THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SINGLE DNA MOLECULE ANALYSIS

NEW TOOLS FOR MEDICAL DIAGNOSIS

MY NYBLOM

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

Single DNA molecule analysis

New tools for medical diagnosis

MY NYBLOM

ISBN 978-91-7905-762-6

© MY NYBLOM, 2023.

Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie nr 5228

ISSN 0346-718X

Department of Life Sciences

Chalmers University of Technology

SE-412 96 Gothenburg

Sweden

Telephone + 46 (0)31-772 1000

Cover:

Two illustrations, a serpentine streamer (top), partly still wrapped as a coil, partly as two bundles and partly stretched out in a transparent straw to enable a message to be read. A DNA molecule (bottom), partly still in a chromosome, partly as chromatin, partly wrapped around two histones and partly stretched out in a nanochannel to enable the underlying DNA sequence, and DNA damage, to be read. DNA created with BioRender.

Printed by Chalmers Digitaltryck

Gothenburg, Sweden 2023

SINGLE DNA MOLECULE ANALYSIS

NEW TOOLS FOR MEDICAL DIAGNOSIS

MY NYBLOM

Department of Life Sciences, Chalmers University of Technology

ABSTRACT

The DNA molecule, the blueprint of life, contains an enormous amount of information. The information is coded by the combination of four bases; adenine, cytosine, guanine, and thymine, that, together with the sugar-phosphate backbones, make up the DNA double helix. There are variants in the human DNA sequence that are related to the onset and progression of disease. Under different conditions the DNA can also be damaged, which if not repaired correctly can result in a shortened life span, rapid ageing and development or progression of a variety of diseases, including cancer. Human disease can also be induced by external factors in our surroundings, such as pathogens. One of the cornerstones in modern medicine has been the use of antibiotics to prevent and treat these pathogenic infections, but the global spread of antibiotic resistance is today one of the largest threats to mankind according to the World Health Organization. One consequence of a large global increase in antibiotic resistance would be that routine surgery or chemotherapy treatment might be considered too perilous, because there are no drugs available to prevent or treat the bacterial infections that are closely connected with these procedures.

Novel techniques are needed to characterize different features of DNA in medicine and diagnostics. Single molecule analysis is one method to unveil different kinds of information from individual biomolecules, such as DNA. This thesis uses fluorescence microscopy to shine light upon such information in single DNA molecules from both humans and bacteria, and with that unveil important biological and medical characteristics of that DNA. It describes one method for identifying and quantifying DNA damage induced by a chemotherapy agent, helping to understanding the processes of DNA damage and repair related to diseases and medical treatments. Another method developed is for rapid identification of bacterial infections, with the classification of bacterial sub-species groups and identification of antibiotic resistance genes on plasmids. The methods have the potential to rapidly provide comprehensive diagnostics information, to optimize either early antibiotic treatment or chemotherapy treatment, and thereby enable future precision medicine management.

Keywords: DNA, optical DNA mapping, nanofluidics, DNA damage, fluorescence microscopy, genetics, antibiotic resistance, single-molecule analysis

LIST OF PUBLICATIONS

This thesis is based on the work presented in the following papers, referred to by roman numerals, and shared first authorship referred to by asterisks (*):

I. Quantifying Mitomycin C Induced DNA Crosslinking Damage
Obed Akwasi Aning, My Nyblom, Albertas Dvirnas, Jens Krog, Erik Fröbrant, Tobias Ambjörnsson, Pegah Johansson, Ola Hammarsten, Fredrik Westerlund
(Manuscript in preparation)

II. Cultivation-Free Typing of Bacteria Using Optical DNA Mapping
Vilhelm Müller*, My Nyblom*, Anna Johnning*, Marie Wrande, Albertas Dvirnas, Sriram KK, Christian G. Giske, Tobias Ambjörnsson, Linus Sandegren, Erik Kristiansson, and Fredrik Westerlund
ACS Infectious Diseases, 2020, 6(5), 1076-1084

III. Strain-level bacterial typing directly from patient samples using optical DNA mapping
My Nyblom*, Anna Johnning*, Vilhelm Müller, Karolin Frykholm, Miriam Robertsson, Gaurav Goyal, Marie Wrande, Albertas Dvirnas, Sriram KK, Christian G. Giske, Tobias Ambjörnsson, Linus Sandegren, Erik Kristiansson, and Fredrik Westerlund
Accepted for publication in Communications Medicine (December 2022)

IV. A simple cut and stretch assay to detect antimicrobial resistance genes on bacterial plasmids by single-molecule fluorescence microscopy
Gaurav Goyal, Elina Ekedahl, My Nyblom, Jens Krog, Erik Fröbrant, Magnus Brander, Tsegaye Sewunet, Teerawit Tangkoskul, Christian G. Giske, Linus Sandegren, Visanu Thamlikitkul, Tobias Ambjörnsson, and Fredrik Westerlund
Scientific Reports, 2022, 12, 9301

CONTRIBUTION REPORT

Below follows a description of my contributions to the papers, where the initials are from authors in the list of publications:

I. MN performed experiments, contributed to the conceptual design of the data analysis software, analyzed data and wrote the paper, all together with OA.

II. MN planned and performed experiments together with VM, analyzed the data, contributed to the conceptual design of the data analysis software and wrote the paper together with VM.

III. MN planned and performed experiments together with VM and KF, analyzed the data together with AJ, contributed to the conceptual design of the data analysis software and wrote the paper together with AJ.

IV. MN performed experiments, contributed to the conceptual design of the data analysis software, contributed to writing the paper and supervised EE together with GG.

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 August 2018 and February 2023 at the Life Sciences Department (named Department of Biology and Biological Engineering until January 2023) at Chalmers University of Technology, under the supervision of Professor Fredrik Westerlund. The research was funded by EU Horizon 2020 program BeyondSeq, The Erling Persson Family Foundation, Barncancerfonden, and Cancerfonden.

My Nyblom

January 2023

ABBREVIATIONS

A	Adenine
AMR	Antimicrobial resistance
BER	Base excision repair
C	Cytosine
Cas	CRISPR associated
CNVs	Copy number variations
CRISPR	Clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
DDR	DNA damage response
DNA	Deoxyribonucleic acid
DSB	Double stranded DNA break
dsDNA	Double stranded DNA
G	Guanine
Gbp	Giga base pairs
gRNA	Guide RNA
HR	Homologous recombination
ICL	Interstrand crosslink
kbp	Kilo base pairs
Mbp	Mega base pairs
MMC	Mitomycin C
MMR	Mismatch repair
NER	Nucleotide excision repair
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining
PAM	Protospacer adjacent motif

PBMC	Peripheral blood mononuclear cell
PCC	Pearson correlation coefficient
PSF	Point spread function
RNA	Ribonucleic acid
SGs	Strain groups
SNPs	Single nucleotide polymorphism
SSBR	Single-strand break repair
ssDNA	Single stranded DNA
ST	Sequence type
T	Thymine
tracrRNA	Trans-activating CRISPR RNA
TPR	True positive rate
U	Uracil
UTI	Urinary tract infection
UV	Ultra-violet

TABLE OF CONTENTS

1 INTRODUCTION

The deoxyribonucleic acid (DNA), the blueprint of life, encodes an enormous amount of information. It is encoded by the combination of four bases; adenine (A), cytosine (C), guanine (G), and thymine (T), that are connected by a sugar-phosphate backbone that forms single stranded DNA (ssDNA). Two ssDNA strands form double stranded DNA (dsDNA), which has a helical structure. The DNA double helix is probably the most recognized chemical structure of all to the public ever since Francis, Crick[1] and Franklin[2,3] unveiled it by X-ray crystallography in 1953.

Between 1990 and 2003 The Human Genome Project was carried out internationally[4], to unveil the base pair code that makes up human DNA and what the genes stand for[5,6]. With the advances of the sequencing technologies, us humans have been able to understand even more about this base pair code and what large parts of that DNA codes for. There are variants in the human DNA sequence as well as damage to the DNA that are related to onset and progression of disease. Furthermore, it is not only our own DNA that has been studied extensively, but the DNA of the things in our surroundings, such as bacteria or viruses that affect our health.

To understand more about the correlations between DNA and disease, and to apply that knowledge in the fields of medicine and diagnostics, novel techniques are needed to explore different features of DNA. One way to address it is through single molecule analysis, that can unveil differences between individual molecules, in what otherwise, in a bulk-measurement, would give an averaged result of the population of molecules. Some DNA molecules are long

and linear, whereas others are short and circular. Some DNA molecules carry specific genes of interest, whereas others do not. Furthermore, under different conditions, the DNA can also be damaged. This thesis uses fluorescence microscopy to shine light upon DNA, from humans as well as bacteria, lit up by different fluorescent probes. The resulting fluorescence patterns can be utilized to unveil the biological consequences of these differences and how this information can be used for medical diagnosis.

The more we learn about the human genome, the more it becomes apparent that nearly all diseases seem to have a genetic compound[7]. Our DNA is constantly damaged, the DNA in each human cell is exposed to on the order of tens of thousands DNA lesions each day[8]. Changes to our DNA can have fatal results, and as a protective mechanism, several cellular systems for detecting, signaling, and repairing DNA damage have been developed. The DNA damage is induced by both endogenous factors, such as metabolic compounds from cellular metabolism, and environmental factors such as ultra-violet (UV) radiation or chemicals from use of tobacco products. People that have an inherited genetic defect in one or more DNA repairing mechanisms, resulting in an increased rate and accumulated number of DNA mutations for the cells, show an increased risk of developing diseases, including cancer. The DNA itself is the main target in multiple cancer treatments, as in radiotherapy or in treatment with chemotherapy agents. However, not all patients respond in the same way to the same type and dose of treatment.

My research in this area has focused on further developing a method to identify and quantify induced DNA damage by conducting fluorescence measurements on single DNA molecules, that are stretched on silanized glass slides, with fluorescent labeled individual DNA damage sites (**Paper I**). This research aims towards developing a tool to help understanding the processes of DNA damage and repair related to diseases and medical treatments. The final aim would be to identify individual differences in the repair profile of the DNA damage repair from a specific drug (or multiple drugs). This could be used as a tool that would determine what type and amount of DNA damage can be formed by a drug and repaired by an individual. In the end, the method would then serve as a tool for personalized therapy.

For the past years, when the work that led up to this thesis was carried out, we have all been living through the Covid-19 pandemic. Pandemics are nothing new, they have happened before in our history, and they will happen again. However, not all pandemics are as rapidly developing as Covid-19. We are currently also living in what is often referred to as a silent pandemic that did not hit worldwide in a few months, but that slowly has been creeping up on us ever since Alexander Fleming discovered penicillin in 1928[9]. Different types of bacteria can cause different acute and harmful diseases to humans. After 1928, antibiotics became a magical quick fix for a lot of diseases as they could be used to treat and prevent bacterial infections. The use of antibiotics also enabled the medical care we have today, where for example surgery without using antibiotics would be extremely precarious. However, the advancement of antibiotics was followed by the uprise of antibiotic resistance. Today the global spread of antibiotic resistance is considered one of the largest threats to mankind by the World Health Organization. It is predicted that in the year of 2050, 10 millions people will die from infections caused by antibiotic resistant bacteria[10]. Already on December 11th of 1945, when Alexander Fleming

gave his Nobel Lecture[11], he said "The time may come when penicillin can be bought by anyone in the shops. Then there is the danger that the ignorant man may easily underdose himself and by exposing his microbes to non-lethal quantities of the drug make them resistant."

Several types of antibiotic resistance are encoded by genes in the bacteria's DNA. They are sometimes located on plasmids, circular DNA molecules that bacteria can share with each other, and thereby quickly spread the antibiotic resistance. To efficiently select the correct antibiotic in clinical settings, rapid identification of the carriage of plasmids harboring antibiotic resistance genes and sub-species identification is important. Unfortunately, current methods for identification of the bacteria and antimicrobial resistance (AMR) routinely used in hospitals are too time consuming. They often requiring one or two overnight cultivations, where some bacteria are not even easily cultured in laboratories.

My research in this area has focused on developing methods for rapid identification of bacterial infections by analyzing bacterial DNA. The aim of the research is to further develop tools for identifying pathogenic bacteria and AMR-gene harboring plasmids, by conducting fluorescence microscopy on single DNA molecules that are fluorescently labeled (**Paper II**, **Paper III** and **Paper IV**). In **Paper II** and **Paper III**, the underlying DNA sequence is reflected by hindered/fluorescently labeled AT/GC regions. In **Paper II** we developed a method where we used long chromosomal fragments for bacterial species identification by optical DNA mapping in nanochannels. The method was tested successfully for six different bacterial species, including both Gram-positive and Gram-negative bacteria, as well as mixes of up to four bacterial species and non-cultured UTI samples. In **Paper III** we further developed this method where we used the long chromosomal fragments for sub-species identification of *Escherichia coli* (*E. coli*) and *Klebsiella pneumoniae* (*K. pneumoniae*) while simultaneously characterizing their plasmids and identifying antibiotic resistance genes with CRISPR-Cas9 mediated cuts. In **Paper IV** we developed a method that combined stretching of fluorescently labeled bacterial plasmids on silanized glass slides with CRISPR-Cas9 mediated cuts for targeted antibiotic resistance genes that could identify antibiotic resistance genes by plasmid linearization. This method was developed to work on a simple fluorescence Zeiss Primo-Star iLED microscope that is used in low- and middle-income countries for diagnosing tuberculosis, in combination with a smartphone camera for imaging, thus being directly applicable to low-resource settings.

2 DNA

2.1 CHEMICAL STRUCTURE OF DNA

The chemical structure of the deoxyribonucleic acid molecule, more commonly known as DNA, is probably the most well recognized chemical structure to the public there is. The double helix (pictured in Figure 1A), unveiled by X-ray crystallography in 1953 by Watson, Crick[1] and Franklin[2,3], is composed of two complementary polynucleotide chains of nucleotides, running in opposite directions, that are held together by the hydrogen bonds between four different nucleobases.

A nucleotide consists of one nucleobase bound to a five-carbon sugar deoxyribose that is bound to a phosphate group. The four nucleobases in DNA are A, C, G, and T. The specific base-pairing is of a purine with a pyrimidine, where A pairs with T with two hydrogen bonds and G pairs with C with three hydrogen bonds[1] (Figure 1B).

Each of the polynucleotide chains has a covalently linked pentose-phosphate backbone that forms a ssDNA molecule. They have a directionality from the phosphate 5' to hydroxyl 3' end of the nucleotides, that gives each ssDNA molecule a directionality. Two ssDNA molecules held together by the base paring form the dsDNA molecule. At neutral pH the bases are uncharged, and the ribose-phosphate backbone is negatively charged. In the DNA double helix, the bases are stacked, shielding their hydrophobic structures inwards, and exposing the polar backbone to the solvent.

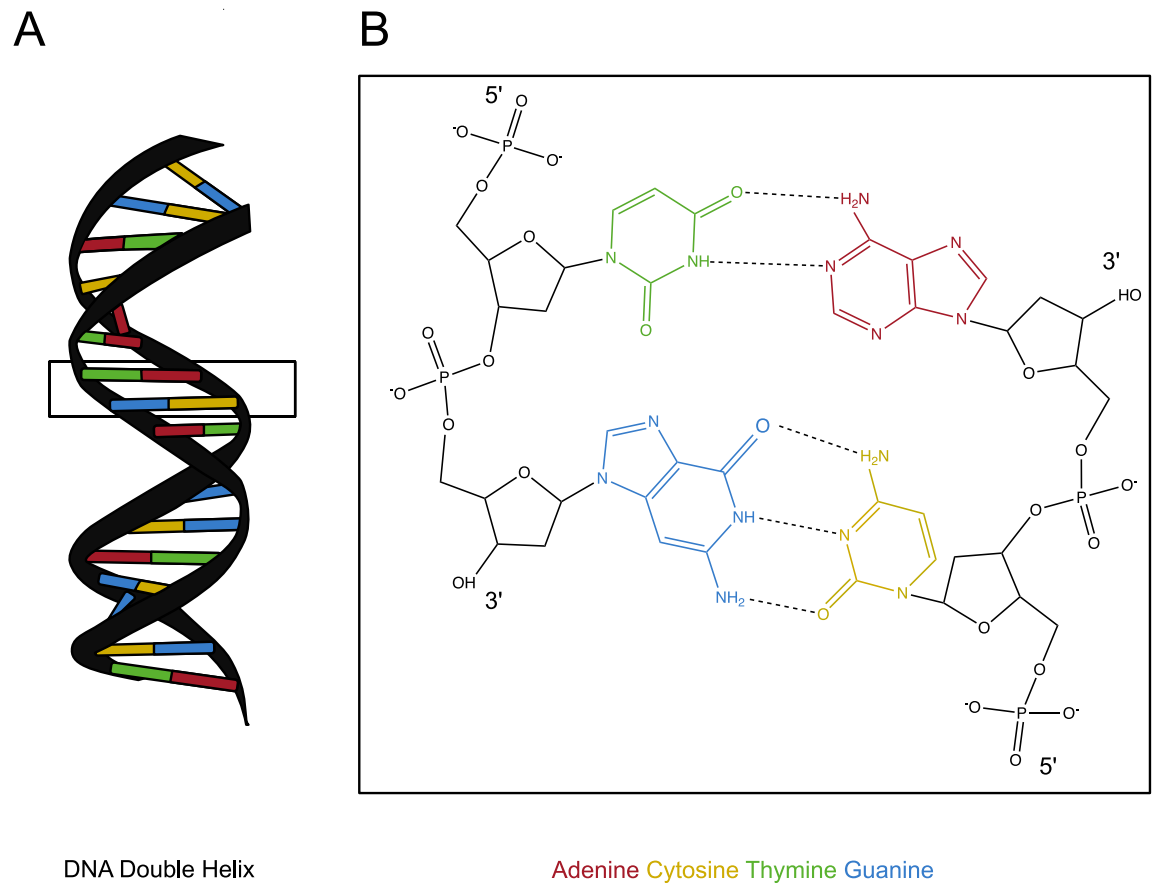

FIGURE 1: IN A, A REPRESENTATION OF THE B-FORM DNA DOUBLE HELIX WITH THE FOUR BASES, ADENINE (RED), CYTOSINE (YELLOW), THYMINE (GREEN) AND GUANINE (BLUE), PICTURED IN COLOUR AND THE SUGAR-PHOSPHATE BACKBONE IN BLACK. IN B, THE CHEMICAL STRUCTURE OF THE SUGAR-PHOSPHATE BACKBONE AND BASE PAIRING OF THE FOUR NUCLEOBASES WITH THE DIRECTIONALITY FROM 5' TO 3' OF THE SSDNA STRANDS MARKED OUT.

In nature, there are three forms of DNA, B-form, A-form, and Z-form, where the B-form (in Figure 1A) is the most common at physiological conditions[1,12-14]. The B-form of DNA is a right-handed helix with a diameter of 20 Å and approximately 10 base pairs per turn[13]. The average separation between the base pairs is 3.4 Å, where the bases are sitting nearly perpendicular to the long axis of the helix facing inward in the hydrophobic center and the charged backbone out towards the solvent[13,15]. As a consequence of the twist, a major and a minor groove of the B-form DNA helix is formed, where the major groove is 11 Å wide and 4 Å deep, and the minor groove is 6 Å wide and 5 Å deep[16]. The A-form DNA is also a right-handed helix with a similar diameter, but more compressed and with the base pairs sitting tilted approximately 20° against the long axis of the helix. Biologically, the A-form is mostly occurring during dehydrated conditions, induced by binding of proteins or in G/C rich regions[17,18]. The radically different Z-form DNA is instead a left-handed helix with a slightly smaller diameter of 18 Å, that has a zig-

zag conformation of the phosphate backbone[13,19]. Biologically, the Z-form of DNA is formed in GC-alternating sequences under high-salt conditions[18].

2.2 BIOLOGICAL FUNCTION

Some say that the meaning of life, the prime objective of every life form, is to pass its genetic material forward to its next generation in an intact and unchanged form. Life thereby depends on the cell's abilities to store, retrieve, and translate the genetic information encoded in the DNA molecule, to both maintain a living organism and to make a new one. The specific base pairing and the duality of the DNA double helix is very neat, meaning that the information is often stored twice. Thereby one of the strands can be used as a scaffold for repair or copying for the other one, something that Watson and Crick[1] understood with the proposal of the chemical structure of the DNA helix: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material".

All information needed for a cell to function is stored in the DNA molecule, while the ribonucleic acid (RNA) and proteins carry out the functions encoded in the DNA molecule. The central dogma of molecular biology (Figure 2), first published by Francis Crick in 1958, explains the flow of genetic information in a biological system[20]. The classical transfers of genetic information (black arrows in Figure 2) are that DNA is replicated to new DNA, DNA is translated into RNA, and RNA is coding for proteins. In addition to these, there are three additional transfers of information possible that have been shown to occur in nature (grey arrows in Figure 2), RNA replication, where RNA is copied to RNA, reverse transcription where DNA is synthesized from an RNA template, and direct translation where a protein is translated from a DNA template. In RNA the bases are the same as in DNA, except for that RNA contains uracils (Us) rather than Ts, and proteins consist of amino acids.

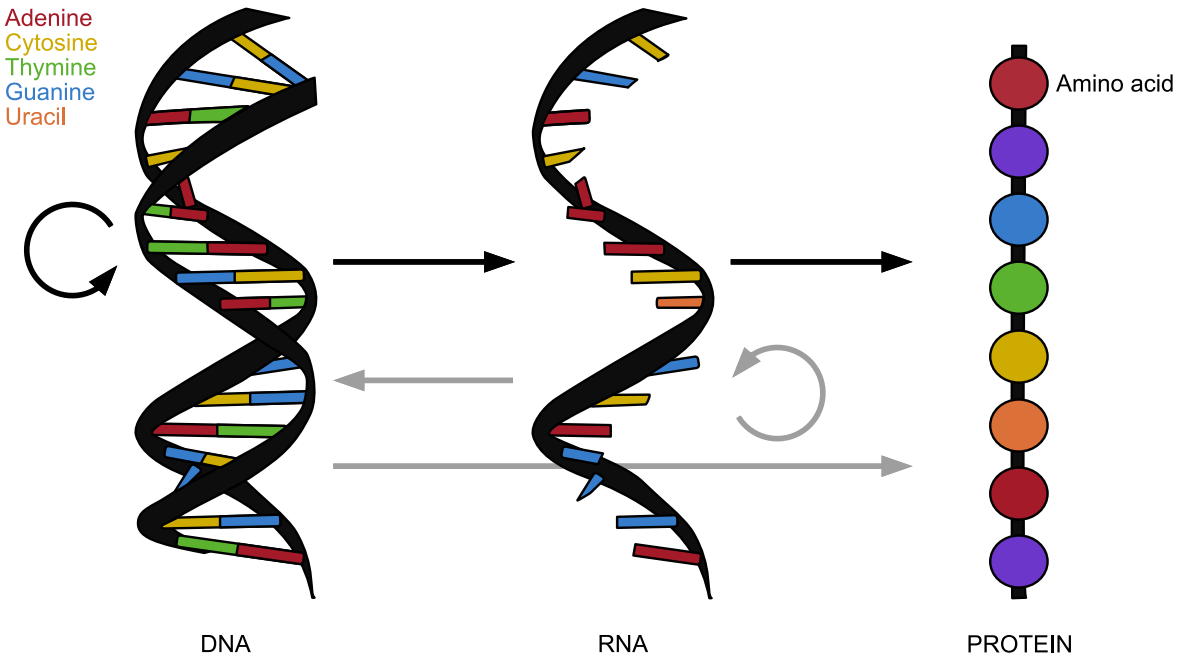

FIGURE 2: THE CENTRAL DOGMA OF MOLECULAR BIOLOGY, WHERE THE CLASSIC TRANSFERS OF GENETIC INFORMATION ARE PICTURED BY BLACK ARROWS AND THE ATYPICAL ROUTES ARE PICTURED BY GREY ARROWS.

2.2.1 THE HUMAN GENOME

Between 1990 and 2003, the Human Genome Project[5,6] was carried out internationally, trying to unveil the base pair code that makes up human DNA and what the genes code for. Ever since the advances of the sequencing technologies took place, humans have been able to study this base pair code and decipher large parts of the DNA code.

The human genome, 3.23 giga base pairs (Gb) long, is contained on 23 pairs of chromosomes. Stretched out, that corresponds to about two meters of DNA in one single cell. In human cells, most of the DNA is found in the nucleus where it is tightly wrapped around histones and bundled up in chromosomes. A small amount of our DNA is found in the mitochondria. The DNA consists of both coding and non-coding regions, where the coding regions are the genes that we today know transcribe for proteins. The non-coding regions, that compose 98-99% of human DNA[21,22], can instead have functional roles, e.g. regulation of gene expression, or functions not yet known.

One of the main goals of human genetics has been to identify the variants in the DNA sequence that are related to the onset and progression of human disease. There are diseases caused by variants in single genes that are inherited, such as Huntington's disease[23]. Other diseases are more complex, where genome wide association studies are needed to identify specific genetic variants across many genomes that can be statistically linked with a disease[24].

However, for almost all human diseases, the risks of an individual developing a certain disease are, to some degree, influenced by genetic variation[7]. Unravelling the connection between the variance in the DNA sequence and disease susceptibility provides a very powerful tool for identifying processes fundamental to disease pathogenesis. Furthermore, the obtained information will help to develop strategies for prevention and treatment. One such example is when a genome wide association study linked the IL-12/IL-23 pathway with development of Crohn's disease[25], that supported clinical trials of drugs targeting that pathway[26].

The variation in the genomic structure can come from alterations, such as mutations or replication errors, leading to a sequence variation of a single nucleotide (also called single nucleotide polymorphism, SNPs). In addition, there are also large scale sequence alterations, such as insertions, deletions, inversions and translocations, that are frequently discovered and are linked to disease[27]. The number of copies of a certain gene, referred to as copy number variation (CNVs), can also vary when comparing human genomes. The sizes of these different human genomic alterations are pictured in Figure 3.

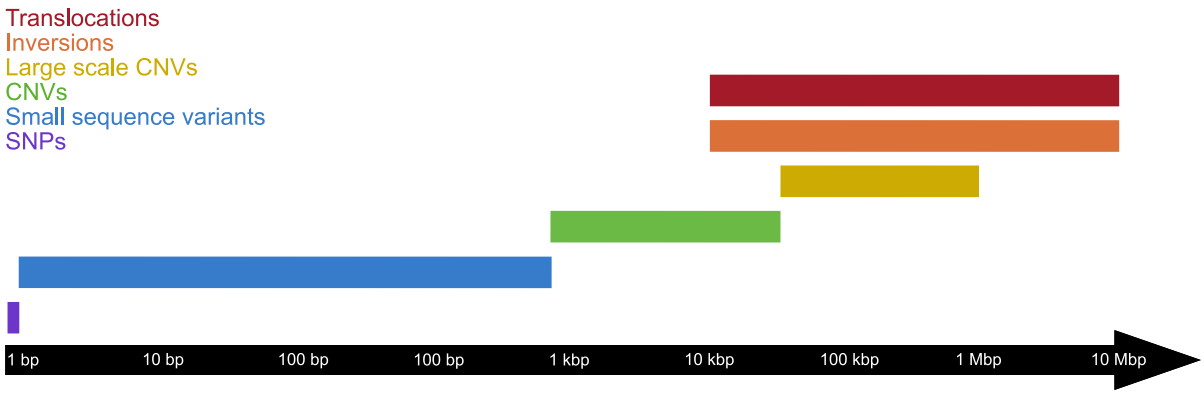

FIGURE 3: SIZES OF HUMAN GENOMIC ALTERATIONS.

Furthermore, it is not only the sequence of the DNA itself that affects what will happen in a cell, there are also epigenetic modifications. Epigenetic modifications include changes to the DNA, such as DNA methylation, that do not change the sequence of the DNA, but still can regulate and turn on or off gene transcription[28]. That DNA methylation had a role in gene regulation was proposed even back in the 1970s[29], and then anormal DNA methylation patterns have been reported for, for example, different types of cancer[28].

DNA DAMAGE AND REPAIR

Our DNA is constantly damaged by both endogenous factors (such as metabolic compounds from cellular metabolism) and exogenous factors, also called environmental factors (such as exposure to UV-light or chemicals). Altogether it has been estimated that the DNA in each human cell is subjected to (on the order of) 10 000 DNA lesions each day[8]. To cope with all of the DNA lesions, the cell has developed a series of smart cellular systems for recognition, signaling and repair of DNA damages. The first DNA repair mechanism, enzymatic photoreactivation[30], was discovered in the late 1940s, even before the discovery that the genetic material was made out of DNA[31]. This mechanism restores two pyrimidine dimers, covalently linked by UV-radiation exposure, to their original nucleotide sequence and structural DNA-form[31]. Since then, many DNA repair mechanisms, as well as other biological responses to DNA damage, have been discovered and studied.

There are various forms of DNA damage, where the different types of damages include base modifications, mismatch of nucleotides, intra- or inter-strand crosslinks, and single or double DNA strand breaks (Figure 4). To cope with these various damages cells have evolved multiple mechanisms to address various types of DNA damage that are called the DNA damage response (DDR). The choice of repair mechanism is largely defined by the type of lesion that is formed, but other factors such as the stage of the cell cycle also play a role[32]. In the DDR, important mechanisms are; mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), single-strand break repair (SSBR), homologous recombination (HR) or non-homologous end-joining (NHEJ)[33].

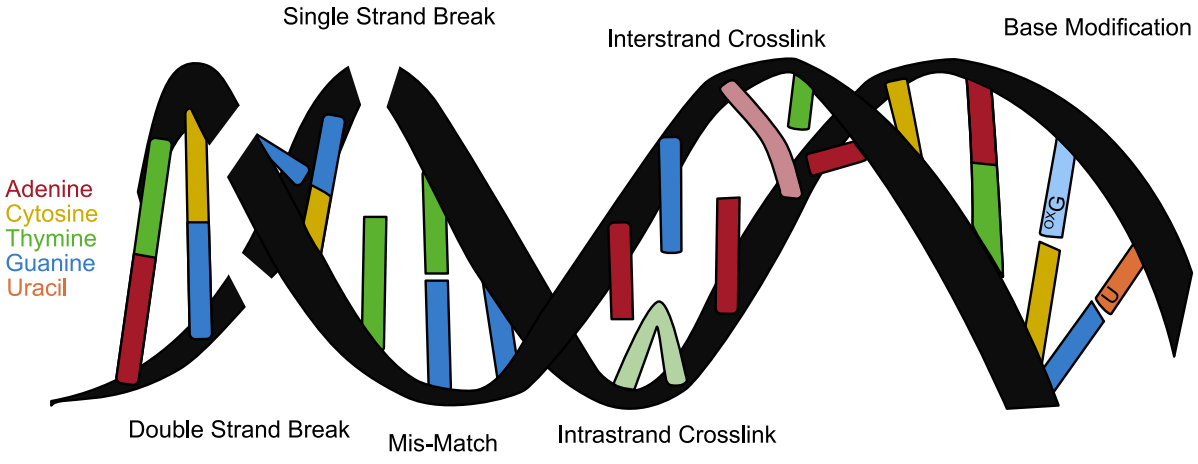

FIGURE 4: VARIOUS FORMS OF DNA DAMAGE EXEMPLIFIED.

Double stranded breaks (DSBs), occurring at a much lower frequency (around 10 each day) than the ssDNA lesions, are repaired by the two major pathways HR and NHEJ[34]. HR relies on recombining DNA ends based on distinct sequence homology[35]. In humans, the sister chromatid is typically required as a template. NHEJ does not require a homologous template, but instead allows joining of any free ends of DNA[34]. NHEJ is typically more error prone but way faster than HR (on the scale that NHEJ is activated to rejoin free ends within a few minutes where HR takes several hours)[36].

Single stranded breaks are repaired by SSBR, where the lesion is acted on by nucleases, DNA polymerases and DNA ligases responsible for processing of the ssDNA ends, filling of the gap and ligation of the DNA backbone[37]. MMR recognizes mismatched DNA bases and replaces them with the correct one by triggering a single-strand incision followed by nucleases, DNA polymerases and DNA ligases acting on the lesion to repair it[38]. Small chemical modification of the DNA bases, such as the oxidation product 8-oxoguanine (^{ox}G) or uracil (U) that is a product after deamination of a C[39], are repaired by BER. In BER, the damaged base is often recognized by a glycosylase enzyme that excises the damaged base, followed by repair by nuclease, DNA polymerase and DNA ligase enzymes[39]. NER is the main pathway for repairing bulky DNA lesions, such as pyrimidine dimers and crosslinks, where it recognizes helix-distorting base lesions. A key aspect of NER is that it excises a 20-30 base oligonucleotide, that then is filled in with nucleotides and ligated by DNA polymerases, associated factors and DNA ligases[27].

Crosslinking agents, such as mitomycin C (MMC)[40], cisplatin[41,42] or 4-hydroperoxy-cyclophosphamide (4HC)[43], have for a long time been used for cancer treatment[44]. They can covalently interact with either one or both strands of the DNA, resulting in different lesions, such as DNA monoadducts, intrastrand crosslinks or interstrand crosslinks (ICLs)[44]. The intrastrand crosslinks, that only involve one of the DNA strands, can be bypassed by some DNA polymerases and are thereby less toxic compared to the ICLs. An ICL is a covalent link between the two DNA strands that prevent separation of the complementary strands of the DNA duplex during replication and transcription. Since ICLs include both DNA strands, the repair process involves multiple the DDR pathways, where HR is thought to collaborate with NER, translesion DNA synthesis and factors responsible for Fanconi anemia for repair[35,45]. Cells that are defective in mechanisms of DDR display an increased sensitivity towards DNA damaging agents, where many of such defects also cause human disease[27].

2.2.2 THE BACTERIAL GENOME

Bacteria are single-cell organisms that belong to the family of prokaryotes, meaning that they do not have membrane enclosed organelles, such as the nucleus. They were one of the first life forms on Earth and have today evolved into a vast variety of bacteria. Bacteria can vary both in shape and size, but are typically a few micrometers in size[46]. The habitats of bacteria vary from extreme environments to living both on us, in us and among us humans. In fact, the number of microbial cells outnumbers the human cells by a factor of ten in the human body[46]. While some bacterial species can contribute positively to human health, some species can invade our body and cause infections that can be life threatening.

An overview of a bacterial cell with its basic components is shown in Figure 5. The cell wall of a bacterium provides structural strength and support to the cell and functions as the separation between the inside of the cell with its cytoplasm and the outside environment, while also providing a semipermeable surface for molecules to pass in and out. Some bacteria also have pili attached on the outside of their cell wall. These pili are hair-like structures that are responsible for both attachment to surfaces, and transfer of DNA between two bacterial cells. The chromosome in prokaryotes is a circular double stranded DNA molecule. The chromosome is compacted with proteins and RNA (responsible for controlling the bacteria's activity and reproduction) in a nucleoid. In addition to the chromosomal DNA, bacteria can also store their genetic information in plasmids, which are smaller, circular DNA molecules. Ribosomes synthesize all the cellular proteins in the cytoplasm of the cell.

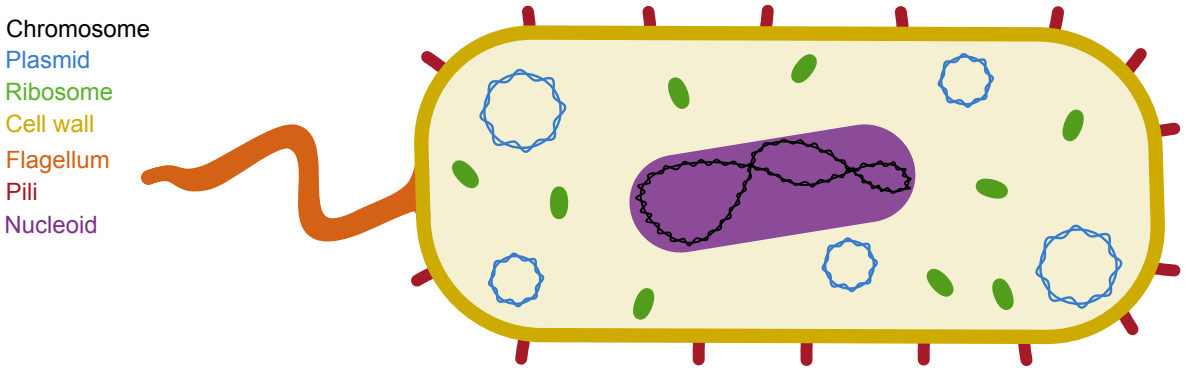

FIGURE 5: AN OVERVIEW OF THE STRUCTURE OF A BACTERIUM AND ITS BASIC COMPONENTS; CHROMOSOME (BLACK), PLASMIDS (BLUE), RIBOSOMES (GREEN), CELL WALL (YELLOW), FLAGELLUM (ORANGE), PILUS (RED) AND NUCLEIOD (PURPLE).

The plasmids are extrachromosomal, double stranded, circular parts of DNA typically on the size range from 2 kb to 1Mb[47] that can exist in copy numbers from one to over a hundred per bacterial cell[48]. Some plasmids can replicate autonomously with the information encoded on the plasmid, meaning that they are replicated and inherited without integration into the host's chromosome. The minimal self-replicating unit of a plasmid is called a replicon, including an origin of replication (ORI), where the replication starts, and proteins that assist in replication and plasmid maintenance. Plasmids are mobile and variable DNA molecules that are shared from one bacterium to another by horizontal gene transfer. The most common mechanism for horizontal gene transfer, between a donor and recipient cell, is conjugation. Conjugation requires that the donor and recipient cells have physical contact via a conjugation pilus, through which the plasmid is transferred[49].

Most bacteria can broadly be classified as either Gram-positive or Gram-negative based on the composition of their cell wall. Due to a thicker cell wall, Gram-negative bacteria are more drug resistant than Gram-positive bacteria, and cause significant morbidity and mortality worldwide[50]. The majority of the bacterial pathogens on the WHO list from 2017 of bacteria where new antibiotics to tackle them are needed urgently are Gram-negatives[51].

CRISPR-CAS9

CRISPR-Cas, a combination of clustered regularly interspaced palindromic repeats (CRISPR) and a Cas (CRISPR associated) protein, constitute an adaptive immune system throughout (~40% of) bacterial and (~90% of) archaeal communities[52,53]. It functions as a "genetic infection memory", where DNA-sequence parts from previous foreign genetic elements (typically bacteriophages or conjugative plasmids) are encoded into CRISPR loci in the host's genome that can be expressed as CRISPR RNAs[53].

CRISPR-Cas9, where Cas9 is an RNA guided DNA endonuclease, has rapidly become an important part of the toolbox of life science research, where it has been used for gene editing, regulation and targeting[54,55]. The CRISPR-Cas9 system (Figure 6) induces a site-specific double stranded cleavage of the DNA target via the guiding RNA sequence on the guide RNA (gRNA). The gRNA consists of a complex of a CRISPR RNA (crRNA), which contains a 20-nucleotide long target-specific sequence responsible for DNA binding, and a trans-activating CRISPR RNA (tracrRNA), responsible for the association to the Cas9 protein. The protospacer adjacent motif (PAM) sequence is a short DNA sequence, 2-6 nucleotides depending on the CRISPR-Cas types in different bacterial species[56], immediately located after the target site in the targeted DNA that is needed for Cas9 cleavage.

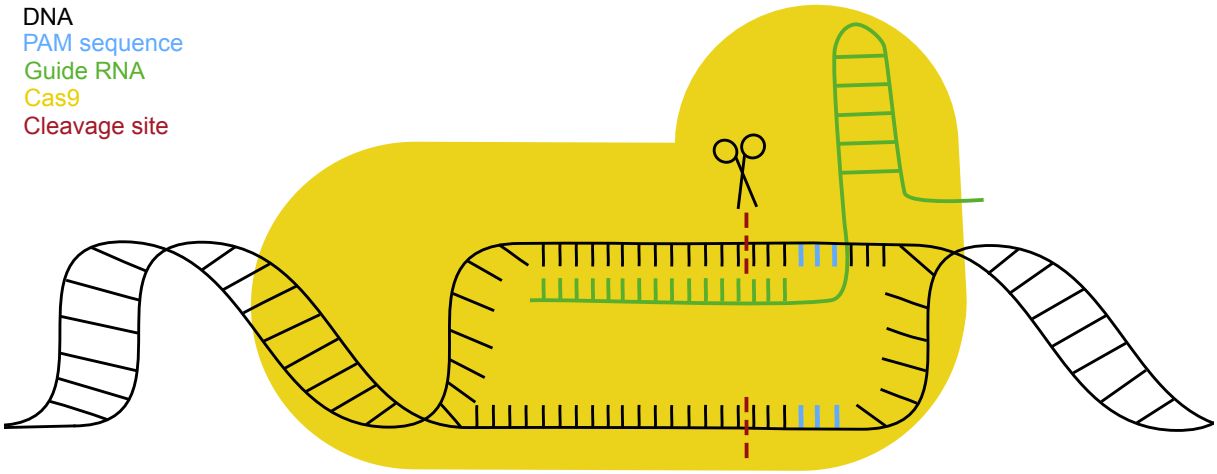

FIGURE 6: CRISPR-CAS9 SYSTEM THAT INDUCES A SITE-SPECIFIC DOUBLE STRAND DNA CLEAVAGE. MAIN COMPONENTS OF THE SYSTEM ARE; TARGET DNA IN BLACK, CLEAVAGE SITE IN RED, ENDONUCLEASE PROTEIN CAS9 IN YELLOW, GUIDE RNA IN GREEN AND PAM SEQUENCE IN BLUE.

ANTIBIOTIC RESISTANCE

Antibiotics are substances that either kill or inhibit the growth of bacteria. They can be derived both naturally from e.g. bacteria and fungi, as well as synthetically. The most recognizable example of an antibiotic drug is penicillin, which is originally naturally produced by *Penicillum* mold. Its antibiotic properties were discovered by Alexander Fleming almost a century ago[9]. Penicillin is a beta-lactamase, which is one of the most common classes of antibiotics, that has been widely used ever since the second world war and still is today.

Antibiotic resistance is the ability of bacteria to withstand the effects of antibiotics. There are different ways that bacteria can resist the effects of antibiotics, classified into intrinsic, adaptive, and acquired resistance. Intrinsic resistance is an inherited characteristic, that does not change over time, that makes the bacterial species unaffected by antibiotics. One example of such a characteristic is the outer membrane that many of the Gram-negative bacteria have, which is non-permeable to certain antibiotics[57]. Adaptive resistance, also called stress response, is a defense system triggered by a sudden change in the environment of the bacteria, for example leading to an overexpression of efflux pumps that leads to a decreased antibiotic concentration within the cell[57]. Finally, acquired resistance is when bacteria acquire resistance via either vertical gene transfer, such as mutations, or by horizontal gene transfer of resistance genes. The resistance genes can be transferred between cells (and species) through genetic mobile elements, such as plasmids or transposons[58]. This further enables fast spread of antibiotic resistance. There are four main mechanisms for acquired resistance (Figure 7). These are: drug target protection or modification, enzymatic inactivation of the drug and decreased accumulation of the drug by either efflux pumps or decreased permeability.

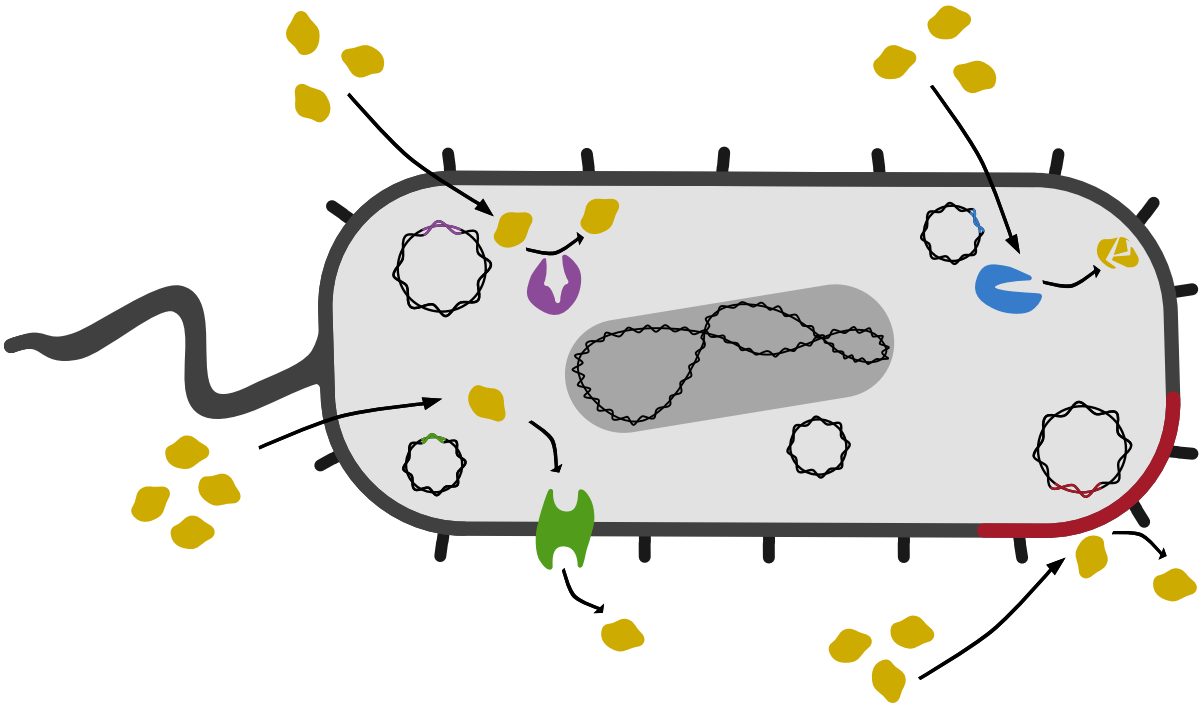

FIGURE 7: FOUR BACTERIAL DEFENCE MECHANISMS FOR ACQUIRED RESISTANCE TO INDIVIDUAL ANTIBIOTICS (OR INDIVIDUAL CLASSES OF ANTIBIOTICS). THE MECHANISMS ARE: DRUG TARGET PROTECTION OR MODIFICATION (PURPLE), ENZYMATIC INACTIVATION OF DRUG (BLUE), DECREASED ACCUMULATION OF DRUG BY EFFLUX PUMPS (GREEN), AND DECREASED ACCUMULATION OF DRUGS BY DECREASED PERMEABILITY (RED). AN ANTIBIOTIC DRUG IS PICTURED IN YELLOW.

Since the discovery of antibiotics in the early 20th century, they have not only saved people from dying of infectious diseases, but also become a cornerstone in our healthcare, in many cases enabling modern medicine and surgery. Most of the antibiotics discovered in the past 60 years have been results from natural production and only a few have been synthetically constructed[59]. There are several different classes of antibiotics, but since the end of the golden era (typically between 1940 and 1962), where most of the antibiotic classes we use as medicines today were discovered and introduced to the market, only two new classes have been introduced. Antibiotic resistance is a globally emerging challenge, and it is considered one of the largest threats to mankind by the World Health Organization. In the year of 2050, 10 million people are estimated to die from infections by antibiotic resistant bacteria[10]. A historical timeline of important events for antibiotic discovery and resistance development is pictured in Figure 8.

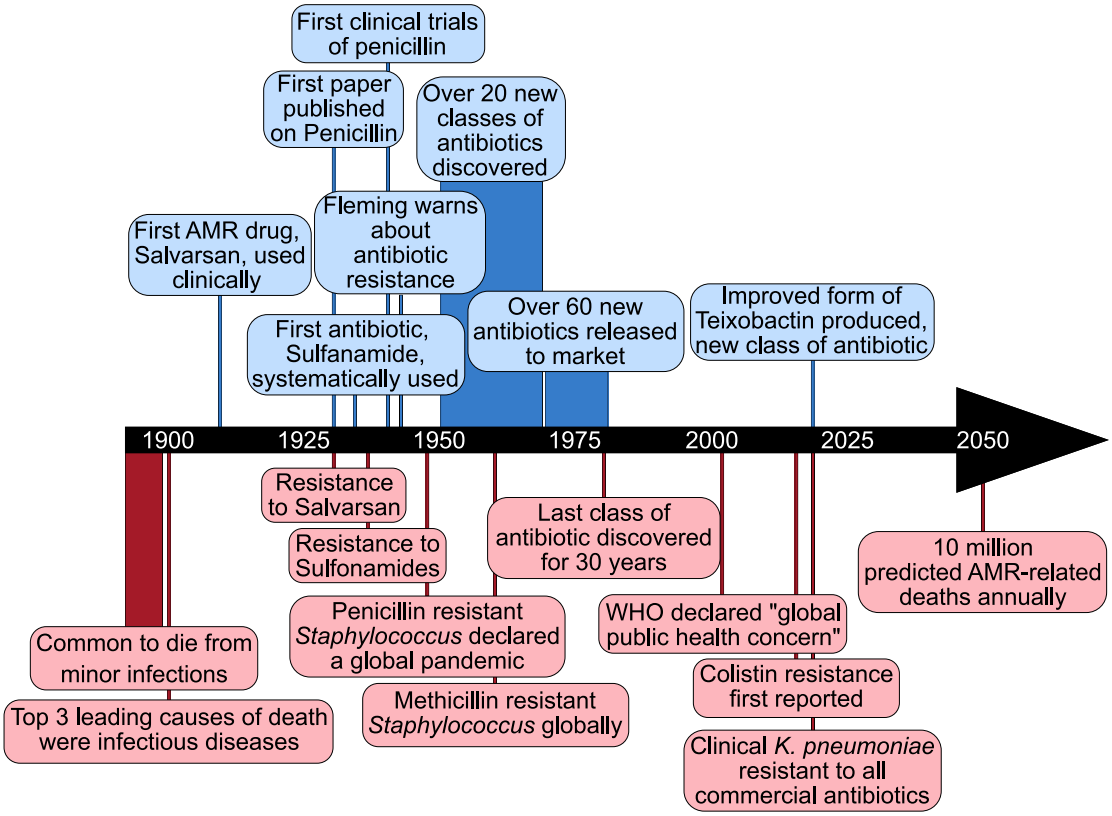

FIGURE 8: HISTORICAL TIMELINE OF IMPORTANT EVENTS FOR ANTIBIOTIC DISCOVERY AND RESISTANCE DEVELOPMENT.

2.3 PHYSICAL PROPERTIES OF DNA

The physical properties of the DNA double helix are unlike those of any other natural or synthetic polymer[60]. The phosphates in the backbone make it one of the most highly charged polymers known and the characteristic stacking of its bases together with the double-helical shape makes it unusually stiff. The energy that it takes to bend a molecule of dsDNA into a circle compared to if it was ssDNA, is about 50 times more[60].

Many length scales play a role when it comes to understanding the thermodynamics of DNA in confinement. Conformations of DNA can be characterized by length scales such as the contour length (L_C), that is the total length of the DNA backbone, and the persistence length (L_P) that provides a measure of its rigidity[18] (Figure 9A). A way to think about L_P is that it characterizes the ability of the DNA to be bent by thermal energy. The L_P of a DNA molecule is influenced by its sequence, by the base pair rigidity and the stacking interactions of the base pairs. A/T rich sequences have a lower persistence length and are therefore more bendable than G/C rich sequences[61]. A purine-pyrimidine base step has a larger stacking area (and is thereby thermodynamically more stable) than purine-purine or pyrimidine-pyrimidine base steps, with the pyrimidine-purine being the least stable[18]. The L_P is also affected by the ionic strength, as

the ions in solution screen the charges along the DNA backbone[62,63]. The persistence length of B-DNA under physiological conditions is approximately 50 nm[64,65].

Free in solution, a DNA polymer will coil to minimize its free energy. The size of the coil can be described by the radius of gyration (R_G). R_G (in Figure 9B) quantifies the root-mean-squared distance between parts of the coil and its center of mass[66].

A channel is often defined by its dimensions, where the channels used for the work in this Thesis are assumed to be infinitely long, but with a restricted width (D_W) and height (D_H). The degree of extension in a channel will depend on the average cross-section (of D_W and D_H) of the channel, D, and the L_P and R_G of the polymer. If D is larger than R_G, the polymer will behave as if it is free in solution (Figure 9C). If D is smaller than R_G, but still larger than L_P, the polymer will be weekly confined and act as a series of blobs, as explained by deGennes[67,68] (Figure 9C). The principle is that the polymer will divide itself into blobs, with diameter D, that are not interacting with each other[64,68]. If D is smaller than L_P, the polymer cannot coil up anymore and will only undulate gently, and will behave according to Odjik[69,70] (Figure 9C). For experiments on DNA molecules confined in nanochannels, such as the ones in this Thesis, the D is often similar to the L_P. This results in a regime between the ones described by deGennes and Odijk, often referred to as the extended deGennes regime[71,72]. In both the deGennes and Odijk regimes, the extension of the DNA is directly proportional to L_C, resulting in that observed features along the contour of the molecule, extended under these regimes, will be directly corresponding to the position in the sequence.

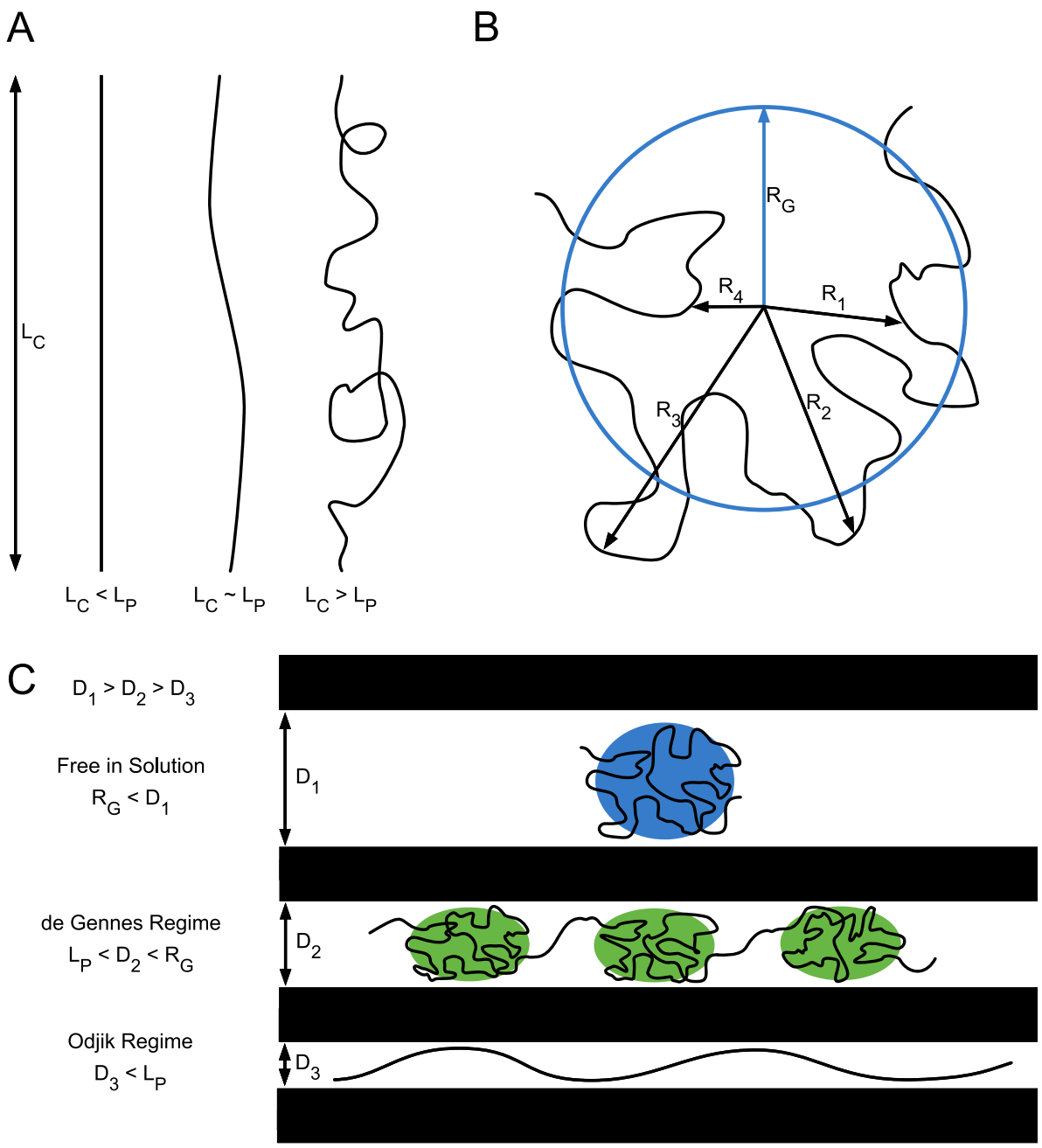

FIGURE 9: PERSISTANCE LENGTH (L_P) OF A POLYMER WITH COUNTOUR LENGTH L_C (A). RADIUS OF GYRATION (R_G) OF A POLYMER (B). THE BEHAVIOR OF A POLYMER UNDER DIFFERENT DEGREES OF CONFINEMENT IN A CHANNEL (C).

3 FLUORESCENCE

3.1 ABSORPTION AND EMISSION OF LIGHT

Light, either visible to the human eye or not, can be described as electromagnetic radiation that propagates in a direction as a waveform. The energy of light is quantized, where the smallest energy unit quantized is called a photon. A photon is a massless and chargeless elementary particle best explained by quantum mechanics that exhibits wave-particle duality, meaning that the behavior of light contains both wave- and particle properties. The type of electromagnetic radiation is usually defined by the wavelength, where light visible to the human eye lies in a range of wavelengths typically between 400-700 nm[73].

The interaction between light and matter will depend on the electronic configurations of the atoms in a molecule. Molecules are atoms held together by chemical bonds. Atoms are, as proposed by Niels Bohr, consisting of a positively charged nucleus and negatively charged electrons that revolves in orbits around the nucleus[74]. This model was further developed to include the atomic orbitals, describing where the electrons orbiting the nucleus can be found by a probability distribution function, where the electron orbits are quantized around the nucleus with energies depending on their angular momentum. For molecules, the orbitals in which the individual electrons orbit are associated with the entire molecule. These orbitals are also quantized and provide different electronic energy levels which the molecules can reside in. The chemical bonds in a molecule can also contain energy by vibrating and rotating, resulting in three main ways that energy levels within a molecule can be expressed: electronic, vibrational, and rotational.

A molecule in a sample can absorb the energy of a photon by a process called absorption, exciting the molecule to a higher molecular state. For a molecule to be excited to a higher state, the energy of the photon needs to match the energy gap between the higher state (S_n) and the ground state of the molecule (S_0). The energy levels and possible transitions between them when a molecule is photoexcited and de-activated can be visualized in a Jablonski diagram, exemplified in Figure 10.

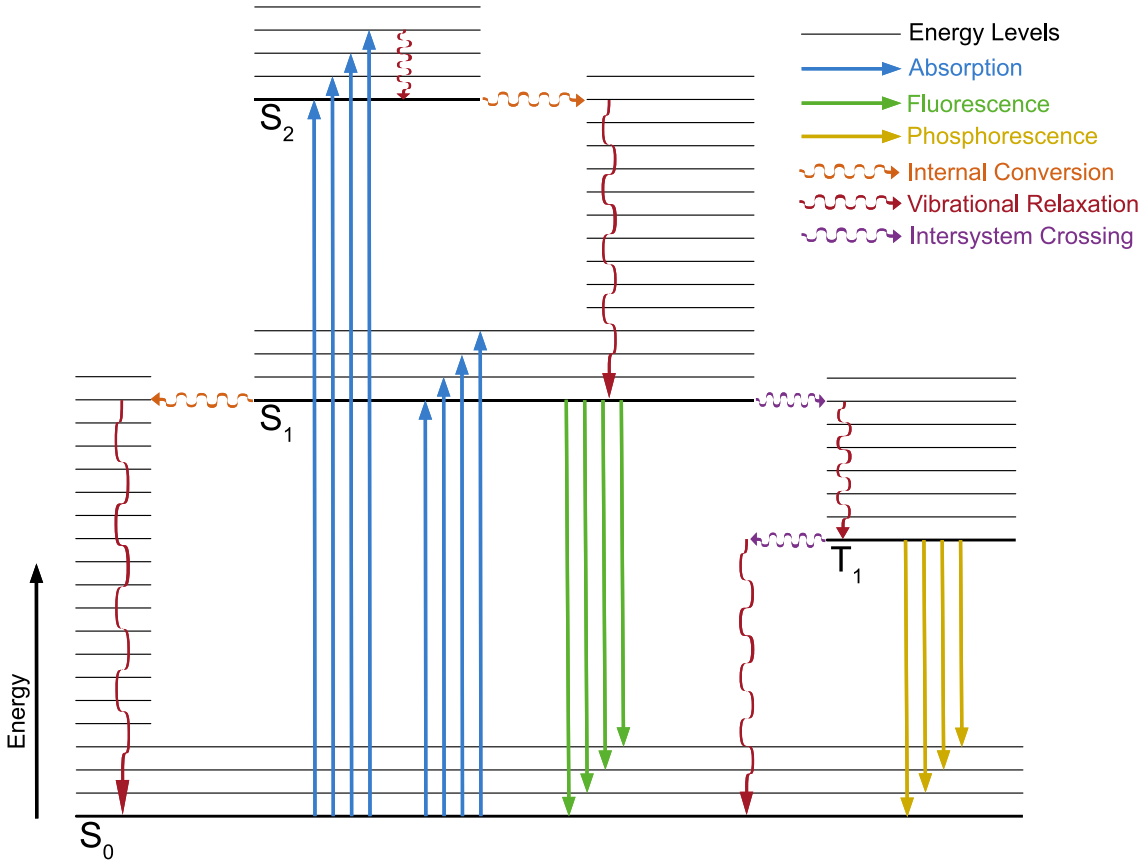

FIGURE 10: A JABLONSKI DIAGRAM WITH THE ENERGY LEVELS AND THE ELECTRONIC TRANSITIONS BETWEEN THEM. SINGLET STATES (S_0, S_1, S_2), A TRIPLET STATE (T_1), AND THE CORRESPONDING VIBRATIONAL STATES ARE PICTURED SHOWN AS HORIZONTAL LINES. THE RADIATIVE PROCESSES, ABSORPTION (BLUE), FLUORESCENCE (GREEN), AND PHOSPHORESCENCE (YELLOW) ARE SHOWN AS VERTICAL ARROWS. THE NON-RADIATIVE PROCESSES INTERNAL CONVERSSION (ORANGE), VIBRATIONAL RELAXATION (RED) AND INTERSYSTEM CROSSING (PURPLE) ARE SHOWN AS WAVY ARROWS.

When a molecule is excited to a higher molecular state, the extra energy added to the system will lead to the reconfiguration of the electrons in the molecule. The excited singlet states (S_n), with paired electrons, and triplet states (T_n), with unpaired electrons, are numbered relative to the ground state (S_0) based on the energy gaps. In addition to the electronic states, there are also several vibrational states related to each electronic state, that can be subdivided further into

rotational states. The difference in energy between these states are for the electronic states typically in the orders of 10^0 eV, where the corresponding number are smaller for the vibrational states (10^{-2} eV) and the rotational states (10^{-4} eV)[75].

As the molecule returns from an excited state to its ground state, excess energy can be dissipated through various processes. These de-activations can be either radiative, by fluorescence or phosphorescence, or non-radiative, by internal conversion, vibrational relaxation, or intersystem crossing. Following absorption, the excited molecule will typically enter a higher vibrational state (V_n) of the excited singlet states (S_1 or S_2). Then, the molecule will relax to the lowest excited single state (S_1) via a combination of vibrational relaxation and internal conversion. The following relaxation, can be assumed by Kasha's rule to only occur from the lowest vibrational level of S_1[76].

Once the molecule is in the lowest vibrational level of S_1, there are three main pathways to return to the ground state: by non-radiative relaxation, fluorescence, or phosphorescence via T_1. The dominating pathway to return to ground state are determined by the structure of the molecule, the surrounding environment, and the temperature. By non-radiative relaxation, the molecule undergoes internal conversion from S_1 to S_0 and then by vibrational relaxation to the lowest level of S_0 (Figure 10), transferring the energy to the surrounding environment. By fluorescence, the molecule relaxes from S_1 to S_0 (Figure 10) while emitting a photon with a wavelength corresponding to the energy difference. Another way for the molecule to relax from S1 is if it undergoes intersystem crossing from S_1 to T_1, with a spin transition for the excited electron, resulting in two electrons with the same spin. The molecule can relax from T_1 to S_0 either by non-radiative intersystem crossing and vibrational relaxation or by phosphorescence (Figure 10) while emitting a photon with a wavelength relative to the energy difference and reversing the spin of the excited electron.

The wavelength of an emitted photon is longer than the wavelength of the photon absorbed during excitation. For fluorescence, this shift in energy is called a Stokes shift, usually given as the difference between the maxima of the absorption and emission spectra[77]. This energy difference is mainly due to energy loss through internal conversion and that the molecule often decays to an excited vibrational level of S_0 during the emission process.

3.2 FLUORESCENT DNA LABELS

Fluorescence has become a primary research tool in both chemistry and biology for sensing applications and visualization. Fluorophores, fluorescent compounds, are divided into intrinsic and extrinsic ones. The intrinsic fluorophores are the "naturally" occurring ones, such as the amino acids tryptophan, tyrosine and phenylalanine[78] or chlorophyll in plants[77]. The extrinsic fluorophores are added to a sample to either add fluorescence properties when there are none, or to change the spectral properties of the original sample. Importantly, labeling a sample with extrinsic fluorophores might affect the physical properties of the molecule of interest[77,78]. The use of fluorophores has transformed both the way that we study biological systems and our understanding at the molecular level.

There are different properties that are desirable for a fluorophore to have depending on the application. In general, there are three main properties that are important for fluorophores in fluorescence spectroscopy and microscopy: fluorescence quantum yield (Φ_f), molar absorptivity (ε), and fluorescence lifetime (τ_f). The brightness of a fluorophore is governed by two factors, Φ_f and ε. At each wavelength, ε governs the ability of a molecule to absorb light, and thereby the number of photons absorbed by a sample. Φ_f instead governs the fraction of absorbed photons that will result in an emission event, the number of emitted photons. The time τ_f, is the time a fluorophore spends in the excited state before emitting a photon and returning to the ground state. A large Stokes shift is useful in the above-mentioned applications since it makes it possible to distinguish the incoming excitation light from the outgoing signal and helps to avoid self-quenching, that would otherwise decrease the outgoing signal.

Fluorescent probes are used for lighting up various compounds that normally are not emitting light in the visible region. Today a very large number of fluorophores exist, they are very customizable since fluorescent probes can be tailored to absorb and emit at almost any wavelength[79]. Nucleotides and nucleic acids are generally nonfluorescent at room temperature in aqueous solvents (fluorescence quantum yield, $\Phi_f<10^{-4}$)[80], making the binding of extrinsic fluorophores crucial for enabling fluorescence studies of them. YOYO-1 for example, that has been used in all projects of this thesis, is a bis-intercalator that binds to dsDNA without sequence specificity and that displays a fluorescence quantum yield of 0.52 when intercalated in dsDNA[81]. YOYO-1 displays a molar absorptivity of 10^5 $M^{-1}cm^{-1}$ at the absorption maximum 491 nm and has a emission maximum at 509 nm[81,82]. Furthermore, another advantageous property of YOYO-1 is that its emission is increased over 1000-fold when intercalated to dsDNA compared to free in aqueous solution[81].

The capability to image more than one type of molecule of interest, and thereby see potential colocalization in the same sample, is a very powerful addition for biological studies. One part of the DNA molecule that can be of interest to image in addition to the whole molecule, are nucleotides. There is a variety of fluorescent nucleotides and nucleobases, e.g. the fluorescent nucleotide aminoallyll-dUTP-Atto-647N (Atto-647N) from Jena Bioscience.

3.3 FLUORESCENCE MICROSCOPY

The human eye can typically resolve an object of the order 200 μm[83]. To use optical lenses for magnification of things that cannot be seen with the naked eye has been done for hundreds of years, and lays the basis for the field of microscopy[84]. Today, the microscopes have been developed to a point where our observations of the microscopic world are only limited by the laws of physics, and more specifically by the diffraction of light.

When a point emitter is imaged by microscopy, its image will be blurred due to diffraction. This pattern in the focal plane is called the point spread function (PSF). The radius of the PSF, as defined by the Abbe diffraction limit, is given by $\frac{\lambda}{2 \cdot NA}$, where λ is the wavelength of the emitted light and NA is the numeric aperture of the objective's lens[85]. The generally accepted criteria for the optical resolution between two PSFs of a microscope without aberrations is given by the Rayleigh criterion[86]. The Rayleigh criterion states that the minimum resolvable separation between two points is given by $0.61 \cdot \frac{\lambda}{NA}$.

Fluorescence microscopy relies on emission to visualize the sample and for samples that do not have fluorescence properties fluorescent probes can be used. In Figure 11 there is a schematic of the main components of an epi-fluorescence microscope, that has been used in all projects of this Thesis.

There are six main components of a fluorescence microscope (Figure 11), a light source, two filters, a dichroic mirror, an objective, and a detector. A light source of some sort is emitting either monochromatic light, such as a laser or an LED, or broad-spectrum light, such as a xenon arc lamp. A filter cube, matching the requirements of the fluorophore of interest, usually consists of an excitation filter, a dichroic mirror, and an emission filter. The excitation filter is blocking the wavelengths from the light source that are not matching the excitation spectrum of the fluorophore. The dichroic mirror reflects the excitation light to the sample at the same time as it transmits the emitted fluorescence from the sample to the detector. The emission filter then allows only the fluorescence from the sample to reach the detector, blocking unwanted traces of excitation light due to backscattering. In the epi-fluorescence microscope, the sample holder is placed on top of the microscope and imaged from below.

In short, the light from a light source is filtered through an excitation filter when entering the filter cube, reflects on the dichroic mirror, goes through an objective, and is focused on the sample. The light from the sample then goes through the objective, is transmitted through the dichroic mirror, filtered through the emission filter and to a detector recording the emitted fluorescent light.

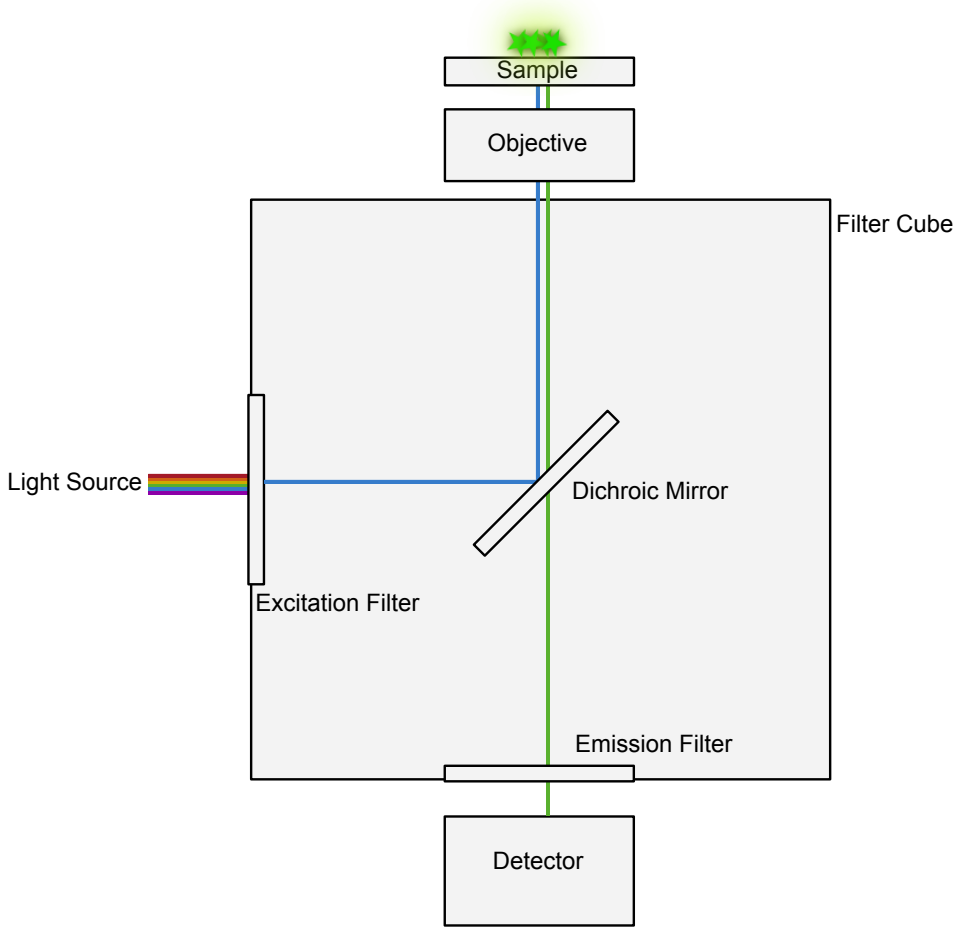

FIGURE 11: A SCHEMATIC OF THE MAIN COMPONENTS OF AN EPI-FLUORESCENE MICROSCOPE.

---

4 UNVEILING GENETIC INFORMATION

4.1 SEQUENCING

The first significant breakthrough in DNA sequencing came in 1977 when Frederik Sanger and Alan R. Coulson published two papers on methods to determine DNA sequences[87,88]. They improved the two earlier DNA sequencing techniques developed by Maxam and Gilbert[89] published in the same year, and Sanger and Coulson's own plus and minus method[90] published 2 years earlier. This would later go on to transform the field of biology as it was the start to providing the DNA code with base-by-base resolution. What later was called Sanger sequencing became more or less the only DNA sequencing method used for the following 30 years. Sanger sequencing used dye termination for sequencing chemistry and today has a very long read length[91,92]. The main advantage of this method that made it the standard until the introduction of Next generation sequencing (NGS), often also referred to as second-generation DNA sequencing, in 2005, was the reduced handling of toxic chemicals and radioisotopes[91,93].

Even if multiple different strategies for NGS exist, such as the systems of Roche/454, Illumina/Solexa, Life/APG and Helicos BioSciences, most of them rely on the same basic principles[94]. These principles include; fragmentation of the DNA into smaller pieces (typically ten to a couple of hundred base pairs in length[95]), PCR-amplification, sequencing of the amplified fragments and data analysis. To achieve a full DNA sequence, the sequences of the fragments, also called reads, need to be assembled. The idea involves random fragmentation and generation of enough copies that can be sequenced, such that there will be an overlap of the sequenced ends, which in turn can be assembled correctly in the analysis.

What was revolutionary in 2005 was the introduction of the sequencing-by-synthesis technology with reduced reaction volume at the same time as the number of sequencing reactions was dramatically increased, leading to decreased costs and more produced data at a shorter time[91]. The third-generation sequencing systems, such as PacBio and Oxford Nanopore,

are instead based on single-molecule analysis with a longer read length (10-25 kb[96] and >100 kb[97], including reported ultra-long reads >2Mbp,[98] respectively) and often have a shorter run time, enabling sequence results in a matter of hours instead of days[91,92].

To be sure of the sequencing results, a single nucleotide should be sequenced multiple times, called coverage or depth. This requires multiple copies of each genome to be generated and read to assemble a sequence of that genome with high accuracy. The sequencing systems are differently prone to errors, requiring that the results are averaged over the coverage in the end. This removes the possibility to unveil potential heterogeneity between cells.

4.2 OPTICAL DNA MAPPING

Optical DNA mapping (ODM) is an umbrella term for methods based on the use of fluorescent dyes and fluorescence microcopy to visualize long-range sequence information of intact DNA molecules[99-101]. The main principle behind ODM is that DNA is labeled in a sequence specific manner, stretched out and imaged. Typically, the read length of the DNA that the sequence information is recorded at is up to several Mbp with approximately 1 kbp resolution[100,101]. One of the main advantages of ODM is the long read length in combination with single molecules being imaged. By using ODM, variations at the single molecule level can be unveiled and averaging across a population of molecules avoided.

4.2.1 STRETCHING STRATEGIES

There are multiple different strategies for stretching individual DNA molecules, including magnetic tweezers[102,103], optical tweezers[104], molecular combing[105,106] and confinement in nanofluidic channels[66]. For ODM, the two most used strategies to obtain information about individual DNA molecules and to be able to record the information along the molecule are by either stretching the DNA molecules on activated surfaces or by stretching the DNA molecules by confinement in a nanochannel. Both these stretching strategies have been used in the projects of this thesis.

STRETCHING ON GLASS

Stretching DNA molecules on an activated surface is for example used in what is called DNA combing or molecular combing, where the ends of stretched out DNA molecules can attach to a functionalized glass surface using a moving air-solution interface[106-108]. There are two important elements that enable the deposition of DNA molecules onto a surface: that the surface is capable of spontaneous DNA capture and a mechanism to extend the DNA molecules. Typically, DNA is deposited onto either a surface with a net positive charge (that attracts the negatively charged DNA backbone) or onto a surface modified with a hydrophobic compound (typically using silane chemistry)[109]. In presence of a flow, the forces at the air-solution

interface will stretch the DNA molecules in the flow direction. The DNA molecules approach full extension as the shear flowrate is increased[110]. Reported stretching on hydrophobic surfaces has been thought to rely on the fraying ends of the DNA to attach to the surface[111], but the binding of circular DNA molecules is also possible[112,113].

A variety of different approaches have been developed to achieve a uniform flow, including retracting a glass slide from a solution[106,114], dragging a droplet over the surface[115], using the gravitational force to drag a droplet over a tilted surface[116], droplet evaporation directly on a surface[117], but a uniform stretching is not guaranteed. In Figure 12, the stretching of DNA molecules by shear-induced stretching force upon pipetting on the corner of an activated coverslip, here a mix of vinyl (hydrophobic) and amino (hydrophilic and positively charged) groups, and an un-modified microscopy slide, is pictured.

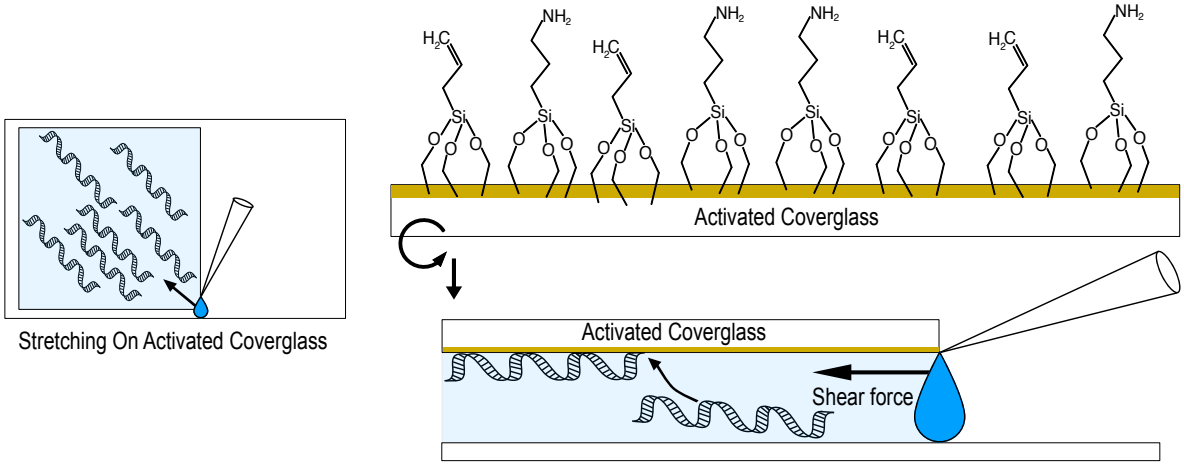

FIGURE 12: ACTIVATION OF GLASS COVER SLIDES IS DONE WITH A MIXTURE OF SILANES WITH AMINE AND VINYL TERMINAL GROUPS. A DNA-CONTAINING DROP IS PIPETTED ON THE CORNER OF AN ACTIVATED GLASS COVER SLIDE ON A MICROSCOPY SLIDE AND STRETCHED BY SHEAR FORCE.

STRETCHING IN NANOFLUIDIC CHANNELS

The other commonly used method for stretching DNA molecules in ODM is by confining them in nanochannels. The confinement of DNA in narrow channels will allow the DNA to attain an extended conformation due to the spatial restrictions[118] and it has been demonstrated that single DNA molecules of genomic length can be uniformly stretched in nanofluidic channels with ~100 nm channel dimensions[119]. The most reliable and used nanofluidic devices are generally fabricated from oxidized silicon (also called silica). For DNA analysis, the negative charge and low surface roughness of the silica prevents adhesion to the channel walls and the hydrophilic nature allows wetting once the buffer is added[100].

A sketch of the nanofluidic device used for the work in this thesis is pictured in Figure 13A and an image of a nanofluidic chip mounted on a holder in Figure 13B. In the nanofluidic devices used, there are hundreds of nanochannels spanning the microchannels, each nanochannel being 500 μm long with a cross section of 100×150 nm^2 (height × width). The chip was fabricated on silicon wafers with traditional semiconductor fabrication processes, such as thermal oxidation, photolithography, E-beam lithography, reactive ion etching and fusion bonding, in a cleanroom[100]. The mounting of the chip on a holder is used for support towards the microscope, placing the chip in contact with the objective and provides access to the loading wells that can be pressurized with nitrogen driven flow.

There are different chip designs that have been developed throughout the years for different applications[100]. Some of these designs include funnel-shaped channels[120], or channels with the width increased in discrete steps along the length (that allows the possibility to study an individual DNA molecule at different degrees of confinement[121]), pre-stretching of the DNA before entering the channels with physical barriers such as pillar structures[122,123], meandering channels[124] for including ultra-long DNA molecules within the field of view of a camera and devices for dynamic studies[125,126] allowing for example exchange of buffers. There are various ways of controlling the DNA within the chip, such as applying a pressure-driven flow or an electrophoretic field[100].

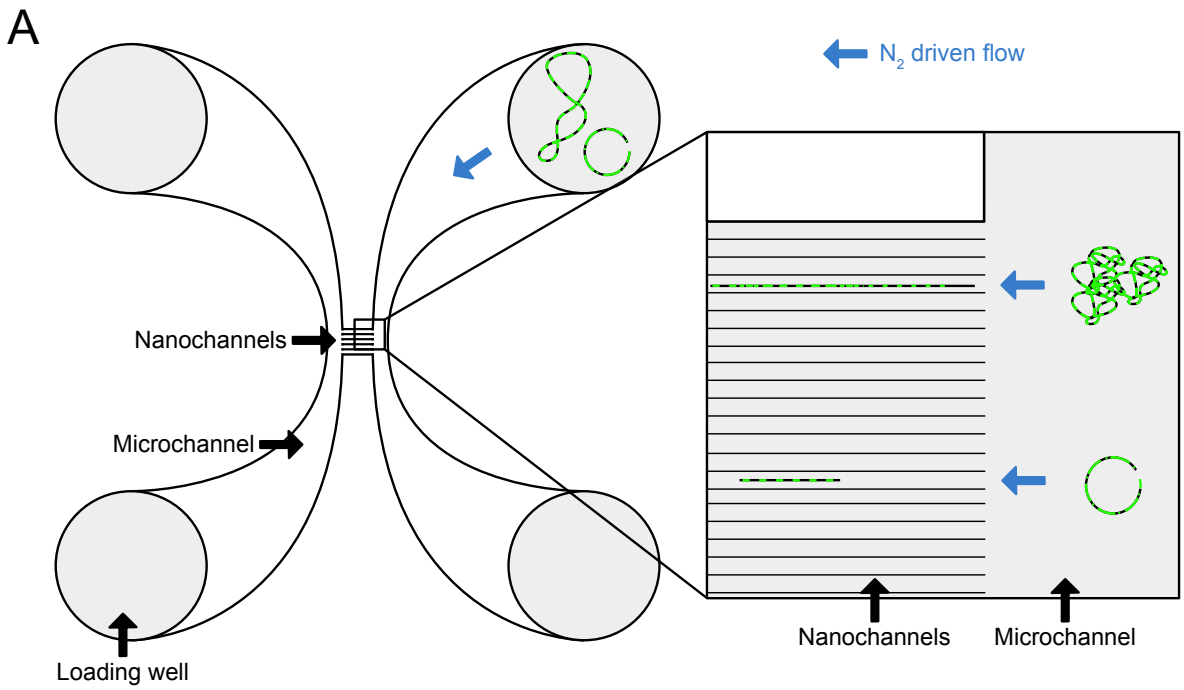

FIGURE 13: IN A, THERE IS AN OVERVIEW OF A NANOFLUIDIC CHIP WITH ITS MAIN FEATURES FOR STRECTHING OF DNA (NOT DRAWN TO SCALE). THE MAIN FEATURES ARE THE NANOCHANNELS SPANNING THE TWO MICROCHANNELS, THAT IN TURN ARE CONNECTING TWO AND TWO OF THE LOADING WELLS THAT CAN BE PRESSURED WITH NITROGEN-DRIVEN FLOW. IN B, THERE IS AN IMAGE OF A NANOFLUIDIC CHIP MOUNTED ON A HOLDER.

4.2.2 LABELING STRATEGIES

In the first optical mapping experiment in Schwartz laboratory, optical DNA mapping with restriction enzymes was used to generate sequence-specific DSBs on DNA molecules[108]. Multiple strategies for labeling the information along the DNA molecule have been developed since the 1990s. Today, the labeling strategies can be divided into two main categories, enzyme-based labeling, and affinity-based labeling (Figure 14).

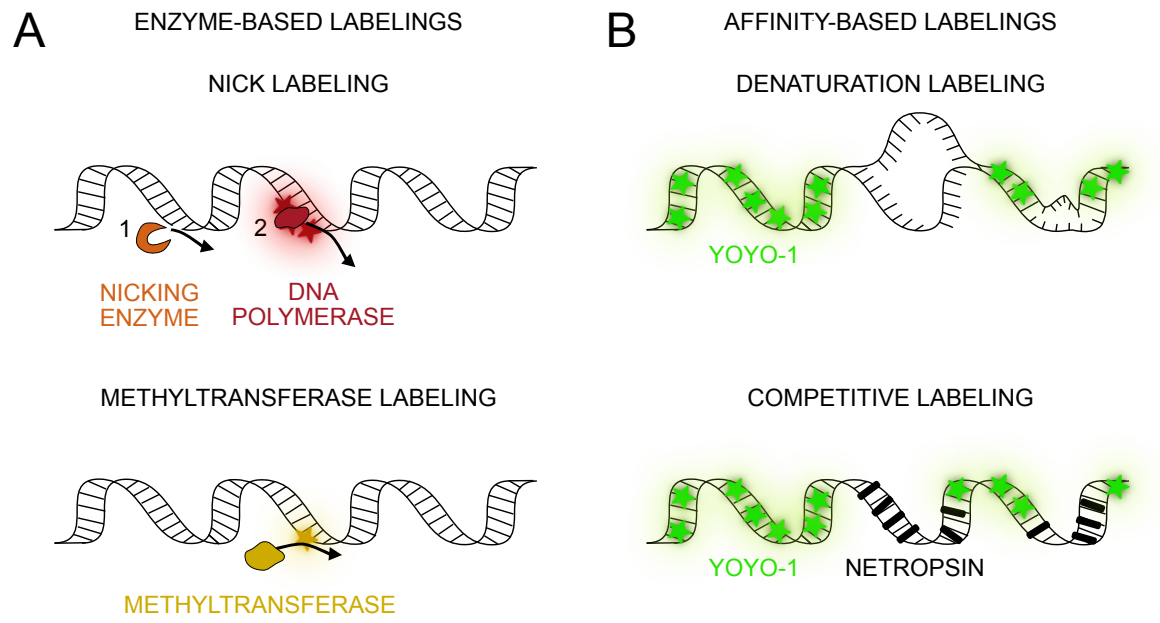

FIGURE 14: THE TWO MAIN TYPES OF LABELINGS IN OPTICAL DNA MAPPING, ENZYME-BASED LABELING (A) AND AFFINITY-BASED LABELING (B).

ENZYME BASED LABELING

Sequence specific labeling of DNA molecules for ODM using enzymes can be achieved in different ways. Nick labeling (red in Figure 14A) uses a nicking enzyme to create a nick in the DNA molecule at a sequence specific site, typically 4-7 bp in size that is specific for the nicking enzyme used[100,127]. Then a DNA polymerase repairs the nick and incorporates fluorescently labeled nucleotides. Another fluorescent DNA binding molecule, such as YOYO-1, can then be used to determine the DNA contour length. Two extensions, nick-flap labeling[128] and dual nick labeling[129], have enabled dual color labeling. To address the otherwise limited specificity of the traditional nicking enzymes, a highly specific Cas9-based nick-labeling strategy have been developed, that provides a designable 20 bp recognition site[130].

The same principle as for the nick labeling has been extended for labeling of DNA lesions. A combination of different bacterial enzymatic enzymes was used for recognition and excision of certain types of DNA damage lesions, prior to incorporation of fluorescent nucleotides in the damage sites by a DNA polymerase as in the nick labeling[131].

Another way to introduce sequence specific labeling is by methyltransferase-based labeling[132,133] (yellow in Figure 14A). A methyltransferase attaches a synthetic co-factor to the DNA molecule at a recognition site, typically 4-6 bp, and that co-factor can be fluorescently labeled. Methyltransferase labeling have been used to generate densely labeled DNA profiles[134-136], called amplitude modulation profiles or DNA fluorocodes, with nanometer resolution[135] (less than 20 bases) for single-molecule characterization of whole genomes.

In addition to the sequence specific information of the genome, there are multiple layers of information that extend beyond the genetic sequence, such as epi-genetic modifications and DNA damage. Epi-genetic modifications such as methylation[137,138] and hydroxy-methylations[139-141] have been successfully labeled and characterized.

AFFINITY BASED LABELING

Sequence specific DNA labeling for ODM can also be achieved without the use of enzymes, bypassing the limited availability of enzymes and corresponding recognition sites, where the other commonly used labeling strategy for ODM is affinity-based labeling. Instead of enzymes, affinity-based labeling relies on the difference between the AT or GC base pairs. Denaturation mapping[142] (top in Figure 14B) uses the difference in melting temperature between the base pairs, as the GC base pairs are connected by three hydrogen bonds and the AT base pairs are connected by two hydrogen bonds, to create a sequence specific emission profile along the DNA molecule. A fluorescent DNA binding molecule, such as YOYO-1, can be used to stain the whole DNA molecule, but as the temperature is carefully raised to a level where only the AT base pairs melt, the YOYO-1 (with a preference for dsDNA) will dissociate at these locations. Another affinity-based labeling for ODM that does not require precise temperature control is competitive binding[143] (below in Figure 14B). In competitive binding, a DNA-binding non-fluorescent molecule, such as netropsin, with a specific affinity for binding to the AT base pairs, blocks the binding of the fluorescent non sequence specific DNA-binding molecule YOYO-1 at these regions. Added in a reasonable range of ratios, the netropsin will block many of the AT regions while YOYO-1 can bind to the GC regions, resulting in a sequence specific emission profile along the DNA molecule.

The different types of labeling have also been combined, obtaining multiple layers of information simultaneously. Two examples are the simultaneous detection of genetic and epi-genetic information for detection of copy number variations and the methylation pattern of individual DNA molecules[138] and simultaneous detection of genetic information and DNA damage by mapping the DNA damage sites induced by the cytotoxic drug etoposide to the human genome[144].

In 2013, Bionano Genomics® commercialized nick labeling in combination with nanofluidic channel-based DNA extension and in 2018 a new labeling concept, direct label and stain, based on a non-nicking enzyme, DLS-1 recognizing a 6bp target site, was commercialized[98,145]. ODM has been applied to a range of applications through Bionano Genomics, as to study genetically linked diseases[122,139,146,147].

5 ORIGINAL WORK

In this chapter, the work on which the Thesis is based will be discussed and summarized. The work has resulted in three types of assays focusing on DNA damage quantification (**Paper I**), bacterial typing (**Paper II and Paper III**) and plasmid AMR detection (**Paper III and Paper IV**). An overview with the key methods, advances and results from the appended papers is provided, followed by a more general discussion. All data analysis tools have been developed in collaboration with Tobias Ambjörnson's research group at Lund University and the bioinformatic tools by Anna Johnning in Erik Kristiansson's research group at Chalmers University of Technology.

5.1 DNA DAMAGE QUANTIFICATION

The first assay is for identifying and quantifying DNA damage induced by a chemotherapy agent (**Paper I**), helping to understanding the processes of DNA damage and repair related to diseases and medical treatments.

Fluorescence imaging after stretching DNA on glass surfaces in combination with DNA damage detection (overview in Figure 15A) has been done previously for DNA damage induced by multiple agents[131,148-152]. In these studies, the labeling of the DNA damage is based on the use of DNA repair enzymes that excise the damaged lesion, followed by the incorporation of fluorescently labeled nucleotides at the damaged sites by a DNA polymerase. Then the relative number of fluorescent "dots" per stretched DNA length (example image in Figure 15B for the stretched DNA molecules in the YOYO-1 channel and DNA damage dots in the dUTP-Atto647N channel) is reported as a measure of the amount of DNA damage.

The large difference in the types of DNA damage that can be formed by treatment for many cancer patients makes it challenging to globally quantify DNA damage levels. Crosslinking agents, such as Mitomycin C (MMC), are commonly used drugs to treat various cancers and DNA crosslinks, including ICLs, are by far the most toxic lesions caused by these drugs. However, in the previous studies that were labeling DNA damage sites and stretching the DNA, it was typically damages that can be repaired by BER enzymes that have been demonstrated. ICLs cannot be repaired by the BER-pathway, so in order to detect ICLs, the enzyme(s) used needed to be changed.

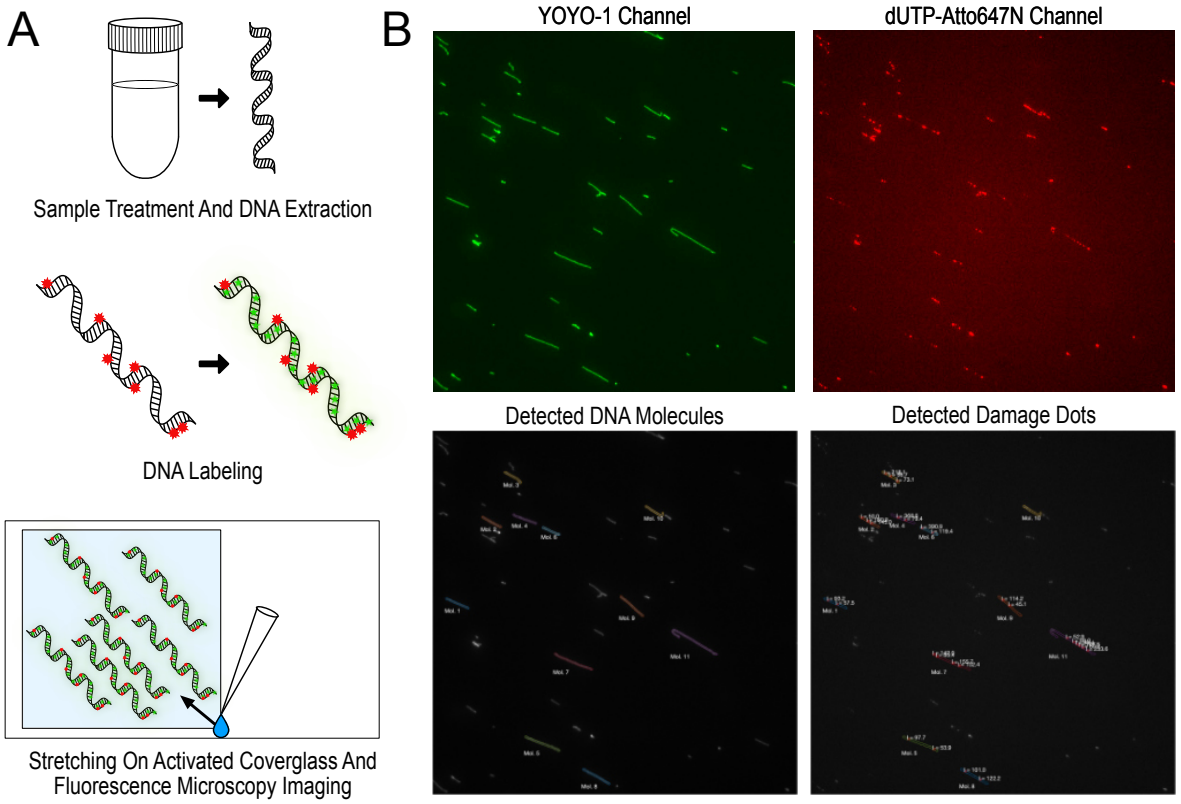

FIGURE 15: SCHEMATIC OVERVIEW OF THE METHOD FOR DETECTING DNA DAMAGE (A) AND AN EXMPLE IMAGE OF THE YOYO-1 CHANNEL WITH DNA MOLECULES, DUTP-ATTO647N CHANNEL WITH DNA DAMAGE DOTS AND THE DETECTED DNA MOLECULES AND DAMAGE DOTS BY THE IMAGE ANALYSIS SDD (B).

The absorption and emission spectra for YOYO-1 and dUTP Atto-647N are plotted in Figure 16. There is little overlap between the spectra from YOYO-1 and Atto647-N making it possible to choose well separated optical filters, and also results in a small chance of having the emission from the dyes bleeding through the channels. This offers the capability to easily image both labels in the same sample.

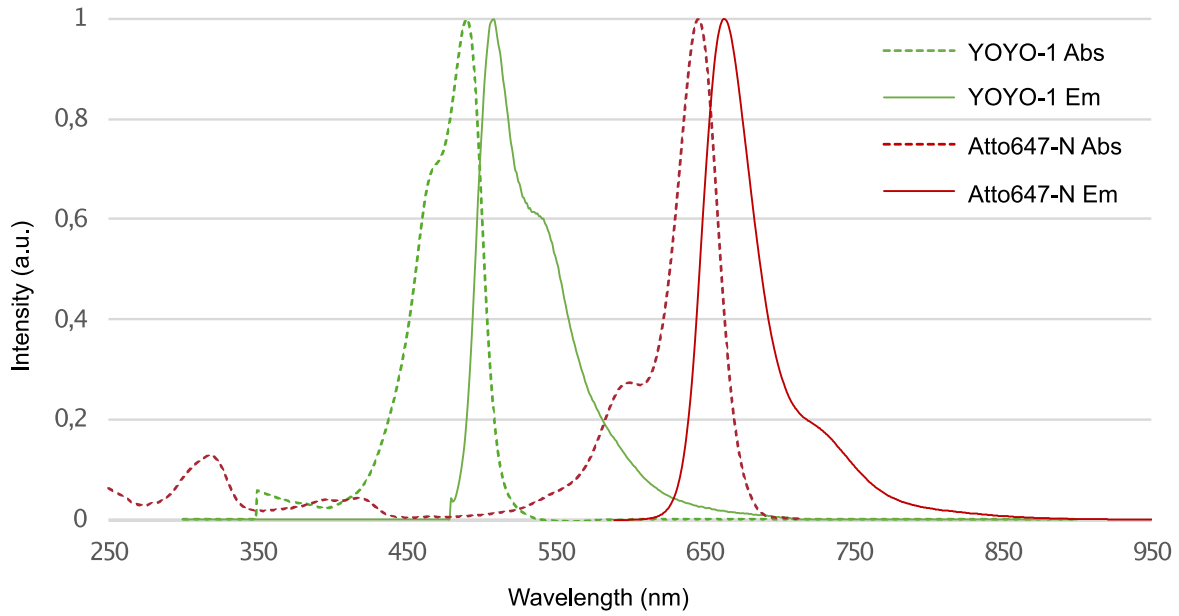

FIGURE 16: ABSORPTION (DASHED LINES) AND EMISSION (SOLID LINES) SPECTRA FOR THE FLUORESCENT LABELS YOYO-1 (GREEN) AND ATTO647-N (RED).

5.1.1 FROM SAMPLE TO QUANTIFIED DAMAGE LEVEL

Blood from healthy individuals were used for the development of the method. Peripheral blood mononuclear cells (PBMCs) were isolated using Lymphoprep density centrifugation, treated with MMC or left untreated and DNA extracted by the GenElute-Mammalian Genomic DNA Miniprep Kit. The DNA damage sites were processed by different bacterial and viral enzymes, or left unprocessed, and a DNA polymerase incorporated fluorescent nucleotides (dUTP-Atto647N or dUTP-Atto550). The DNA was stained with YOYO-1, stretched on activated cover glass slides and imaged by fluorescence microscopy.

To acquire a quantifiable measure of the DNA damage level, the MATLAB-based image software SDD_dots was developed. In short, the length of the stretched DNA in each image was estimated in pixels and the total number of fluorescent nucleotide dots (patches) along each single DNA molecule was counted to quantify the dots/pixel for each sample. A conversion factor obtained from stretching lambda-DNA (48 502 bp) on activated coverslips in the same buffer conditions was used to convert the DNA length per pixel to DNA length in Mbp[149]. An example image for the SDD_dots software detection can be seen in Figure 15B, where the whole DNA stained with YOYO-1 can be seen in the YOYO-1 channel, the detected DNA molecules by SDD_dots in detected DNA molecules, the DNA damage labeled by the incorporation of fluorescent nucleotides in the dUTP-Atto647N channel and the detected damage dots by SDD_dots in detected damage dots.

5.1.2 KEY RESULTS

It has been reported that the first step in crosslink repair is the excision of the ICL by a nuclease complex[153,154]. The enzyme Endo VII was chosen for repair recognition of crosslinks since it was previously shown to cleave DNA containing cisplatin adducts[155], that has a broadly similar mechanism of action as MMC and 4HC. We found that the enzyme Endo VII is important for detecting DNA damage induced by MMC. The increase in detected DNA damage per DNA length for Endo VII in MMC treated sampled, compared to untreated samples, was much higher compared to the other enzymes tested (quantified levels in Figure 17A and image example in Figure 17B). This suggested that Endo VII is crucial for recognizing the DNA damage induced by MMC, and with DNA crosslinks as the main toxic lesion caused by MMC it is likely that Endo VII can be used to detect DNA crosslinks. In the damage induced by MMC and 4HC, an endonuclease, such as Endo VII might be needed to pre-process the ICL damage site to enable incorporation of the fluorescent nucleotides that are the DNA damage reporters.

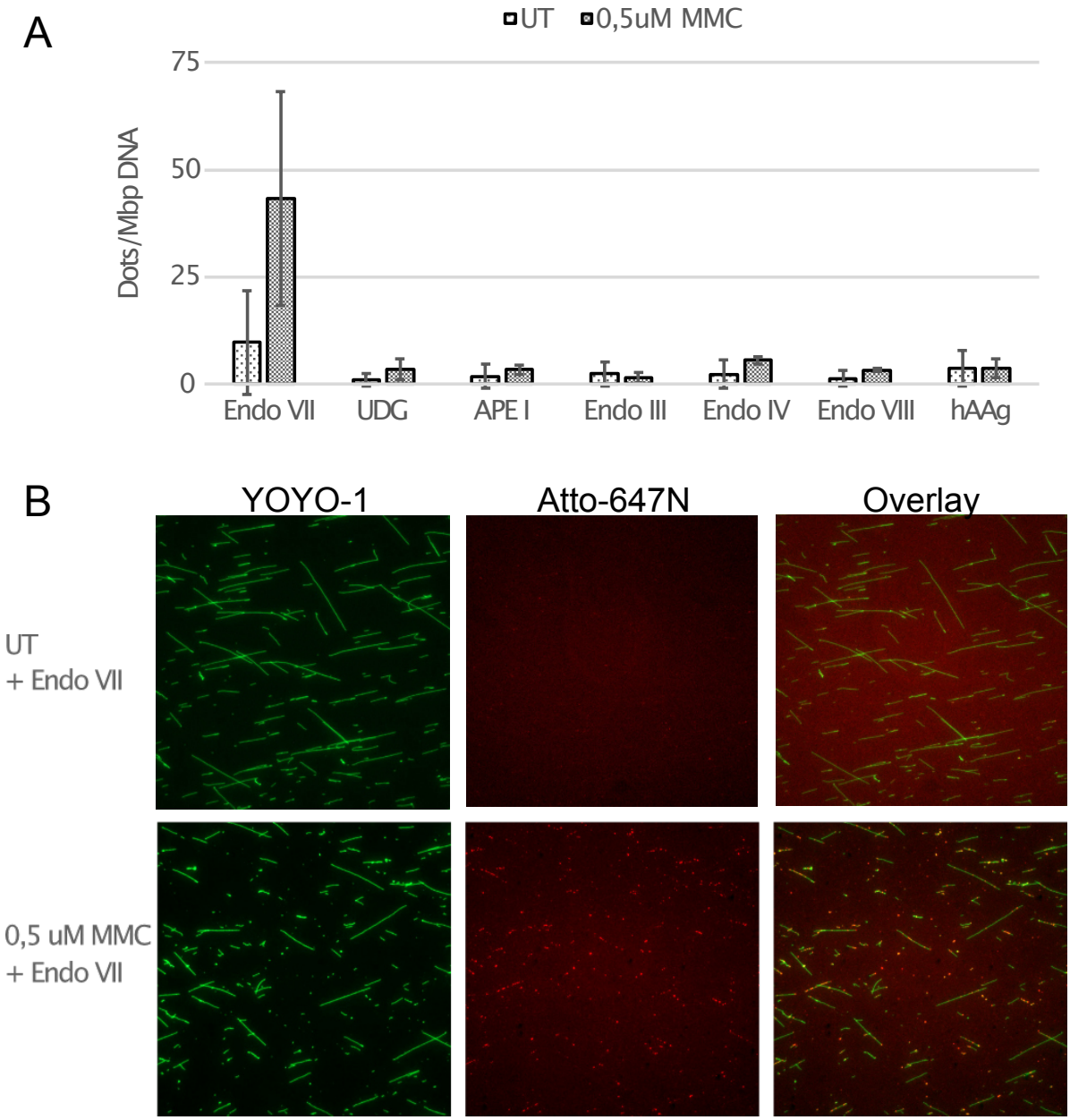

FIGURE 17: IN A, THE CALCULATED LEVELS OF DNA DAMAGE DOTS/MBP DNA FOR MMC INDUCED DAMAGE RECOGNIZED BY 7 ENZYMES, ENDO VII, UDG, APE I, ENDO III, ENDO IV, ENDO VIII AND HAAG. ERROR BARS REPRESENT STANDARD DEVIATION CALCULATED FROM TWO REPEATS. IN B THERE ARE EXAMPLE IMAGES FOR MMC INDUCED DAMAGE RECOGNIZED BY ENDO VII. THE TWO CHANNELS OF THE IMAGE, YOYO-1 AND ATTO-647N ARE INCLUDED BOTH SPLIT AND AS AN OVERLAYED IMAGE.

To simultaneously quantify both the damage identified by Endo VII as well as the global damage level we used a previously published dual labeling protocol[151] comparing DNA damage repair recognized by a combination of BER enzymes (UDG, hAAG and Endo IV) to only Endo VII. The dual labeling protocol was based on two consecutive labeling runs (overview in Figure 18A), where we first ran the BER-enzyme damage recognition with incorporation of dUTP-

Atto550 and then the Endo VII damage recognition with incorporation of dUTP-Atto647N. The importance of Endo VII in the detection of MMC-induced damage was also suggested in the dual labeling results (exemplified zoomed overlay of the three channels in Figure 18B, quantified levels in Figure 18C), where the BER-enzymes still detected a number of damage sites, but when treated with MMC the majority of the increase in recognized DNA damage is by Endo VII repair. This suggests that MMC, mainly induces damages that can be detected by Endo VII, compared to the BER-enzymes.

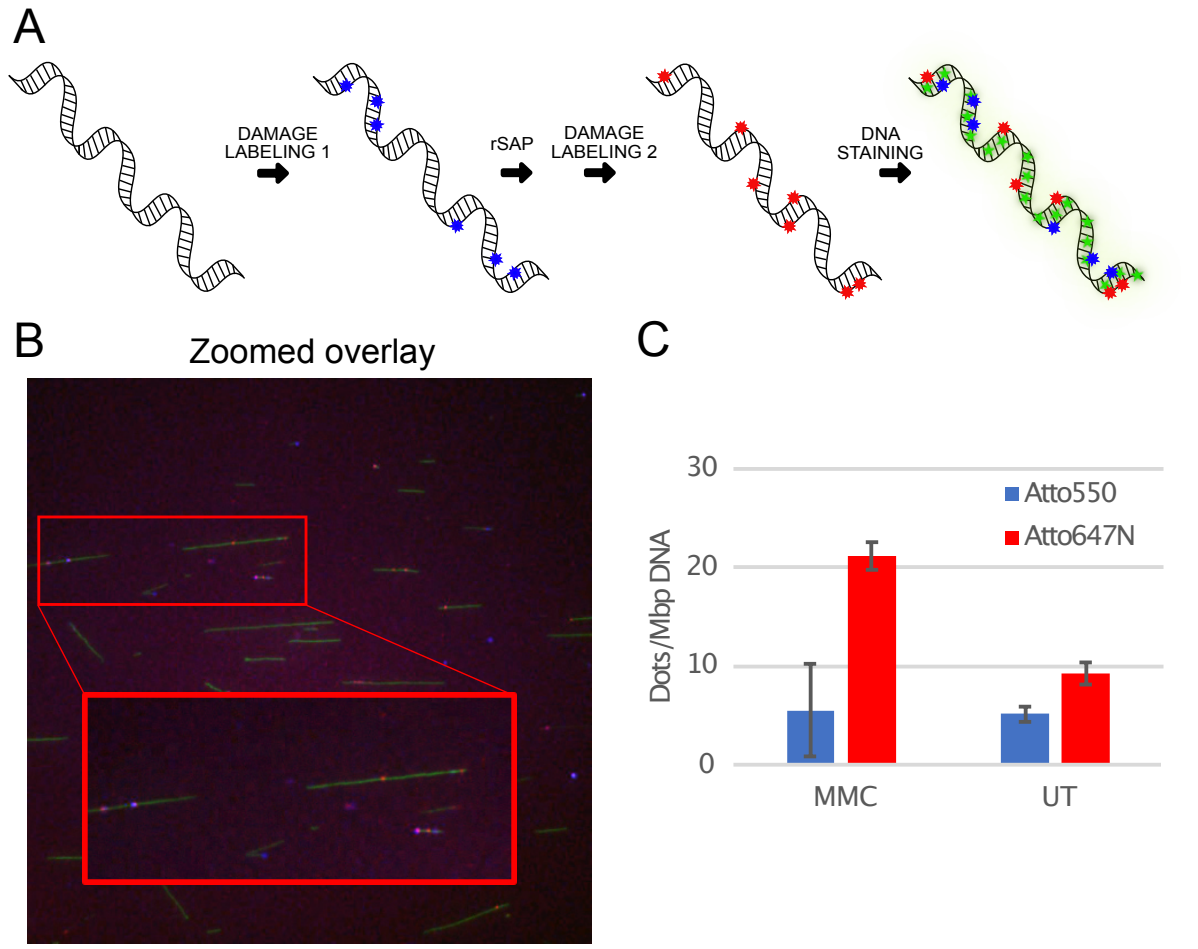

FIGURE 18: IN A, A SCHEMATIC OVERVIEW OF THE DUAL LABELING FOR DNA DAMAGE. AN OVERLAY OF THE THREE CHANNELS WITH DUAL LABELED STRETCHED DNA, WITH A ZOOMED IN PART (C). THE RED RECTANGLE IS THE REGION THAT IS ZOOMED IN. THE CALCULATED LEVELS OF DNA DAMAGE DOTS/MBP DNA FOR EITHER MMC TREATED SAMPLES OR LEFT UNTREATED BEFORE DUAL LABELING. THE DUAL LABELING WAS FIRST RUN WITH BER-ENZYMES AND LABELED WITH ATTO-550 AND THEN RUN WITH ENDOVII AND LABELED WITH ATTO-647N (C). ERROR BARS REPRESENT STANDARD DEVIATION CALCULATED FROM TWO REPEATS.

A future possible use of this developed assay is to use it to screen patient responses to specific drugs and their overall response to treatment, enabling detection of sensitive patients before start of a specific treatments or adjusting/changing the dose or drug to avoid severe side effects. However, comparing experiments of MMC treated DNA and untreated DNA, recognized by Endo VII there is a large variation between experiments. There might be other non-enzyme specific DNA damages recognized by the method, requiring a pre-repairing step as done in the dual labeling to get rid of the non-specific damages, caused by for example handling of the DNA. The stretching and detection software used is also of great importance, and it is hard to compare results other than trends between the different studies as the number of damage sites per DNA length detected by different softwares and with different settings, such as length boundaries and edge detection exclusion. Previous papers using the same method for identification of other types of induced DNA damages have also reported a large inter-individual variation of the DNA damages recognized[148-150], however if so, comparisons with another method for detecting MMC induced DNA damage verifying these individual differences would be needed. The need of a stable output of the method, and to identify the variation factor(s), is crucial both for research purposes and potential future clinical applications.

5.2 BACTERIAL TYPING

The second assay developed is for rapid identification of bacterial infections from the intensity profile of chromosomal fragments. The identification was done by a applying a discriminative approach to individual DNA fragments for bacterial typing, first at species resolution (**Paper II**) and then of sub-species strain groups (**Paper III**). An experimental intensity profile was deemed discriminative at a given resolution if all the high-scoring matches to a reference database belonged to the same species (**Paper II**) or strain group (SG) (**Paper III**).

Bacterial identification can often be time-consuming and depend on cultivation of the bacterial pathogen, which can introduce biases. Rapid and correct identification is crucial for both patient outcomes, to minimize the spread of disease, and to control outbreaks. The assay (overview in Figure 19) is based on matching optical maps of large DNA molecules to a reference database of theoretical intensity profiles of known bacterial genomes and applying a discriminative approach for the single DNA fragments.

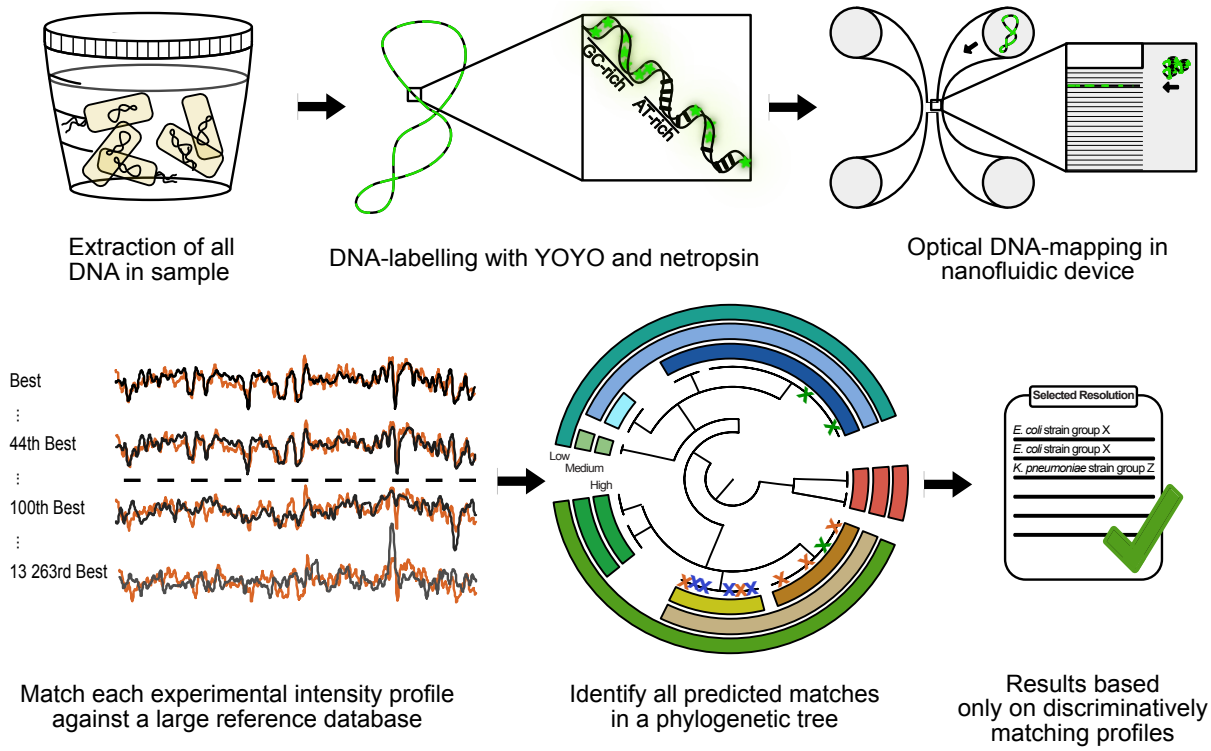

FIGURE 19: METHOD OVERVIEW FOR BACTERIAL IDENTIFICATION FROM LONG CHROMOSOMAL FRAGMENTS, USED IN PAPER II AND PAPER III.

5.2.1 DNA EXTRACTION AND LABELING

The DNA was extracted by lysis of the bacteria in agarose plugs to minimize shearing of the DNA. The extraction was initially done based on CHEF genomic DNA kit (BIO-RAD) and later adapted, inspired by Matushek et al[156]. The adapted protocol for extraction by plug lysis was designed to work for both Gram-positive and Gram-negative bacteria, to obtain DNA fragments long (> 100 kb) for subsequent labeling and analysis, and reduce the extraction time. The DNA staining, by competitive binding of netropsin and YOYO-1 to achieve a sequence specific emission intensity profile[143], does not depend on primers or probes, and, therefore, works for any bacterial species and is invulnerable to allelic variation or the isolate type.

5.2.2 FROM IMAGE TO INTENSITY PROFILE

The labeled DNA was loaded on a nanofluidic chip in a buffer with added beta-mercaptoethanol, to reduce the probability of photocleavage of DNA[157]. To record the intensity profiles, each DNA molecule was confined in a nanofluidic channel with a profile of 100 x 150 nm^2 (height x width) and imaged by fluorescence microscopy. In all nanofluidic experiments, lambda-DNA (48 502 bp), was included before DNA staining as an internal size reference due to the small variations in stretching of the DNA molecules that can depend on differences in ionic strength and effect of individual channels.

The imaged DNA molecules were then converted into intensity profiles (Figure 20). Typically, 20-100 time frames (T_1-T_N, N=20-100), were recorded for each DNA molecule while extended in a nanofluidic channel. The edges of the DNA molecule were detected in each time frame and a one-dimensional image of the molecule for each time frame was stacked underneath each other, creating a kymograph with the extension of the molecule in one direction and the time in the other direction. The kymographs were aligned to account for the fluctuations of the molecule and drift in the channel[158] and averaged to obtain the mean value of intensity in each column versus the molecule extension. The intensity profiles were normalized, to account for experimental differences resulting in an intensity difference between molecules, with the average value set to 0 and the standard deviation to 1. This resulted in an intensity profile of the recorded DNA molecule where the high peaks represent the bright GC-rich regions of the DNA sequence, stained with YOYO-1, and the valleys represent the dark AT-rich regions, stained mainly with netropsin (Figure 20).

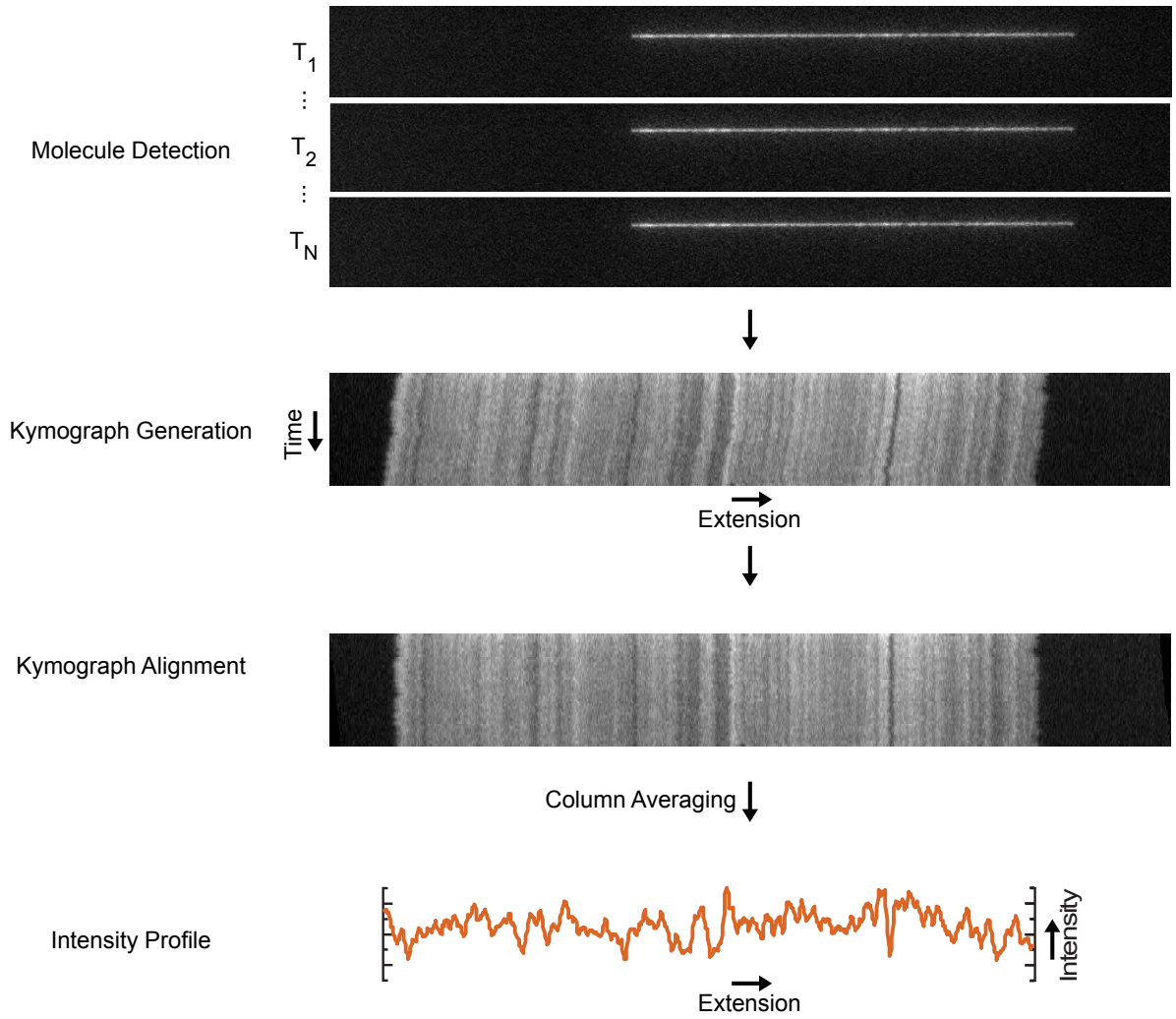

FIGURE 20: PROCESSING EXAMPLE WHERE A SERIES OF IMAGES ARE USED TO GENERATE AN INTENSITY PROFILE VIA AVERAGING OF A KYMOGRAPH.

5.2.3 MATCHING TO REFERENCE DATABASE

To compare the similarity of two intensity profiles, a Pearson Correlation Coefficient (PCC) was calculated. The PCC value varies between 1 to -1, where a value of 1 implies that the two profiles are identical and a value of -1 that the two profiles are mirror images of each other. The PCC was calculated by keeping one profile at a fixed position and the other profile shifted one pixel at a time, and then repeating the procedure for the flipped version of the profile. The maximal PCC (Cmax) value and the positions of the intensity profiles for that similarity match was then saved.

Theoretical intensity profiles for comparison with a previously known DNA sequence can be generated[159,160]. They are generated considering the likelihood of YOYO-1 and netropsin binding to a four bp site along the DNA sequence, taking the resolution of the experimental setup and the ratio of concentrations into consideration. The reference databases in **Paper II** and **Paper III** were based on the complete bacterial genomes RefSeq (downloaded on October

16th 2018, or on July 17th 2019) excluding sequences shorter than 500 kb or with the word plasmid in the FASTA headers. Resulting in databases that respectively consisted of 10 310 sequences belonging to 2355 different bacterial species (**Paper II**) or 13 263 sequences belonging to 2 697 different bacterial species (**Paper III**).

After an intensity profile was generated for each recorded DNA fragment, they were matched using every possible start position to all theoretically generated intensity profiles in the reference database. For each of the experimental intensity profiles, the database matches were filtered based on three criteria. First, short experimental intensity profiles were discarded (length < L_{min}). Then, the best matches were selected (a C_{max} within the range max(C_{max}) to max(C_{max}) – C_{diff}), and if all of the highest-scoring matches matched to the same group at the taxonomic resolution examined, the intensity profile was deemed discriminative. Lastly, the discriminative intensity profiles with sufficiently high-scoring matches (max(Cmax) > Cthresh) were reported back to the user. After exploring the parameter space, the parameters were set to: L_{min} = 250 px~125 kb, C_{diff} = 0.05 and C_{thresh} = 0.5. An example of an experimental intensity profile matched against theoretical intensity profiles in the reference database in **Paper II** is shown in Figure 21. In Figure 21 it is also included where the stringency cutoff, C_{diff}, between the matching of the experimental intensity profile and the theoretical intensity profiles in the reference database. The experimental intensity profile is from an *E. coli* and out of the exemplified theoretical intensity profiles, the best, 10th best and the 44th best matching theories are also *E. coli*.

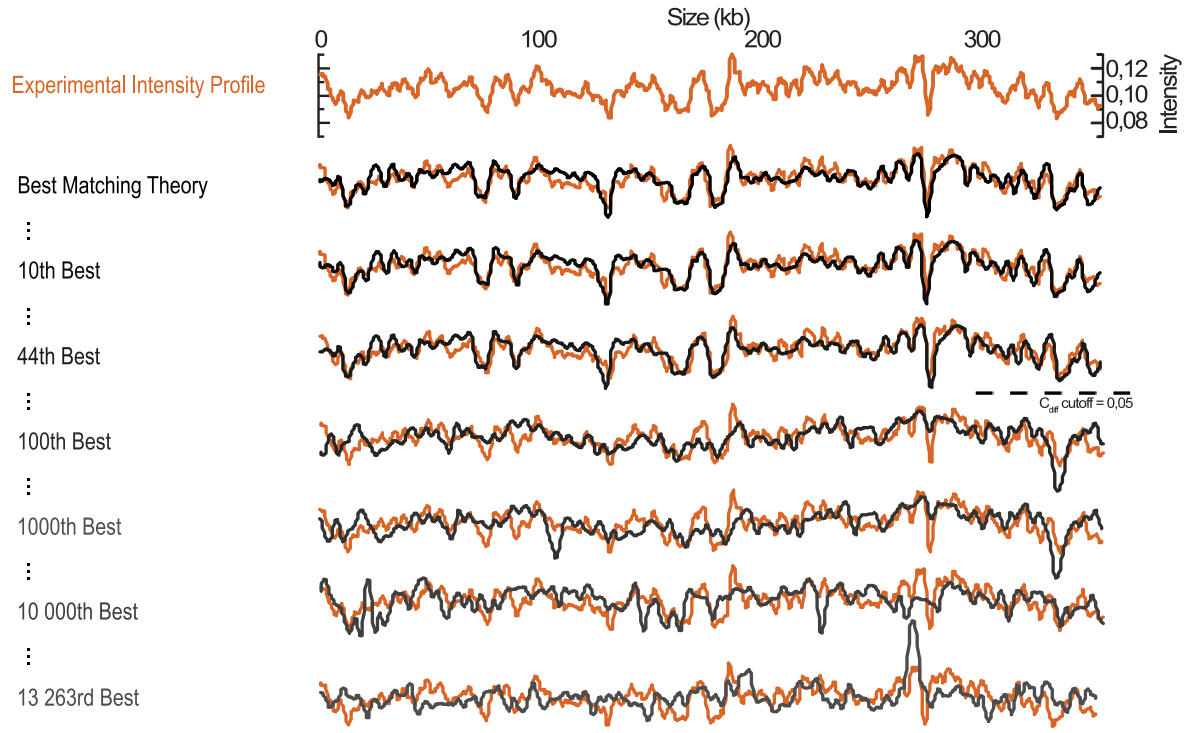

FIGURE 21: EXAMPLE MATCHING OF AN EXPERIMENTAL INTENSITY PROFILE TO A REFERENCE DATABASE WITH THEORETICAL INTENSITY PROFILES.

The strain groups (SGs), taxonomic groups at a higher resolution than species, were based on phylogenetic trees generated from a core genome alignment and, therefore, reflect the evolutionary relationship between the strains or subtypes of each species. The SGs were formed by cutting each tree at a given branch length, which defined the size and diversity of the SGs and thereby the taxonomic resolution. A short branch length created many SGs and, consequently, an SG scheme of a high taxonomic resolution and a long branch length created fewer SGs and a SG scheme of a lower taxonomic resolution. The taxonomical resolutions used for sub-species typing in **Paper III** were SG_{High}, SG_{Medium} and SG_{Low}. For *E. coli* all reference genomes of the globally disseminated sequence type (ST) 131 formed their own SG at SG_{High} and for *K. pneumoniae* the clinically important clonal complexes CC14, CC147, CC258, and CC307 all formed distinct SGs, except for the SG containing CC258, which also contained a single additional genome at SG_{High}.

Only the highest-scoring matches between an experimental profile and the reference profiles in the database were retained. An experimental profile was deemed discriminative at a given resolution if all the high-scoring reference matches belonged to the same SG. Only discriminative profiles were used for the identification, and the remaining profiles were discarded since they were deemed uninformative at the taxonomic resolution of interest.

5.2.4 KEY RESULTS

The data analysis parameters were optimized to maximize the performance of the assay based on all data collected (in **Paper II**) for cultured isolates.

The performance for bacterial species typing was assessed by measuring the sensitivity – estimated by the average true positive rate (TPR) – and the average proportion of discriminative profiles for different stringency thresholds (defined by the parameter C_{diff}) in the database matching (Figure 22A). By increasing the stringency, the sensitivity generally increased for all resolutions, while the average proportion of discriminative profiles decreased, giving a trade-off between a obtaining a high sensitivity and including as much data as possible. To obtain a high TPR, without reducing the throughput, the C_{diff} was set to 0.05.

The effect of the fragment length was assessed by measuring the TPR and the average proportion of barcodes of different lengths that were discriminative (Figure 22B). The length of the molecule had a significant effect on the performance of the method, where the TPR increased drastically for molecules up to 250 px, but then remained the same for increasing lengths. The proportion discriminative fragments also increased with the profile length, making the profiles needed for a reliable identification fewer with increased profile lengths.

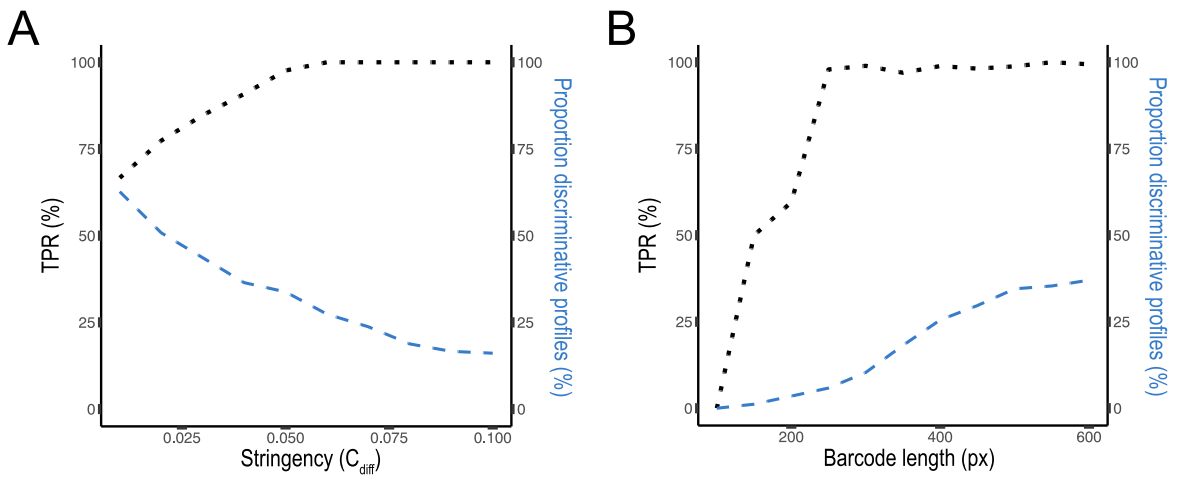

FIGURE 22: PARAMETER EVALUATION ON ALL DATA IN PAPER II FOR SPECIES LEVEL BACTERIAL IDENTIFICATION. THE EFFECT OF STRINGENCY (C_{DIFF}) ON THE TRUE POSITIVE RATE (TPR) AND PROPORTION DISCRIMINATIVE PROFILES AT THE SPECIES RESOLUTION IN A. THE EFFECT OF BARCODE LENGTH ON THE TPR AND PROPORTION DISCRIMINATIVE PROFILES AT THE SPECIES RESOLUTION IN B.

The species typing assay (in **Paper II**) was validated with a range of samples. Six different bacterial species were tested, among them both Gram-positive and Gram-negative bacteria, both as single isolates and in different mixes (Figure 23). Overall, the assay performed well, with an overall TPR of 99% and identifying only the correct species for all samples examined. The very few fragments giving a false positive discriminative match were distributed among various species and samples.

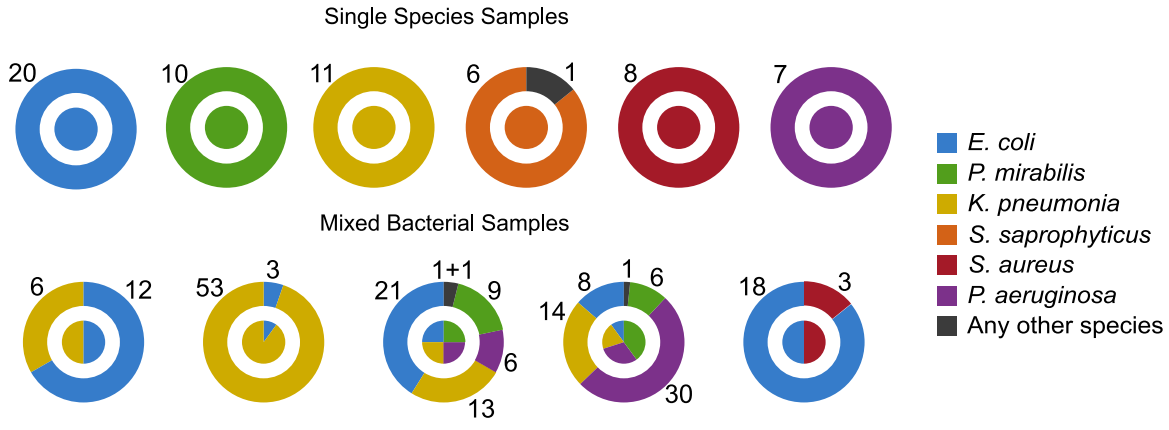

FIGURE 23: KEY RESULTS BACTERIAL IDENTIFICATION AT THE SPECIES LEVEL FOR SINGLE SPECIES SAMPLES AND MIXED BACTERIAL SAMPLES. MARKED ARE THE RATIOS OF THE DISCRIMINATIVE FRAGMENTS DETECTED (OUTER CIRCLE), THE RATIO OF BACTERIA IN THE SAMPLE (INNER CIRCLE) AND THE NUMBER OF DISCRIMINATIVE DNA FRAGMENTS FOR EACH TYPE OF BACTERIA IN THE SAMPLES (NUMBERS).

For the sub-species typing in **Paper III**, different taxonomical resolutions were examined and cultivated pure *E. coli* (Figure 24A) and *K. pneumoniae* (Figure 24B) samples were used for parameter optimization. The method performance was assessed by measuring the sensitivity – estimated by the average TPR and the average proportion of discriminative profiles at different stringency thresholds (C_{diff}). The results for the effect of stringency (C_{diff}) on TPR and proportion discriminative profiles, for all examined taxonomical resolutions (species, SG_{Low}, SG_{Medium} and SG_{High}) for *E. coli* and *K. pneumoniae*, show similar trends to the ones for the species typing. By increasing the stringency, the sensitivity generally increases for all resolutions, while the average proportion of discriminative profiles decreases. There is, consequently, a trade-off between obtaining high sensitivity and including as many profiles as possible – which ultimately comes down to the total number of DNA fragments analyzed. In situations where data is abundant, the stringency of the matches to the reference database can be increased to ensure an even higher TPR. On the other hand, if time is of the essence and/or data is scarce, the stringency can be lowered, or a lower taxonomic resolution can be used and, thus, the method is adaptable to different clinical settings and applications.

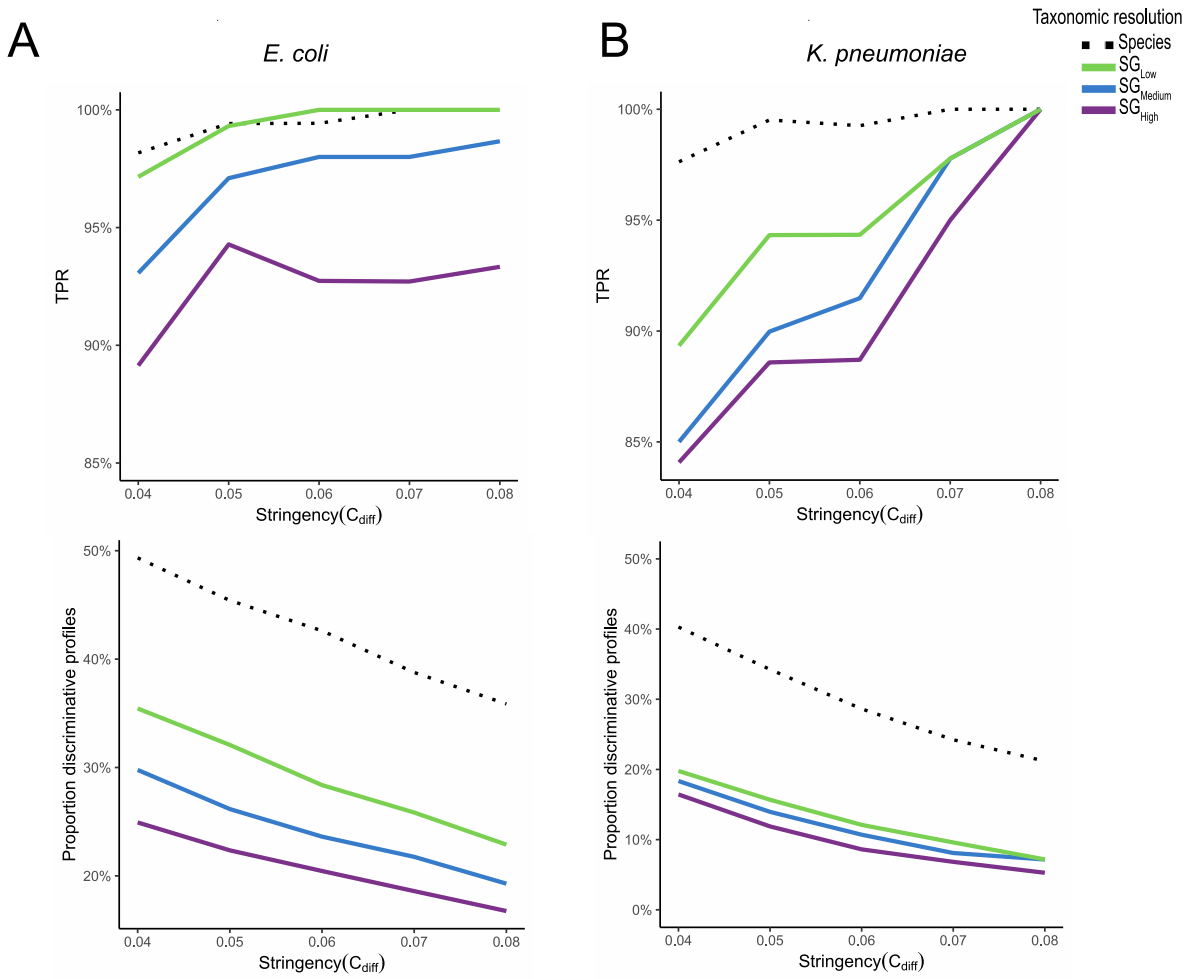

FIGURE 24: PARAMETER EVALUATION FOR E. COLI (A) AND K. PNEUMONIAE (B) SAMPLES. THE EFFECT OF STRINGENCY (C_{DIFF}) ON THE TRUE POSITIVE RATE (TPR) AND PROPORTION DISCRIMINATIEV FRAGMENTS AT FOUR DIFFFFERENT TAXONOMICAL RESOLUTIONS, SPECIES, SG_{LOW}, SG_{MEDIUM} AND SG_{HIGH} IN PAPER III.

The sub-species typing was done for 25 pure cultures of *E. coli* bacteria from various sequence types (Figure 25A), 9 pure cultures of *K. pneumoniae* bacteria from various sequence types (Figure 25B) and 6 uncultured urine samples from patients with *E. coli* urinary tract infections (UTIs) (Figure 25C). Using a set stringency of C_{diff}=0.05 resulted in an average TPR of 99.3/97.1/94.3% for *E. coli* and 94.3/90.0/88.6% for K. *pneumoniae* at SG_{Low}/SG_{Medium}/SG_{High}, respectively, which can be compared to 99.4% (*E. coli)* and 99.5% (*K. pneumoniae*) at the species level. However, there is a variation between the samples, both in sensitivity and the number of discriminative profiles obtained, where certain samples were comparatively more challenging to type. Furthermore, the assay was successfully applied to both mixes and uncultured urine samples.

For the uncultured urine samples (Figure 25C), all except one sample (U1, ST69) had a TPR ≥90% for SG_{Medium}. For U1, two barcodes were discriminative to the correct SG and three fragments to three other non-related SGs. That some samples are more challenging to type is possibly due to fewer genomes from that SG represented in the reference database or from false positives caused by matches against chromosomal mobile genetic elements.

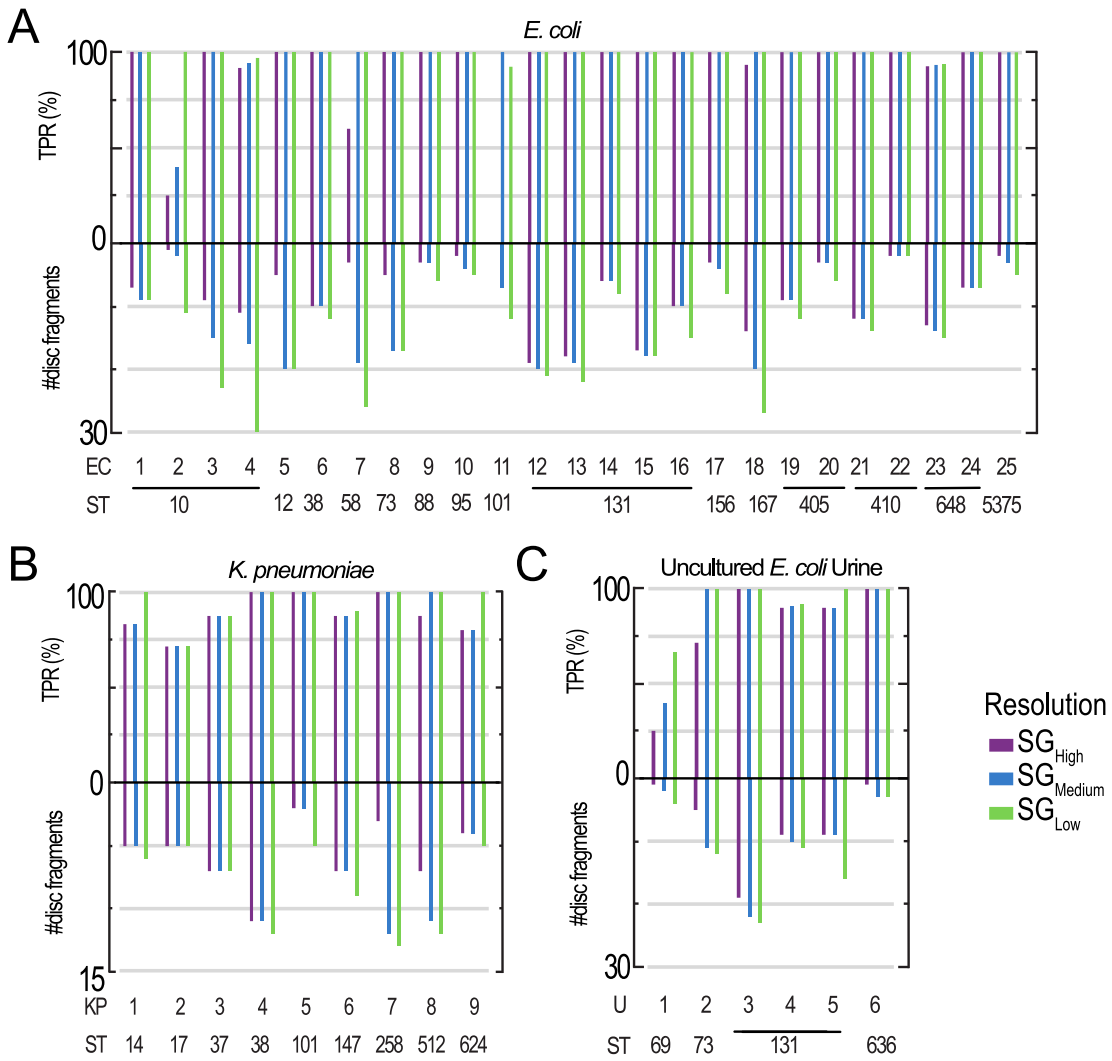

FIGURE 25: KEY RESULTS WHEN IDENTIFYING E. COLI (A), K. PNEUMONIAE (B) AND UNCULTURED UTI URINE (C) SAMPLES AT THE BACTERIAL STRAIN-GROUP LEVEL AT THREE DIFFERENT TAXINOMICAL RESOLUTIONS; SG_{HIGH}, SG_{MEDIUM} AND SG_{LOW}.

To further explore the resolution of the method, we investigated if different clonal lineages of *E. coli* ST131 could be distinguished. *E. coli* ST131 is a globally spread high-risk lineage that has rapidly emerged and spread since its discovery and is associated with extensive AMR[161,162]. We used an ultra-high SG resolution, that divided the 61 ST131 reference genomes in the database into five SGs (Figure 26A). The resulting sensitivity was high with an average TPR of 94% and six of the nine samples had a TPR of 100% (Figure 26B). This shows that the assay can accurately type bacteria, even at a taxonomic resolution that is higher than the traditional multi-locus sequencing type level.

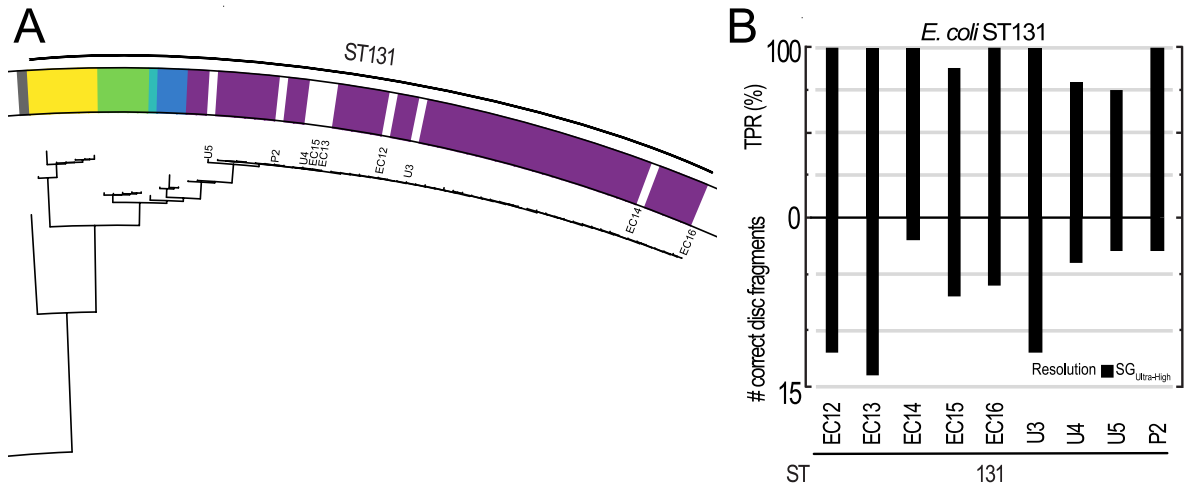

FIGURE 26: THE ST131-BRANCH OF THE PHYLOGENETIC TREE WITH THE SUB-ST131 GROUPS (A) AND $SG_{ULTRA-HIGH}$ RESULTS FOR THE SUB-ST131 IDENTIFICATION OF NINE E. COLI ST131 SAMPLES (B).

In **Paper III**, the time it took from the patient sample to an obtained typing result was around 7 h: 4 h DNA extraction, 40 min DNA processing and labeling, 1 h nanofluidic imaging, and 1 h data analysis, which could be further optimized. Furthermore, the method is adaptable to different clinical settings and applications. In situations where there is a lot of data, the stringency of the matches to the reference database can be increased to ensure an even higher TPR, but if time is of the essence or there is less data, either the stringency can be lowered, or a lower taxonomic resolution can be used.

Compared to other methods routinely used in hospitals for distinguishing infecting bacteria on the species level, such as phenotypic tests and mass spectrometry, or together with more high-resolution methods like multi-locus sequence typing and whole-genome sequencing, they all require one or two overnight cultivations of the bacteria to obtain pure samples and a sufficient amount of DNA or protein for analysis[163-165]. Furthermore, there are some bacteria that cannot easily be cultured in laboratories. Therefore, in comparison, the time to reach a complete resistance profile and hence a correct diagnosis is limited by the growth rate of the bacterium and typically takes at least 24 h[166,167].

A clear benefit of the ODM-based typing presented in **Paper II** and **Paper III**, is that each intensity profile is analysed individually. Therefore, the method easily identified even mixed polymicrobial samples. Another benefit is the small amount of DNA needed for analysis, that enabled the typing directly from uncultured urine samples.

5.3 PLASMID AMR DETECTION

The third assay developed is for identification of antibiotic resistance genes on plasmids, using CRISPR-Cas9 targeting of AMR-genes to induce specific cuts to the plasmid. We used this CRISPR-Cas9 targeting of different beta-lactamase genes for two of the projects, **Paper III** and **Paper IV.**

The genes making bacteria resistant are often located on plasmids, that can be transferred between both bacterial strains and species. Therefore, effective methods to characterize bacterial plasmids and detect resistance genes can assist in tackling the emerging spread of resistance, for example, during outbreaks in hospitals. In **Paper III** we, simultaneously to the bacterial typing, used CRISPR-Cas9 induced cleavage to target AMR-genes and aligned the intensity profile of the plasmids to differentiate the targeted cut from light induced breaks (Figure 27A). In **Paper IV** we used the fact that the CRISPR-Cas9 induced cleavage of targeted AMR-genes in a plasmid, will linearize the plasmid (Figure 27A). Hence, both the length and intensity would change for a fluorescently labeled plasmid stretched on glass slides if it is linearized.

5.3.1 PLASMID ALIGNMENT AND CONSENSUS GENERATION

In **Paper III**, the lengths of the short DNA fragments (<250 pixels, ~135 kb), extracted from the time-averaged kymographs, were used for detecting multiple copies of the same plasmid and clustering them. Intact circular plasmids were identified from groups with high emission intensity and similar extensions. The corresponding linear fragments, at approximately double the extension, were then extracted and grouped based on their length. For a given length group, the associated intensity profiles were then aligned and clustered based on similarity to one another, calculating PCCs. A typical similarity score threshold for the clusters was set to 0.75. The presence of the target gene was verified by analyzing whether most DSBs occurred at the same position in a similarity cluster. That a majority of the DBSs were located at the same position indicated that the cuts were caused by CRISPR-Cas9 at the target gene location and not induced by mechanical forces or photo nicking, and thereby verified the presence of the targeted gene.

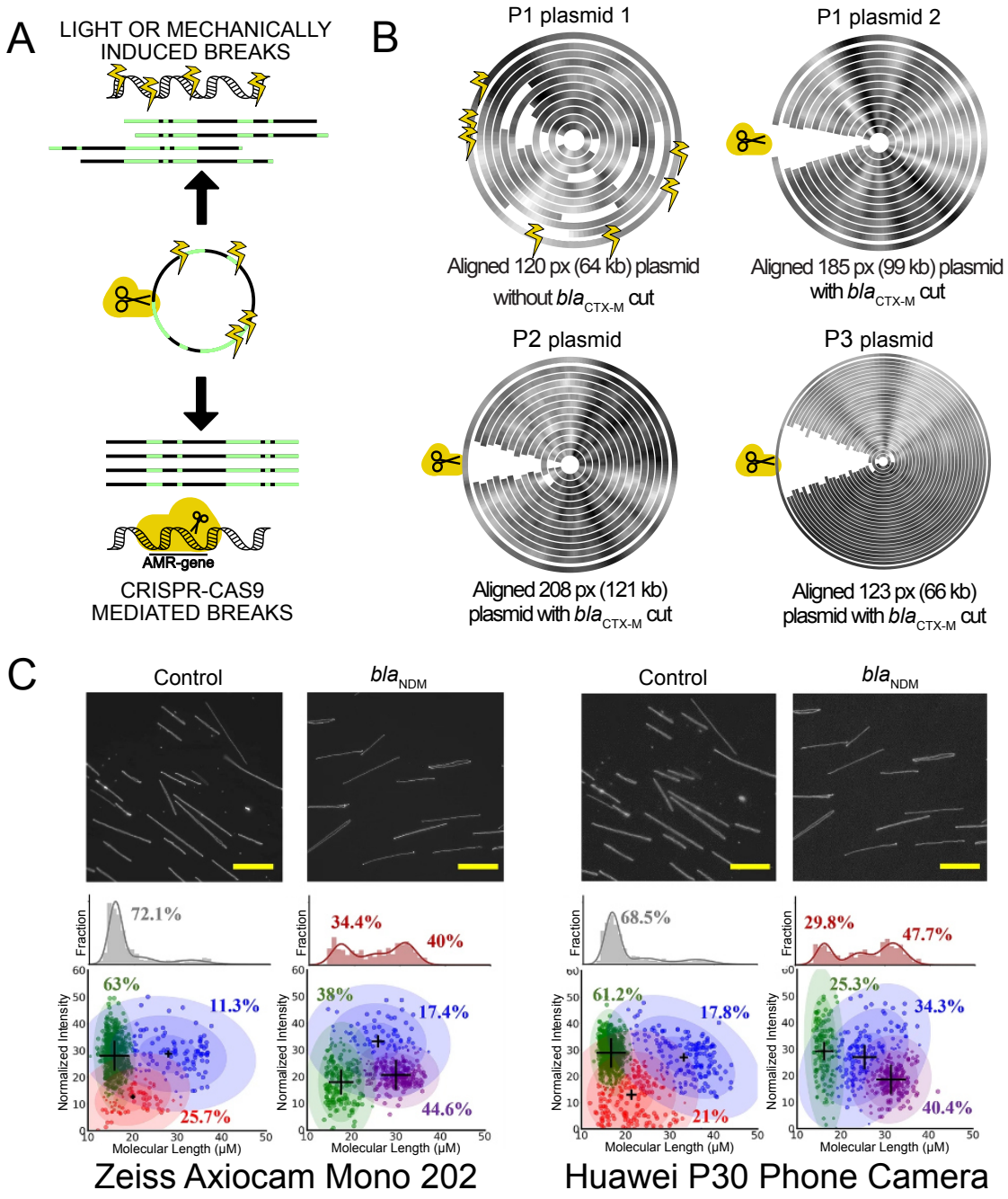

FIGURE 27: A SCHEMATIC OVERVIEW OF A PLASMID CUT BY CRISPR-CAS9 OR LIGHT INDUCED BREAKAGE (A). IN THE CIRCULAR PLOTS FOR SAMPLES P1, P2 AND P3 IN B, EACH RIBBON REPRESENTS INDIVIDUAL DNA FRAGMENTS AND THEIR BRIGHTNESS CORRESPONDS TO THE INTENSITY PROFILE. THE PROFILES HAVE BEEN ALIGNED AND THE CONSENSUS OF ALL THE PROFILES IS INCLUDED AS THE OUTERMOST RIBBON. LINEARIZATION AT THE SAME POSITION INDICATES A CUT BY CRISPR-CAS9 AND VERIFIES THE PRESENCE OF THE TARGET GENE. DNA IMAGES, CORRESPONDING LENGTH HISTOGRAMS AND CLUSTER PLOTS (B) FOR THE CONTROL AND TREATED WITH CRISPR-CAS9 TARGETING BLA_{NDM} USING EITHER THE AXIOCAM MONO-202 CAMERA OR AN HUAWEI P30 SMARTPHONE CAMERA.

5.3.2 KEY RESULTS

In **Paper III**, chromosomal DNA and plasmids from three samples containing both, were recorded in the same experiment. The long DNA fragments (>250 pixels, ~135 kb), probably chromosomal, were used for bacterial sub-species typing and the shorter DNA fragments (<250 pixels, ~135 kb), potentially plasmid DNA, were used for AMR-gene detection on plasmids. The simultaneous detection was tested for three different samples (P1-P3), where a plasmid harboring *bla*$_{\text{CTX-M}}$ group 1 or 9 genes was detected in each sample (Figure 27B). In P1 two plasmids of different lengths were found, where P1 plasmid 1 was not cut by the CRISPR-Cas9 but linearized randomly by light or mechanical shearing. In P1 plasmid 2, P2 plasmid and P3 plasmid, linearization at the same position can be seen for the 99 kb P1 plasmid 2 (12/12 profiles linearized at the same position), 121 kb P2 plasmid (12/13) and for the 66 kb P3 plasmid (26/28). This verified the presence of the target gene on one plasmid each of the samples.

One key step in limiting the global spread of AMR is the rational use of antibiotics. This necessitates the development of novel techniques that can diagnose AMR infections, as well as confirm the spread of resistant bacteria in hospital settings so that proper antibiotic therapy can be prescribed, and appropriate infection prevention and control measures can be employed. Since the healthcare infrastructure, social hygiene practices and economic status of a country are directly correlated to the prevalence of AMR-genes, the diagnostic technique also needs to be fast, simple, and cost-effective to be suitable in low-resource settings. Current methods for plasmid analysis either provide limited information or are expensive and challenging to implement in low-resource settings.

In **Paper IV**, we used CRISPR-Cas9 induced linearization of plasmids harboring targeted AMR-genes to detect bacteria carrying antibiotic resistance (Figure 27C). When stained with a fluorescent DNA dye, YOYO-1, and stretched out on a surface, a non-cut plasmid appears with approximately half the length of a linearized version of the same type of plasmid. The non-cut plasmid will also be approximately double as light-intense as a linearized version of the same type of plasmid. We developed the assay to be performed on simple fluorescence microscopes, with the setup already present in many African countries via a tuberculosis diagnosis program and where the images could be acquired by mounting a smartphone camera on the eyepiece of the microscope.

The principle of the demonstrated assay is general, and any gene encoded on a plasmid can be identified by using the corresponding gRNA for CRISPR-Cas9 targeting. Beta-lactamases are one of the most important agents that makes Gram-negative bacteria drug resistant and continuous mutations make them remarkably diverse. As an example, a plasmid harboring a *bla*$_{\mathrm{NDM}}$-gene was detected by comparing the distribution of the lengths and intensities for the detected plasmids in a control sample or with CRISPR-Cas9 targeting *bla*$_{\mathrm{NDM}}$-gene (Figure 27C). In both length histograms, there was a similar shift from shorter molecule lengths to longer ones in presence of *bla*$_{\mathrm{NDM}}$ targeting CRISPR-Cas9. In the cluster plots the molecules were separated in two dimensions, based on molecule length and intensity. In presence of *bla*$_{\mathrm{NDM}}$ targeting CRISPR-Cas9, a short length-high intensity cluster is decreased, and a long length-low intensity cluster is formed, verifying the plasmid linearization by CRISPR-Cas9 cleavage and the presence of *bla*$_{\mathrm{NDM}}$, Furthermore, the assay provides a measurement of the number and sizes of plasmids in the sample.

6 CONCLUDING REMARKS

Even if set as goals for the Human Genome Project, to unveil the base pair-code that makes up the blueprint of human life, to understand what the genes stand for and how variants are coupled to human onset and progression of disease, there are still regions and functions of our DNA that are unknown. Today, we know what large parts of the DNA codes for, and not only when it comes to our own human DNA, but also the DNA of the organisms in our surroundings, such as bacteria or viruses, that also affect our health.

This thesis set out to describe the development of single DNA molecule analysis tools for medical diagnosis, and why it is important that these tools are developed. The assays presented in this thesis are all studying stretched-out DNA optically, with fluorescence microscopy.

The first assay for identifying and quantifying induced DNA damage in **Paper I** can provide information about the DNA damage levels induced by certain drugs and shine light upon the damage lesions that are induced. Furthermore, it could help to link the processes of DNA damage and repair related to diseases and medical treatments. The main challenge to address in the future is where the variation of recognized DNA lesions between experiments originates from, if it is an effect introduced by the method or has a biological relevance. If the variation is understood, the method could be used as a tool to determine what type and amount of DNA damage can be formed by a drug and repaired by an individual. In the end, that method would then serve as a tool for personalized therapy or to help in the development of new drugs.

Now imagine a world where the modern healthcare with routine surgery or chemotherapy treatment is considered too perilous, because there are no drugs available to prevent or treat bacterial infections. Unless new antibiotics and therapeutics are developed, that will become a reality. Furthermore, in a potential post-antibiotic era, where antibiotic resistance is currently a major global healthcare problem in the 21st century[168], there will be an even greater need for continuous surveillance, rapid diagnosis and real time tracking of infectious diseases.

Putting it in a very current perspective, it includes tracking of new emerging infectious diseases implemented in the routine care producing genomic data, that can reveal critical epidemiological aspects of an outbreak and allows the dynamics of an epidemic, to be followed and actually shared between labs and countries as brought up in 2017 after the Ebola and Zika epidemics[169] and in the Covid-19 pandemic. They are just some of the few examples where pathogens spread undetected in a population before being identified, which dramatically increased their societal impact. Even if the opportunities for global surveillance of infectious diseases, with the development of new methods for rapid identification and tracking, are huge, I believe it takes worldwide decision makers as FAO, WHO or OECS together with political decisions are required to enable sharing of data on a global scale to make a difference[170].

Accurate identification and characterization of the causative pathogen is crucial to make well-informed decisions in the clinic, to efficiently treat and prevent bacterial infections. Current methods often rely on cultivation of the bacteria, making them time-consuming. Therefore, there is a need for rapid and accurate tools for identification and characterization of bacteria. The assays in **Paper II** and **Paper III** were shown to accurately type bacteria, both at the species level and for higher taxonomical resolutions than species. Hence, it has the potential to rapidly provide comprehensive diagnostics information, thereby optimizing early antibiotic treatment and enable future precision medicine management. For the ODM bacterial typing to be a reliable alternative for clinical diagnostics, it would benefit from improving the DNA extraction protocol by enabling extraction from faecal and blood samples. Increasing the throughput from multiplexing[171] and automated nanofluidic experiments would also simplify the method and speed it up significantly.

Antibiotic resistance rapidly spreads, between bacteria and bacterial species, through horizontal gene transfer of plasmids harboring resistance genes. The assays in **Paper III** and **Paper IV** enabled identification of antibiotic carriage on plasmids by detection of CRISPR-Cas9 induced cleavage. In **Paper III** the plasmid characterization and AMR detection was done simultaneously as the bacterial typing. Ideally, detection of AMR-genes, not only on plasmids, but on chromosomes would also be performed to obtain a full resistance profile of the sample.

Since AMR prevalence is correlated to healthcare infrastructure, social hygiene practices and the economics of a country, there is a need for development of new methods that are applicable in resource limited settings. The assay in **Paper IV** was designed to be used with simple fluorescence microscopes, with the setup already present in many African countries via a tuberculosis diagnosis program. The simplicity of the assay enables the potential for it to be developed into a tool for plasmid microbiology, epidemiology, and AMR-gene detection.

To conclude, the work presented in this thesis demonstrates how optical DNA mapping of single molecules provides a versatile tool that enables to shine light upon the amount and type of DNA damage lesions that are induced by certain chemotherapy drugs, enable rapid pathogen identification and plasmid AMR-gene detection.

7 ACKNOWLEDGEMENTS

First and foremost, I must thank my main supervisor, **Fredrik Westerlund**, for taking me in all those years ago and introducing me to a corner of the deep world of DNA analysis. Thank you, not only for the science that you have brought me, but also for being a nice friend, for being there when times were not always easy, at times for just stating that you have been as lost as I was while looking at some results and not least, for your great taste in music. And I am so sorry for scaring you at that Christian Kjellvander concert.

My examiner **Pernilla Wittung Stafshede**, if you would not have been around with your positive energy and never-ending support, I highly doubt that I would finish this PhD at all.

Erik Kristiansson and **Anna Johnning** for not only the great bioinformatic work during the collaborations and continuous points on antibiotic resistance throughout all my PhD years, but for you being both great people and scientists.

Tobias Ambjörnsson and his group at Lund, especially **Albertas Dvirnas**, for all the friendly meetings, for the never-ending questions about how things are done or how we can change the image analysis for yet another improvement or application.

Linus Sandegren and **Christian G Giske** for all nice discussions, the answers that you have given and for sharing your insight and "real world" experiences of the medical setting for yet another academic.

Pegah Johansson, **Ola Hammarsten**, the hematology staff and the rest of the people at Klinisk Kemi at Sahlgrenska for all the help in the lab, access to blood and sharing your knowledge from just the basics of cell culturing to the big field of DNA damage.

All other co-authors, thanks for all discussions and help going forward with the various projects.

All former and current people of the Westerlund group, at Chemical Biology and at Physical Chemistry, it has been a pleasure to get to know you. Thank you for all the help in the labs, for the floorball games, for the nice retreats, for venting and sharing yet another fruit, coffee, lunch, fika or afterwork together.

To the one that in some sort of way started this journey of mine so many years ago, **Per Paulsrud**. Thank you for sparking an interest. Your enthusiasm and eccentric explanations live on with so many of your students. Hopefully all the bus trips to Chalmers with the science students still brings a smile to your face, the memory of when we went there together, my first trip to Chalmers of many, still makes me laugh. Finally, I really hope that you see some improvements from that old lab report, “sönderdelning av natriumvätekarbonat”, of mine.

To all my friends, to those that came into my life years and years ago and that stayed a chosen family, and the ones that I have gotten to know these last years, thank you for being there. For all dinners, concerts, coffee-tastings, mushroom forest walks, badminton or boardgames and for showing that work is merely work.

To my beloved family, **Gun-Marie**, **Kerstin**, **Kim**, **Linnea**, **Caroline** and my nephews, thank you for your shared enthusiasm for the things about my work that is so abstract. Thank you, **Gun-Marie,** for sharing your stories of when you went with Ubban and Sten to Chalmers, you of course only for the parties. Thank you for sharing yet another Chalmers-party with me so many years later. Thank you all, for the joy that the structure of a DNA molecule brings and for every unclear image of lines and dots that I have showed you throughout the years. Thank you for asking for yet another explanation that explains how bacteria in a culture are just like if us humans would live life sitting in a hot tub full of soup. Finally, thank you for all the love I could ask for.

To my beloved partner **Måns**, thank you for staying a very stable part throughout these years. All our cold swims probably helped keeping me (somewhat) sane. Thank you for your kindness, weirdness, laughter, all adventures, and the everyday happiness that you bring. I am extremely grateful for that, and that I get to share my days with you.

8 REFERENCES

1 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

2 Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica* **6**, 673-677 (1953).

3 Franklin, R. E. & Gosling, R. G. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172**, 156-157 (1953).

4 Venter, J. C. *et al.* The sequence of the human genome. *Science (New York, N.Y.)* **291**, 1304-1351 (2001).

5 Gibbs, R. A. The human genome project changed everything. *Nature Reviews Genetics* **21**, 575-576 (2020).

6 Nurk, S. *et al.* The complete sequence of a human genome. *Science (New York, N.Y.)* **376**, 44-53 (2022).

7 Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179-189 (2020).

8 Hoeijmakers, J. H. DNA damage, aging, and cancer. *New England Journal of Medicine* **361**, 1475-1485 (2009).

9 Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. *British journal of experimental pathology* **10**, 226-236 (1929).

10 World Health Organization. Antimicrobial resistance: global report on surveillance. (2014).

11 Fleming, S. A. (Nobel Prize Outreach AB NobelPrize.org, 2023).

12 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).

13 Heinemann, U. & Roske, Y. Symmetry in nucleic-acid double helices. *Symmetry* **12**, 737 (2020).

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

15 Bloomfield, V. A. C., Donald M. & Tinoco, I. *Nucleic Acids: Structure, Properties, and Functions*. (University Science Books, 2000).
16 Stofer, E. & Lavery, R. Measuring the geometry of DNA grooves. *Biopolymers: Original Research on Biomolecules* **34**, 337-346 (1994).
17 Lu, X.-J., Shakked, Z. & Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *Journal of molecular biology* **300**, 819-840 (2000).
18 Harteis, S. & Schneider, S. Making the bend: DNA tertiary structure and protein-DNA interactions. *International journal of molecular sciences* **15**, 12335-12363 (2014).
19 Dickerson, R., Drew, H., Conner, B., Kopka, M. & Pjura, P. in *Cold Spring Harbor symposia on quantitative biology.* 13-24 (Cold Spring Harbor Laboratory Press).
20 Crick, F. H. On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138-163 (1958).
21 Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC research notes* **12**, 1-5 (2019).
22 Omenn, G. S. Reflections on the hupo human proteome project, the flagship project of the human proteome organization, at 10 years. *Mol Cell Proteomics* **20** (2021).
23 Kremer, B. *et al.* A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. *New England Journal of Medicine* **330**, 1401-1406 (1994).
24 Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods Primers* **1**, 1-21 (2021).
25 Wang, K. *et al.* Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *The American Journal of Human Genetics* **84**, 399-405 (2009).
26 Moschen, A. R., Tilg, H. & Raine, T. IL-12, IL-23 and IL-17 in IBD: immunobiology and therapeutic targeting. *Nature Reviews Gastroenterology & Hepatology* **16**, 185-196 (2019).
27 Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078 (2009).
28 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* **17**, 487-500 (2016).
29 Holliday, R. & Pugh, J. E. DNA Modification Mechanisms and Gene Activity During Development: Developmental clocks may depend on the enzymic modification of specific bases in repeated DNA sequences. *Science (New York, N.Y.)* **187**, 226-232 (1975).
30 Kelner, A. Photoreactivation of ultraviolet-irradiated Escherichia coli, with special reference to the dose-reduction principle and to ultraviolet-induced mutation. *J Bacteriol* **58**, 511-522 (1949).
31 Friedberg, E. C. A brief history of the DNA repair field. *Cell research* **18**, 3-7 (2008).
32 Lord, C. J. & Ashworth, A. The DNA damage response and cancer therapy. *Nature* **481**, 287-294 (2012).
33 Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Molecular cell* **40**, 179-204 (2010).
34 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. *Annual review of biochemistry* **79**, 181 (2010).
35 Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell research* **18**, 99-113 (2008).

36 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).
37 Caldecott, K. W. DNA single-strand break repair and spinocerebellar ataxia. *Cell* **112**, 7-10 (2003).
38 Jiricny, J. The multifaceted mismatch-repair system. *Nature reviews Molecular cell biology* **7**, 335-346 (2006).
39 David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941-950 (2007).
40 Roh, D. S. *et al.* DNA cross-linking, double-strand breaks, and apoptosis in corneal endothelial cells after a single exposure to mitomycin C. *Investigative ophthalmology & visual science* **49**, 4837-4843 (2008).
41 Jordan, P. Molecular mechanisms involved in cisplatin cytotoxicity. *Cellular and Molecular Life Sciences CMLS* **57**, 1229-1235 (2000).
42 Fichtinger-Schepman, A. M. J., Van der Veer, J. L., Den Hartog, J. H., Lohman, P. H. & Reedijk, J. Adducts of the antitumor drug cis-diamminedichloroplatinum (II) with DNA: formation, identification, and quantitation. *Biochemistry* **24**, 707-713 (1985).
43 Murata, M., Suzuki, T., Midorikawa, K., Oikawa, S. & Kawanishi, S. Oxidative DNA damage induced by a hydroperoxide derivative of cyclophosphamide. *Free Radical Biology and Medicine* **37**, 793-802 (2004).
44 Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. *Nature reviews cancer* **11**, 467-480 (2011).
45 Hashimoto, S., Anai, H. & Hanada, K. Mechanisms of interstrand DNA crosslink repair and human disorders. *Genes and Environment* **38**, 1-8 (2016).
46 Wassenaar, T. M. *Bacteria: The Benign, the Bad, and the Beautiful*. (Wiley-Blackwell, 2011).
47 Kothari, A. *et al.* Large circular plasmids from groundwater plasmidomes span multiple incompatibility groups and are enriched in multimetal resistance genes. *MBio* **10**, e02899-02818 (2019).
48 Madigan, M. T., Martinko, J. M. & Parker, J. *Brock biology of microorganisms*. Vol. 11 (Pearson Prentice Hall Upper Saddle River, NJ, 2006).
49 Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**, 472-482 (2015).
50 Breijyeh, Z., Jubeh, B. & Karaman, R. Resistance of gram-negative bacteria to current antibacterial agents and approaches to resolve it. *Molecules* **25**, 1340 (2020).
51 Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *World Health Organization* (2017).
52 Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.)* **327**, 167-170 (2010).
53 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genetics* **11**, 181-190 (2010).
54 Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)* **346**, 1258096 (2014).
55 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
56 Rawashdeh, O., Rawashdeh, R. Y., Kebede, T., Kapp, D. & Ralescu, A. Bio-informatic analysis of CRISPR protospacer adjacent motifs (PAMs) in T4 genome. *BMC Genomic Data* **23**, 1-10 (2022).
57 Amábile-Cuevas, C. F. *Antibiotics and antibiotic resistance in the environment*. (CRC Press/Balkema is, 2016).

58 Stokes, H. W. & Gillings, M. R. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS microbiology reviews* **35**, 790-819 (2011).
59 Walsh, C. *Antibiotics: actions, origins, resistance*. (American Society for Microbiology (ASM), 2003).
60 Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423-427 (2003).
61 Nelson, H., Finch, J. T., Luisi, B. F. & Klug, A. The structure of an oligo (dA)· oligo (dT) tract and its biological implications. *Nature* **330**, 221-226 (1987).
62 Odijk, T. Polyelectrolytes near the rod limit. *Journal of Polymer Science: Polymer Physics Edition* **15**, 477-483 (1977).
63 Skolnick, J. & Fixman, M. Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules* **10**, 944-948 (1977).
64 Reisner, W., Pedersen, J. N. & Austin, R. H. DNA confinement in nanochannels: physics and biological applications. *Reports on progress in physics* **75**, 106601 (2012).
65 Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of λ-phage DNA. *Science (New York, N.Y.)* **265**, 1599-1600 (1994).
66 Dorfman, K. D., King, S. B., Olson, D. W., Thomas, J. D. P. & Tree, D. R. Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching. *Chemical Reviews* **113**, 2584-2667 (2013).
67 Daoud, M. & De Gennes, P. G. Statistics of macromolecular solutions trapped in small pores. *J. Phys. France* **38**, 85-93 (1977).
68 De Gennes, P.-G. & Gennes, P.-G. *Scaling concepts in polymer physics*. (Cornell university press, 1979).
69 Odijk, T. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules* **16**, 1340-1344 (1983).
70 Odijk, T. Similarity applied to the statistics of confined stiff polymers. *Macromolecules* **17**, 502-503 (1984).
71 Wang, Y., Tree, D. R. & Dorfman, K. D. Simulation of DNA Extension in Nanochannels. *Macromolecules* **44**, 6594-6604 (2011).
72 Werner, E. & Mehlig, B. Confined polymers in the extended de Gennes regime. *Physical review E* **90**, 062602 (2014).
73 Atkins, P. & Jones, L. Chemical Principles: The Quest for Insight. (2004).
74 Bohr, N. I. On the constitution of atoms and molecules. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **26**, 1-25 (1913).
75 Atkins, P., De Paula, J. & Friedman, R. *Quanta, matter, and change: a molecular approach to physical chemistry*. (Oxford University Press, USA, 2009).
76 Kasha, M. Characterization of electronic transitions in complex molecules. *Discussions of the Faraday Society* **9**, 14-19 (1950).
77 Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Springer, 2006).
78 Cheng, Z., Kuru, E., Sachdeva, A. & Vendrell, M. Fluorescent amino acids as versatile building blocks for chemical biology. *Nature Reviews Chemistry* **4**, 275-290 (2020).
79 Umezawa, K., Nakamura, Y., Makino, H., Citterio, D. & Suzuki, K. Bright, color-tunable fluorescent dyes in the visible− near-infrared region. *Journal of the American Chemical Society* **130**, 1550-1551 (2008).
80 Sauer, M., Hofkens, J. & Enderlein, J. *Handbook of fluorescence spectroscopy and imaging: from ensemble to single molecules*. (John Wiley & Sons, 2010).
81 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).

82 Glazer, A. N. & Rye, H. S. Stable dye–DNA intercalation complexes as reagents for high-sensitivity fluorescence detection. *Nature* **359**, 859-861 (1992).
83 Donnelly, W. J. & Roorda, A. Optimal pupil size in the human eye for axial resolution. *JOSA A* **20**, 2010-2015 (2003).
84 Kriss, T. C. & Kriss, V. M. History of the operating microscope: from magnifying glass to microneurosurgery. *Neurosurgery* **42**, 899-907 (1998).
85 Meschede, D. *Optics, light and lasers: the practical approach to modern aspects of photonics and laser physics*. (John Wiley & Sons, 2017).
86 Rayleigh, L. XXXI. Investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **8**, 261-274 (1879).
87 Sanger, F. *et al.* Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**, 687-695 (1977).
88 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **74**, 5463-5467 (1977).
89 Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**, 560-564 (1977).
90 Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology* **94**, 441-448 (1975).
91 Schuster, S. C. Next-generation sequencing transforms today's biology. *Nature methods* **5**, 16-18 (2008).
92 Bertelli, C. & Greub, G. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. *Clinical Microbiology and Infection* **19**, 803-813 (2013).
93 Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
94 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genetics* **11**, 31-46 (2010).
95 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature biotechnology* **26**, 1135-1145 (2008).
96 Hon, T. *et al.* Highly accurate long-read HiFi sequencing data for five complex genomes. *Scientific data* **7**, 1-11 (2020).
97 Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology* **36**, 338-345 (2018).
98 Deschamps, S. *et al.* A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping. *Nature communications* **9**, 1-10 (2018).
99 Muller, V. & Westerlund, F. Optical DNA mapping in nanofluidic devices: principles and applications. *Lab on a Chip* **17**, 579-590 (2017).
100 Frykholm, K., Müller, V., Sriram, K., Dorfman, K. D. & Westerlund, F. DNA in Nanochannels–Theory and Applications. *Quarterly Reviews of Biophysics*, 1-83 (2022).
101 Neely, R. K., Deen, J. & Hofkens, J. Optical mapping of DNA: single-molecule-based methods for mapping genomes. *Biopolymers* **95**, 298-311 (2011).
102 Smith, S. B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science (New York, N.Y.)* **258**, 1122-1126 (1992).
103 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).

104 Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science (New York, N.Y.)* **271**, 795-799 (1996).
105 Bensimon, A. *et al.* Alignment and sensitive detection of DNA by a moving interface. *Science (New York, N.Y.)* **265**, 2096-2098 (1994).
106 Michalet, X. *et al.* Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science (New York, N.Y.)* **277**, 1518-1523 (1997).
107 Nazari, Z. E. & Gurevich, L. Molecular combing of DNA: methods and applications. *Journal of Self-assembly and Molecular Electronics (same)* **1**, 125-148 (2013).
108 Schwartz, D. C. *et al.* Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. *Science (New York, N.Y.)* **262**, 110-114 (1993).
109 Deen, J. *et al.* Combing of genomic DNA from droplets containing picograms of material. *ACS nano* **9**, 809-816 (2015).
110 Ladoux, B. & Doyle, P. Stretching tethered DNA chains in shear flow. *EPL (Europhysics Letters)* **52**, 511 (2000).
111 Allemand, J., Bensimon, D., Jullien, L., Bensimon, A. & Croquette, V. pH-dependent specific binding and combing of DNA. *Biophysical journal* **73**, 2064-2070 (1997).
112 Benke, A., Mertig, M. & Pompe, W. pH-and salt-dependent molecular combing of DNA: experiments and phenomenological model. *Nanotechnology* **22**, 035304 (2010).
113 Goyal, G. *et al.* A simple cut and stretch assay to detect antimicrobial resistance genes on bacterial plasmids by single-molecule fluorescence microscopy. *Scientific reports* **12**, 1-12 (2022).
114 Kaykov, A., Taillefumier, T., Bensimon, A. & Nurse, P. Molecular combing of single DNA molecules on the 10 megabase scale. *Scientific reports* **6**, 1-9 (2016).
115 Otobe, K. & Ohtani, T. Behavior of DNA fibers stretched by precise meniscus motion control. *Nucleic acids research* **29**, e109-e109 (2001).
116 Oshige, M. *et al.* A new DNA combing method for biochemical analysis. *Analytical biochemistry* **400**, 145-147 (2010).
117 Jing, J. *et al.* Automated high resolution optical mapping using arrayed, fluid-fixed DNA molecules. *Proceedings of the National Academy of Sciences* **95**, 8046-8051 (1998).
118 Tegenfeldt, J. O. *et al.* Micro-and nanofluidics for DNA analysis. *Analytical and bioanalytical chemistry* **378**, 1678-1692 (2004).
119 Tegenfeldt, J. O. *et al.* The dynamics of genomic-length DNA molecules in 100-nm channels. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10979-10983 (2004).
120 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).
121 Gupta, D. *et al.* Experimental Evidence of Weak Excluded Volume Effects for Nanochannel Confined DNA. *ACS Macro Letters* **4**, 759-763 (2015).
122 Lam, E. T. *et al.* Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature biotechnology* **30**, 771-776 (2012).
123 Cao, H., Tegenfeldt, J. O., Austin, R. H. & Chou, S. Y. Gradient nanostructures for interfacing microfluidics and nanofluidics. *Applied Physics Letters* **81**, 3058-3060 (2002).
124 Freitag, C. *et al.* Visualizing the entire DNA from a chromosome in a single frame. *Biomicrofluidics* **9**, 044114 (2015).

125 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).
126 Öz, R., Sriram, K. & Westerlund, F. A nanofluidic device for real-time visualization of DNA–protein interactions on the single DNA molecule level. *Nanoscale* **11**, 2071-2078 (2019).
127 Jo, K. *et al.* A single-molecule barcoding system using nanoslits for DNA analysis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2673-2678 (2007).
128 Das, S. K. *et al.* Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Research* **38**, e177 (2010).
129 Hastie, A. R. *et al.* Rapid Genome Mapping in Nanochannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome. *PLoS ONE* **8**, e55864 (2013).
130 McCaffrey, J. *et al.* CRISPR-CAS9 D10A nickase target-specific fluorescent labeling of double strand DNA for whole genome mapping and structural variation analysis. *Nucleic Acids Research* **44**, e11 (2016).
131 Zirkin, S. *et al.* Lighting up individual DNA damage sites by in vitro repair synthesis. *Journal of the American Chemical Society* **136**, 7771-7776 (2014).
132 Klimasauskas, S. & Weinhold, E. A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends in biotechnology* **25**, 99-104 (2007).
133 Lukinavičius, G. *et al.* Targeted Labeling of DNA by Methyltransferase-Directed Transfer of Activated Groups (mTAG). *Journal of the American Chemical Society* **129**, 2758-2759 (2007).
134 Grunwald, A. *et al.* Bacteriophage strain typing by rapid single molecule analysis. *Nucleic Acids Research* **43**, e117 (2015).
135 Neely, R. K. *et al.* DNA fluorocode: A single molecule, optical map of DNA with nanometre resolution. *Chemical Science* **1**, 453-460 (2010).
136 Vranken, C. *et al.* Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. *Nucleic Acids Research* **42**, e50-e50 (2014).
137 Lim, S. F. *et al.* DNA methylation profiling in nanochannels. *Biomicrofluidics* **5**, 34106-341068 (2011).
138 Sharim, H. *et al.* Long-read single-molecule maps of the functional methylome. *Genome research* (2019).
139 Levy-Sakin, M. *et al.* Toward single-molecule optical mapping of the epigenome. *ACS nano* **8**, 14-26 (2014).
140 Michaeli, Y. *et al.* Optical detection of epigenetic marks: sensitive quantification and direct imaging of individual hydroxymethylcytosine bases. *Chemical Communications* **49**, 8599-8601 (2013).
141 Gabrieli, T. *et al.* Epigenetic Optical Mapping of 5-Hydroxymethylcytosine in Nanochannel Arrays. *ACS nano* **12**, 7148-7158 (2018).
142 Reisner, W. *et al.* Single-molecule denaturation mapping of DNA in nanofluidic channels. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13294-13299 (2010).
143 Nyberg, L. K. *et al.* A single-step competitive binding assay for mapping of single DNA molecules. *Biochemical and biophysical research communications* **417**, 404-408 (2012).
144 Müller, V. *et al.* Enzyme-free optical DNA mapping of the human genome using competitive binding. *Nucleic Acids Research* **47**, e89-e89 (2019).

145 Formenti, G. *et al.* SMRT long reads and Direct Label and Stain optical maps allow the generation of a high-quality genome assembly for the European barn swallow (Hirundo rustica rustica). *Gigascience* **8**, giy142 (2019).
146 Barseghyan, H. *et al.* Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome medicine* **9**, 1-11 (2017).
147 Jaratlerdsiri, W. *et al.* Next generation mapping reveals novel large genomic rearrangements in prostate cancer. *Oncotarget* **8**, 23588 (2017).
148 Singh, V. *et al.* Quantifying DNA damage induced by ionizing radiation and hyperthermia using single DNA molecule imaging. *Translational oncology* **13**, 100822 (2020).
149 Singh, V., Johansson, P., Lin, Y.-L., Hammarsten, O. & Westerlund, F. Shining light on single-strand lesions caused by the chemotherapy drug bleomycin. *DNA repair* **105**, 103153 (2021).
150 Singh, V. *et al.* Quantification of single-strand DNA lesions caused by the topoisomerase II poison etoposide using single DNA molecule imaging. *Biochemical and biophysical research communications* **594**, 57-62 (2022).
151 Torchinsky, D., Michaeli, Y., Gassman, N. R. & Ebenstein, Y. Simultaneous detection of multiple DNA damage types by multi-colour fluorescent labelling. *Chemical Communications* **55**, 11414-11417 (2019).
152 Gilat, N. *et al.* From Single-molecule to genome-wide mapping of DNA lesions: Repair-assisted damage detection sequencing. *Biophysical reports* **1**, 100017 (2021).
153 McHugh, P. J., Spanswick, V. J. & Hartley, J. A. Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *The lancet oncology* **2**, 483-490 (2001).
154 Niedernhofer, L. J. *et al.* The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Molecular and cellular biology* **24**, 5776-5787 (2004).
155 Murchie, A. I. & Lilley, D. M. T4 endonuclease VII cleaves DNA containing a cisplatin adduct. *Journal of molecular biology* **233**, 77-85 (1993).
156 Matushek, M. G., Bonten, M. J. & Hayden, M. K. Rapid preparation of bacterial DNA for pulsed-field gel electrophoresis. *J Clin Microbiol* **34**, 2598-2600 (1996).
157 Åkerman, B. & Tuite, E. Single-and double-strand photocleavage of DNA by YO, YOYO and TOTO. *Nucleic acids research* **24**, 1080-1090 (1996).
158 Noble, C. *et al.* A Fast and Scalable Kymograph Alignment Algorithm for Nanochannel-Based Optical DNA Mappings. *PLoS ONE* **10**, e0121905 (2015).
159 Nilsson, A. N. *et al.* Competitive binding-based optical DNA mapping for fast identification of bacteria--multi-ligand transfer matrix theory and experimental applications on Escherichia coli. *Nucleic Acids Research* **42**, e118 (2014).
160 Dvirnas, A. *et al.* Facilitated sequence assembly using densely labeled optical DNA barcodes: A combinatorial auction approach. *PLOS ONE* **13**, e0193900 (2018).
161 Nicolas-Chanoine, M.-H., Bertrand, X. & Madec, J.-Y. Escherichia coli ST131, an intriguing clonal group. *Clinical microbiology reviews* **27**, 543-574 (2014).
162 Decano, A. G. & Downing, T. An Escherichia coli ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. *Scientific reports* **9**, 1-13 (2019).
163 Boulund, F. *et al.* Typing and Characterization of Bacteria Using Bottom-up Tandem Mass Spectrometry Proteomics. *Mol Cell Proteomics* **16**, 1052-1063 (2017).
164 Sloan, A., Wang, G. & Cheng, K. Traditional approaches versus mass spectrometry in bacterial identification and typing. *Clin Chim Acta* **473**, 180-185 (2017).

165 Li, W., Raoult, D. & Fournier, P.-E. Bacterial strain typing in the genomic era. *FEMS microbiology reviews* **33**, 892-916 (2009).
166 Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature biotechnology* **37**, 783-792 (2019).
167 van Belkum, A. *et al.* Innovative and rapid antimicrobial susceptibility testing systems. *Nature Reviews Microbiology* **18**, 299-311 (2020).
168 Alanis, A. J. Resistance to antibiotics: are we in the post-antibiotic era? *Archives of medical research* **36**, 697-705 (2005).
169 Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics* **19**, 9-20 (2018).
170 Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics* **13**, 601-612 (2012).
171 Kk, S. *et al.* A Parallelized Nanofluidic Device for High-Throughput Optical DNA Mapping of Bacterial Plasmids. *Micromachines* **12**, 1234 (2021).