematic illustration of the unfolding event in* ***(c)*** *demonstrating how the circular DNA complexes detach from each other upon breaking and unfolding of one molecule. Vertical and horizontal scale bars correspond to 5 s and 3 μm, respectively.*

6.5 The End-Joining Activity of MRX in Yeast HR

The main role of the MRX complex in yeast is to initiate and execute end resection together with Sae2, which is homologous to the human CtIP protein[164, 165]. Although this feature of MRX has been well characterized, it is still not fully clear whether or not it possesses any specific DNA end-joining properties. In **Paper IV**, the previously developed nanofluidics assay (further discussed in section 6.1 and 6.4) was used to study the end-joining activity of MRX and its MR and Xrs2 components *in vitro*.

Xrs2 is Indispensable for the DNA End-Joining Activity of MRX

The MR and Xrs2 components of the MRX complex were individually characterized with respects to their abilities to promote annealing of the 12-nt complementary overhangs of λ-DNA. It is evident from the scatterplot that the full MRX complex promotes annealing of the complementary DNA ends (Figure 36a). The fraction of circles and concatemers, generated in the presence of MRX, was found to be 24% for both DNA conformations at a concentration equivalent to one protein complex per 57

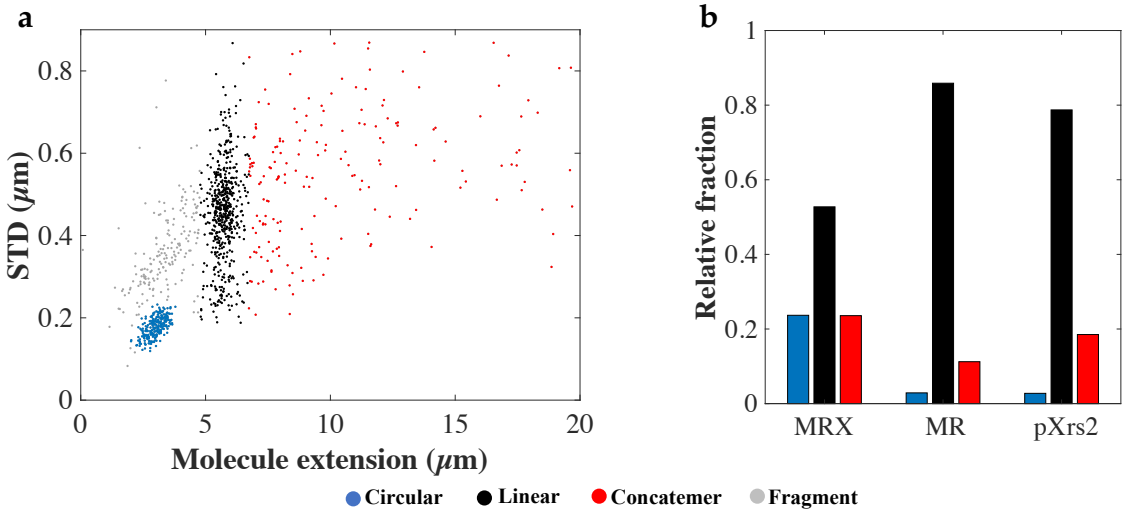

Figure 36: ***(a)*** *Scatterplot for λ-DNA in the presence of MRX at a concentration equivalent to one protein complex per 57 bp DNA (N = 1228).* ***(b)*** *Relative fractions of circular, full-length linear and concatemerized DNA-protein complexes at a concentration equivalent to one protein per 57 bp DNA (MRX) and 50 bp DNA (MR and phosphorylated Xrs2).*

bp DNA. This is in stark contrast to CtIP, which generated a significantly larger fraction of circles, while the number of concatemers was almost negligible. The circularization of λ-DNA by CtIP was attributed to a DNA-bridging ability, which appears to be heavily reduced for MRX. Quantification of the different DNA conformations in the presence of MR complex revealed a significant decrease in the number of annealed molecules, suggesting that the Xrs2 component is indispensable for joining of the complementary ends (Figure 36b). This is, however, in opposition to previous data on the endonucleolytic activity, where MRX and MR are essentially indistinguishable[82]. An end-annealing activity, similar to that of MR, was observed for the Xrs2 component alone, although, there appears to be a slightly higher fraction of concatemers at the same DNA-protein ratio (Figure 36b). The higher fraction of concatemers suggests that Xrs2 alone potentially possesses an end-joining activity, possibly through direct interactions with the free DNA ends. This feature is compromised for MR. Although a slightly lower fraction of concatemers was observed for Xrs2 compared to MRX, it is likely that Xrs2 plays an important role in directing the full MRX complex to the ends to initiate the resection by direct end-interactions. This is consistent with previous biochemical evidence on the end-binding activity of Xrs2[77]. No apparent DNA bridging activity was observed for the MRX complex, nor for the MR and Xrs2 components. However, these preliminary results highlight a potentially critical role for Xrs2 in directing the MRX complex to the site of repair to initiate processing of the broken DNA ends.

The DNA End-Joining Activity of MRX is Not Enhanced by Sae2

To investigate whether the Sae2-MRX interaction has any implications on the DNA end-joining activity, λ-DNA was mixed with both proteins simultaneously and the interactions were characterized by nanofluidics. Quantification of the circular and concatemerized DNA-protein complexes for MRX and Sae2, separately and in combination, showed no additive or synergistic effects (Figure 37). Sae2 at a concentration equivalent to one protein per 10 bp DNA, generated 42% concatemers, which is in agreement with previous nanofluidics experiments[71]. Upon combination of Sae2 and MRX at a concentration equivalent to one Sae2 per 10 bp DNA and one MRX complex per 100 bp DNA, the end-joining activity was significantly reduced, yielding only 20% concatemers. This may indicate that the end-binding ability of the Xrs2 component in the MRX complex, or the Sae2 tetramer is compromised when both proteins are mixed. Concludingly, these preliminary results could potentially hint on the sequential order by which the different protein complexes interact with DNA during the initial stages of HR.

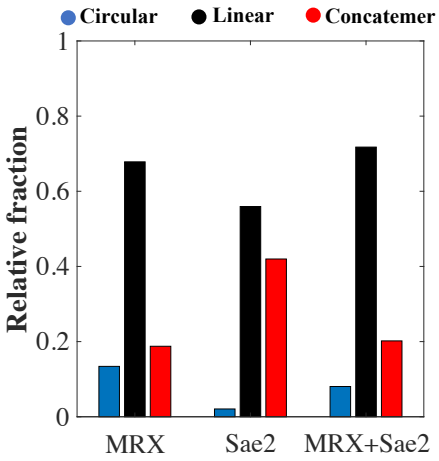

Figure 37: *Relative fractions of circular, full-length linear and concatemerized DNA-protein complexes at a concentration equivalent to one protein per 100 bp DNA (MRX), 10 bp DNA (Sae2) and the combination of MRX and Sae2 at the respective DNA: protein ratios.*

7 Concluding Remarks – *Outlook*

It is of paramount importance to map the numerous complex molecular interactions that mediate DSB repair *in vivo*, to obtain better understanding on the development of for instance cancer, and how such disorders can be treated. The swiftly emerging field of single-molecule techniques has allowed for detailed characterization of features that would typically be omitted in traditional bulk assays. The aim of the original work, upon which this Thesis is based, is to highlight the recent developments within the field of single-molecule nanofluidics, and further demonstrate how this particular technique can be employed to study DSB repair *in vitro*. In **Paper I**, a novel nanofluidics design was presented that allows for dynamic control of the local environment within the nanofluidic channels *in situ*. The new generation nanofluidic device was successfully used to manipulate confined and stretched DNA molecules on-demand and in real time, demonstrated by a series of proof-of-concept experiments. The current design allows for up to two analytes to be sequentially added

to confined DNA molecules. By adding more loading reservoirs and reorganizing the microfluidic channels accordingly, more analytes may be loaded on the chip, which will consequently allow for mapping of more complex biochemical systems, involving multiple interacting components.

DSB repair by NHEJ in bacteria was studied in **Paper II**, using single-molecule analysis in combination with traditional bulk assays. Mechanistic features of the homodimeric Ku complex were investigated, primarily with regards to the formation of DNA end-synapses. The results demonstrate the importance of the conserved core structure of the Ku homodimer for stabilizing annealing of single-strand overhangs. Furthermore, it was shown that the Ku C-terminal is necessary for the recruitment of LigD to the site of repair, which is mediated through interactions with the LigD polymerase domain. To further understand the dynamics of NHEJ and how the process has evolved among the different domains of life, thorough studies on the multi-functional LigD is essential and may explain potential reasons for the major difference in the number of NHEJ components required for DSB repair in bacteria and humans.

In **Paper III-IV**, protein complexes involved in the initial stages of human and yeast HR were studied to gain insight on their individual end-joining activities. An assay, based on nanofluidics, was developed and used in **Paper III** to characterize the DNA bridging ability of the human tetrameric CtIP complex. The assay allows for distinguishing concatemerized and circularized DNA molecules, resulting from protein-mediated end-annealing. In this particular study, the circularization of DNA was attributed to a previously unknown, structure dependent bridging feature of the CtIP. In **Paper IV**, the yeast MRX complex was characterized by the same means to show the importance of the Xrs2 component for DNA end-joining activity. To gain better insights on the pre-synaptic events in HR, further studies on the joint activity of the human MRN complex and CtIP, are essential. Important aspects to study include, for instance how the components interact with each other and in combination with DNA. Studying the interactions under more *in vivo*-like conditions may provide additional interesting insights on HR-mediated DSB repair.

The original work presented in this Thesis has demonstrated the usefulness of single-molecule nanofluidics tools to study biomolecular interactions *in vitro*. The technique is particularly advantageous for studying reactions, where free and unobstructed DNA ends are required, for instance in DSB repair. The recent developments to this technique have enabled mapping of complex biochemical systems, thereby complementing the existing plethora of analytical tools with new important functionalities.

8 Acknowledgements

I would like to express my sincere gratitude to the following people for their direct or indirect support and contributions to all the work behind this Thesis.

First and foremost, my supervisor **Fredrik**, for your great support and guidance throughout my entire time in your research group. I deeply appreciate your trust in me to make my own decisions and freely explore different fields of science, which has allowed me to grow into an independent scientist.

My examiner **Pernilla** and my co-supervisors **Elin** and **Björn**, for your valuable support and advice and for always being there whenever needed. Björn, I am truly grateful for our many and long scientific discussions and all your tricky questions to challenge my hypotheses and ideas, which I believe have shaped me into a critical thinker.

All current and former members of the Westerlund group, and in particular; **Ville**, for being a great friend. Even though we were mostly in constant disagreement with everything, I really enjoined working with you all these years. **Rajhans**, for always being the happy guy in the lab. It is always fun being around you. **Gaurav**, for all the interesting discussions. **Sriram**, for being very responsive, in particular with the nanofabrication. **Sune**, **My** and **Vandana** for great company in the lab.

My former master thesis students **Hanna** and **Nora** as well as project workers **Hilda, Margareta** and **Firat** for your great and hard work.

Petr and his team members **Sean**, **Ilaria** and **Giordano** at IRB in Bellinzona, Switzerland for the great collaboration on the human and yeast HR projects. This also includes **Erik** at the department of Mathematical Sciences, Chalmers, for helping out with the data analysis. I would also like to thank all co-authors for their valuable contributions to the bacterial NHEJ project, in particular **Terrence**, **Jinglong** at IBENS in Paris, France and **Raphael** at I2BC in Paris, France.

Mauro, thank you for letting me visit your lab at CRCM in Marseille, France and for taking time to teach me new analytical methods. Although short, it was a great experience for me.

Jesper, **Emelie** and **Gaurav**, thank you very much for proof-reading parts of this Thesis.

All colleagues at the divisions of **Chemical biology** and **Physical chemistry** for creating such a nice and friendly work environment. A special thanks to **Emelie** and **Maria** for all of your pep-talks and support. It really meant a lot to me.

9 References

1. Crick, F.H.C. in Symposia of the Society for Experimental Biology, Vol. 12, Edn. 138-63 8 (1958).

2. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).

3. Hoeijmakers, J.H.J. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374 (2001).

4. Lieber, M.R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry* **79**, 181-211 (2010).

5. Rich, T., Allen, R.L. & Wyllie, A.H. Defying death after DNA damage. *Nature* **407**, 777-783 (2000).

6. Rothkamm, K., Krüger, I., Thompson, L.H. & Löbrich, M. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Molecular and cellular biology* **23**, 5706-5715 (2003).

7. van Oijen, A.M. Cutting the forest to see a single tree? *Nature Chemical Biology* **4**, 440-443 (2008).

8. Persson, F. & Tegenfeldt, J.O. DNA in nanochannels—directly visualizing genomic information. *Chemical Society Reviews* **39**, 985-999 (2010).

9. Tegenfeldt, J.O. et al. The dynamics of genomic-length DNA molecules in 100-nm channels. *Proceedings of the National Academy of Sciences* **101**, 10979 (2004).

10. Tegenfeldt, J.O. et al. Micro- and nanofluidics for DNA analysis. *Analytical and Bioanalytical Chemistry* **378**, 1678-1692 (2004).

11. Frykholm, K., Nyberg, L.K. & Westerlund, F. Exploring DNA–protein interactions on the single DNA molecule level using nanofluidic tools. *Integrative Biology* **9**, 650-661 (2017).

12. Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology* **278**, 274-288 (2005).

13. Avery, O.T., MacLeod, C.M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *The Journal of experimental medicine* **79**, 137-158 (1944).

14. Watson, J.D. & Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737-738 (1953).

15. Conner, B.N., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R.E. The molecular structure of d(ICpCpGpG), a fragment of right-handed double helical A-DNA. *Nature* **295**, 294-299 (1982).

16. Wang, A.H.J. et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680-686 (1979).

17. Bloomfield, V.A., Crothers, D.M. & Tinoco, I. Nucleic Acids: Structures, Properties and Functions. (University Science Books, 2000).

18. Dickerson, R.E. et al. The anatomy of A-, B-, and Z-DNA. *Science* **216**, 475 (1982).

19. Kratky, O. & Porod, G. Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas* **68**, 1106-1122 (1949).

20. Rivetti, C., Walker, C. & Bustamante, C. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility11Edited by D. Draper. *Journal of Molecular Biology* **280**, 41-59 (1998).

21. Odijk, T. On the ionic-strength dependence of the intrinsic viscosity of DNA. *Biopolymers: Original Research on Biomolecules* **18**, 3111-3113 (1979).

22. De Gennes, P.-G. & Gennes, P.-G. Scaling concepts in polymer physics. (Cornell university press, 1979).

23. Reisner, W., Pedersen, J.N. & Austin, R.H. DNA confinement in nanochannels: physics and biological applications. *Reports on Progress in Physics* **75**, 106601 (2012).

24. Odijk, T. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules* **16**, 1340-1344 (1983).

25. Wang, Y., Tree, D.R. & Dorfman, K.D. Simulation of DNA Extension in Nanochannels. *Macromolecules* **44**, 6594-6604 (2011).

26. Dai, L., van der Maarel, J. & Doyle, P.S. Extended de Gennes Regime of DNA Confined in a Nanochannel. *Macromolecules* **47**, 2445-2450 (2014).

27. Brochard-Wyart, F. & Raphael, E. Scaling theory of molten polymers in small pores. *Macromolecules* **23**, 2276-2280 (1990).

28. Temin, H.M. & Mizutami, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213 (1970).

29. Pogue, G.P., Huntley, C.C. & Hall, T.C. Common replication strategies emerging from the study of diverse groups of positive-strand RNA viruses. (Springer Vienna, 1994).

30. McCarthy, B.J. & Holland, J.J. Denatured DNA as a direct template for in vitro protein synthesis. *Proceedings of the National Academy of Sciences* **54**, 880-886 (1965).

31. Woodcock, C.L. & Ghosh, R.P. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2**, a000596 (2010).

32. Garcia, H.G. et al. Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers* **85**, 115-130 (2007).

33. Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes & Development* **7**, 1291-1308 (1993).

34. Seeman, N.C., Rosenberg, J.M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proceedings of the National Academy of Sciences* **73**, 804 (1976).

35. Lerman, L.S. Structural considerations in the interaction of DNA and acridines. *Journal of molecular biology* **3**, 18-IN14 (1961).

36. Richards, A.D. & Rodger, A. Synthetic metallomolecules as agents for the control of DNA structure. *Chemical Society Reviews* **36**, 471-483 (2007).

37. Rye, H.S. et al. Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications. *Nucleic Acids Research* **20**, 2803-2812 (1992).

38. Günther, K., Mertig, M. & Seidel, R. Mechanical and structural properties of YOYO-1 complexed DNA. *Nucleic Acids Research* **38**, 6526-6532 (2010).

39. Glazer, A.N. & Rye, H.S. Stable dye–DNA intercalation complexes as reagents for high-sensitivity fluorescence detection. *Nature* **359**, 859-861 (1992).

40. Reuter, M. & Dryden, D.T.F. The kinetics of YOYO-1 intercalation into single molecules of double-stranded DNA. *Biochemical and Biophysical Research Communications* **403**, 225-229 (2010).

41. Larsson, A., Carlsson, C., Jonsson, M. & Albinsson, B. Characterization of the Binding of the Fluorescent Dyes YO and YOYO to DNA by Polarized Light Spectroscopy. *Journal of the American Chemical Society* **116**, 8459-8465 (1994).

42. Spielmann, H.P., Wemmer, D.E. & Jacobsen, J.P. Solution Structure of a DNA Complex with the Fluorescent Bis-Intercalator TOTO Determined by NMR Spectroscopy. *Biochemistry* **34**, 8542-8553 (1995).

43. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715 (1993).

44. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194-1217 (2013).

45. Hoeijmakers, J.H.J. DNA Damage, Aging, and Cancer. *New England Journal of Medicine* **361**, 1475-1485 (2009).

46. Alberts, B. et al. Molecular Biology of The Cell. (Garland Science, 2008).

47. Feijó Delgado, F. et al. Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells. *PloS one* (2013).

48. Perrett, D. From 'protein' to the beginnings of clinical proteomics. *PROTEOMICS – Clinical Applications* **1**, 720-738 (2007).

49. Sanger, F. The terminal peptides of insulin. *The Biochemical journal* **45**, 563-574 (1949).

50. Pauling, L., Corey, R.B. & Branson, H.R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* **37**, 205 (1951).

51. Rafferty, J.B. et al. Crystal Structure of DNA Recombination Protein RuvA and a Model for Its Binding to the Holliday Junction. *Science* **274**, 415 (1996).

52. Küppers, R. & Dalla-Favera, R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580-5594 (2001).

53. Valko, M. et al. Free radicals and antioxidants in normal physiological functions and human disease. *The International Journal of Biochemistry & Cell Biology* **39**, 44-84 (2007).

54. Friedberg, E.C., Walker, G.C., Siede, W. & Wood, R.D. DNA repair and mutagenesis. (American Society for Microbiology Press, 2005).

55. Lin, Z., Kong, H., Nei, M. & Ma, H. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences* **103**, 10328-10333 (2006).

56. Moynahan, M.E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews Molecular Cell Biology* **11**, 196-207 (2010).

57. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell cycle* **7**, 2902-2906 (2008).

58. Fabre, F. Induced intragenic recombination in yeast can occur during the G1 mitotic phase. *Nature* **272**, 795-798 (1978).

59. Heyer, W.-D., Ehmsen, K.T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annual review of genetics* **44**, 113-139 (2010).

60. Sartori, A.A. et al. Human CtIP promotes DNA end resection. *Nature* **450**, 509-514 (2007).

61. Nicolette, M.L. et al. Mre11–Rad50–Xrs2 and Sae2 promote 5′ strand resection of DNA double-strand breaks. *Nature structural & molecular biology* **17**, 1478 (2010).

62. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annual Review of Biochemistry* **77**, 229-257 (2008).

63. Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell research* **18**, 99-113 (2008).

64. Schaeper, U., Subramanian, T., Lim, L., Boyd, J.M. & Chinnadurai, G. Interaction between a cellular protein that binds to the C-terminal region of adenovirus E1A (CtBP) and a novel cellular protein is disrupted by E1A through a conserved PLDLS motif. *Journal of Biological Chemistry* **273**, 8549-8552 (1998).

65. Wong, A.K.C. et al. Characterization of a carboxy-terminal BRCA1 interacting protein. *Oncogene* **17**, 2279-2285 (1998).

66. Yun, M.H. & Hiom, K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* **459**, 460-463 (2009).

67. Deshpande, R.A. et al. DNA-dependent protein kinase promotes DNA end processing by MRN and CtIP. *Science Advances* **6**, eaay0922 (2020).

68. Anand, R., Ranjha, L., Cannavo, E. & Cejka, P. Phosphorylated CtIP functions as a co-factor of the MRE11-RAD50-NBS1 endonuclease in DNA end resection. *Molecular cell* **64**, 940-950 (2016).

69. Andres, S.N., Li, Z.M., Erie, D.A. & Williams, R.S. Ctp1 protein–DNA filaments promote DNA bridging and DNA double-strand break repair. *Journal of Biological Chemistry* **294**, 3312-3320 (2019).

70. Wilkinson, O.J. et al. CtIP forms a tetrameric dumbbell-shaped particle which bridges complex DNA end structures for double-strand break repair. *Elife* **8**, e42129 (2019).

71. Öz, R. et al. Phosphorylated CtIP bridges DNA to promote annealing of broken ends. *Proceedings of the National Academy of Sciences* **117**, 21403-21412 (2020).

72. Davies, O.R. et al. CtIP tetramer assembly is required for DNA-end resection and repair. *Nature structural & molecular biology* **22**, 150-157 (2015).

73. Paull, T.T. & Gellert, M. Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes & development* **13**, 1276-1288 (1999).

74. Williams, R.S. et al. Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**, 97-109 (2008).

75. Dolganov, G.M. et al. Human Rad50 is physically associated with human Mre11: identification of a conserved multiprotein complex implicated in recombinational DNA repair. *Molecular and cellular biology* **16**, 4832-4841 (1996).

76. Johzuka, K. & Ogawa, H. Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in Saccharomyces cerevisiae. *Genetics* **139**, 1521-1532 (1995).

77. Trujillo, K.M. et al. Yeast xrs2 binds DNA and helps target rad50 and mre11 to DNA ends. *Journal of Biological Chemistry* **278**, 48957-48964 (2003).

78. Schiller, C.B. et al. Structure of Mre11–Nbs1 complex yields insights into ataxia-telangiectasia–like disease mutations and DNA damage signaling. *Nature structural & molecular biology* **19**, 693 (2012).

79. Hopfner, K.-P. et al. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**, 562-566 (2002).

80. Wang, H. et al. The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR–mediated double-strand break repair. *PLoS genet* **9**, e1003277 (2013).

81. Liang, J., Suhandynata, R.T. & Zhou, H. Phosphorylation of Sae2 mediates forkhead-associated (FHA) domain-specific interaction and regulates its DNA repair function. *Journal of Biological Chemistry* **290**, 10751-10763 (2015).

82. Oh, J., Al-Zain, A., Cannavo, E., Cejka, P. & Symington, L.S. Xrs2 dependent and independent functions of the Mre11-Rad50 complex. *Molecular cell* **64**, 405-415 (2016).

83. Moore, J.K. & Haber, J.E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and cellular biology* **16**, 2164-2173 (1996).

84. Bétermier, M., Bertrand, P. & Lopez, B.S. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* **10**, e1004086 (2014).

85. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA repair* **7**, 1765-1771 (2008).

86. Wilson, J.H., Berget, P.B. & Pipas, J.M. Somatic cells efficiently join unrelated DNA segments end-to-end. *Molecular and cellular biology* **2**, 1258-1269 (1982).

87. Pitcher, R.S., Brissett, N.C. & Doherty, A.J. Nonhomologous end-joining in bacteria: a microbial perspective. *Annual Review of Microbiology* **61**, 259-282 (2007).

88. Smith, G.C.M. & Jackson, S.P. The DNA-dependent protein kinase. *Genes & development* **13**, 916-934 (1999).

89. Gottlieb, T.M. & Jackson, S.P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131-142 (1993).

90. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M.R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V (D) J recombination. *Cell* **108**, 781-794 (2002).

91. Davis, A.J. & Chen, D.J. DNA double strand break repair via non-homologous end-joining. *Translational cancer research* **2**, 130 (2013).

92. Koch, C.A. et al. Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *The EMBO journal* **23**, 3874-3885 (2004).

93. Hammel, M. et al. XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *Journal of biological chemistry* **286**, 32638-32650 (2011).

94. Andres, S.N. et al. A human XRCC4–XLF complex bridges DNA. *Nucleic acids research* **40**, 1868-1878 (2012).

95. Weller, G.R. et al. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* **297**, 1686-1689 (2002).

96. Bowater, R. & Doherty, A.J. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet* **2**, e8 (2006).

97. Zhu, H. & Shuman, S. Gap filling activities of Pseudomonas DNA ligase D (LigD) polymerase and functional interactions of LigD with the DNA end-binding Ku protein. *Journal of Biological Chemistry* **285**, 4815-4825 (2010).

98. de Vega, M. The minimal Bacillus subtilis nonhomologous end joining repair machinery. *PLoS One* **8**, e64232 (2013).

99. Ayora, S. et al. Double-strand break repair in bacteria: a view from Bacillus subtilis. *FEMS microbiology reviews* **35**, 1055-1081 (2011).

100. Downs, J.A. & Jackson, S.P. A means to a DNA end: the many roles of Ku. *Nature Reviews Molecular Cell Biology* **5**, 367-378 (2004).

101. Mimori, T. et al. Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap. *The Journal of clinical investigation* **68**, 611-620 (1981).

102. Koike, M., Shiomi, T. & Koike, A. Dimerization and nuclear localization of Ku proteins. *Journal of Biological Chemistry* **276**, 11167-11173 (2001).

103. Sawada, M. et al. Ku70 suppresses the apoptotic translocation of Bax to mitochondria. *Nature cell biology* **5**, 320-329 (2003).

104. Fewell, J.W. & Kuff, E.L. Intracellular redistribution of Ku immunoreactivity in response to cell-cell contact and growth modulating components in the medium. *Journal of Cell Science* **109**, 1937-1946 (1996).

105. Mimori, T. & Hardin, J.A. Mechanism of interaction between Ku protein and DNA. *Journal of Biological Chemistry* **261**, 10375-10379 (1986).

106. Walker, J.R., Corpina, R.A. & Goldberg, J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-614 (2001).

107. de Vries, E., van Driel, W., Bergsma, W.G., Arnberg, A.C. & van der Vliet, P.C. HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *Journal of molecular biology* **208**, 65-78 (1989).

108. Postow, L. et al. Ku80 removal from DNA through double strand break–induced ubiquitylation. *Journal of Cell Biology* **182**, 467-479 (2008).

109. van den Boom, J. et al. VCP/p97 extracts sterically trapped Ku70/80 rings from DNA in double-strand break repair. *Molecular cell* **64**, 189-198 (2016).

110. Feng, L. & Chen, J. The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. *Nature structural & molecular biology* **19**, 201 (2012).

111. McGovern, S. et al. C-terminal region of bacterial Ku controls DNA bridging, DNA threading and recruitment of DNA ligase D for double strand breaks repair. *Nucleic acids research* **44**, 4785-4806 (2016).

112. Acuña, A.U., Amat-Guerri, F., Morcillo, P., Liras, M. & Rodríguez, B. Structure and Formation of the Fluorescent Compound of Lignum nephriticum. *Organic Letters* **11**, 3020-3023 (2009).

113. Herschel, J.F.W. On a case of superficial colour presented by a homogeneous liquid internally colourless. *Philisopical Translation of the Royal Society of London* **135**, 143-145 (1845).

114. Atkins, P.W. & Jones, L.L. Chemical Principles - The Quest for Insight. (W. H. Freeman and Company, 2010).

115. Bohr, N. On the constitution of atoms and molecules. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **26**, 476-502 (1913).

116. Atkins, P., De Paula, J. & Friedman, R. Quanta, matter, and change: a molecular approach to physical chemistry. (Oxford University Press, 2009).

117. Lakowicz, J.R. Principles of Fluorescence Spectroscopy. (Springer, 2010).

118. Carlsson, C., Larsson, A., Jonsson, M., Albinsson, B. & Norden, B. Optical and photophysical properties of the oxazole yellow DNA probes YO and YOYO. *The Journal of Physical Chemistry* **98**, 10313-10321 (1994).

119. Premaratne, S., Swenson, S.D., Mandel, M. & Mower, H.F. Detection of small quantities of double stranded DNA by enhanced fluorescence of substituted aldehyde modified polydeoxynucleotide-ethidium bromide complexes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1219**, 422-424 (1994).

120. Pohl, F.M., Jovin, T.M., Baehr, W. & Holbrook, J.J. Ethidium Bromide as a Cooperative Effector of a DNA Structure. *Proceedings of the National Academy of Sciences* **69**, 3805 (1972).

121. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology* **21**, 86-89 (2003).

122. Cole, N.B. Site-specific protein labeling with SNAP-tags. *Current protocols in protein science* **73**, 30.31.31-16 (2013).

123. Heim, R., Cubitt, A.B. & Tsien, R.Y. Improved green fluorescence. *Nature* **373**, 663-664 (1995).

124. Shimomura, O., Johnson, F.H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* **59**, 223-239 (1962).

125. Shaner, N.C., Steinbach, P.A. & Tsien, R.Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905-909 (2005).

126. Hillisch, A., Lorenz, M. & Diekmann, S. Recent advances in FRET: distance determination in protein–DNA complexes. *Current Opinion in Structural Biology* **11**, 201-207 (2001).

127. Donnelly, W.J. & Roorda, A. Optimal pupil size in the human eye for axial resolution. *Journal of the Optical Society of America A* **20**, 2010-2015 (2003).

128. Schwartz, D.C. & Cantor, C.R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67-75 (1984).

129. Stenberg, E., Persson, B., Roos, H. & Urbaniczky, C. Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. *Journal of colloid and interface science* **143**, 513-526 (1991).

130. Greenleaf, W.J., Woodside, M.T. & Block, S.M. High-resolution, single-molecule measurements of biomolecular motion. *Annual review of biophysics and biomolecular structure* **36**, 171-190 (2007).

131. Zimmermann, R.M. & Cox, E.C. DNA stretching on functionalized gold surfaces. *Nucleic Acids Research* **22**, 492-497 (1994).

132. Granéli, A., Yeykal, C.C., Prasad, T.K. & Greene, E.C. Organized Arrays of Individual DNA Molecules Tethered to Supported Lipid Bilayers. *Langmuir* **22**, 292-299 (2006).

133. Smith, S.B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **258**, 1122 (1992).

134. Smith, S.B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **271**, 795 (1996).

135. Binnig, G., Quate, C.F. & Gerber, C. Atomic force microscope. *Physical review letters* **56**, 930 (1986).

136. Ohnesorge, F. & Binnig, G. True Atomic Resolution by Atomic Force Microscopy Through Repulsive and Attractive Forces. *Science* **260**, 1451 (1993).

137. Bustamante, C. et al. Circular DNA molecules imaged in air by scanning force microscopy. *Biochemistry* **31**, 22-26 (1992).

138. Dufrêne, Y.F. et al. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nature Nanotechnology* **12**, 295-307 (2017).

139. Möller, C., Allen, M., Elings, V., Engel, A. & Müller, D.J. Tapping-Mode Atomic Force Microscopy Produces Faithful High-Resolution Images of Protein Surfaces. *Biophysical Journal* **77**, 1150-1158 (1999).

140. Putman, C.A.J., Van der Werf, K.O., De Grooth, B.G., Van Hulst, N.F. & Greve, J. Tapping mode atomic force microscopy in liquid. *Applied physics letters* **64**, 2454-2456 (1994).

141. Neuman, K.C. & Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods* **5**, 491-505 (2008).

142. Wang, J.L. et al. Dissection of DNA double-strand-break repair using novel single-molecule forceps. *Nature Structural & Molecular Biology* **25**, 482-487 (2018).

143. Ashkin, A. Acceleration and trapping of particles by radiation pressure. *Physical review letters* **24**, 156 (1970).

144. Ashkin, A. Trapping of atoms by resonance radiation pressure. *Physical Review Letters* **40**, 729 (1978).

145. Ashkin, A., Dziedzic, J.M., Bjorkholm, J.E. & Chu, S. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters* **11**, 288-290 (1986).

146. Heller, I., Hoekstra, T.P., King, G.A., Peterman, E.J.G. & Wuite, G.J.L. Optical Tweezers Analysis of DNA–Protein Complexes. *Chemical Reviews* **114**, 3087-3119 (2014).

147. Candelli, A., Wuite, G.J.L. & Peterman, E.J.G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA–protein interactions. *Physical Chemistry Chemical Physics* **13**, 7263-7272 (2011).

148. Williams, M.C. Optical tweezers: measuring piconewton forces. *Biophysics Textbook Online* (2002).

149. Wang, M.D. et al. Force and Velocity Measured for Single Molecules of RNA Polymerase. *Science* **282**, 902 (1998).

150. Brower-Toland, B.D. et al. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proceedings of the National Academy of Sciences* **99**, 1960 (2002).

151. Blumberg, S., Pennington, M.W. & Meiners, J.-C. Do Femtonewton Forces Affect Genetic Function? A Review. *Journal of Biological Physics* **32**, 73-95 (2006).

152. Persson, F., Utko, P., Reisner, W., Larsen, N.B. & Kristensen, A. Confinement Spectroscopy: Probing Single DNA Molecules with Tapered Nanochannels. *Nano Letters* **9**, 1382-1385 (2009).

153. Duan, C., Wang, W. & Xie, Q. Review article: Fabrication of nanofluidic devices. *Biomicrofluidics* **7**, 026501 (2013).

154. Fornander, L.H. et al. Visualizing the Nonhomogeneous Structure of RAD51 Filaments Using Nanofluidic Channels. *Langmuir* **32**, 8403-8412 (2016).

155. Frykholm, K. et al. Probing Physical Properties of a DNA-Protein Complex Using Nanofluidic Channels. *Small* **10**, 884-887 (2014).

156. Zhang, C. et al. A nanofluidic device for single molecule studies with in situ control of environmental solution conditions. *Lab on a Chip* **13**, 2821-2826 (2013).

157. Henkin, G. et al. Manipulating and Visualizing Molecular Interactions in Customized Nanoscale Spaces. *Analytical Chemistry* **88**, 11100-11107 (2016).

158. Öz, R., Kk, S. & Westerlund, F. A nanofluidic device for real-time visualization of DNA–protein interactions on the single DNA molecule level. *Nanoscale* **11**, 2071-2078 (2019).

159. Khanna, K.K. & Jackson, S.P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature genetics* **27**, 247-254 (2001).

160. Jiang, K. et al. Annealing of ssDNA and compaction of dsDNA by the HIV-1 nucleocapsid and Gag proteins visualized using nanofluidic channels. *Quarterly reviews of biophysics* **52** (2019).

161. Alizadehheidari, M. et al. Nanoconfined circular and linear DNA: Equilibrium conformations and unfolding kinetics. *Macromolecules* **48**, 871-878 (2015).

162. Sharma, R., Sriram, K.K., Holmstrom, E.D. & Westerlund, F. Real-time compaction of nanoconfined DNA by an intrinsically disordered macromolecular counterion. *Biochemical and Biophysical Research Communications* (2020).

163. Nyberg, L.K. et al. Rapid identification of intact bacterial resistance plasmids via optical mapping of single DNA molecules. *Scientific reports* **6**, 30410 (2016).

164. Cannavo, E. et al. Regulatory control of DNA end resection by Sae2 phosphorylation. *Nature communications* **9**, 1-14 (2018).

165. Cannavo, E., Reginato, G. & Cejka, P. Stepwise 5′ DNA end-specific resection of DNA breaks by the Mre11-Rad50-Xrs2 and Sae2 nuclease ensemble. *Proceedings of the National Academy of Sciences* **116**, 5505-5513 (2019).