1 | INTRODUCTION

What a book a devil's chaplain might write on the clumsy, wasteful, blundering, low, and horribly cruel work of nature!
- Charles Darwin

The above quote of Charles Darwin was used in correspondence with his friend Joseph Hooker. It expresses fascination towards small, undirected changes that occur in Nature, while the resulting evolution seems almost directed. In this introduction, I will first describe how Darwin came to express his fascination for Nature and formulate a theory of evolution. Next, I will illustrate how the same quote could be used for describing processes at the smallest scale of life. In particular, I will explain how this applies to an essential process in a cell: DNA replication, which is at the main focus of this thesis. Finally, I will introduce the relatively new field of single-molecule biophysics to the essential basics for understanding of the work presented here.

1.1 Theory of evolution by Charles Darwin

In 1859 Darwin published his book "On the origin of species" [34]. In this book, Darwin explained the diversity of life on Earth with a theory of evolution. The English naturalist and geologist came to the evolution theory following observations he made more than 20 years before, during his voyage with the HMS Beagle to South America. There he had noted significant differences in wildlife due to spatial separation. For example, the Andes separates East from West of the continent or islands lead to isolated ecosystems. These different environments caused otherwise closely related species to be slightly different. Visiting the Galapagos Islands near the end of the voyage undoubtedly contributed greatly to the inception of his evolution theory.

On the Galapagos Islands, separated from each other by only tens of kilometers, Darwin observed a great variety of mockingbirds. At first, these mockingbirds resembled those on the mainland of Chile. At the same time, Darwin noted significant differences between the birds from the various islands. Also other flora and fauna drew Darwin’s attention to variety. For example, someone on the islands informed Darwin that ”that the tortoises differed from the different islands, and that he could with certainty tell from which island any one was brought” [34]. At first, Darwin had not paid attention to this statement.

Back in England, he had an illustrator make sketches of the different specimens he brought (Figure 1.1). Using these illustrations, Darwin related the differences in species to the geography and isolation of the islands. This triggered him to challenge the concept of stable species with a theory of one species evolving into another species. Darwin
hypothesized that population growth resulted in pressure on the population, causing "survival of the fittest". Individuals in the population would have slightly different traits and therefore different rates of survival and reproduction. Natural selection would select those that were fittest to the selective conditions to produce progeny with the same traits. Multiple adaptations could then lead to the coming into existence of a new species.

Despite that the concept of one species evolving into a new species was absolutely new at the time, it was fairly quickly accepted. Natural selection as the sole driving force of evolution, however, was not. A missing link in the theory of evolution was the mechanism of evolution or how the variety of traits would be heritable for progeny. Darwin proposed a unifying theory, pangenesis, in which all traits of the parents could be transferred to offspring. Around the same time, George Mendel conducted experiments with pea plants and showed that certain traits were inheritable according to clear patterns [122]. His work at first did not receive much attention. At the beginning of the 20th century, however, it was realized that it appeared to be the missing link for Darwin’s theory of evolution. Mendel’s seminal work still forms the basis of modern genetics.

During the 1920s and 1930s, the works of Darwin and Mendel were finally combined in one grant theory, the modern evolutionary synthesis, stating that evolution occurs in small steps in random direction. Each step has a low probability of occurring, but is at the same time non-reversible. Very rarely, a small change creates an advantage to the selective condition. Consequently, this adaptation creates stronger progeny and

Figure 1.1
Illustration of two different mockingbirds found by Darwin made by John Gould [59].
is maintained in the population. All these low-probability advantageous adaptations together can explain the big differences found between species around the world.

1.2 Heritable traits

The physical basis of the heritability of traits, however, was still missing. After the combination of theories by Darwin and Mendel, it was profoundly clear that traits were transferred to progeny. The term gene was posed to represent one heritable unit. In the 1940s, deoxyribonucleic acid (DNA) was recognized to be the heritable substance [8]. A monumental step that founded the field of molecular biology and biophysics came with the discovery of the structure of DNA in 1953. Francis Crick and James Watson, with support from Rosalind Franklin and Maurice Wilkins, published the structure of DNA [176]. DNA was readily recognized to be a long polymer which harbored genes, the heritable units. In a way, DNA contains the "blueprint" of life: all information a cell might need is described in the DNA. The mechanism of actual inheritance was still unclear. In order to create progeny with the same traits, the DNA has to be transferred to the progeny. To understand the mechanism of this process, we first have to take a closer look at the DNA molecule.

1.2.1 Deoxyribonucleic acid

DNA is a polymer built of only four different nucleotides: Adenine, Thymine, Cytosine and Guanine. Each nucleotide consists of a sugar ring, a phosphate group and a base (Figure 1.2). The backbone of the polymer is formed by the alternating sugar and phosphate group. Nucleotides only differ in the composition of the base, which is either a purine (A and G) or a pyrimidine (T and C). DNA has two polymers—so-called strands—that form the famous double-helical structure. The two strands are connected to each other by hydrogen bonds that cause base-to-base interaction. These interactions are exclusively formed by with the complementary nucleotide: A only pairs with T, whereas C only pairs with G (Figure 1.2).

An interesting feature of DNA is that the alternating structure of the backbone causes polarity of a strand. The five carbon atoms of the sugar group are named 1’ to 5’. The phosphate group is at the 5’-end of the sugar, whereas the OH-group is located at the 3’-position of the sugar ring. Hence, the DNA direction is either 5’-3’ or 3’-5’. The two strands are oriented in antiparallel fashion so that the 5’-end of one strand is aligned with 3’-end of the complementary strand.

The complementary base formation will prove essential for the copying of DNA. It means that the sequence of only one strand of DNA is enough to know how the missing strand should look like. As we will read shortly, Nature uses this feature to make a copy of DNA for the progeny.
The structure of DNA is a double helix made of two strands. A strand is a long polymer of nucleotides. The backbone of a strand consists of an alternating sugar ring and phosphate group. The two strands, with opposite directions, interact with each other by base pairs that are directed inwards. These base pairs are exclusively formed between complementary bases. Adenine (A) pairs with thymine (T) and guanine (G) forms a base pair with cytosine (C). An A-T base pair forms two hydrogen bonds, while G-C forms three which makes the latter a more stable base pair. The interaction between the strands causes the location of the strands to not be fully symmetrical. As a result the space between the strands along the DNA has two different distances: the major groove and the minor groove. Image credit: Wikipedia.
1.2.2 From DNA to RNA to protein

DNA harbors the blueprint of life. This means that genes contain a code that can be read and translated into necessary actions. In a cell, proteins perform most functions. Many are involved in the metabolism of nutrients, while others take care of the well-being of the cell. When a specific protein is needed, a signal is given by the cell. Next, the information is located on the DNA and transcribed into ribonucleic acid (RNA). RNA is similar to DNA, but has one additional oxygen atom in the sugar ring and consists of only one strand. The messenger RNA (mRNA) that is created by an enzyme, RNA polymerase, is an exact copy of the DNA of the gene.

In a subsequent step, the mRNA is translated by a large molecular complex, the ribosome. mRNA is read in three-nucleotide blocks, codons, that code for one of twenty amino acids. The ribosome reads the mRNA per codon and synthesizes a polymer of amino acids. These amino acids are folded into specific structures and eventually form a protein. These three distinct steps form the central dogma of molecular biology, as formulated by Francis Crick. The flow of information is unidirectional. At present, the unidirectionality of the flow is still acknowledged, although there is more and more evidence of communication in the opposite direction as well.

There are many steps between the reading of stored information and the actual protein. Slight changes during the production of a protein could cause protein-to-protein variation. These steps include, for example, the reading of the information and synthesis into mRNA, the reading of the mRNA and synthesis of amino acid chain, and the folding into the mature protein. Therefore, not every protein is necessarily the same, even though it has been made from the same blueprint. Consequently, not every cell in an organism can be the same, let it alone that it would be possible for two individuals of one species to be exactly the same! Again, this illustrates the randomness in Nature—even though the blueprint is the same, small modifications and alterations in the subsequent processes can cause differences. In the next section, I will focus on the heritability of DNA. More precisely, I will focus on the process of duplication of DNA. First, an overview of the overall process is given. Eventually, remaining questions in DNA replication are briefly discussed.

1.3 DNA replication

In the past fifty years, molecular biology has greatly contributed to our understanding of DNA and the mechanism of heritability. DNA contains all information needed for making cell components. An essential step is therefore the transfer of this information to a new cell: a cell needs to copy its DNA when it is going to duplicate and create a daughter cell. Quickly after the structure DNA was discovered, it was found that DNA is copied in a semi-conservative manner—DNA in the daughter cells have one original, parental strand, while the other strand is newly synthesized [123]. As described in the previous section, this process takes advantage of the complementary structure of DNA:
the parental strand is used as a template so that at each position the complementary base can be inserted for the new strand.

**Figure 1.3**

**DNA replication** For DNA replication, the double-stranded DNA helix is first unwound by a *DNA helicase* into the two single strands of DNA. These can be used as a template for synthesizing the complementary strands. This is done by another enzyme, *DNA polymerase* (DNAp). DNAp reads the nucleotide of the template strand and adds the complementary nucleotide to the newly synthesizing strand. Addition of a nucleotide occurs only in one direction (5’ to 3’). Consequently only one strand (leading strand) is synthesized continuously, while the other strand is synthesized in Okazaki fragments. After some DNA is unwound, a primer is synthesized to form a launch pad for DNAp. Next, this enzyme catalyzes the addition of nucleotides in the 5’ to 3’ direction until the fragment is finished.

The process of DNA replication requires that the two strands of DNA are first separated into two single strands of DNA (Figure 1.3). These single strands can then be used to synthesize new complementary strands. An enzyme, *DNA helicase*, is needed to break the bonds of double-stranded DNA and unwind the double helix. The single-stranded DNA (ssDNA) is then converted to double-stranded DNA nucleotide by nucleotide. A replicative *DNA polymerase* (DNAp) enzyme catalyzes the addition of nucleotides to the new strand. It does so by selecting the correct, complementary nucleotide from a pool of all four nucleotides. It achieves an astonishing speed and accuracy as we will see in the next section. It is important to note that DNAp can only add nucleotides in the 5’-3’ direction. The antiparallel orientation of the two DNA strands therefore poses a problem to continuous replication of both at the same time: unwinding by a helicase only allows one of the two strands to be synthesized continuously by one DNAp. The other strand is synthesized in short fragments by a second DNAp (Figure 1.3). Each of these so-called Okazaki fragments requires a launch pad for DNAp in the form of primer. A primer is synthesized by *DNA primase* and consists of a few ribonucleotides instead of deoxyribonucleotides. While the second DNAp synthesizes a fragment, the helicase continues downstream, unwinding more DNA. The ssDNA of the discontinuous strand is exposed until a primer is made to start DNA synthesis of the fragment. Single-stranded *DNA binding proteins* (SSB) bind to protect the vulnerable ssDNA.
1.3 DNA replication

DNA helicase, DNA polymerase, DNA primase and single-stranded binding proteins constitute the minimal replication machinery or replisome found in almost all living organisms. In the next sections, I will provide more detail on DNAp and DNA helicase.

1.3.1 Catalyzing DNA synthesis: DNA polymerase

Before books were printed, ancient scribes were recruited to copy a text. A scribe was supposed to copy a text letter by letter, preferably without errors. Since texts with obvious errors would be discarded or deemed invalid, this was a time-consuming task. Despite the utmost care, ancient manuscripts are shown to be poised with errors. As these copied texts themselves would also serve as a template, errors accumulated over time, leading to contradictory passages.

The replication of DNA by DNAp is similar to the copying of a text by a scribe. DNAp copies DNA letter by letter or rather nucleotide by nucleotide. The occurrence of errors has an accumulative effect—an error in the DNA will be passed on to all offspring in subsequent reproductions. Luckily, DNAp has some intricate mechanisms to ensure high fidelity!

The DNAp enzyme resembles a right hand, and opens up to select the complementary nucleotide. Closing of the "fingers" results in the addition of the nucleotide, upon which it moves down the template one step. DNAp has some astonishing specifications with regard to speed and accuracy. The enzyme catalyzes the addition of complementary nucleotides with a speed of up to 500 nucleotides per second. To appreciate this speed we can scale this to real life.

Assume the lines separating highway lanes to be the single-stranded DNA template. These lines are approximately 10 meters apart. Each line has one of four colors, representing the nucleotides found in DNA. A complementary line has to be drawn alongside, while standing on a truck. Now the enzyme’s speed compares to drawing these complementary lines while the truck is traveling at a speed of more than 15,000 km per hour! This amazing speed makes the accuracy even more surprising. While racing along the template, only once every million additions DNAp selects a nucleotide that is not complementary. Stuntingly, the actual errors in a DNA copy are even fewer. DNAp possesses the ability to sense whether the last few added nucleotides were correct. A non-complementary nucleotide causes the double helix to be a little distorted—the nucleotide does not fit properly. When the mistake is sensed, the DNAp removes the last incorporated nucleotide(s), a process that is called proofreading. DNAp restarts the synthesis a few steps back, but now the synthesized DNA is error-free. As a result, the two copies of DNA will be practically without of any error.

1.3.2 Helicase

Access to ssDNA is required for DNAp to synthesize the new complementary strands. So before DNAp can copy DNA, the double helix needs to be unwound. DNA helicase is the enzyme that is essential in most organisms to unwind the double helix. Replicative helicases form a central unit in the replisome. They bind DNA polymerases and
could potentially organize and control DNA replication. One of the most interesting challenges DNA replication faces is that of synchronization. Only one of the two strands can be synthesized continuously. One strand is synthesized discontinuously, in so-called Okazaki fragments. Each of these fragments require initiation of DNA synthesis. Intuitively, this would lead to slower overall synthesis of this strand. The build-up of vulnerable ssDNA is undesirable. In one way or another, the replisome manages to synchronize DNA synthesis. This topic is studied and discussed in more detail in Chapter 5.

1.4 Single-molecule methods and experiments

Much of the knowledge introduced so far has been achieved in classical biological studies. In the classical way, an isolated protein or enzyme of interest is put in a small vial, together with the necessary components for the reaction. After a certain time has passed, the reaction is ended. The resulting product can be detected and quantified. By varying reaction conditions such as protein concentration, temperature and/or time, functional properties of the isolated protein can be studied. A drawback of these bulk experiments is the large quantity of proteins in the reaction—hence the name bulk experiment. Even in the smallest reactions billions of proteins are present. This means that the product of the reaction is an average of the action of all proteins. Such experiments are perfect for studying overall functional properties, but incapable of resolving individual properties of a protein.

Today, individual properties of proteins can be studied using single-molecule techniques. Specific questions about the enzyme’s mechanism and dynamics can thus be answered that would be averaged out in bulk studies. One of these techniques, used for the work presented in this thesis is optical tweezers (Chapter 2). Optical tweezers allow single proteins to be studied by manipulating a single DNA molecule at a time. More details of optical tweezers and their use can be found in Chapter 2. As described in the following chapters, manipulation of single DNA molecules allows single DNAP enzymes or single replisomes to be examined. By observing one single enzyme at a time, I was able to resolve properties of enzymes that have not been observed before. The dynamics of single enzymes revealed that an enzyme’s action is not as deterministic as previously thought—the activity of the enzyme is actually much more stochastic. Instead of progressing down to its goal without hesitation, the path is bumpy, noisy and far from straight. The balance is only slightly tilted towards positive duplication of the DNA. The stochasticity in DNA replication requires that more energy is needed, than the minimum that would be theoretically required, to reach an accurate copy of DNA.

By incepting the ideas of evolution, Charles Darwin immediately noted the wastefulness of Nature, as his quote showed. Observing the stochasticity of the smallest machines in Nature now reveals a similar wastefulness occurs at the smallest scale of life, DNA replication. The waste of energy is apparently required to efficiently balance accuracy and speed needed for sustainable life on Earth.
May a reader of this thesis marvel at the complexity of Nature.