The DNA chain and its sense

The DNA chain and its sense

The DNA chain forms as phosphate groups are shared by subsequent nucleotides, creating phosphodiester bonds (see Figure 4) between them. This sharing is possible when the phosphate group attaches itself to the 3' carbon atom of a pentose sugar and to the 5' atom of the next pentose. The polymer thus formed can incorporate unlimited nucleotides, but it will always retain a sense, marked by the free hydroxyl group at the 3' position of the first nucleotide and the free phosphate group at the 5' position of the last nucleotide. To simplify, biologists typically talk of 3'-5' (antisense) and 5'-3' (sense) senses, and use the later as the default standard for denoting nucleic acid sequences.

The double helix

Although Levene made great insights into the nature of DNA, he sustained the view, readily accepted at the time, that DNA was a repetitive sequence with residual structural functions in the nucleus and that proteins, with their superior ~20-letter alphabet, were the most probable carriers of genetic information. This point of view was also grounded on the fact that Levene found even numbers of all four bases in DNA, leading him to believe it had to be a repetitive sequence. But, although there are indeed many repetitive structures in genomes, he could not be more off the mark. It was not until Oswald Avery proved beyond doubt that DNA was the recipient of hereditary information [Avery1944], that efforts were set on the determination of the exact structure of the DNA molecule, leading to the famous double helix depicted in the 1953 article by Watson & Crick [Watson1953].

Figure 5 - A double stranded DNA sequence that can be read both as TG or CA.
*Watson & Crick* correctly theorized that two DNA sequences might arrange themselves in a double-helix structure if they are complementary. This is due to the fact that complementary purines and pyrimidines can interlock via hydrogen bonds if they come together correctly. Complementary purine-pyrimidine pairs, as described by *Watson & Crick*, can be created between adenine and thymine (involving two hydrogen bonds) and cytosine and guanine (involving three hydrogen bonds). By definition, a complementary double helix of DNA will always be readable in two different senses (see Figure 5), corresponding to the 5’-3’ and 3’-5’ of whichever of the two strands is considered to be the master one. With their hypothesis, Watson & Crick did provide a feasible explanation of DNA replication (each strand acting as a template for a new complementary strand) and they also uncovered the secret of *Levene’s* base number evenness (sc. base pairing).

**Figure 6** - Representation of the three-dimensional structure of DNA-A double helix.

Complementary strands do not only provide the means for replication. The difference in stress provoked by succeeding triple and double hydrogen bonds, among other factors, is the origin of the torsional forces that twist DNA into its double-helix structure and help package DNA information into chromosomes. The double helix (see Figure 6) can attain different configurations (A, B or Z) depending on its composition and on the
surrounding environment. The vast majority of living-cells DNA is in a B-helix configuration.

**Biochemical and chemo-physical properties of DNA**

**Biochemical properties**

Being a polymeric macromolecule, DNA presents many interesting biochemical properties, which are mainly related to its secondary and tertiary configurations. Thus, single-stranded DNA and RNA molecules can present well-defined hairpins, bulges, internal loops and even cross-junctions (see Figure 7) that will make them interact distinctly with different chemical substances. Besides, the extensive co-evolution of nucleic acids and proteins within living cells has provided a complex set of biochemical operators for the manipulation of DNA that is in full use in living cells and which comprises a readily tapped source of tools for molecular geneticists.

![Figure 7 - Typical secondary configurations of DNA](image)

*Figure 7 - Typical secondary configurations of DNA*
Secondary and tertiary DNA structures and bends provide the means for active recognition sites of either protein or RNA enzymes, that is, specific sequences of DNA into which enzymes bind and act. This also provides geneticists with the means for cutting (using cleavage or restriction enzymes), pasting (with ligase enzymes) and even copying (with polymerase-type enzymes) specific sections of DNA. The specificity of an enzyme is defined by its binding site and the length of DNA sequence it recognizes. Thus, a whole range of enzymes and target DNA sites is available for use in molecular genetics: from very specific, sequence-dependant, enzymes to low-specificity, general use, enzymes. This vast armamentarium of biochemical tools can be used in many ways for very different analytical purposes. For example, restriction enzymes are a prime element of DNA-fingerprinting techniques (DNA is cleaved at specific recognition sites and the differences in cleaved-fragment sizes generate a unique fingerprint), while polymerase-type enzymes are the main prerequisite of most amplification techniques (see p.37).

**Chemo-physical properties**

Equally, or even more, interesting to the molecular geneticist are the chemo-physical properties of DNA. Although they do not provide the means for specific manipulation and targeting of DNA sequences, the chemo-physical properties of DNA yield a very interesting set of reactions that can be promptly exploited by molecular geneticists.

**Chemical properties**

A main chemical property of DNA is its selective response to different pH environments. For example, a highly acid medium (pH 1) provokes the hydrolysis of phosphodiester bonds, leaving free bases, sugar and phosphates, while a mildly acid medium (pH 4) will result in the selective breakage of the glycosidic bonds between sugars and purines, allowing for the removal of purines and leaving apurinic acid. Conversely, basic mediums change the polarity of groups involved in the hydrogen bonds, leading to total DNA denaturation (separation of the double strand into two single strands) above pH 11.3.

Salt concentration can also widely affect the nature of DNA. For example, since the phosphates of the DNA backbone have a net negative charge, DNA immersed into distilled water will spontaneously denature into single
stranded DNA due to charge repulsion. On the contrary, salts that dissociate into ions will tend to stabilize the double-helix structure of DNA above a 0.2 M salt concentration. This fact is important, since the strength of the DNA double helix will change its melting or denaturation temperature ($T_m$). In this same sense, the presence of hydrophobic substances, like methanol, will lower the $T_m$ of DNA (weakening the double-helix structure) and thus tend to allow the dissolution of its component bases into the solvent.

**Chemo-physical properties**

DNA also presents a wide range of behaviors in changing physical environments. An already mentioned property, and very important for the amplification of DNA by the polymerase chain reaction (PCR), is the response of DNA to high temperatures. As thermal energy increases, the frequency of hydrogen bond breaking also increases. At high temperatures (typically between 93 and 97 ºC), DNA reaches its denaturing temperature ($T_m$) and the double helix splits into two separate single strands of DNA. It is important to note that, since adenine and thymine form a double hydrogen bond while cytosine and guanine form a triple hydrogen bond (see p.26), the denaturing temperature of DNA will vary depending on the %G+C content of the DNA chain. In other words, the $T_m$ of a DNA chain may be used to infer its %G+C content and vice versa.

This is made possible by another physical property of DNA: UV absorption. DNA bases absorb maximally light of 260 nm, but the degree of light absorption will vary depending on the secondary and tertiary configuration of DNA. For example, free bases absorb 1.60 units at 260 nm (hyperchromic DNA), while single-stranded DNA absorbs only 1.37 units and double-stranded DNA just 1.00 units. This makes it possible to gauge the concentration of DNA in a solution by UV spectrophotometry and also to estimate its nature (single-stranded vs. double-stranded) and thus experimentally assess its denaturing temperature. On the other hand, intense UV light or high-energy particle exposition causes DNA to denature and hydrolyze, leading to efficient surface sterilization techniques.

DNA local composition does also affect its secondary and tertiary structures. Breathing, or the partial denaturation of a DNA fragment, is known to occur commonly in A+T rich areas (with less tough hydrogen bonds) and is an important cofactor in the action of DNA-unwinding
proteins which are, in turn, essential for the action of endonucleases and polymerases (i.e. important aspects of cell division and DNA-repair mechanisms). The %G+C content of a DNA molecule will also affect its density, since G+C pairs are denser than A+T ones. Therefore, %G+C content can also be estimated by density-assessing techniques such as ultra-centrifugation.

Almost all of these properties (%G+C content, salt concentration, etc.) play a key role in defining the renaturation behavior of DNA. Renaturation is the process by which two complementary single-stranded DNA molecules come together into a double-stranded molecule if certain conditions are met, and it is the simplest and supposed replication system of the earliest living beings. Temperature, salt concentration and %G+C content (among other factors) will vary the specificity of the renaturation process ([Wetmur1968], the exactness of the complementary match between the two single-stranded molecules required to glue them together), a fact that is of crucial importance in hybridization (see p.32) and amplification assays (see p.37).

Finally, an already mentioned physical property of DNA (its net negative electrical charge) is essential for yet another analytical technique: electrophoresis (see p.34). The negative charge of the phosphate in each nucleotide does not only provide the DNA molecule with a net negative charge, but also with a linear charge/length ratio. Provided that secondary and tertiary configurations are knocked down (typically by the addition of urea in the medium), linear DNA fragments can be separated on the basis of their charge/length ratio by applying a high electric field in a gel, sieve-like, medium.

1.2.2. ANALYTICAL TECHNIQUES

As seen above, the very special nature of the DNA molecule generates an extensive supply of biochemical and chemo-physical properties that can be exploited by molecular geneticists in order to create a myriad of different analytical techniques. But, although the analytical techniques that operate at the biochemical level (ligation, cleavage, etc.) play a very important role in today’s biotech world, they do not add by themselves a specific complexity layer in the integration of analytical techniques within microsystems. That is, the production of a microsystem for ligation requires almost the same micro-mechanical and micro-fluidic devices (electro-osmotic and fluidic pumps, channels, filters and reservoirs) as the production of a microsystem
for cleavage; the main difference between the two systems being the addition to the solution of one or the other type of enzymes. For this reason, these techniques will not be described here in detail and the reader is referred to the literature ([Ausubel1988], [Old1994]) for further documentation on them.

Therefore, within the general framework of this doctoral work, the following section analyzes and briefly describes four analytical techniques based on the chemo-physical properties of DNA that have been successfully implemented in microsystems using very different design approaches and that are of the utmost importance in current biotechnology methods: hybridization, electrophoresis, amplification and sequencing.

**Figure 8 -** Classical Southern blot analysis: (1) denatured DNA is transferred from the slab gel to a nylon membrane; (2) the radioactive probe is inserted in a solution containing the nylon membrane and left to incubate; (3) after washing away unbound probes, the nylon membrane is placed next to a X-ray film in a light-tight container; the result is called the auto-radiograph of the electrophoresis gel.

**Hybridization**

As already seen (see p. 30), annealing (or renaturation) is a natural process of DNA in which a single-stranded DNA molecule attaches itself to its complementary counterpart. In its basic approach (see Figure 8), called Southern blotting after its developer, Ed Southern [Southern1975], hybridization makes clever use of annealing to obtain analytical results. Although Southern blotting involves as a starting point the transfer of DNA from an electrophoresis gel into a nitrocellulose membrane, hybridization can be generalized into a broader scheme, in which many different substrates, detection and transfer techniques may be used.

Typically (see Figure 9), a hybridization assay requires that one or more target DNA become attached to a membrane (e.g. nitrocellulose) or surface (as in DNA-chips) at known positions.
Once the target DNA has been fixed onto the membrane or surface, probe DNA (usually a single locus probe) is labeled with an isotope or a fluorescent dye. After labeling, the probe DNA is introduced in the solution containing the membrane or surface onto which the target DNA has been fixed and left at a certain hybridization temperature for a specified time. Both parameters (temperature and time) will vary depending on the length and composition of the probe and target DNA molecules (i.e. larger molecules → longer times/lower temperatures). As mentioned earlier (see p.30), this will affect the specificity of the renaturation process, which can also be controlled by modulating other parameters (such as salt concentration).

If some of the target DNA's are complementary to the probe DNA, renaturation (hybridization) will take place. After the specified time has elapsed, the unbound probe DNA is washed away and detection of the bound (hybridized) probes takes place. Target sites with bound molecules will reveal positive identification between probe and target (see Figure 10).

This basic scheme of the hybridization technique can be further elaborated in a number of ways. For example, RNA (instead of DNA) can be used for blotting in what has come to be jokingly known as Northern blotting. In
addition, different probes with different dyes may be introduced at the same
time, or multiple target sites may be deposited or grown at different
surface/membrane locations (as is the case of DNA-chips). Although
hybridization is a powerful technique for DNA analysis, in the last decade
the complexity and relative unreliability of its usual schemes (such as
Southern blot) had relegated it to a second plane. For example, in a
classical application of hybridization (parental and forensic tests),
analogous PCR-related techniques were long used due to their easier set-up
methods and increased reliability. This trend has changed in the last few
years, with the outbreak of DNA massive-hybridization chips, which has
repositioned hybridization at the foremost front of DNA-analysis techniques.

Electrophoresis

Traditionally seen as a form of chromatography, electrophoresis has evolved
into an independent field in the last forty years [Andrews1986] and has
acquired an outstanding importance in the domain of molecular genetics. In
general terms, electrophoresis is a separation technique in which
electrically active molecules are separated according to their different
charge/mass ratios by introducing them in a buffer solution and applying a
DC voltage across it. In this way, electrophoresis can separate many
different substances. For instance, in a classical electrophoresis
experiment, diverse proteins can be separated by means of their net
electrical charge/mass ratios: positively charged proteins will migrate
towards the anode and negatively charged proteins towards the cathode, at
a different speed depending on their charge/mass ratio; all the while,
electrically neutral proteins will remain at a standstill.

DNA electrophoresis

As seen before (see p.30), the active electrical charge of a DNA molecule is
located at the phosphate groups of nucleotides. Since all nucleotides make
an equal contribution in terms of mass and net charge to the DNA
molecule, the result is that a DNA molecule always has a constant
charge/mass ratio. Therefore, DNA (and certain families of proteins) cannot
be separated by means of conventional electrophoresis, as all DNA
molecules would migrate towards the cathode at the same speed. To
overcome this problem, slab gel electrophoresis was introduced into the
biotechnology world in 1975 ([O’Farrell1975], [Klose1975]).
In its traditional form (see Figure 11), slab gel electrophoresis makes use of an additional substance (mainly acrylamide or agarose gel) to create a sieve-like medium through which the DNA molecules must migrate. If urea (or another basic substance) is added to the medium, DNA will lose its secondary and tertiary structures and become linear. Thus, DNA is separated by means of its linear charge/length ratio, instead of its constant charge/mass ratio. Statistically, longer DNA fragments will become stacked in the gel meanders for longer periods of time than shorter DNA fragments, which will travel faster across the gel.

![Slab gel electrophoresis diagram](image)

**Figure 11 - Slab or submarine gel electrophoresis**

Typically, visible dyes (bromophenol and methylene blue, acridine orange, etc.) are deposited into the wells together with the sample DNAs in order to ease sample loading and to visually assess the run length. Usually, too, a ladder (a solution containing different strains of DNA of known sizes and concentration) is also run in the gel to gauge sample DNA size and concentration. After enough separation has been achieved, the gel slab can be inspected under UV light (if ethidium bromide or fluorescent labels have been added) or else transferred to a nitrocellulose or nylon membrane (see p.32) for blotting.

Traditional slab gel electrophoresis can be modified in a number of ways. For example, the size and distribution of gel pores can be modified by changing the solute concentration (in agarose or acrylamide gels) or its polymerization method (in acrylamide gels [O’Connell1977]). The purity of the solvent will also determine the amount of electro-osmosis and will thus affect on resolution. The strength and nature (constant vs. pulsating or alternating) of the electric potential is also a key factor in electrophoresis. Alternating potentials [Carle1986] are typically used to separate very large fragments of DNA that, otherwise, would remain stuck in the gel (the crisscross movement induced by an alternating field promotes the
untangling of the long DNA chains), while higher potentials will boost resolution and reduce analysis times [Andrews1986]. Additionally, different gels with different buffers may be linked together ([Ornstein1964], [Davis1964]), in what is called discontinuous gel electrophoresis.

Even though slab gel electrophoresis is a widely used and powerful technique, it has profound limitations in resolution due to the own nature of its substrate: the slab. Electrophoresis resolution depends mainly on two different factors: diffusion and Joule effects, which both lead to band broadening and loss of resolution.

![Figure 12 - Band broadening due to diffusion](image)

Diffusion is basically the entropy-driven property of solutes to spread in any given medium. In this sense, electrophoresis bands tend to diffuse and broaden, ultimately making detection impracticable because of an excessively low concentration of the solute. Hence, diffusion contradicts the intuitive notion that longer separation lengths provide better resolution. In fact, increased lengths typically yield longer analysis times and lower resolution.

![Figure 13 - Band broadening and disruption due to local Joule effects](image)

Considering the problem of diffusion, the obvious solution is to avail shorter separation lengths by increasing the applied field. Larger fields lead to shorter analysis times, minimizing thus diffusion and providing optimum
resolutions. Unfortunately, large fields also induce thermal dissipation (or Joule) effects. If the substrate (as in the case of slab gels) is unable to effectively and uniformly dissipate the generated heat, Joule effects will tend to occur locally and non-homogeneously, provoking band broadening and disruption, and leading to poor resolution assays.

**Capillary electrophoresis**

To overcome these resolution problems and attain lower analysis times, alternative substrates with better thermal properties can be devised. A typical alternative is to conduct electrophoresis in gel-filled capillaries [Li1992]. The use of capillary electrophoresis has been traced back to 1886 [Compton1988], although the modern use of this technique dates back to 1967, when Hjerten [Hjerten1967] used Ø3mm capillaries to conduct high-field electrophoresis. Thin capillaries allow the use of higher voltages because they have a superior volume/area ratio and, hence, air/liquid convection heat dissipation becomes more efficient. The use of capillaries, combined with low-thermal resistivity materials (such as silicon), has been effectively exploited for high-resolution electrophoresis in electrophoresis-chips (see p.51).

**Amplification**

Just a few years ago, in microbiology textbooks, the term amplification was as invariably linked to the acronym PCR, which stands for *polymerase chain reaction*, as it is still linked in the electronics field to the concepts of transistor and vacuum tubes. Indeed, it is hard to assess the impact of the invention of PCR in the biotechnology world, but it seems clear that the revolution in genetics that has taken place in the last decade would have been unthinkable without the fortunate invention, in 1983, of the polymerase chain reaction by Nobel laureate Kary Mullis [Mullis1990].

Amplification is a key concept in molecular genetics, since most cells have a unique copy of their genetic material. In the case of many bacteria and other organisms of easy culture, this lack of material can be compensated by extensive culture, but, even so, the complexity and unreliability of DNA extraction and purification will often lead to poor and contaminated results. In order to make feasible and reliable studies of genetic material, scientists must have the means to exponentially increase the number of original DNA/RNA copies in a sample, with low error rates, good specificity and
efficient yields. Such means are readily provided by the PCR and other related techniques described below.

**Polymerase chain reaction (PCR)**

Due to its relevance to this doctoral work, the exact nature and the underlying concepts of the polymerase chain reaction will be discussed in detail in a later section (see p.62). What follows below is a schematic overview of the reaction, intended to give some insight into the other amplification techniques described hereafter.

As its name implies, PCR is an iterative chain reaction, with three clearly separate steps that are repeated in a cyclic manner.

![Figure 14 - The basic PCR cycle: melting, annealing and extension.](image)

In the first of these steps (melting or denaturation, see Figure 15), temperature is raised to 94-96 °C, in order to denature the initial double-stranded DNA into single-stranded molecules.

![Figure 15 - Sample double-stranded DNA melting (denaturing) at 94-96 °C.](image)

After denaturation has taken place, temperature is lowered to the whereabouts of 40-70 °C (depending on primer length and composition), to allow for annealing (or hybridization) of selected *primers* to the single-stranded DNA molecules (see Figure 16). Primers are short DNA sequences...
that are complementary to a *starter* region of the DNA template and that serve as attachment points to the DNA-polymerase enzyme.

![Figure 16](image1.png)  
**Figure 16** - Annealing of complementary primers to sample single-stranded DNA.

Once the primers are set in, the polymerase attaches to the primer-template complex and begins to extend the complementary strand using available nucleotides (see Figure 17). For reasons that will be clarified later (see p.67), this extension step usually takes place at 72 °C.

![Figure 17](image2.png)  
**Figure 17** - Extension of the complementary strand with DNA-polymerase.

After the initial double-stranded molecule has been denatured, annealed to primers and extended, two double-stranded replicas of a portion of the initial DNA sample will have been produced. Iterating the same procedure (melting-annealing-extension) again will produce four double-stranded replicas. Ultimately, the process will ideally yield $2^n$ replicas of a targeted section of the original sample, where $n$ is the number of completed iterations. Considering, as it will be seen (p.64), that $n$ typically lies in the 25-35 range and discarding errors, primer-dimer occurrences and other nuisances, the ultimate number of replicas after a complete PCR analysis should be $2^{25-35}$ (roughly $10^{9-11}$); that is, between 33 million and 34 billion* replicas: more than enough to conduct further analytical procedures.

* For consistence with the writing language, all here-referred billions are American ($10^9$) billions.