BIOSYNTH®

Nucleosides, Nucleotides & Nucleic Acids Toolbox

Part 2 -Nucleic Acids
The Basis of Genes
Biological Significance
Chemistry
Analysis
Applications

Table of Contents

Introduction

This eBook (Toolbox Part II) follows on from the handbook Nucleosides, Nucleotides and their Analogs (Toolbox Part I); focused on Nucleic Acids and their structure, analysis, and synthesis aspects. This handbook focuses on the story of the oligomers, including DNA and RNA, our molecules of inheritance. This creates a pair of complementary handbooks telling the complete story across the building blocks to complex oligomers.

Nucleic acids are the carriers of genetic information in all living organisms. This handbook describes the origin of the concept of the gene. We follow the history of the isolation and characterisation of nucleic acids through to the recognition of their fundamental role as carriers of genetic information. The key discoveries around the structure and function of DNA and RNA are described as are the roles played by these extraordinary molecules. Key developments in the sequencing and synthesis of nucleic acids are explained together with practical applications, particularly in healthcare, resulting from this work.

About the Authors

Chris Lawson

Chris has a Ph.D. in chemistry from Edinburgh University (1969) working with Sir Dai Rees on the double helical structure of polysaccharides. He then spent 20 years in senior scientific positions at Tate & Lyle leading teams in the development of specialty products from sucrose (biopolymers, surfactants, high-intensity sweeteners). In 1989, he formed Dextra Laboratories with a colleague and grew the company into a producer of glyco-chemicals for pharma/biotech research. In 2007 they sold Dextra to The New Zealand Pharmaceutical Company and Chris formed Glycomix Ltd., concentrating on polysaccharide products and technology. He sold Glycomix to Carbosynth in 2016 and now works for the company as a senior advisor for carbohydrates. He has published over 40 papers, patents, and book chapters.

John Birch

John has a Ph.D. in microbiology from the University of London where he worked under the supervision of Professor John Pirt on factors required for the growth of mammalian cells in culture. He held a lectureship in the University before moving to Tate & Lyle where he worked with his co-author Chris Lawson. He subsequently worked for the pharmaceutical company G.D. Searle before taking up senior scientific management positions at the biotechnology company Celltech and then Lonza. Most of his career has been involved with the development of process technology and he has more than 80 publications. He was awarded the Donald Medal by The Institution of Chemical Engineers for outstanding services in the field of Biochemical Engineering and has held visiting professorships at University College London and the Universities of Oxford and Reading.

About Biosynth

Where Chemistry meets Biology, Products meet Services and Innovation meets Quality, Biosynth is at the Edge of Innovation.

With an unrivaled research product portfolio of over a million products and end-to-end manufacturing services, we are science led and customer focused to solve problems, taking pride in delivering products and projects that others cannot. Our expertise and capability runs across Chemicals, Peptides and Key Biologics all from one trusted partner.

Biosynth's mission is to supply critical raw materials and services to the pharmaceutical and diagnostic industries. Biosynth strives to de-risk its customer's supply chains.

Biosynth is an innovative life sciences reagents, custom synthesis and manufacturing services company. We are by scientists, for scientists, securing supply chains with consistent quality, across the globe. We manufacture and source a vast range of chemical and biochemical products, being experts in small molecules, peptides and key biological raw materials. We provide a full range of products and services to support life science research and development, with more than a million products in our research catalog and hundreds of complex manufacturing service projects. We have facilities across three continents and a rapid global distribution network. Our R&D resources and production facilities at both ISO 90001 and GMP, are modern and versatile, allowing us to produce chemicals on the milligram to ton scale, with peptides at mg to multikilogram scale.

Areas of Focus

Chemicals

- Route Scouting and Optimization
- Custom Chemical Synthesis
- Large Scale and GMP Production
- Biocatalysis and Green Chemistry
- Small Molecules and Polymers

Peptides

- Discovery and Optimization
- Custom Peptide Synthesis
- Multi Kg Scale and GMP Production
- Complex, Long, Labeled and Bridged
- Novel Technologies

Biologics

- Custom Antibodies
- Antigens and Antibodies for IVD
- Plasma, Biospecimens and Controls
- Epitope Mapping
- Enzyme Development and Production

Research Products — More than 1M Products in Catalog Across All Areas

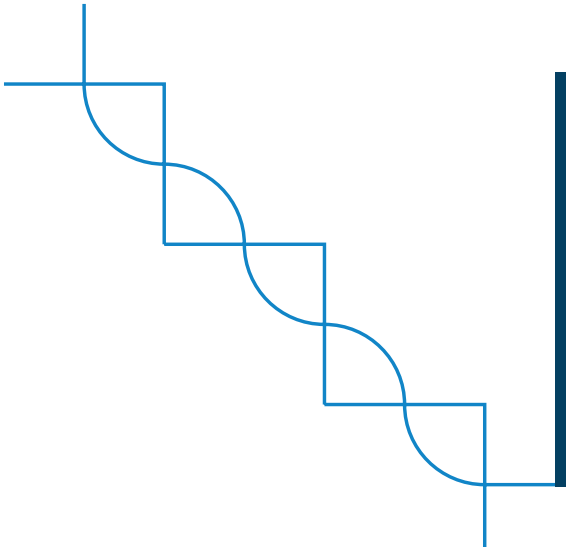
Bioconjugation — Production and Analysis

Chemical Manufacturing Capabilities

- Route Scouting and Optimization
- Custom Synthesis
- Large Scale Manufacturing
- GMP Manufacturing
- Quality Control and Quality Assurance
- End-to-end CDMO Services
- Custom Filling and Packaging
- Sourcing
- Biocatalysis Projects
- Bioconjugation Chemistry

Biosynth History

Biosynth was founded in 1966 by Hans Spitz, and has grown to a global products and services business. Biosynth merged with Carbosynth in 2019 to form Biosynth Carbosynth, creating a world leader in carbohydrates, nucleosides and enzyme substrates. In 2022 the company returned to the Biosynth brand, and embarked on a period of expansion and acquisition. The Biosynth Group acquired a number of companies from then on, including vivitide, Pepscan, Aalto Bio Reagents, EUCODIS Bioscience, Fitzgerald Industries International, Cambridge Research Biochemicals (CRB), celares, Pepceuticals, and VIO Chemicals. Biosynth is owned, amongst others, by KKR, Ampersand and senior management.

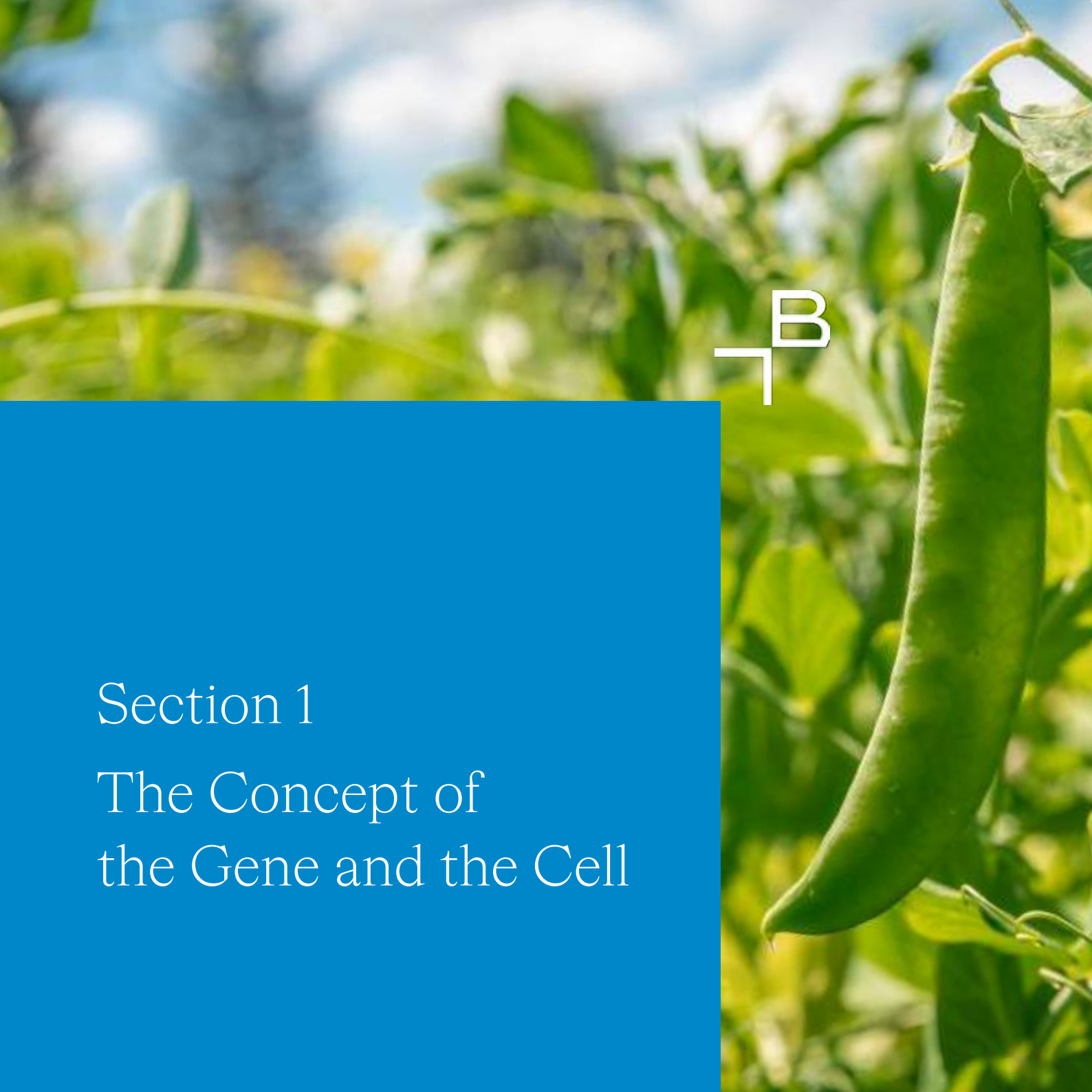


Order products online at

www.biosynth.com

Place an order, make a product inquiry or ask about services at

sales@biosynth.com



Section 1

The Concept of the Gene and the Cell

1.1 Introduction

The field of genetics started with Gregor Mendel, an Austrian monk born in 1822, who discovered the basic principles of heredity; his work in breeding experiments with pea plants with different characteristics, such as the shape or colour of the seed, was revolutionary. For example, he showed that when a yellow pea and a green pea were cross-bred, their offspring always produced yellow seeds, but in the next generation, the green peas reappeared at a ratio of one green to three yellow (Figure 1).

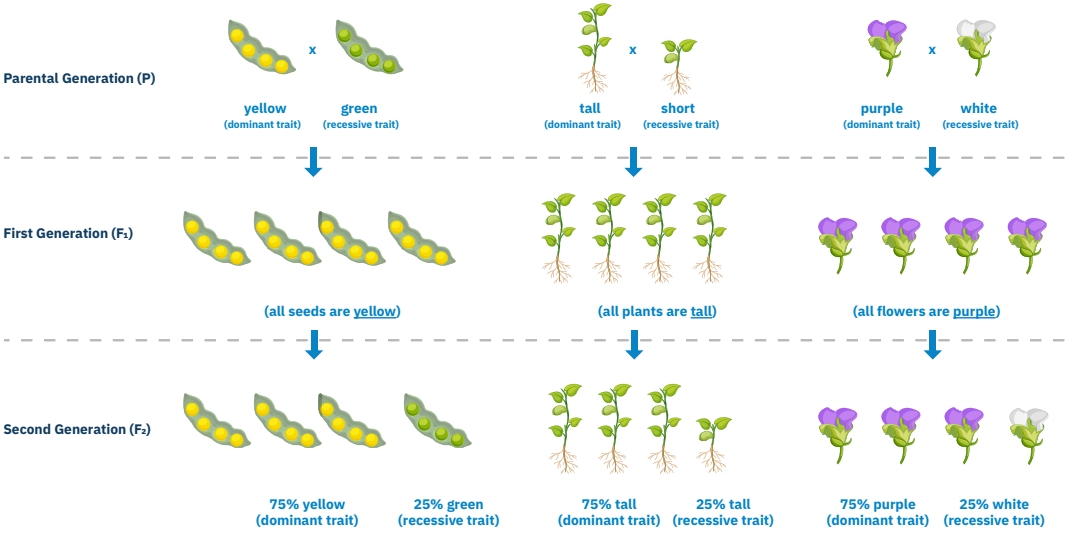


Figure 1: The Mendel pea experiment.

Mendel coined the terms “recessive” and “dominant” to explain his findings. In the example above, the green trait, which had vanished in the first generation, was recessive, and the yellow was dominant. He published his work in 1866, demonstrating the actions of invisible “factors” - now called genes - in predictably determining the traits of an organism’s offspring (Mendel, 1866). Until this time, an organism’s features were thought to arise by a general blending of the characteristics of the parents.

The profound significance of Mendel’s work was not recognised until the turn of the 20th century, more than three decades later, when the German botanist Carl Correns verified Mendel’s laws of heredity using a different experimental model (Correns, 1900).

Gregor Mendel

Carl Correns

Although the concept of a physical basis for heredity was becoming accepted, the nature of the genetic material was still unknown.

By the mid to late 1800s, improved microscopy, particularly the use of recently developed dyes, significantly added to the knowledge of cell structure and function. Evidence, based on observing cell division, suggested that nuclei in cells might be linked to heredity, and researchers started to elucidate the structures involved. In the late 1800s, the German biologist Walther Flemming was using dyes to study cell nuclei and identified structures in cell nuclei, which he called chromatin (Flemming, 1875). These structures would later be called chromosomes, and the work of Flemming and others established how chromosomes split during cell division to provide copies to each daughter cell. In 1902, the American geneticist Walter Sutton discovered that male and female gametes (egg and sperm cells) contain half the number of chromosomes as normal body cells. He deduced that following fertilisation of the egg, the chromosome number in cells doubled with half the chromosomes coming from each parent. This was clear evidence for how traits could be physically inherited from both parents (Sutton 1903)

In 1910, evidence for genes being on chromosomes was strengthened by the work of Thomas Hunt Morgan in the USA. He studied an eye colour trait in fruit flies and was able to show that the inheritance of the trait was sex-linked, and the data strongly indicated that the gene for the trait was on the X chromosome, which is present as two copies in females but only one in males (Morgan, 1915).

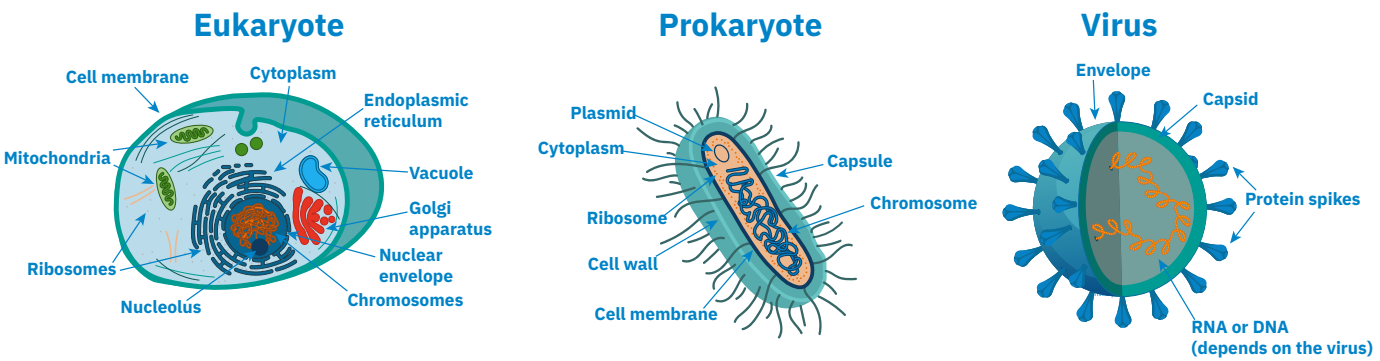
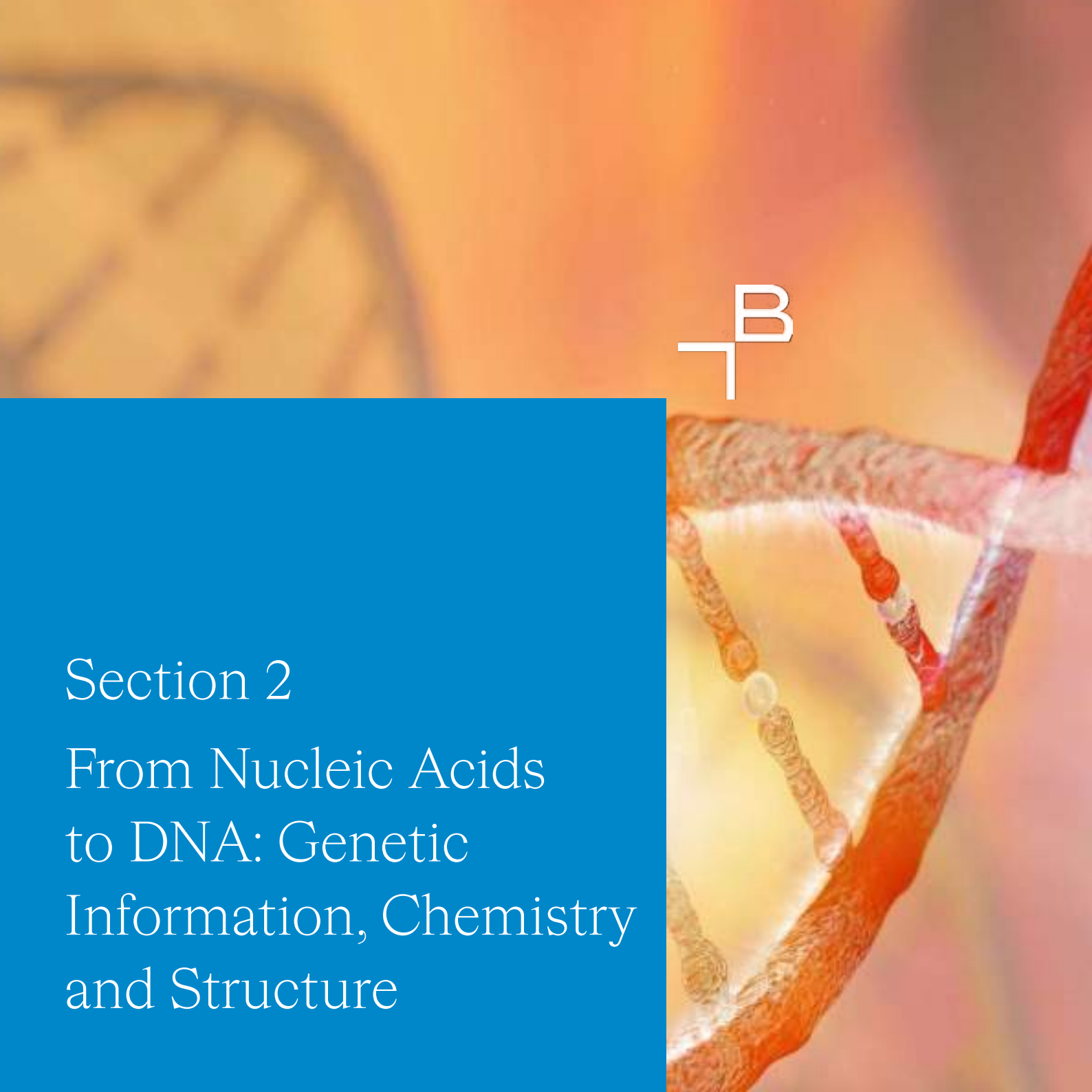


Figure 2: The three types of cell that store genetic information: eukaryote cell (left), prokaryote (centre), and virus (right).

We now know that how genetic material is stored in cells depends on the type of organism. Higher organisms, such as animals, plants, and fungi, are known as eukaryotes, and their key distinguishing feature is the presence of a nucleus in their cells containing multiple chromosomes. An intriguing feature of eukaryotic cells is that they also carry genetic information outside the nucleus in mitochondria. In plants, DNA is found in chloroplasts as well as in mitochondria. The DNA in these organelles is in a single circular chromosome as in bacteria, and it is thought that mitochondria and chloroplasts might have a microbial origin in their distant past. Sequencing of mitochondrial DNA has proven useful in establishing ancestral relationships because it passes unchanged from generation to generation via the female egg cell. It was mitochondrial DNA sequencing that allowed the remains of King Richard III, found in a car park in Leicester, UK, to be formally identified. A comparison of DNA from his skeleton matched the sequence of a known descendant. A second class of organisms, typified by bacteria, are known as prokaryotes. These are typically about ten times smaller than a human cell. They do not have nuclei, and the cell typically has a single circular chromosome. Bacteria also frequently contain small circular DNA “plasmids”. These molecules can often pass between bacteria and sometimes carry genes for antibiotic resistance. They are used in recombinant DNA technology as vehicles for transferring genes for useful proteins into bacteria.

Even smaller than bacteria, we have viruses (Figure 2). They are typically ten to a hundred times smaller than bacteria. Unlike eukaryotes and prokaryotes, viral genomes are quite varied and can be based on RNA as well as DNA, and they may be single or double-stranded. The coronavirus that causes COVID-19 is a single stranded RNA virus. As well as acting as a template for replication, it also serves as mRNA for synthesising viral proteins.



Section 2

From Nucleic Acids to DNA: Genetic Information, Chemistry and Structure

2.1 Early Work

In addition to progress based on cytology and genetics, advances were also being made in understanding the chemical composition of substances, which would ultimately be shown to be relevant to genetics.

In 1869, the Swiss physiological chemist Friedrich Miescher first identified what he called “nuclein” inside the nuclei of human white blood cells.

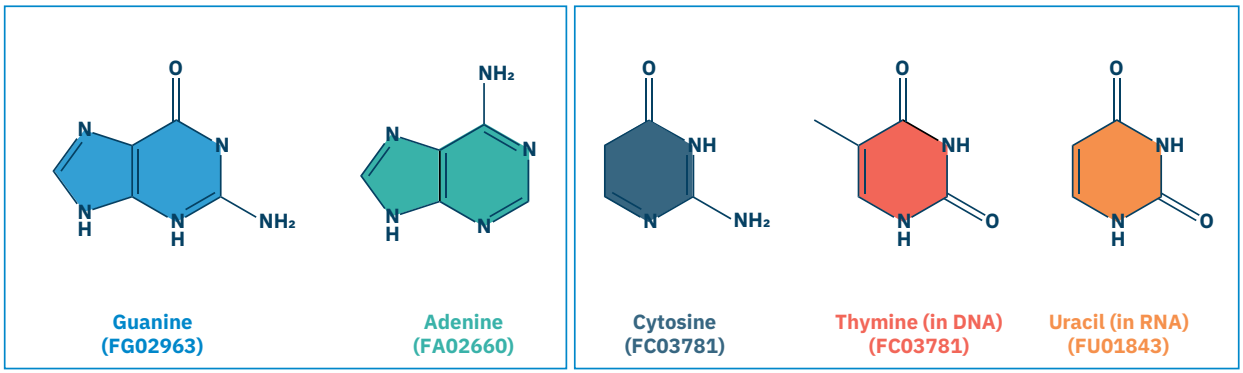
Miescher realised he had discovered a new substance (Dahm, 2008). Sensing the importance of his findings, Miescher wrote, “It seems probable to me that a whole family of such slightly varying phosphorous-containing substances will appear, as a group of nucleins, equivalent to proteins” (Wolf, 2003). Richard Altmann, discovering that nucleins were acidic, renamed them nucleic acids (Altmann, 1889).

Friedrich Miescher

Albrecht Kossel

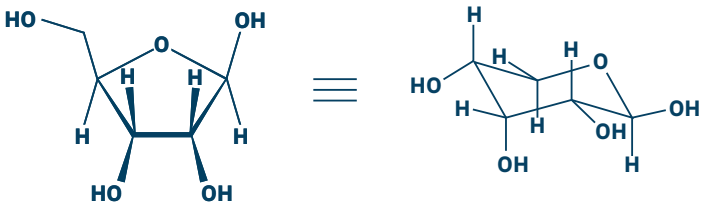
Between 1885 and 1901, the German physiologist Albrecht Kossel (Jones, 1953) isolated and described the five purine and pyrimidine bases that are present in nucleic acids: adenine, cytosine, guanine, thymine, and uracil (Figure 3). In 1893, he also identified a reducing sugar in nuclein. His work earned him the Nobel Prize in 1910. Kossel speculated, incorrectly as we now know, that there were probably four nucleic acids, each containing just one of the bases.

Figure 3: The five purine and pyrimidine bases.



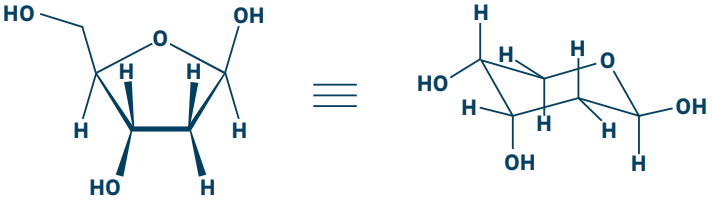
2.2 Nucleosides and Nucleotides

Very significant progress was made in our understanding of nucleic acid structure in the first two decades of the 20th century, thanks, particularly to the work of Phoebus Levene at the Rockefeller Institute for Medical Research in the USA (Frixione & Ruiz-Zamarripa, 2019; Levene, 1900; Levene 1935;). He was the first to use what are now standard terms, i.e., nucleosides, nucleotides, ribonucleic acids (RNA), and deoxyribonucleic acids (DNA). He identified D-ribose as the sugar in RNA in 1909, and in 1929, he went on to discover a second sugar, 2-deoxy-D-ribose, in DNA (Figure 4).



D-Ribose

(MR05196)



2-Deoxy-2-ribose

(MD04977)

Figure 4: The sugars in nucleic acids.

Levene also proposed that the bases were in a linear complex (phosphoric acid - carbohydrate - base). By 1935, he had provided an accurate molecular structure for DNA and a close-to-accurate structure for RNA (Figure 5). He incorrectly postulated that the DNA was made up of repeating units of the four nucleotides in what was known as the tetranucleotide hypothesis. It eventually became clear that this could not be the case when it was shown that the proportions of the bases were not constant for all DNA. It's interesting to note that Levene did not believe that nucleic acids were carriers of genetic information.

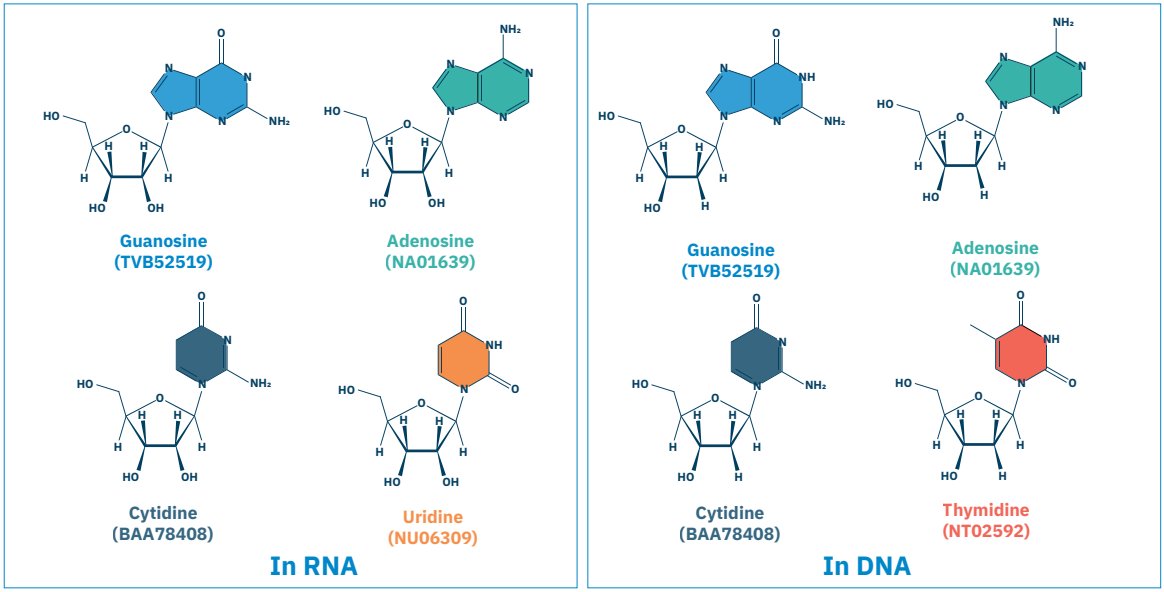


Figure 5: Nucleotides in DNA and RNA.

2.3 Chargaff's Rules

In 1952 Erwin Chargaff in the USA established that DNA from any organism has a 1:1 stoichiometric ratio of purine and pyrimidine bases and, more specifically, that the amounts of guanine and cytosine are equivalent as are the amounts of adenine and thymine (Elson 1952; Chargaff 1952). These became known as Chargaff's rules, and as we shall see later this was fundamental to the eventual understanding of base pairing in the DNA double helix. He also found that the proportions of the four bases varied from species to species supporting the conclusion that DNA carried genetic information.

2.4 Evidence for DNA as the Carrier of Genetic Information

On the face of it DNA appeared to have rather a simple structure to form the basis of genetic material and it still seemed possible that proteins, with their apparently more complex chemistry were more likely candidates.

In 1928, the British bacteriologist Frederick Griffith working on the possibility of creating a vaccine against pneumonia found (almost by accident) a "transforming principle" which could be transferred between bacteria. He showed that material from a killed culture of virulent bacteria when added to a non-virulent strain converted it to virulence in mice (Griffith, 1928). In 1944 Oswald Avery and colleagues at The Rockefeller Institute for Medical Research purified DNA from the virulent strain of bacteria and established that it was most likely the 'transforming principle' rather than protein (Avery, 1944).

In 1952 Alfred Hershey and Martha Chase used bacteriophage (a type of virus that infects bacteria) labelled with radioactive sulphur (which went into protein) or phosphorus (which went into DNA) and analysed infected cells. The bacteriophage transferred the labelled DNA, not protein, into bacteria. This showed that DNA was the likely carrier of genetic information (Hershey, 1952).

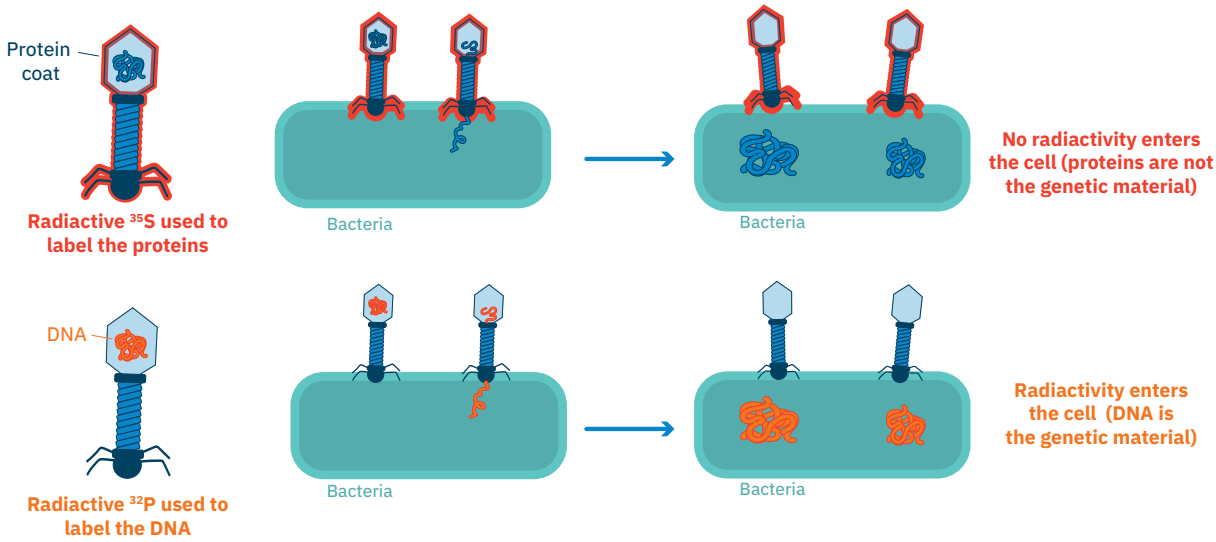


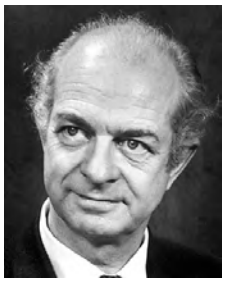
Figure 6: A schematic of the Hershey-Chase experiment that demonstrated that DNA is the carrier of genetic information.

2.5 Watson, Crick and the Double Helix

In the early 1950s, James Watson and Francis Crick joined an MRC unit in Cambridge headed by Max Perutz and commenced work on the tertiary structure of DNA.

Several groups were already working on the structure of DNA and some suggestions had been put forward. For example, Michael Creeth, a Ph.D. student at the then University College Nottingham, - who had recently discovered with JM Gulland and DO Jordan the existence of hydrogen bonds between the bases in DNA - correctly proposed a structure with two chains which had the nucleobases inside the molecule joined by hydrogen bonds with sugar and phosphate moieties pointing out (Creeth, 1947), but he did not identify the helical nature of the molecule and in his model, the nucleobases were not continuous – they were in blocks. The eminent chemist Linus Pauling, working at CalTech in California incorrectly proposed a helical three-chained structure (Pauling, 1953).

Max Perutz



Linus Pauling

Watson and Crick realised that two irregular sequences of bases could be packed in the centre of a helix if a purine always bonded with a pyrimidine. Also, hydrogen-bonding would be satisfied if adenine always bonded with thymine and guanine could bond only with cytosine in agreeing with Chargaff's rules. X-ray studies by Rosalind Franklin, working with Maurice Wilkins at Kings College London, supported a double helix structure which was a key feature of Watson and Crick's eventual model (Figure 7) (Watson, 1953).

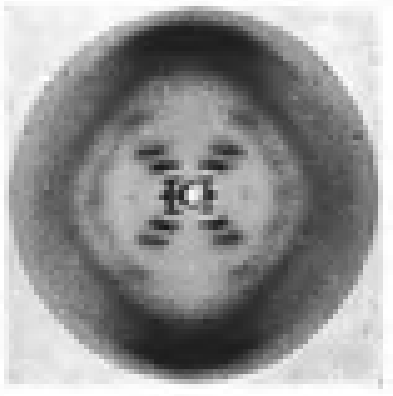
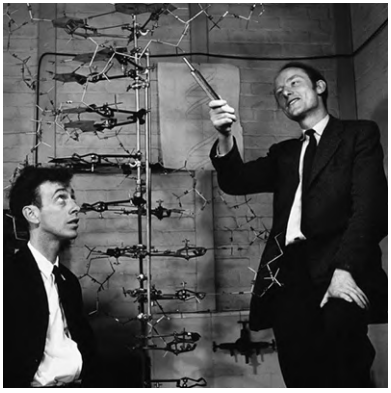
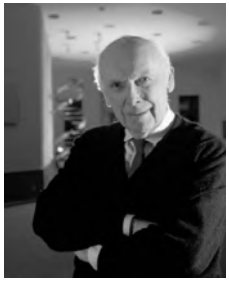


Figure 7: Franklin's X-ray picture of DNA (above) and Watson and Crick's model (below).



Watson & Crick with their DNA Model

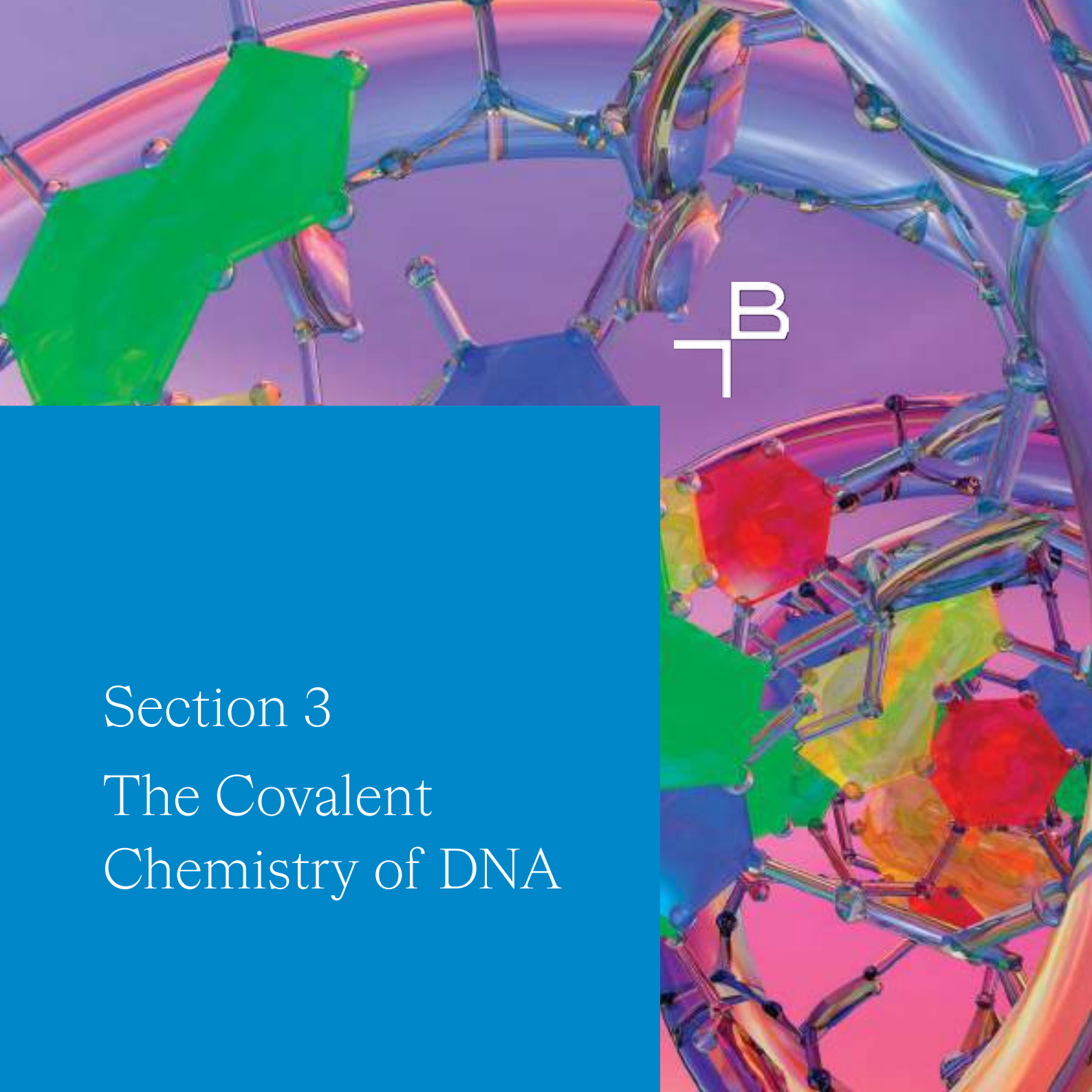
Francis Crick



James Watson

Rosalind Franklin

Maurice Wilkins



B

Section 3

The Covalent Chemistry of DNA

3.1 Understanding the DNA Chain Structure

An important feature of a DNA chain is the polarity conferred by its nucleotide subunits. The 5’ phosphate group on each ribose residue can be considered as a protruding knob on one side of the molecule and the 3′ hydroxyl on the other as a hole with each completed chain formed by interlocking knobs and holes with all subunits lined up in the same orientation. Moreover, the two ends of the chain are easily distinguishable, as one has a 3′ hydroxyl and the other a 5′ phosphate. Chain polarity is then referred to as the 3′- end and the other as the 5′- end.

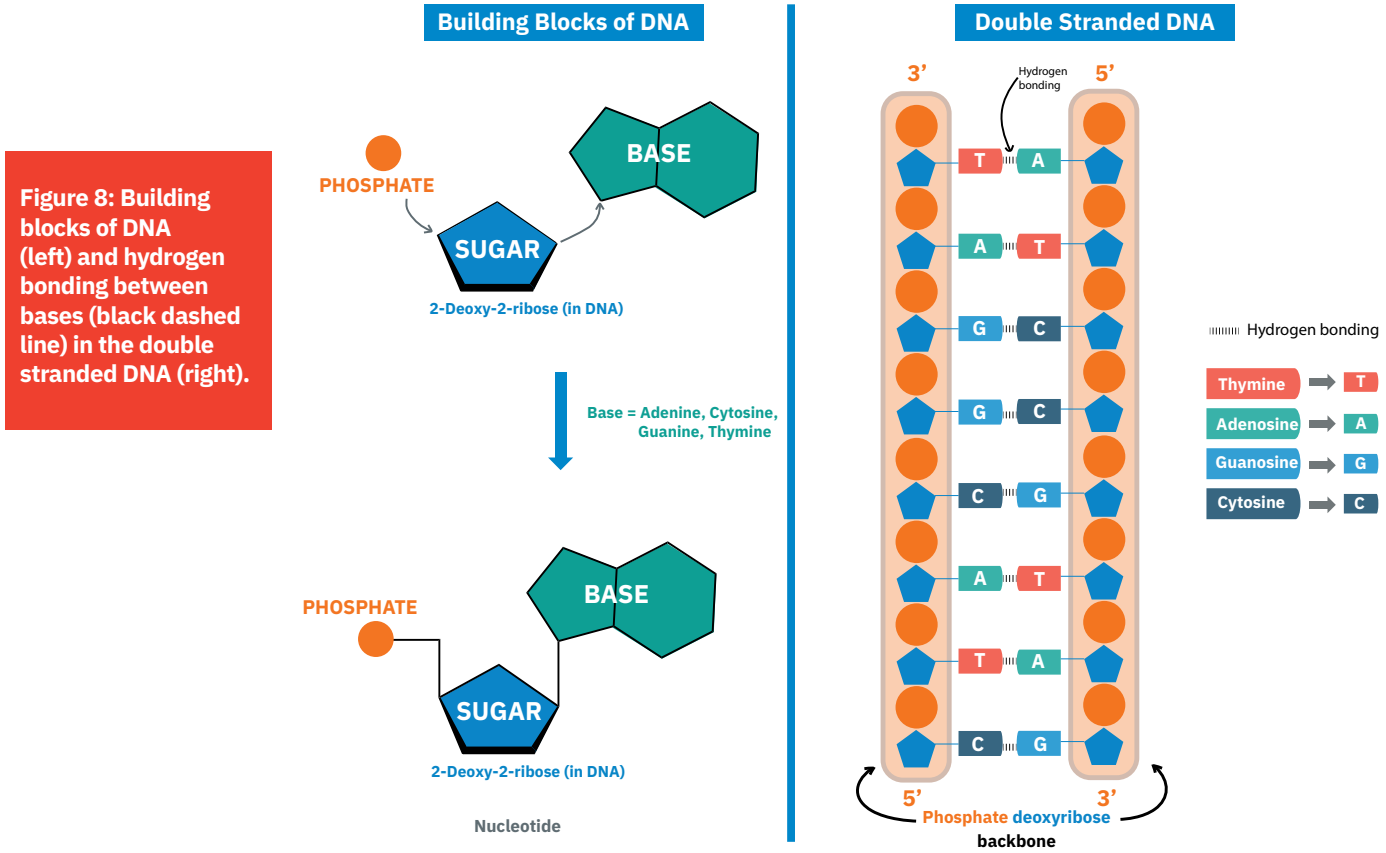


Figure 8: Building blocks of DNA (left) and hydrogen bonding between bases (black dashed line) in the double stranded DNA (right).

The two chains of the DNA double helix are held together by hydrogen bonding between the bases on the different strands, where all the bases are on the inside of the double helix, and the sugar-phosphate backbones are on the outside (Figure 8). In each case, a bulkier two-ring base is paired with a single-ring base (a pyrimidine); A always pairs with T, and G with C (Figure 9). This complementary base-pairing enables the base pairs to be packed in the energetically most favourable arrangement in the interior of the double helix. In this arrangement, each base pair is of similar width, thus holding the sugar-phosphate backbones an equal distance apart along the DNA molecule. To maximise the efficiency of base-pair packing, the two sugar-phosphate backbones wind around each other to form a double helix, with one complete turn every ten base pairs (Figure 9). The detailed covalent chemistry of the base pairing linking through to the sugar-phosphate backbone is shown in Figure 9 and 10.

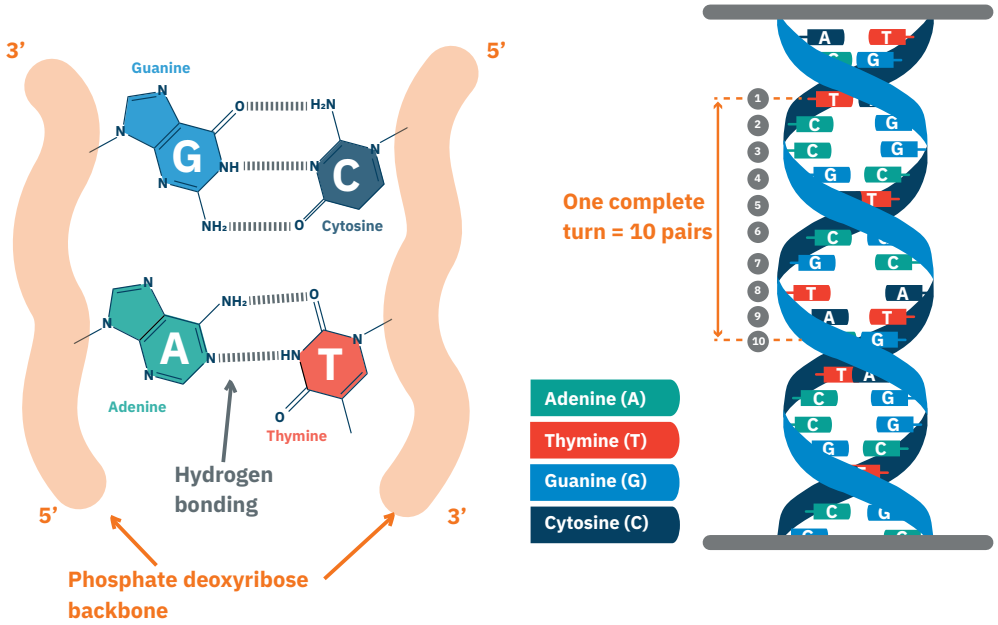


Figure 9: Base pairing, where A pairs with T and G with C via two and three hydrogen bonds respectively.

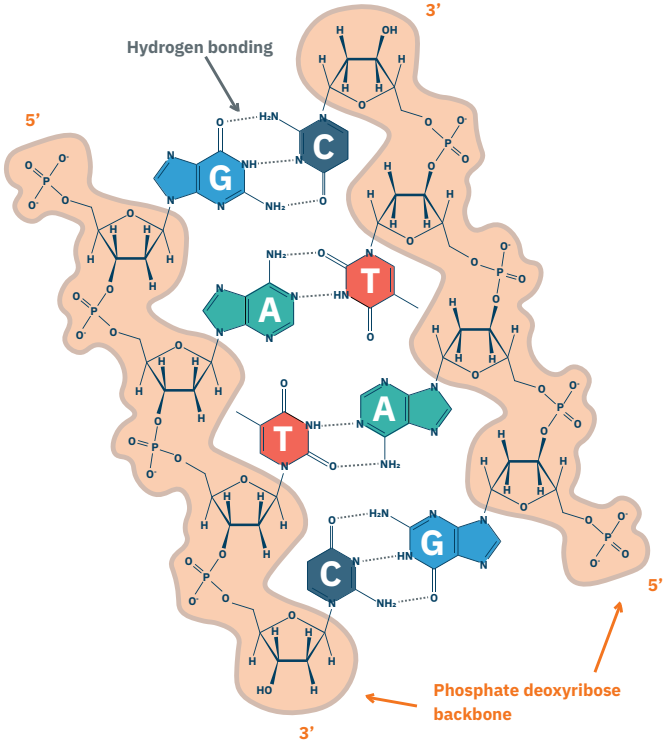


Figure 10: The covalent structure of DNA.

3.2 Nucleic Acids as Polymers

Initially, DNA was believed to be quite a small molecule but, by the mid-1930s, Casperssen and Hammarsten, working in Sweden had shown that it was a polymer. As improved methods have become available to measure the size of these molecules, it has become clear that they can be exceptionally large. Each chromosome in the cell nucleus of a higher (eukaryotic) organism contains two strands of DNA known as chromatids. These strands of DNA have molecular weights on average of 10^{11} (Gross-Bellard, 1973). The DNA strands are wound around protein complexes of eight histone proteins to create nucleosomes with stretches of DNA joining the nucleosomes in a repeating structure (Figure 11). The resultant condensed fibre is further coiled, creating a 30nm diameter chromatin fibre, which forms loops that are compressed and folded to generate what we recognise as a chromatid. The consequence of all this condensation in size is that the DNA in a human cell is reduced from a total length of around 1.8 m to about 0.09 mm. The DNA is held tightly in place on the protein, mostly by the electrostatic attraction resulting from the positive charge on the histones and the negative charge on the DNA.

In simpler (prokaryotic) organisms such as bacteria, which do not have nuclei, there is usually a single circular chromosome which is compacted into a nucleoid structure by a combination of supercoiling and combination with specific proteins and RNA molecules. The molecular weight of their chromosomal DNA is substantially less than in Eukaryotes (around 2.5 x 10^8, for *Escherichia coli*) (Massie, 1965). Prokaryotes do not contain histones.

RNA molecules are substantially smaller, with molecular weights ranging from tens of thousands to two million depending on the type (Genthner, 1985).

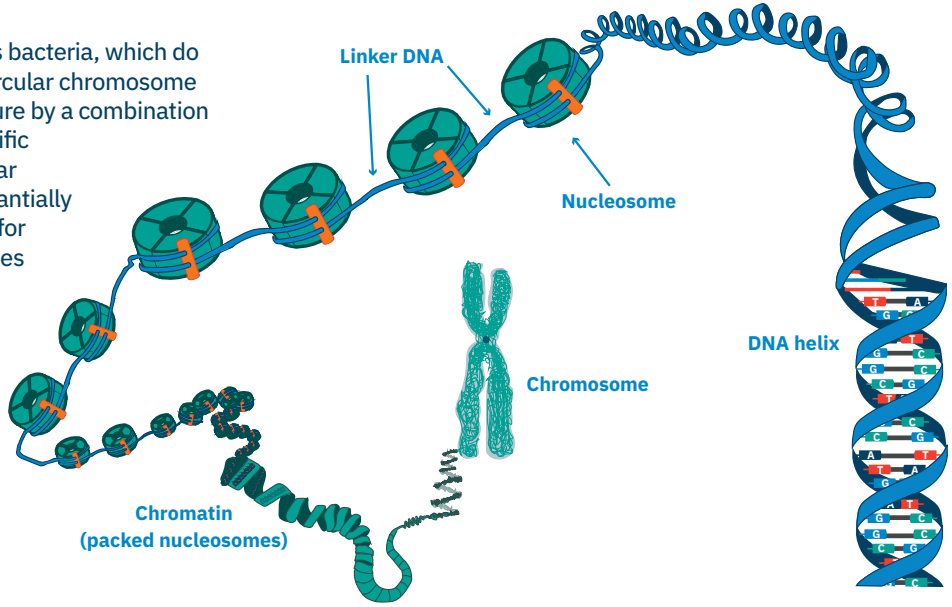
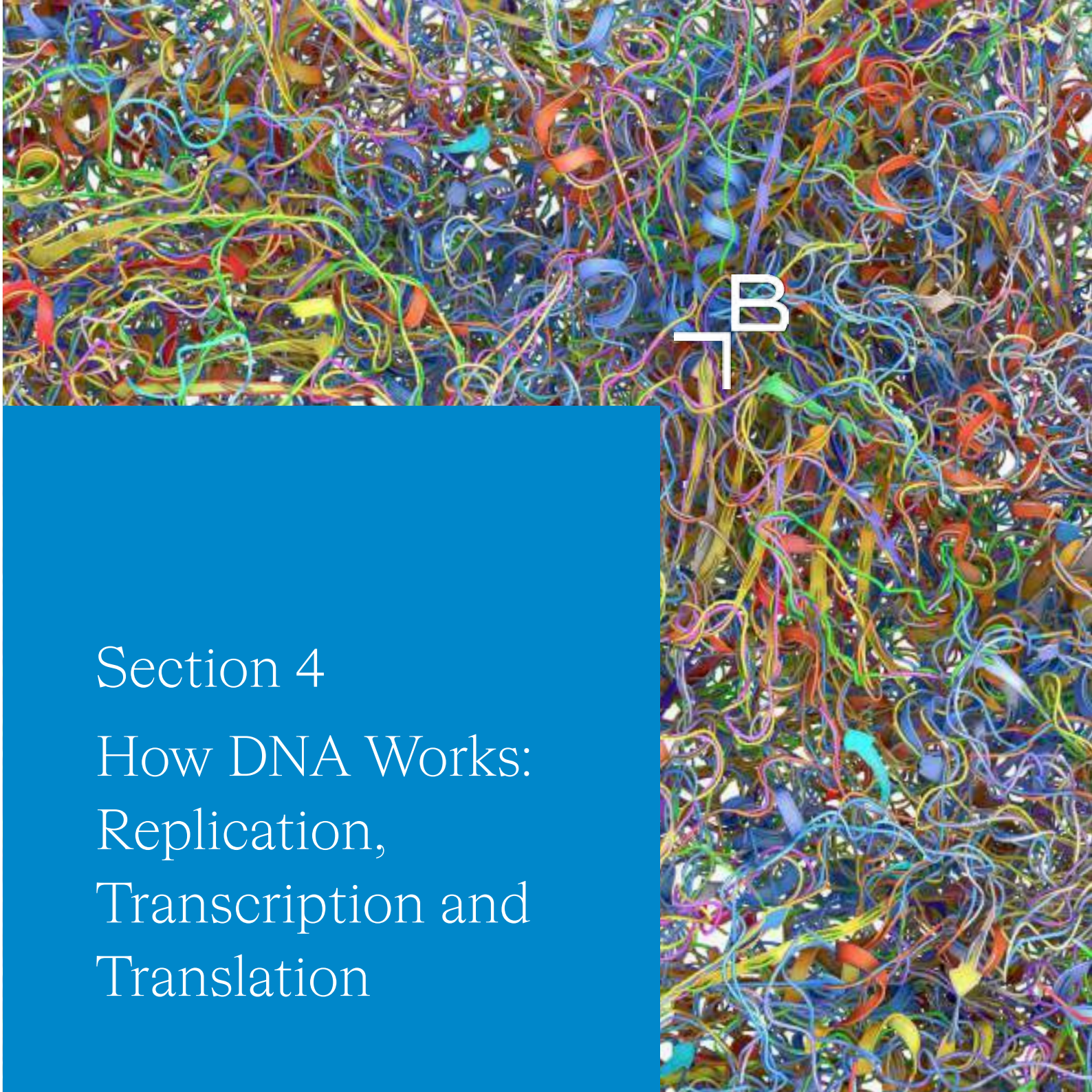


Figure 11: Schematic representation of the DNA packing around the nucleosomes.

3.3 Stability of Nucleic Acids

DNA is a quite stable molecule and this is demonstrated by its persistence in nature under appropriate conditions. This has allowed biologists to isolate and sequence the genomes from extremely ancient animal and plant remains including, for example, million year old mammoth remains (Miller, 2008) and a weevil that was more than 100 million years old (Cano, 1993).

In contrast, the ribose component of RNA makes the latter less stable than DNA. Ribose has a lower stability than deoxyribose because the additional hydroxyl group on the pentose ring makes it more prone to hydrolysis.



Section 4

How DNA Works: Replication, Transcription and Translation

4.1 Introduction

Whilst we now recognise the DNA double helix as one of the great discoveries of the 20th century, this was not so obvious at the time. Although DNA was a very strong contender to be the carrier of genetic information, this had not yet been proven and there were still significant questions about how it might function. It would take a decade of intense research to confirm the role of DNA and how it works (Crick, 1961).

4.2 DNA Replication

A carrier of genetic information must be able to replicate itself. Watson and Crick realised that the helical structure suggested a basis for how DNA might separate and copy each strand. In 1958, Meselson and Stahl, working at Caltech in the USA, confirmed Watson and Crick's prediction. By growing bacterial cultures in media containing first the heavier ^{15}N isotope of nitrogen, followed by ^{14}N and then separating the resulting DNA species by density gradient centrifugation, they were able to show that the double helix does indeed separate to allow each strand to act as a template to create a new strand (Figure 12). Each new double helix contains one parent strand and one complementary daughter strand (Meselson, 1958).

Around the same time, a landmark breakthrough was made by Arthur Kornberg in 1956 with the discovery of the enzyme DNA polymerase (Friedberg, 2006; Kornberg, 1956). This enzyme produces replicas of DNA templates by catalysing the addition of deoxyribonucleotides to the growing strand of DNA. His discovery, for which he won the Nobel Prize, is fundamental to our understanding of gene replication. DNA polymerases have become important tools, for example, in nucleic acid synthesis and the application of PCR technology (Figure 13).

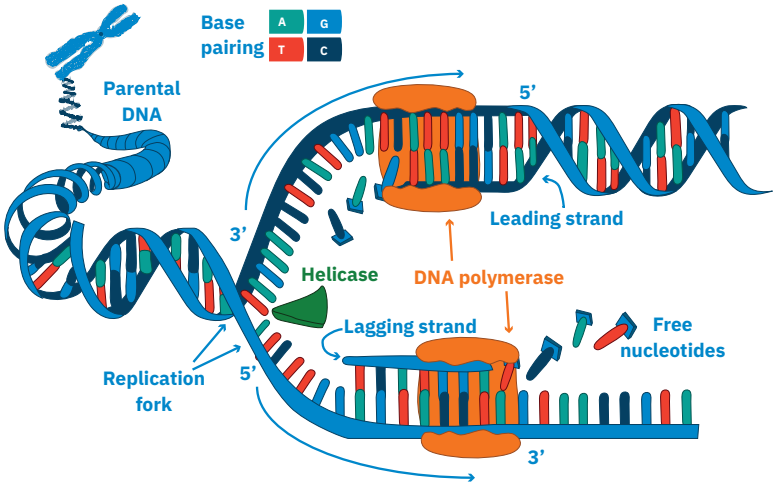
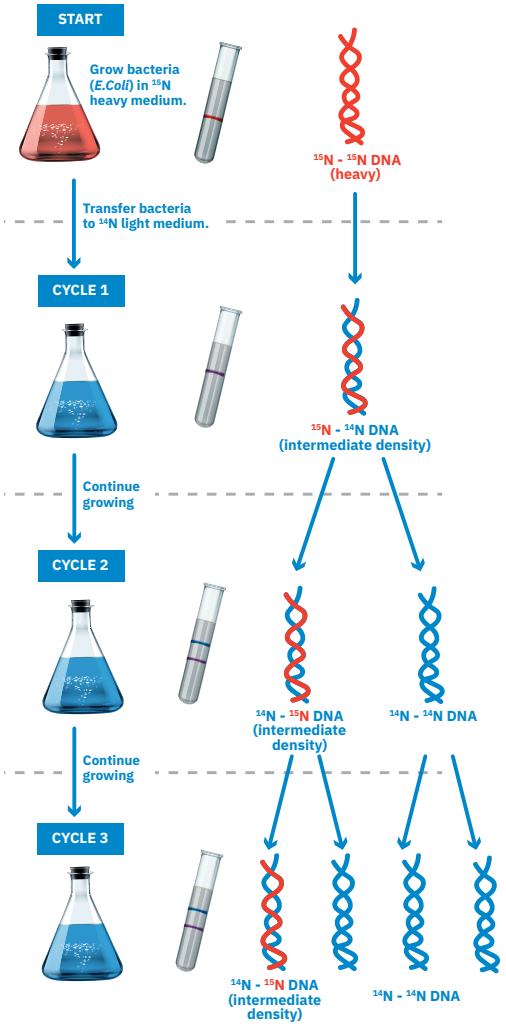


Figure 13:Replication via DNA polymerase.

Figure 12: The Meselson experiment.



4.3 From DNA to Protein: Transcription and Translation

Two other big questions were how exactly genetic information might be stored in DNA and how that information is used to generate the tens of thousands of different proteins found in a cell. Following on closely from their double helix paper, Watson and Crick suggested in 1953 that the sequence of bases in DNA provided the basis of genetic information.

It initially wasn't clear whether DNA might act directly as a template for protein synthesis, as some proposed, or by some indirect mechanism possibly based on RNA. By 1959, it was established that proteins are synthesised on ribosomes, which are particles containing protein and RNA and were discovered by George Palade in 1955 (Palade, 1975). Since ribosomes are in the cytoplasm of a cell and DNA is almost entirely contained in the nucleus of cells in higher organisms, it seemed increasingly likely that there was an intermediate step between DNA and protein synthesis, possibly involving a "messenger" RNA.

The answers to these questions came in 1961 with a series of ground breaking discoveries. Nirenberg and Mathaei, working at the National Institutes of Health in the USA, were conducting experiments to investigate the potential role of a messenger RNA in protein synthesis. Using cell-free extracts of the bacterium *E. coli*, they added a synthetic RNA made of just uracil and discovered that a polyphenylalanine protein was produced (Matthael, 1961; Nirenberg, 1961). This was the first definitive evidence that RNA and not DNA was the template for protein synthesis and also the first step in breaking the code; a uracil sequence coded for phenylalanine. Nirenberg won the Nobel Prize for his work on the genetic code. Around the same time, two papers were published in Nature, which described the isolation from cells of a transitory type of RNA that had the properties expected of a "messenger" and became known as messenger RNA or mRNA. (Brenner, 1961; Gros, 1961).

The nature of the genetic code was clarified when Crick and colleagues reported on experimental work with mutants of a bacteriophage (virus that infects bacteria), which indicated that each amino acid in a protein is coded for by specific triplets (codons) of nucleotides (Crick, 1961).

Whilst a given triplet can only code for one amino acid, they concluded that conversely, there was probably redundancy in the code, i.e., an amino acid could be coded for by more than one triplet. Gradually, the codons for all twenty amino acids were identified. A major player in this effort was Har Gobind Khorana, who used chemical synthesis to create short DNA segments and then an enzyme to produce repeating sequences of these segments. (Khorana, 1965). By expressing these sequences in bacterial cell-free extracts he determined which sequences, i.e. codons, coded for which amino acids. He confirmed that there is more than one codon for most amino acids. Khorana received the Nobel Prize for this work. Only two amino acids have a single codon: methionine and tryptophan. Even though there are a few rare examples of additional codons being used, the genetic code is universal for all living organisms, suggesting a common ancestor.

The frequency with which particular codons are used varies between different organisms and even between genes expressing a specific protein. The choice of codons may act as a control mechanism influencing the rate at which proteins are produced (Komar, 2016). In producing biopharmaceutical proteins in cells, using recombinant DNA techniques, it is common to optimise codons to maximise protein productivity (Mauro, 2014; Sorensen, 1989).

The picture was becoming clear: DNA in the cell's nucleus is transcribed into a short-lived messenger RNA (mRNA) that passes out of the nucleus and interacts with ribosomes, translating the mRNA message into protein. Perhaps not surprisingly the details of this process turned out to be more complex than this brief description suggests.

4.4 Transcription

The stretch of DNA coding for a specific protein is copied, in a process known as transcription, to generate a complementary single-stranded messenger RNA (mRNA) (Figure 14). Start and stop codons define the limits of transcription which is carried out by the enzyme RNA polymerase. This enzyme binds to the starting site on the coding strand of DNA in a process known as initiation. The helix is unwound, RNA is synthesized until the end of the gene sequence is reached, and a termination step occurs. In DNA, only one strand is coding. This begs the question of why is DNA double-stranded if one of the strands is non-coding? The answer seems to be that the second strand plays an important role in correcting errors that can occur during DNA replication (Straus, 2018).

A surprising discovery was made independently in 1977 by Philip Sharp at MIT and Richard Roberts at the Cold Spring Harbor Laboratory, earning them Nobel Prizes (Berget, 1977; Chow, 1977). They found that RNA, as initially transcribed, is often longer than the expected coding sequence. This is a common phenomenon in eukaryotes. It became apparent that this "pre-mRNA" contains sequences called introns, which are not needed for protein expression and are spliced out to leave just the protein-coding sequences (exons), the ends of which are rejoined. The splicing is carried out by large ribonucleoprotein structures called spliceosomes, which are made up of proteins and small nuclear RNAs which have a catalytic activity responsible for the splicing. Some introns are self-splicing. As described later, introns can have significant roles and are often the precursors for RNAs that regulate protein expression levels. It was then discovered that pre-mRNA transcripts can be spliced in alternative ways resulting in different proteins being expressed from an original DNA sequence (Nilsen, 2010). Two to several thousand different mRNAs can be produced depending on the particular gene. This mechanism for producing multiple mRNA transcripts from a single gene helps to explain the puzzling observation in humans that more than 90,000 proteins are generated from around 25,000 protein-coding genes (Gordon, 2015). This may be a mechanism for making more efficient use of the cell's DNA.

In addition to splicing, other modifications to the RNA are also made. For example at the 5'-end of mRNA in eukaryotes a guanosine cap is commonly added and then methylated. This cap is important for gene translation. At the 3'-end a polyadenine tail is frequently added. This is believed to protect against degradation by exonuclease enzymes and may play a role in transport from the nucleus.

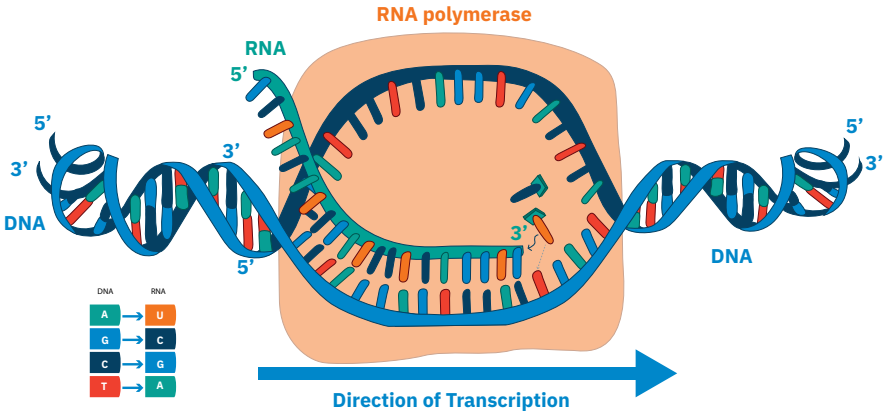


Figure 14: Transcription.

4.5 Translation

Proteins make up the majority of the macromolecules found in cells. Therefore it is not surprising that much of a cell's resource is taken up in translation, the process that converts the information in mRNA into protein.

4.5.1 Ribosomes

After transcription, mRNA passes to ribosomes, where protein synthesis takes place. Ribosomes are made up of about 60 % ribosomal RNA (rRNA) and 40 % protein. rRNA, which is transcribed from DNA, makes up about 80 % of the total RNA in a cell. There are four different rRNAs in eukaryotes and about 80 ribosomal proteins. The ribosomal RNAs and proteins are arranged in two subunits, one large and one small. In 2009, Venkatraman Ramakrishnan, Thomas Steitz and Ada Yonath won the Nobel Prize for elucidating the structure of the ribosome at an atomic level (Williamson, 2009). The ribosome moves along the mRNA as amino acids are added to the growing protein chain. One mRNA can be read by multiple ribosomes operating at the same time. The ribosome is an example of an RNA acting as a catalyst or "ribozyme". mRNAs may be used more than once but can also be degraded by enzymes back to the constituent nucleotides.

4.5.2 Transfer RNA (tRNA)

Apart from mRNA and the RNA in the ribosomes (ribosomal RNA, rRNA), a further class of RNA is required; transfer RNA (tRNA). Crick had predicted the existence of tRNAs as being necessary as adaptor molecules to transport specific amino acids to the ribosome. They were discovered by Zamecnik and Hoagland in 1958 (Hoagland, 1958).

A series of transfer RNAs (tRNAs), specific for each amino acid codon, transfer the amino acids to the ribosome in the sequence dictated by mRNA to generate the relevant protein. Each tRNA has a corresponding enzyme that links it to the relevant amino acid (Figure 15). In 1965, Holley and colleagues determined the primary structure of the 75 nucleotide alanine tRNA (Holley, 1965). tRNAs have a tri-clover structure, and three regions of the molecule recognise respectively ribosomal RNA (rRNA), a complementary sequence on mRNA and an amino acid.

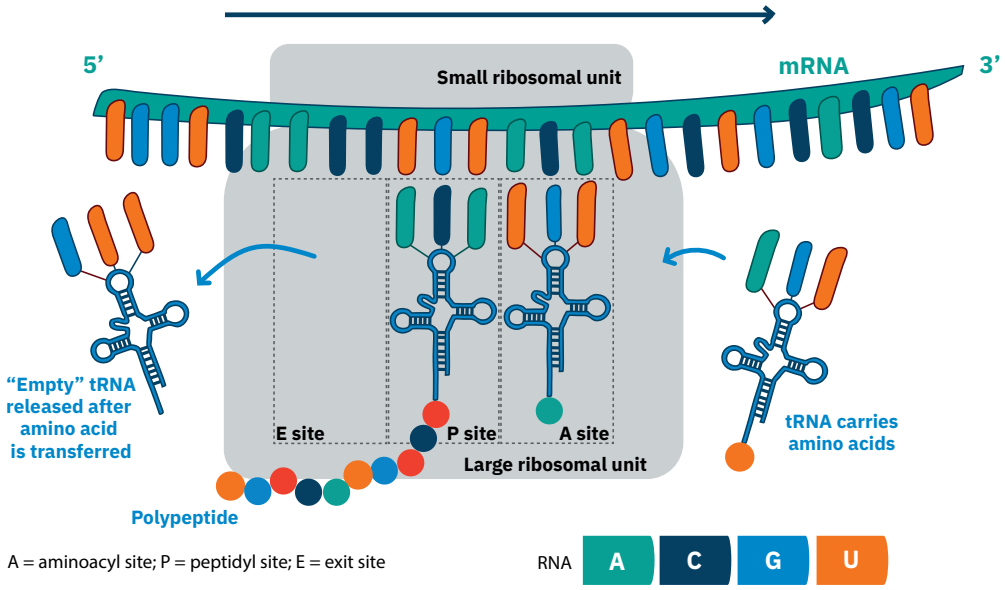
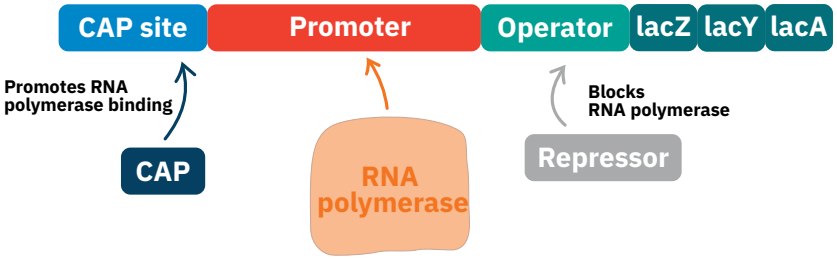


Figure 15: Simplified translation diagram. (1) The A site accepts the new amino acid; (2) the P site holds and transfers the amino acid to grow the polypeptide chain; (3) the E site allows release of the empty tRNA (without the amino acid).

4.6 Regulation of Transcription and Translation

1961 wasn't just the year when the genetic code was cracked and the processes of transcription and translation began to be understood; it also heralded the beginning of our understanding of how gene functions are regulated. This was the year when Nobel Prize winners Francois Jacob and Jacques Monod, working at the Pasteur Institute in Paris discovered the first genetic regulation system (Monod, 1952). Their work focused on the bacterium *E. coli,* particularly the genes involved in the lactose metabolism. This nutrient can be used as a carbon and energy source. By studying mutants of *E.coli* which varied in the way that lactose was metabolised, they discovered that the genes for the proteins required to metabolise lactose were grouped along a sequence of DNA which they called an operon (Figure 16). In addition to these "structural" genes, they also identified two closely associated regulatory genes which essentially acted as on/off switches for transcription of the whole operon. These switches are turned on or off by proteins which can detect the presence or absence of lactose and glucose (a nutrient used in preference to lactose if present) (Jacob, 1961).

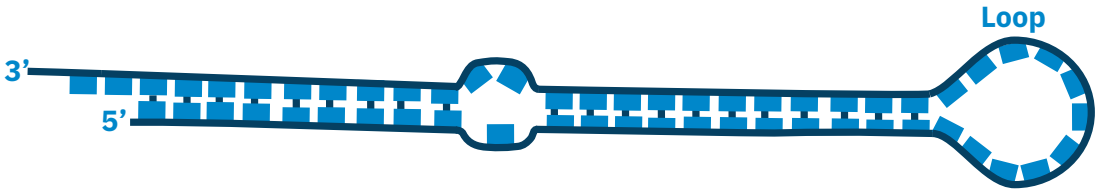
Figure 16: The lac operon which requires the genes LacZ, LacY and LacA to be intact and in order for lactose metabolism to occur.



For eukaryotic organisms, the situation is much more complex. While regulation is mostly at the transcription level in prokaryotes, it is far more complicated in eukaryotes and can be at many levels. In animals, plants and fungi, there is the huge additional complication of controlling genetic expression to allow an egg, seed or spore to develop into a differentiated organism. The creation and maintenance of all the different cell types in an organism, all sharing a common genome, is a result of complex "epigenetic" regulatory processes. (Jaenisch, 2003).

Whilst, as in bacteria, proteins can be involved in regulation, what perhaps came as something of a surprise was the important role of several different types of RNA. Whereas the proportion of DNA coding for proteins in bacteria is typically around 90 %, in eukaryotes, it is amazingly small (less than 2 %). Some of this non(protein)-coding DNA is transcribed into transfer and ribosomal RNAs, but the role of the very large remainder was a mystery and was commonly referred to as junk DNA. It turns out that it is far from being junk. During the 1990s, it became clear that at least some of these regions play a key role in regulating gene activity. Some non-coding DNA is transcribed to produce small RNA molecules (sRNA), which have a regulatory function. The most common of these are microRNA (miRNA) and small interfering RNA (siRNA). miRNA was discovered in 1993 by Lee and colleagues (Lee, 1993) and functions by reducing the efficiency of mRNA translation. In the years following their discovery, more than 2000 different miRNAs have been identified in human cells. They are small single-stranded molecules about 22 nucleotides in length, which are folded back on themselves to produce hairpin structures (Figure 17). A significant proportion (around 40 %) of miRNA genes are found in introns.

Figure 17: Pre-miRNA structure.

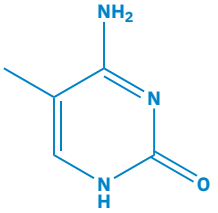


Small interfering RNAs (siRNA) are double-stranded molecules comprising 20 to 24 base pairs which were discovered by Fire in 1998 (Fire, 1998). They usually function by binding to complementary sites on mRNA causing cleavage. siRNAs are more specific in their action than miRNAs. Whilst individual miRNAs can regulate a large number of different mRNAs, siRNAs are very target specific. This has resulted in widespread interest in their therapeutic potential.

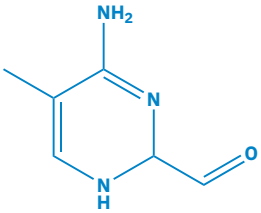
Not all small RNAs act to reduce gene expression. Some small double stranded RNAs can activate transcription (Li, 2006). In addition to the small RNAs there are also long (greater than 200 nucleotides) non-coding RNAs which have important roles in gene regulation (Statello, 2021).

4.6.1 Chemical Modifications

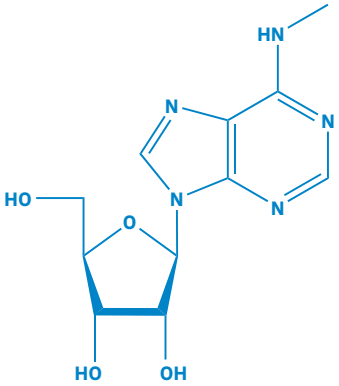
In addition to these regulatory molecules, there also chemical modifications made to both DNA and RNA which can influence transcription and translation. A large number of modifications are seen in rRNA, tRNA and mRNA. In DNA, cytosine can be methylated to generate 5-methylcytosine. This usually occurs when cytosine is next to guanosine and appears to be linked to gene silencing. In RNAs, N6-methyladenosine is the most common modification and 5-methylcytosine is also seen in a variety of RNA molecules. Inappropriate methylation is associated with some diseases including cancers. Demethylation of methylated bases can also occur. Other modifications of cytosine include; 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxycytosine. In addition to methylation a number of other chemical modificatons can occur. For example, sulphur modifications are found in tRNAs on uridine (2-thiouridine, 4-thiouridine, geranyl-2-thiouridine) and on ribose (phosphorothioate) which act in different ways as enhancers. In addition to changes to RNA and DNA (Figure 18), chemical modifications may also be made to histones as part of gene regulation.



5-methylcytosine (FM63889)



5-formylcytosine (FF175440)



N6-methyladenosine (NM32281)

Figure 18: Some examples of typical chemical modifications to cytosine and adenosine.

Section 5

Isolation and Purification

5.1 DNA and RNA Isolation Techniques: A Brief Overview

Natural, genomic DNA can be isolated by initially freeing it from cells or tissue by mechanical or chemical disruption. This is then followed by steps to remove contaminants such as proteins, for example by selective precipitation. The DNA can then be precipitated with alcohol.

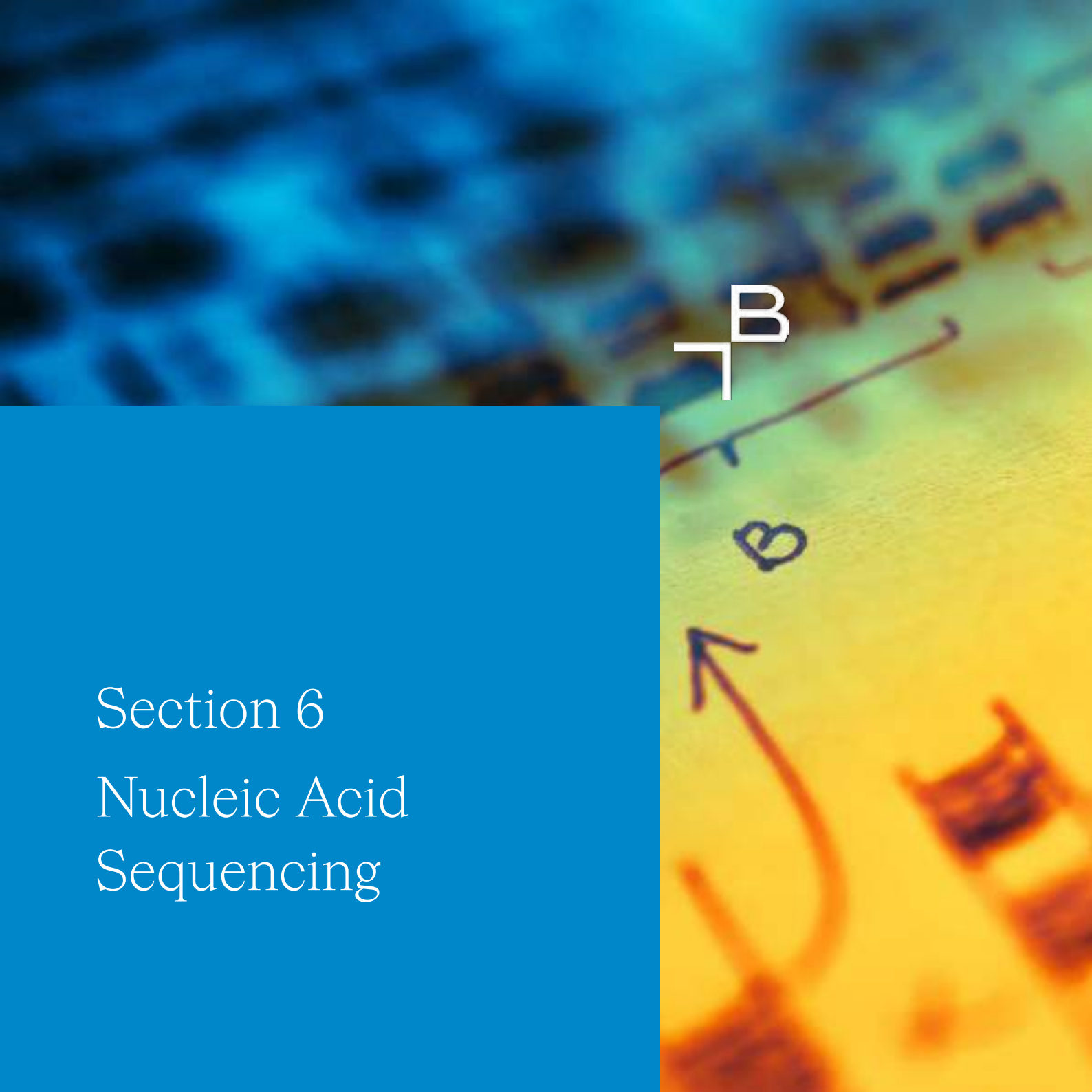
An alternative simpler but lower yielding process uses Chelex resin to directly bind cellular contaminants leaving DNA in solution. DNA can also be recovered using solid phase extraction by adsorption onto silica in the spin column method or onto magnetic beads coated with silica. Several of these systems have been automated (Vogelstein, 1979).

RNA isolation from cells was originally based on phenol/chloroform extraction. In more recent times, the silica based spin column and magnetic bead methods have become available (Table 1). Detection and quantitation of DNA and RNA can be done by UV absorbance spectrophotometry or by fluorimetry using fluorescent dyes that bind specifically to nucleic acids.

The development of purification methods since around 1955 are listed in Table 1.

Table 1 DNA Purification Methods 1955-1998

Date	Method	Reference
1955	Size Exclusion Chromatography (SEC)	Lathe, 1955
1956	Ion Exchange Chromatography (IEC)	Peterson, 1956
1957	Density gradient centrifugation	Meselson, 1957
1968	Affinity Chromatography	Cuatrecasas, 1968
1979	Alkaline Extraction	Bimboim, 1979
1979	Silica Matrices	Vogelstein, 1979
1988	Salting Out	Miller, 1988
1990	CTAB extraction	Doyle, 1990
1998	Phenol-Chloroform	Barker, 1998
1998	Phenol-Chloroform	Hawkins, 1998; Boom, 1990



Section 6

Nucleic Acid Sequencing

6.1 Introduction

Nucleic acid sequencing is the process of determining the order of nucleotides in DNA or RNA. Sequencing is making a fundamental contribution to our understanding of relationships between organisms, how biological systems function and how they evolve not just in the long term but also, for example, in the case of viruses, in the short term. In the recent COVID-19 pandemic, sequencing has made it possible to identify mutations in the virus as new variants arise. This, in turn, informs measures that can be taken such as the development of new vaccines.

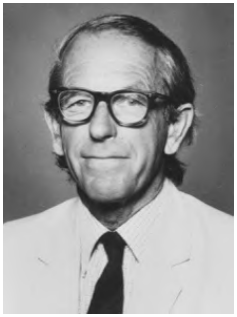
Publication of the human genome sequence and advances in sequencing technology has profoundly impacted our understanding of disease at a genetic level. The analysis of data from large numbers of genomes is invaluable for identifying potential correlations with disease. Genome sequencing is already being used as a diagnostic tool to detect genetic variations known to be associated with specific diseases such as cancer. The UK government has announced a programme to obtain the full genome sequence of 100,000 newborn babies to see if this is useful in identifying and treating genetically based conditions.

6.2 Sequencing Technologies

As the nature of the genetic code became clear, the next challenge was determining the linear sequence of nucleotides along specific genes (Heather, 2016). Initially, the methods available could only provide small sequences. Efforts focused on RNA where known molecules of relatively short length, such as tRNAs, were available. In addition, RNAase enzymes which cut RNA at different specific nucleotide sites were available and could be used to assist sequencing. Such enzymes were not available for DNA. In 1965, Robert Holley at Cornell University published the 77 nucleotide sequence of the yeast alanine tRNA. He determined the sequence using RNAases, and by comparing the resulting fragments, he was able to deduce the sequence (Dudock, 1969; Holley, 1965). This was the first nucleic acid to be fully sequenced, and his work led to a Nobel Prize in 1968. Shortly afterwards, George Brownlee, working with Fred Sanger in Cambridge, determined the sequence of 5S ribosomal RNA from *E. coli*, using RNAases to generate fragments that could be compared (Brownlee, 1967).

In 1976, Walter Fiers at the University of Ghent completed the sequencing of a bacteriophage's 3,569 nucleotide RNA genome. This was the first time a complete genome had been sequenced and was accomplished using RNAase enzymes, followed by separation of the resultant oligonucleotides using electrophoresis and chromatography (Fiers, 1976).

By the late 1970s, better methods for sequencing were developed by Sanger at Cambridge and Allan Maxam and Walter Gilbert at Harvard. Maxam and Gilbert developed techniques for specifically cleaving radiolabelled DNA at each of the four bases by chemical means and then separating the fragments by gel electrophoresis. Sequence could be deduced from the patterns of radioactive bands on the gels. The Maxam and Gilbert method was the first to be published (Maxam, 1977) and was widely used. However, the Sanger method became preferred because it used fewer toxic chemicals and was more straightforward, particularly when it came to automating the method. The Sanger method, which has a high level of accuracy, is still widely used.



Frederick Sanger

Sanger had been working on the sequencing of proteins for which he was awarded a Nobel Prize, before turning his attention to nucleic acids. In 1977, he published a method known as the chain termination or dideoxy method, which became the most widely used technique in the following decades mainly because of its accuracy and ease of use compared with other methods, and he received his second Nobel Prize for this work (Sanger, 1977). The current Sanger method (Figure 19) is based on the use of a DNA polymerase enzyme that initiates DNA replication in the presence of the four deoxynucleotides in DNA and a DNA primer which binds next to the sequence of interest. DNA transcription is terminated by adding a fluorescently labelled dideoxyribonucleotide version of the four natural nucleotides to the reaction. So, for example, if the polymerase is adding nucleotides and the next required in the sequence is cytosine, the addition of dideoxycytosine will prevent extension beyond that point. Adding the dideoxy derivatives of the other nucleotides in separate reactions terminates elongation at different positions in the sequence. Analysis of the nucleotide specific termination fragments, which vary in length, allows the sequence to be deduced.

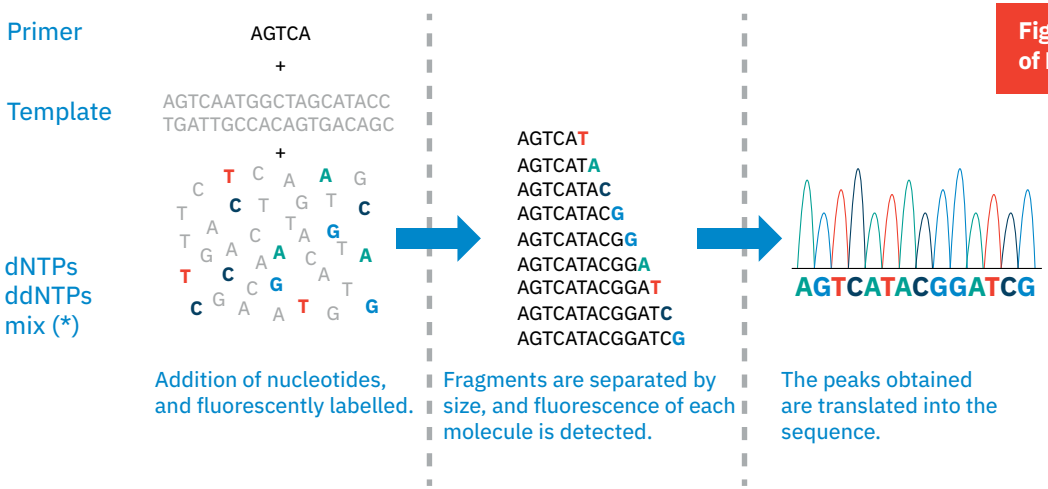


Figure 19: An overview of DNA sequencing.

* dNTPs (deoxynucleotide triphosphates) and ddNTPs (dideoxynucleotide triphosphates) are nucleotides used in the Sanger method. They serve as building blocks for new DNA.

6.3 Automation

The next big step forward in sequencing was automation, and a commercially available system based on Sanger sequencing was introduced by Applied Biosystems in 1987. This increased throughput from tens to thousands of bases per day, and much more ambitious projects could be undertaken including the sequencing of the whole human genome. Improvements in the automated systems included the use of fluorophore-labelled dideoxynucleotides instead of radioactive labels and the use of capillary electrophoresis for analysis (Smith, 1986). The fluorometric method uses fluorophores of different colours to label each base and allowed the termination reactions to be carried out in a single vessel. Capillary electrophoresis gave improvements in speed and resolution compared with existing electrophoretic methods. By the mid 1990s, improvements to automated systems meant that millions rather than thousands of bases could be sequenced per day. In recent years, second generation systems have been commercialised which now make it possible to sequence billions of bases per day (Figure 20).

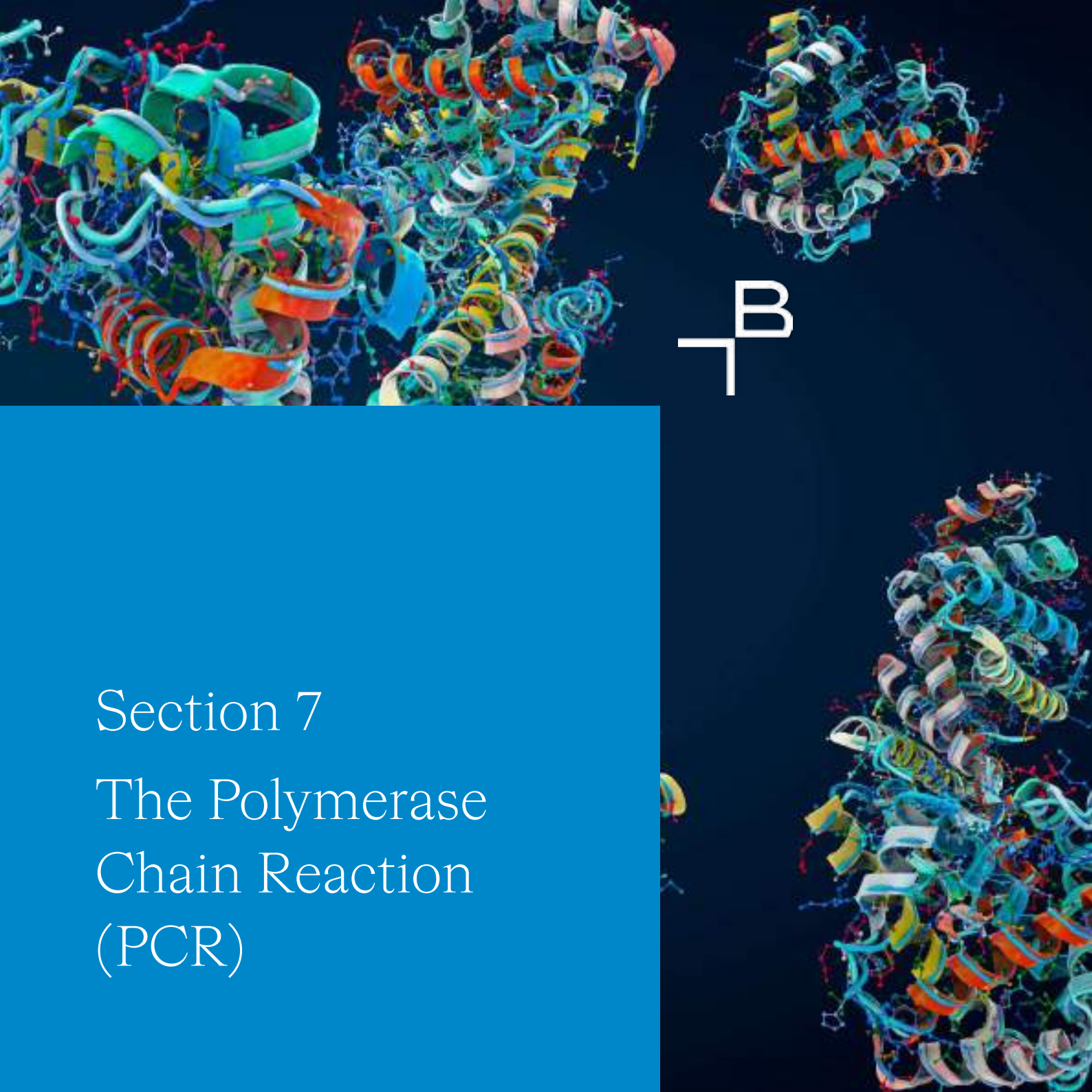
Figure 20: Example of automated DNA sequencing.

Examples include systems from Illumina and Oxford Nanopore. The Illumina system uses massively parallel sequencing using fluorescently labelled reversible DNA chain terminators. The technology detects single bases as they are added to the DNA. The Oxford Nanopore system is a very different technology. The device has a membrane containing pores each of which acts as an electrode. As a sample of DNA or RNA is passed through the membrane, the electrode detects changes in current. These changes vary depending on the sequence of nucleotides and can be interpreted to provide that sequence.

6.4 The Human Genome Project

The Human Genome Project was an international research project with the goal of determining the complete sequence of nucleotides in the genes of the human genome. In the public sector, most of the sequencing was carried out by the International Human Genome Sequencing Consortium, (IHGSC) funded by governments at research centres in several countries (Zwart, 2015). In the private sector, Celera Genomics carried out a sequencing programme in parallel. The international programme started in 1990 and much of the sequence had been completed by 2001 when draft sequences were published by the IHGSC and by Celera in Science and Nature (IHGSC, 2004; Lander, 2001; Venter, 2001). By 2003, the international project was completed, having sequenced about 90% of the genome. By the beginning of 2022, the remaining gaps had been sequenced. This was a remarkable achievement when one considers the complexity of the task - the human genome contains three billion base pairs. In the intervening years since the genome project, sequencing technology has improved to the point where a human sequence can be obtained in a day. The first Guinness world record for the fastest human genome sequencing (5hr 2min) was established in 2022 at Stanford using the Oxford Nanopore system (Gorzynski, 2022). In addition, the costs of sequencing have come down. Whereas the thirteen year human genome project cost around three billion dollars, a sequence can now be carried out for less than $1,000. Sequencing efficiency has improved to the point where the Sanger Institute could report in 2022 that it had sequenced 243,633 human genomes in the previous 3.5 years. In total, they had sequenced 21 quadrillion (21×10^{15}) letters of DNA.

Publication of the human genome sequence and developments in sequencing technology are having a significant impact on the understanding of disease at a genetic level. Genome sequencing is now used regularly as a diagnostic tool to detect genetic variations known to be associated with specific diseases such as cancer and with specific health risks.



Section 7

The Polymerase Chain Reaction (PCR)

7.1 Introduction

The polymerase chain reaction has become one of the key techniques in molecular biology, earning its inventor, Kary Mullis, the Nobel Prize in chemistry (Mullis, 1990). The technique uses a polymerase enzyme to amplify a gene sequence of interest to provide enough DNA for sequencing or other applications.

The first step in PCR is to heat the sample to separate the two strands of DNA. Short strands of "primer" DNA are added, which attach to the desired complementary regions on the separated strands. The reaction mixture includes the four nucleotides required to make DNA and a heat resistant DNA polymerase enzyme which adds the nucleotides in the appropriate sequence to generate double-stranded DNA, using the primer as a starting point. This sequence of heating and replication is repeated through multiple cycles usually using automated systems.

7.2 Quantitative PCR (qPCR)

It is possible to make PCR quantitative by monitoring amplification in real time by using fluorescent probes that bind to double stranded DNA or labelled DNA probes that bind to specific sequences in the DNA. The method is made quantitative by calibration against known concentrations of a standard DNA.

7.3 Reverse Transcriptase PCR (rtPCR)

Standard PCR technology can only be used to amplify DNA but the technique can be modified for use with RNA by use of a reverse transcriptase enzyme to convert the RNA into DNA. It is then possible to use either standard or quantitative PCR methods.

7.4 Applications of PCR

PCR technology is now widely used in many applications and it played an important role in the Human Genome Project described earlier. In molecular biology, the qPCR method can be used to measure the number of copies of a gene in a genome, or using rtPCR, it is possible to measure levels of expression of mRNA transcripts. As an example, this can be useful when studying the effect of different environmental conditions on the expression of specific genes.

In healthcare, it is used in the diagnosis of some cancers and genetic disorders as well as the detection of viral infections, where it can detect the relevant virus even when it is present at very low levels. For example, rtPCR is being widely used in the diagnosis of COVID-19 along with lateral flow testing, which detects viral protein, not DNA. The PCR method is more sensitive but requires specialist facilities, unlike the lateral flow method which can be used in the home. In addition to detecting a number of viral infections, qPCR is also used to monitor the progress of infection and of treatment (Watzinger, 2006).

PCR has applications in other fields, for example in agriculture for identifying plant pathogens, and in the food industry for detecting pathogens and food adulteration (Bohme, 2019). It is also used to monitor water quality.

PCR is also very useful in identity testing (Saad, 2005), for example, in forensic science, where DNA from a crime scene can help to identify a suspect (McCord, 2019). The DNA of an individual is different from that of every other (except for identical twins). In 1984, Alec Jeffreys at the University of Leicester developed a genetic fingerprinting method that could distinguish between the DNA of different individuals. He cut DNA into fragments using restriction enzymes, which are found in bacteria, and cleave DNA at specific sites. The resulting fragments are then separated by electrophoresis, giving patterns which are unique to an individual. In practice, the amounts of DNA found at a crime scene are often too small to analyse directly, and PCR amplification is required to provide enough material for fingerprinting. This fingerprinting approach has other applications, for example in paternity testing and population ancestry.

Section 8

Nucleic Acid Synthesis

8.1 Introduction

As discussed in section 4, Arthur Kornberg had discovered an enzymatic route to synthesise replicas of DNA in the 1950s. Still, there was also growing interest in how DNA might be synthesised chemically. The first problem to be resolved was how to couple nucleotides, which was addressed by the chemist Lord Todd, working at Cambridge University (Figure 21). Lord Todd won a Nobel Prize for his work on the synthesis of nucleotides and was the first to synthesise a dinucleotide (Brown, 2000; Michelson and Todd, 1955; Michelson, 1961). Todd and his team did this by condensing 3'-0-acetylthymidine with thymidine 3'-(benzyl phosphochloridate) 5'-(dibenzyl phosphate) followed by removal of the protecting groups.

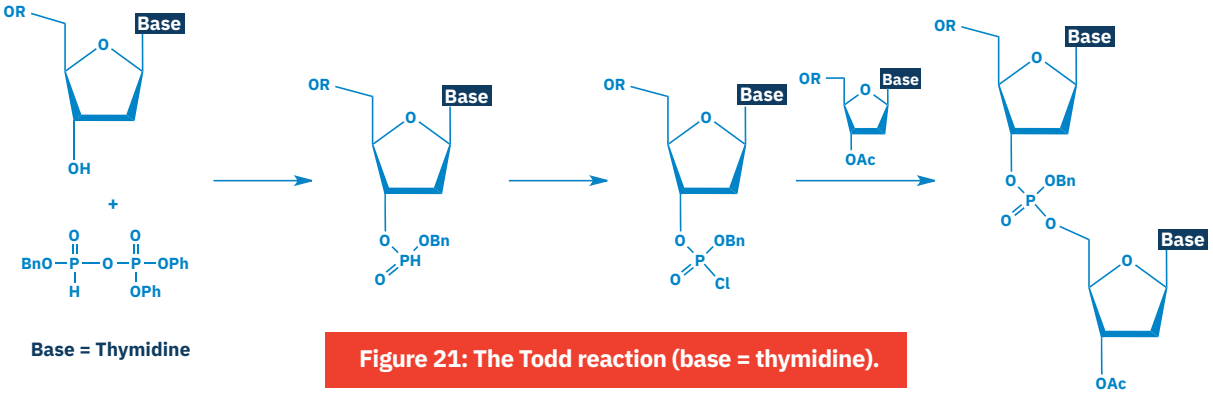


Figure 21: The Todd reaction (base = thymidine).

Further pioneering work using phosphodiester chemistry was carried out by Har Govind Khorana (Khorana, 1957) (Figure 22), who had worked with Todd and who as noted earlier did much of the synthetic work which established the basis of the genetic code.

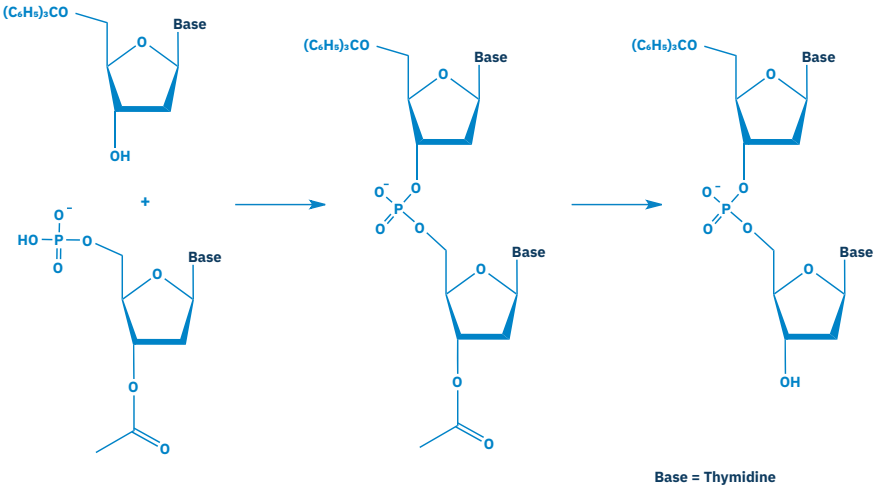


Figure 22: Khorana phosphodiester synthesis (base = thymidine).

In 1970, now working at MIT his team synthesised the DNA coding for the 77 nucleotide yeast alanine transfer RNA (Agarwal, 1970; Gilham, 1958; Gilham, 1959; Khorana, 1972; Khorana, 1979). This was the first synthesis of a complete gene. It involved the synthesis of 15 segments and then use of a ligase enzyme to link the segments.

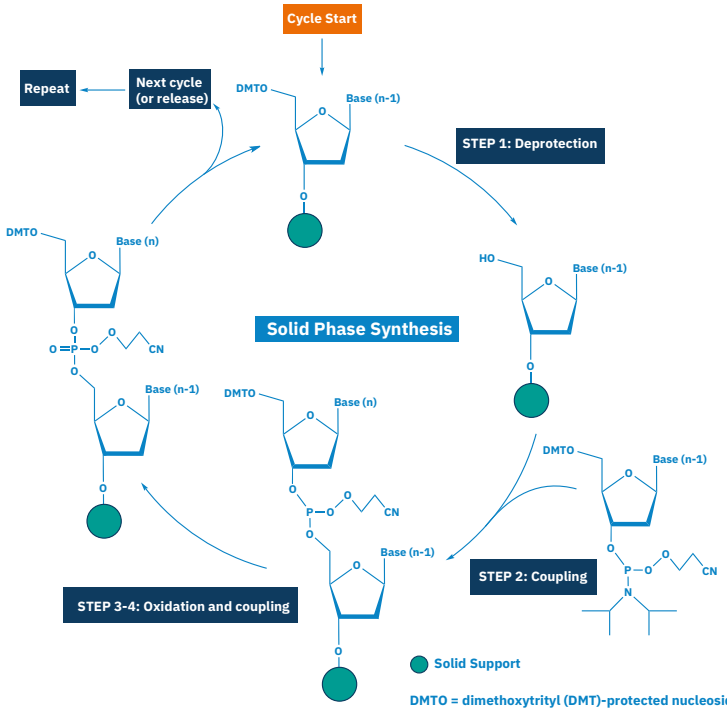
8.2 Solid Phase Synthesis

The next major step came with the development of solid phase synthesis, initially on polystyrene (Letsinger and Mahadevan, 1965; Letsinger and Mahadevan, 1966), and subsequently, in the early 1980s, on controlled pore glass using phosphoramidite chemistry (Caruthers, 2011; Caruthers, 2013; Matteucci and Caruthers, 1981). This solid phase technology (Figure 23) was hugely important because it allowed for stepwise addition of protected nucleosides without the need to separate the intermediates after each addition. It is still the standard method used today. This technology lent itself to automation and commercial machines were available from the 1980s.

Marvin H Caruthers

Oligonucleotides are typically purified by HPLC after release from the support. Synthesis is typically limited to oligonucleotides of about 200 nucleotides in length. Larger nucleic acids are made by synthesising subunits and using enzymes to couple them into the final structure.

In 1977, five years after Khorana's synthesis of the tRNA gene, the gene for somatostatin, a small protein containing just fourteen amino acids, was synthesised (Itakura, 1977) and expressed in *E. coli* cells by the fledgling biotechnology company Genentech. This was the first human protein to be produced in a bacterium. In 1978, they went on to synthesise the genes for the two chains of insulin and express them in *E.coli.* (Goeddel). This was the first therapeutic protein produced by recombinant DNA technology. It was approved by the FDA in 1982 and marketed by Eli Lilly, who developed the manufacturing process (Johnson, 1983). The production of recombinant therapeutic proteins, including monoclonal antibodies, has now become a very significant part of the pharmaceutical industry. Nucleic acid synthesis is now used in a range of applications in addition to recombinant protein production. Examples include the synthesis of oligonucleotides for therapeutic applications, as probes for detecting complimentary DNA or RNA, and as primers for use in sequencing and amplification of nucleic acids.



In addition to the synthesis of individual genes, it is now possible to synthesise whole genomes. The first genome to be synthesised was from a bacteriophage (Smith, 2003).

By 2008, a complete bacterial chromosome had been synthesised (Gibson, 2008). This work was undertaken at the J. Craig Venter Institute in the USA. The genome contained 582,970 base pairs claimed to be the largest chemically defined structure synthesised in the lab. The genome was synthesised as smaller components, which were then linked together. In 2010, this work went one step further when the same group synthesised a bacterial genome, which was then transferred to a bacterium lacking a genome to create a living cell capable of replication (Gibson, 2010).

The ability to synthesise genomes also makes it possible to make changes which give insights into how genomes function (Analuru, 2014).

Figure 23: Solid phase synthesis.

Section 9

Applications of Nucleic Acids

9.1 Diagnostics

As already discussed, techniques such as PCR and genome sequencing are important tools in diagnosing and monitoring a number of diseases, such as viral infections and cancer. Other techniques include Fluorescence *in situ* hybridisation (FISH) and comparative genomic hybridisation (CGH). FISH uses fluorescently labelled nucleic acid probes to detect genetic markers, usually for particular cancers, in histology samples (Levsky, 2003). CGH is a method for analysing changes in gene copy number in DNA extracted from the relevant tissue. Loss or gain of genetic material can be associated with a number of diseases. Fluorescently labeled DNA from test material is hybridised with normal DNA labelled with a second fluorophore, and differences in fluorescence indicate differences in copy number. The method is also used in microarray format.

There is a strong interest in developing diagnostic tools that can detect disease-related DNA in blood or other fluid samples that can be easily obtained, and especially having techniques that can detect cancer at an early stage. In a recent example Karan Budhraja and colleagues described a method using sequencing of DNA fragments from blood together with machine learning techniques to detect cancer (Budhraja, 2023).

Diagnostic methods are also becoming increasingly efficient particularly as sequencing becomes faster and more cost effective. For example, in a recent publication, scientists at Cambridge and Oxford Universities describe a technique using DNA "nanobaits" to simultaneously identify multiple viruses using the Oxford Nanopore system (Boscovic, 2023).

9.2 Gene Therapy

Recombinant DNA technology has made possible the generation of therapeutic proteins such as insulin and monoclonal antibodies (Smith, 2017). A more significant challenge, which is now beginning to deliver results, is the use of nucleic acids directly as therapeutic agents for treating various conditions including genetic diseases, cancer and infectious diseases.

The first category of these "gene therapies" involves treatments for diseases characterised by failure to produce an essential protein due to a faulty gene. One example is the failure to produce the Factor IX protein in patients with Haemophilia B. A treatment is now available where the gene for the coagulation Factor IX is transported into cells where it is able to generate the protein.

Other examples of this type of gene therapy are shown in Table 2. A key challenge has been how to get DNA into cells, and the approach taken is to use a viral vector. The vectors most used are based on a modified adeno-associated virus (AAV). This single-stranded DNA virus is non-pathogenic and is replication defective.

It is also possible to use this approach to introduce genes, not to substitute for a faulty protein-expressing gene but to express other proteins which have a therapeutic effect. In a recent example, the AAV vector has been used to introduce the gene for interferon alpha-2b in patients with bladder cancer (Table 2). Interferon is part of the body's natural response to fighting cancer (Mokkapati, 2022).

Table 2 Approved Gene Therapies for Protein Expression

Product	Disease	Protein Expressed	Approval (FDA)
Luxturna	Retinal dystrophy	RPE65 enzyme	2017
Zynteglo	Beta thalassemia	Beta globin	2022
Zolgensma	Spinal muscular atrophy	SMN protein	2019
Hemgenix	Haemophilia B	Factor IX	2022
Adstiladrin	Bladder cancer	Interferon alpha-2b	2022

9.2.1 Therapies Based on Gene Silencing or Activation

As understanding of the genetic disorders underlying a significant number of diseases has increased, it becomes increasingly apparent that in many diseases, the underlying causes can be more complex than simply not producing a particular protein. Sometimes genes are inappropriately expressed, and an approach to treatment is to down-regulate this expression using synthetic oligonucleotides. A number of approaches have been taken based on the use of DNA and RNA. Two types of nucleic acid have been developed that can be used to target transcription or translation specifically: single-stranded DNA or RNA antisense oligonucleotides (ASOs) and double-stranded RNAs (RNAi). ASOs are short synthetic DNA sequences of about 13-25 nucleotides that block or modify mRNA translation. RNAs (RNAi) molecules have already been discussed in the context of controlling gene expression. These include both small interfering RNAs (siRNA) and microRNAs (miRNA), which have been evaluated as a means to block translation. A significant number of these DNA and RNA therapeutics are now licensed (Table 3), and many more are being assessed.

In most cases, the approach has been to down-regulate gene expression, but there are instances where it can be useful to do the opposite. There are small activating RNAs (saRNA) that have the same basic structure as siRNas but up-regulate specific gene promoter sequences. There is interest in their therapeutic potential, particularly in cancer.

There is also interest in using antisense drugs as antivirals; one example, Bepiroversen is in clinical trials. It targets mRNAs produced by the Hepatitis B virus inhibiting the production of viral proteins (Yuen, 2021).

9.3 Aptamers

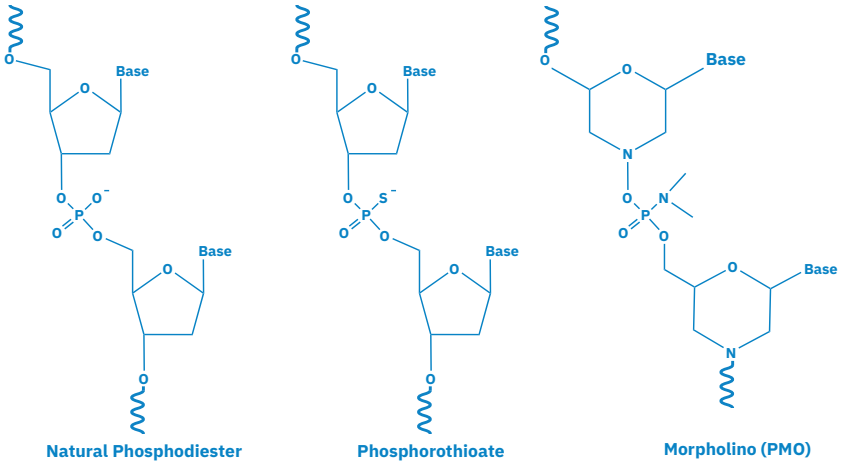
These are small (20-100 nucleotide) single-stranded DNA or RNA molecules that behave like chemical antibodies (Kulabhusan, 2020; Zhou, 2018). Aptamers recognising a wide range of target molecules, including proteins, can be selected from pools of random oligonucleotides. They can have advantages over monoclonal antibodies in some situations, and there is interest in them as therapeutics, diagnostics and research reagents. One aptamer, Macugen, was licensed in 2004 for use in macular degeneration. It is a pegylated 28 nucleotide RNA oligonucleotide which inhibits an enzyme, vascular endothelial growth factor.

There are major technical challenges with oligonucleotide drugs, particularly concerning finding effective means of delivering them to the relevant cells or tissue, getting them into cells and avoiding degradation by nuclease enzymes. To address this, a wide range of chemical modifications to the nucleic acid structure have been developed (Roberts, 2020). Several modifications to the ribose and phosphate components have been described. In the case of phosphate, it was found that incorporating sulphur to create a phosphorothioate linkage with ribose reduces degradation. Likewise, several ribose modifications were described which also increased stability. For a number of antisense oligonucleotides (ASOs), the whole ribose phosphate backbone is replaced by a backbone made up of methylene morpholone rings linked by phosphoroamidate groups (Figure 24). This confers significantly increased stability and reduced off-target effects. This type of oligonucleotide is often referred to as a morpholino and several are now approved (Table 3). DNA antisense oligonucleotides have been developed in which the DNA is flanked at both ends with short RNA-like sequences. These “gapmers” have increased resistance to nuclease degradation and higher affinity for the target RNA. The approved product Tegsedi is an example of a gapmer.

Table 3 Approved Oligonucleotide Drugs (FDA)

Drug	Disease target	Chemical	Approval
Macugen	Macular degeneration	Aptamer	2004
Exondys 51	Duchenne muscular dystrophy	Morpholino	2016
Spinraza	Spinal muscular atrophy	Antisense	2016
Defitelio	Veno-occlusive disease	Oligonucleotide	2016
Onpattro	Type of polyneuropathy	siRNA	2018
Tegsedi	Type of polyneuropathy	Antisense	2018
Givlaari	Acute hepatic porphyria	siRNA	2019
Vyondis 53	Duchenne muscular dystrophy	Morpholino	2019
Viltepso	Duchenne muscular dystrophy	Morpholino	2020
Oxlumo	Hyperoxaluria	siRNA	2020
Amondys 45	Duchenne muscular dystrophy	Antisense	2021
Leqvio	Hypercholesteraemia	siRNA	2021
Amvuttra	Transthyretin cardiac amyloidosis	siRNA	2022

Figure 24: Oligonucleotide modifications.



Several approaches have been taken to improve the delivery of oligonucleotides. Lipid conjugation or incorporation in lipid nanoparticles have been used to enhance uptake into cells, and there is also interest in using conjugates incorporating cell-penetrating peptides (Klabenkova, 2021). Conjugation to N-acetylgalactosamine has been used to target oligonucleotides into the liver. A further approach to site-specific targeting is conjugating the oligonucleotide to an antibody (Dugal-Tessier, 2021). Antibody oligonucleotide conjugates are also of interest as potential imaging and detection agents (Dovgan, 2019).

9.4 Vaccines

Over the last few decades, there has been much interest in using DNA and RNA vaccines against infectious diseases. The concept is to use DNA or RNA coding for protein antigens that will generate an immune response. Development of these vaccines accelerated dramatically in response to the COVID-19 pandemic caused by the SARS-CoV-2 virus and two RNA-based vaccines were developed by Moderna and Pfizer/BioNTech. Work is in progress to develop vaccines against other infectious diseases, including influenza (Arevalo, 2022), malaria (Hayashi, 2022) and against some cancers (Lorentzen, 2022). mRNA vaccines are typically encapsulated in lipid nanoparticles (Hou, 2021). This formulation stabilises the RNA and helps its transfer into cells where it is translated into antigen proteins recognised by the body's immune system.

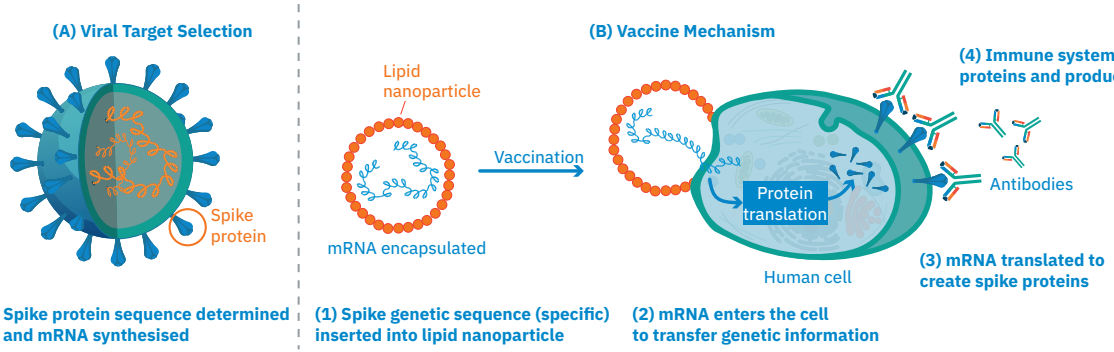


Figure 25: A general mechanism for a mRNA vaccine development.

9.5 CRISPR Technology

Another exciting new technology called CRISPR (clustered regularly interspaced short palindromic repeats) has become available making it possible to add, delete, or change DNA sequences much more precisely and efficiently than previous methods. It was originally discovered in bacteria, where it acts as a kind of immune system protecting against viral infection, and there is now strong interest in its use in several areas, including human therapy and agricultural crop improvement. The technique uses a short RNA sequence which attaches to a chosen target sequence in DNA or RNA. An associated enzyme, CAS9, cuts the DNA at this sequence. It is then possible to add, delete or replace sequences at this site. For example, this technology could be used as a therapy to replace a faulty gene or to prevent the translation of an RNA coding for a disease-related protein (Morelli, 2023).

9.6 CAR T-Cell Therapy

In the last few years, this new type of gene therapy has generated a lot of excitement in the field of cancer. The therapy is based on taking patients' T cell lymphocytes and genetically modifiying them to express a chimeric antigen receptor (CAR). The modified cells are then grown in culture and infused into the patient. The CAR protein is selected for its ability to bind to specific proteins in the relevant cancer cells allowing the T cells to attach to the cancer cells and kill them. By the end of 2022, there were six therapies approved by the FDA (Table 4). There are a large number of CAR-T cell therapies in development, and much effort is going into improving the technology, not least in developing universal CAR T- cell therapies. Making a universal therapy would be significantly less complex and, therefore, less costly (Feins, 2019). Research is also ongoing to extend this type of therapy beyond haematological cancers.

Table 4 Licensed CAR T-cell Therapies

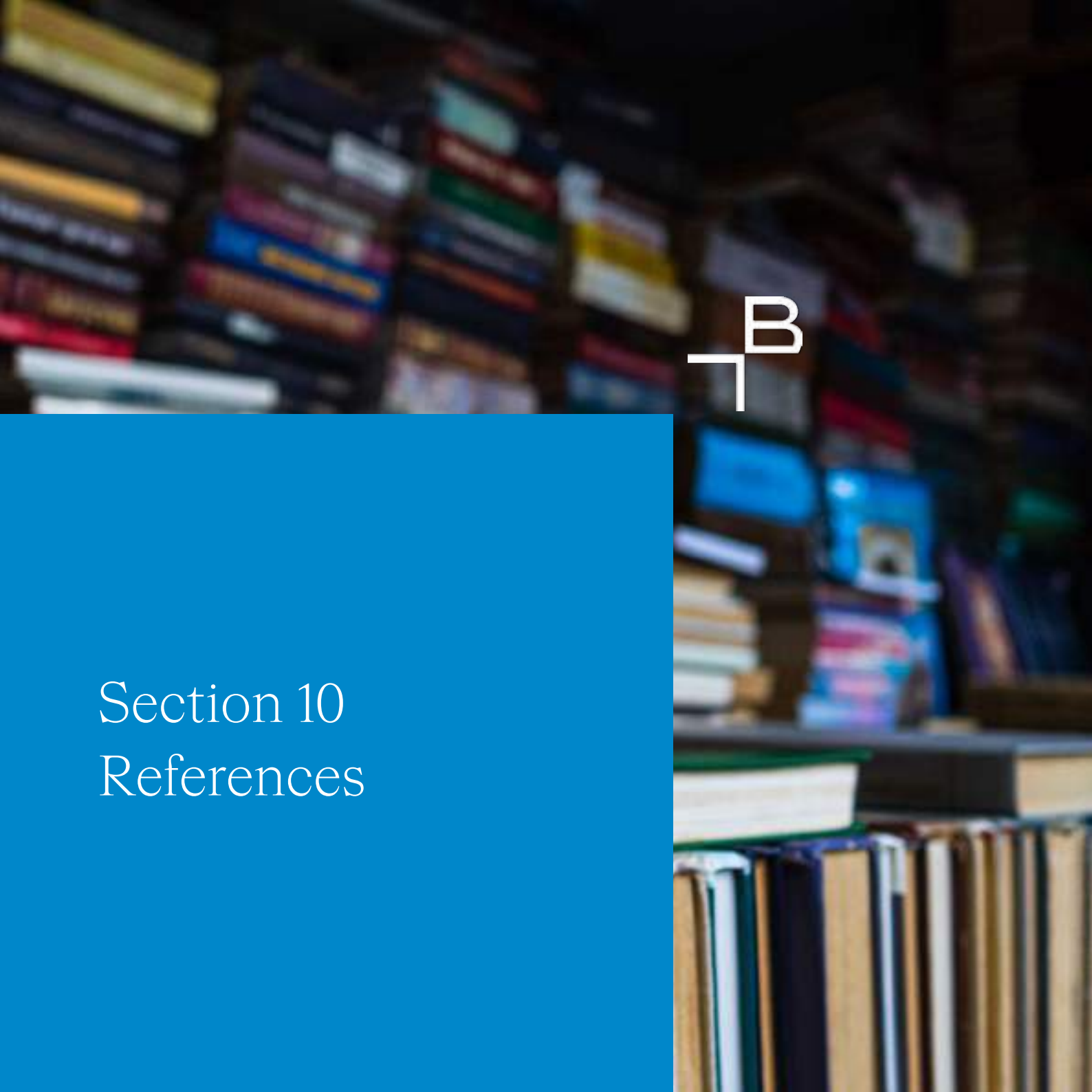
Product	Disease Target	Approval (FDA)
Kymriah	Acute lymphoblastic leukaemia	2017
Abecma	Multiple myeloma	2021
Tecartus	Mantle cell myeloma, acute lymphoblastic leukaemia	2021
Breyanzi	Non-Hodgkins lymphoma	2022
Yescarta	Non-Hodgkins lymphoma, follicular lymphoma	2022
Carvykti	Multiple myeloma	2022

9.7 Food and Agriculture

Genetic modification has been used in crop breeding programs to introduce genes that, for example, improve crop yields, product quality and the nutritional properties of plants (Table 5) (Fang, 2016; Philips, 2008). In addition to the examples given, research is ongoing to improve crop yields, for example, by improving the efficiency of photosynthesis (Simkin, 2019) and stress tolerance. The recent development of CRISPR-Cas technology, as in human therapeutic applications, is a significant step forward and makes it possible to carry out very precise gene editing (Zhu, 2020).

Table 5 Benefits from Genetically Modified Plants

Modification	Plant	Benefit
Herbicide resistance	Soybean, maize, rapeseed, sugar beet, alfalfa	Increased yield (allows the use of weed killers)
Insect pest resistance	Soybean, maize, cotton	Increased yield, reduced use of insecticides
Viral disease resistance	Papaya, squash	Better yield, less fruit damage
Reduced Polyunsaturated fats	Soybean	Reduced off flavours, better shelf life
Production of omega-3 and other fatty acids	Soybean	Nutritional benefits in foods and farmed fish
Reduced polyphenol oxidase	Potatoes, apples	Discoloration reduced (customer acceptance)



Section 10
References

Agarwal, K. L., Büchi, H., Caruthers, M. H., Gupta, N., Khorana, H. G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U. L., H., J., Sgaramella, V., & Weber, H. (1970). Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature,* 227(5253), 27–34.

Altmann R. (1889). Ueber nucleinsäuren. *Archiv für anatomie und physiologie, 1,* 524–536.

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., Cai, Y., Zeller, K., Agmon, N., Han, J. S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., Chandrasegaran, S. (2014). Total Synthesis of a Functional Designer *Eukaryotic Chromosome. Science* (New York, N.Y.), 344(6179), 55.

Arevalo, C. P., Bolton, M. J., Le Sage, V., Ye, N., Furey, C., Muramatsu, H., Alameh, M. G., Pardi, N., Drapeau, E. M., Parkhouse, K., Garretson, T., Morris, J. S., Moncla, L. H., Tam, Y. K., Fan, S. H. Y., Lakdawala, S. S., Weissman, D., & Hensley, S. E. (2022). A multivalent nucleoside-modified mRNA vaccine against all known influenza virus subtypes. *Science* (New York, N.Y.), 378(6622), 899–904.

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *The Journal of experimental medicine,* 79(2), 137–158.

Barker, K., (1998). Phenol-Chloroform Isoamyl Alcohol (PCI) DNA Extraction. *At the Bench.*

Berget, S. M., Moore, C., & Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA*. *Proceedings of the National Academy of Sciences,* 74(8), 3171–3175.

Bessman, M. J., Kornberg, A., Lehman, I. R., & Simms, E. S. (1956). Enzymic synthesis of deoxyribonucleic acid. *Biochimica et biophysica acta,* 21(1), 197–198.

Birnboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic acids research,* 7(6), 1513–1523.

Böhme, K., Calo-Mata, P., Barros-Velázquez, J., & Ortea, I. (2019). Review of Recent DNA-Based Methods for Main Food-Authentication Topics. *Journal of agricultural and food chemistry,* 67(14), 3854–3864.

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of clinical microbiology,* 28(3), 495–503.

Bošković, F., Zhu, J., Tivony, R., Ohmann, A., Chen, K., Alawami, M. F., Đorđević, M., Ermann, N., Fairhead, M., Howarth, M., Baker, S., & Keyser, U. F. (2023). Simultaneous identification of viruses and viral variants with programmable DNA nanobait. *Nature Nanotechnology,* 18(3), 290–298.

Brenner, S., Jacob, F., & Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature,* 190, 576–581.

Brown D. M., Kornberg H. (2000). Alexander Robertus Todd, O.M., Baron Todd of Trumpington. 2 October 1907 - 10 January 1997, *Biographical Memoirs of Fellows of the Royal Society,* 46, 515–532.

Brownlee, G. G., Sanger, F., & Barrell, B. G. (1967). Nucleotide Sequence of 5S-ribosomal RNA from Escherichia coli. *Nature,* 215(5102), 735–736.

Budhraja, K. K., McDonald, B. R., Stephens, M. D., Contente-Cuomo, T., Markus, H., Farooq, M., Favaro, P. F., Connor, S., Byron, S. A., Egan, J. B., Ernst, B., McDaniel, T. K., Sekulic, A., Tran, N. L., Prados, M. D., Borad, M. J., Berens, M. E., Pockaj, B. A., LoRusso, P. M., Bryce, A., Murtaza, M. (2023). Genome-wide analysis of aberrant position and sequence of plasma DNA fragment ends in patients with cancer. *Science translational medicine, 15*(678), eabm6863.

Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., & Poinar, G. O., Jr (1993). Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature,* 363(6429), 536–538.

Caruthers M. H. (2011). A brief review of DNA and RNA chemical synthesis. *Biochemical Society transactions,* 39(2), 575–580.

Caruthers M. H. (2013). The chemical synthesis of DNA/RNA: our gift to science. *The Journal of biological chemistry,* 288(2), 1420–1427.

Chargaff, E., Lipshitz, R., & Green, C. (1952). Composition of the desoxypentose nucleic acids of four genera of sea-urchin. *The Journal of biological chemistry,* 195(1), 155–160.

Chow, L. T., Gelinas, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell,* 12(1)*, 1–8.*

Correns C. (1900). G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde, *Berichte der Deutschen Botanischen Gesellschaft,* 18*,* 158–168.

Creeth, J. M., (1947), Some physico-chemical studies on nucleic acids and related substances, Ph.D. *Dissertation, University College Nottingham.*

Crick, F. H., Barnett, L., Brenner, S., & J., R. (1961). General Nature of the Genetic Code for Proteins. *Nature,* 192(4809), 1227–1232.

Cuatrecasas, P., Wilchek, M., & Anfinsen, C. B. (1968). Selective enzyme purification by affinity chromatography. *Proceedings of the National Academy of Sciences of the United States of America,* 61(2), 636–643.

Dahm R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human genetics,* 122(6), 565–581.

Dovgan, I., Koniev, O., Kolodych, S., & Wagner, A. (2019). Antibody-Oligonucleotide Conjugates as Therapeutic, Imaging, and Detection Agents. *Bioconjugate chemistry,* 30(10), 2483–2501.

Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus,* 12*,* 13–15.

Dudock, B. S., Katz, G., Taylor, E. K., & Holley, R. W. (1969). Primary structure of wheat germ phenylalanine transfer RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 62(3), 941–945.

Dugal-Tessier, J., Thirumalairajan, S., & Jain, N. (2021). Antibody-Oligonucleotide Conjugates: A Twist to Antibody-Drug Conjugates. *Journal of clinical medicine,* 10(4), 838.

Elkin, C., Kapur, H., Smith, T., Humphries, D., Pollard, M., Hammon, N., & Hawkins, T. (2002). Magnetic bead purification of labeled DNA fragments for high-throughput capillary electrophoresis sequencing. *BioTechniques,* 32(6), 1296–1302.

Elson, D., & Chargaff, E. (1952). On the desoxyribonucleic acid content of sea urchin gametes. *Experientia,* 8(4), 143–145.

Fang, J., Zhu, X., Wang, C., & Shangguan, L. (2016). Applications of DNA Technologies in Agriculture. *Current genomics,* 17(4), 379–386.

*F*eins, S., Kong, W., Williams, E. F., Milone, M. C., & Fraietta, J. A. (2019). An introduction to chimeric antigen receptor (CAR) T-cell immunotherapy for human cancer. *American journal of hematology,* 94(S1), S3–S9.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., & Ysebaert, M. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature,* 260(5551), 500–507.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature,* 391(6669), 806–811.

Flemming, W. (1875). Studien über die entwicklungsgeschichte der najaden. *Denkschriften der Kaiserlichen Akademie der Wissenschaften / Mathematisch-Naturwissenschaftliche Class, 71,* 81–212.

Friedberg E. C. (2006). The eureka enzyme: the discovery of DNA polymerase. *Nature reviews. Molecular cell biology,* 7(2), 143–147.

Frixione, E., & Ruiz-Zamarripa, L. (2019). The "scientific catastrophe" in nucleic acids research that boosted molecular biology. *The Journal of biological chemistry,* 294(7), 2249–2255.

Genthner, F.J., Hook, L.A. & Strohl, W.R. (1985) Determination of the Molecular Mass of Bacterial Genomic DNA and Plasmid Copy Number by High-Pressure Liquid Chromatography. *Applied and Environmental Microbiology,* 50, 1007–1013.

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* (New York, N.Y.), *319*(5867), 1215–1220.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* (New York, N.Y.), 329(5987), 52–56.

Gilham P.T. & Khorana H.G. (1958). Studies on polynucleotides. I. A new general method for the chemical synthesis of the C5"-C3" internucleotidic linkage. Syntheses of deoxyribo-dinucleotides, *Journal of the American Chemical Society, 80,* 6212–6222.

Gilham P. T. & Khorana H.G. (1959). Studies on polynucleotides. V. Stepwise synthesis of oligonucleotides. Syntheses of thymidylyl-(5'-3')-thymidylyl-(5' – 3')-thymidine and deoxycytidylyl-(5' – 3')-deoxyadenyl-(5' – 3')- thymidine2. J*ournal of the American Chemical Society,* 81, 4647–4650.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., & Riggs, A. D. (1979). Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America,* 76(1), 106–110.

Gordon, S. P., Tseng, E., Salamov, A., Zhang, J., Meng, X., Zhao, Z., Kang, D., Underwood, J., Grigoriev, I. V., Figueroa, M., Schilling, J. S., Chen, F., & Wang, Z. (2015). Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing. *PLOS ONE*, 10(7), e0132628.

Gorzynski, J. E., Goenka, S. D., Shafin, K., Jensen, T. D., Fisk, D. G., Grove, M. E., Spiteri, E., Pesout, T., Monlong, J., Baid, G., Bernstein, J. A., Ceresnak, S., Chang, P. C., Christle, J. W., Chubb, H., Dalton, K. P., Dunn, K., Garalde, D. R., Guillory, J., Knowles, J. W. Ashley, E. A. (2022). Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting. *The New England journal of medicine,* 386(7), 700–702.

Griffith F. (1928). The Significance of Pneumococcal Types. *The Journal of hygiene,* 27(2), 113–159.

Gros, F., Hiatt, H., Gilbert, W., Kurland, C. G., Risebrough, R. W., & Watson, J. D. (1961). Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. *Nature,* 190, 581–585.

Gross-Bellard, M., Oudet, P., & Chambon, P. (1973). Isolation of high-molecular-weight DNA from mammalian cells. *European journal of biochemistry,* 36(1), 32–38.

Hayashi, C. T., Cao, Y., Clark, L. C., Tripathi, A. K., Zavala, F., Dwivedi, G., Knox, J., Alameh, M., Lin, P. J., Tam, Y. K., Weissman, D., & Kumar, N. (2022). MRNA-LNP expressing PfCSP and Pfs25 vaccine candidates targeting infection and transmission of Plasmodium falciparum. *Npj Vaccines,* 7(1), 1–14.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics,* 107(1), 1–8.

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology,* 36(1), 39–56.

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., & Zamecnik, P. C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of biological chemistry,* 231(1), 241–257.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., & Zamir, A. (1965). Structure of a ribonucleic acid. *Science* (New York, N.Y.)*,* 147(3664), 1462–1465.

Hou, X., Zaks, T., Langer, R., & Dong, Y. (2021). Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials,* 6(12), 1078–1094.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., ... International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature,* 409(6822), 860–921.

IHGSC (International Human Genome Sequencing Consortium) (2004), Finishing the euchromatic sequence of the human genome, *Nature,* 431, 931–945.

Itakura K., Hirose T., Crea R., Riggs A. D., Heyneker H.L., Bolivar F., & Boyer H.W. (1977). Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science,* 198, 1056–1063.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology,* 3, 318–356.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics,* 33(3), 245–254.

Johnson I. S. (1983). Human insulin from recombinant DNA technology. *Science* (New York, N.Y.), 219(4585), 632–637.

Jones M. E. (1953). Albrecht Kossel, a biographical sketch. *The Yale journal of biology and medicine,* 26(1), 80–97.

Khorana H. G., Razzell W. E., Gilham P. T., Tener G. M. & Pol E. H. (1957). Synthesis of dideoxyribonucleotides. *Journal of the American Chemical Society,* 79, 1002–1003.

Khorana, H. G., Jacob, T. M., Moon, M. W., Narang, S. A., & Ohtsuka, E. (1965). Studies on polynucleotides.xlii. The synthesis of deoxyribopolynucleotides containing repeating nucleotide sequences. Introduction and general considerations. *Journal of the American Chemical Society,* 87, 2954–2956.

Khorana, H. G., Agarwal, K. L., Büchi, H., Caruthers, M. H., Gupta, N. K., Kleppe, K., Kumar, A., Otsuka, E., RajBhandary, U. L., Van de Sande, J. H., Sgaramella, V., Terao, T., Weber, H., & Yamada, T. (1972). Studies on polynucleotides. 103. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. *Journal of molecular biology,* 72(2), 209–217.

Khorana, H. G. (1979). Total Synthesis of a Gene. *Science,* 203(4381), 614–625.

Klabenkova, K., Fokina, A., & Stetsenko, D. (2021). Chemistry of Peptide-Oligonucleotide Conjugates: A Review. *Molecules (Basel, Switzerland),* 26(17), 5420.

Komar A. A. (2016). The Yin and Yang of codon usage. *Human molecular genetics,* 25(R2), R77–R85.

Kulabhusan, P. K., Hussain, B., & Yüce, M. (2020). Current Perspectives on Aptamers as Diagnostic Tools and Therapeutic Agents. *Pharmaceutics,* 12(7), 646.

Lathe, G. H., & Ruthven, C. R. (1955). The separation of substances on the basis of their molecular weights, using columns of starch and water. *The Biochemical journal,* 60(4), xxxiv.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* 75(5), 843–854.

Letsinger, R. L., & Mahadevan, V. (1965). Oligonucleotide synthesis on a polymer support. *Journal of the American Chemical Society,* 87, 3526–3527.

Letsinger, R. L., & Mahadevan, V. (1966). Stepwise synthesis of oligodeoxyribonucleotides on an insoluble polymer support. *Journal of the American Chemical Society,* 88(22), 5319–5324.

Levene P. A. (1900). On The Preparation of Nucleic Acids 1. *Journal of the American Chemical Society,* 22(6), 329–331.

Levsky, J. M., & Singer, R. H. (2003). Fluorescence in situ hybridization: past, present and future. *Journal of cell science,* 116(14), 2833–2838.

Li, L., Okino, S. T., Zhao, H., Pookot, D., Place, R. F., Urakami, S., Enokida, H., & Dahiya, R. (2006). Small dsRNAs induce transcriptional activation in human cells. *Proceedings of the National Academy of Sciences,* 103(46), 17337–17342.

Lorentzen, C. L., Haanen, J. B., Met, Ö., & Svane, I. M. (2022). Clinical advances and ongoing trials on mRNA vaccines for cancer treatment. *The Lancet. Oncology,* 23(10), e450–e458.

Massie, H. R., & Zimm, B. H. (1965). Molecular weight of the DNA in the chromosomes of E. coli and B. subtilis. *Proceedings of the National Academy of Sciences of the United States of America,* 54(6), 1636–1641.

Matthaei, J. H., & Nirenberg, M. W. (1961a). Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America,* 47(10), 1580–1588.

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America, 47*(10), 1588–1602.

Matteucci M. D., & Caruthers M. H. (1981). Synthesis of deoxyoligonulceotides on a polymer support. *Journal of the American Chemical Society,* 103, 3185–3191.

Mauro, V. P., & Chappell, S. A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends in molecular medicine,* 20(11), 604–613.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 74(2), 560–564.

McCord, B. R., Gauthier, Q., Cho, S., Roig, M. N., Gibson-Daw, G. C., Young, B., Taglia, F., Zapico, S. C., Mariot, R. F., Lee, S. B., & Duncan, G. (2019). Forensic DNA Analysis. *Analytical chemistry,* 91(1), 673–688.

Mendel G., (1866), Versuche über pflanzen-hybriden. *Verhandlungen des Naturforschenden Vereines, Abhandlungen, Brünn, 4,* 3–47.

Meselson, M., Stahl, F. W., & Vinograd, J. (1957). Equilibrium sedimentation of macromolecules in density gradients. *Proceedings of the National Academy of Sciences,* 43(7), 581–588.

Meselson, M., & Stahl, F. W. (1958). The replication of dna in escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America,* 44(7), 671–682.

Michelson A. M., & Todd A.R. (1955). Nucleotides part XXXII, Synthesis of a dithymidine dinucleotide containing a 3':5' internucleotidic linkage. *Journal of The Chemical Society,* 26, 32–2638.

Michelson A. M. (1961). Chemistry of the nucleotides. *Annual Review of Biochemistry,* 30, 133–164.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research,* 16(3), 1215.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lander, E. S., Knight, J. R., Irzyk, G. P., Fredrikson, K. M., Harkins, T. T., Sheridan, S., Schuster, S. C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature,* 456(7220), 387–390.

Mokkapati, S., Narayan, V. M., Manyam, G. C., Lim, A. H., Duplisea, J. J., Kokorovic, A., Miest, T. S., Mitra, A. P., Plote, D., Anand, S. S., Metcalfe, M. J., Dunner, K., Jr, Johnson, B. A., Czerniak, B. A., Nieminen, T., Heikura, T., Yla-Herttuala, S., Parker, N. R., Schluns, K. S., McConkey, D. J. Dinney, C. P. (2022). Lentiviral interferon: A novel method for gene therapy in bladder cancer. *Molecular therapy oncolytics,* 26, 141–157.

Monod, J., Pappenheimer, A., & Cohen-Bazire, G. (1952). La cinétique de la biosynthèse de la β-galactosidase chez E. Coli considérée comme fonction de la croissance. *Biochimica et Biophysica Acta,* 9, 648–660.

Morelli, K. H., Wu, Q., Gosztyla, M. L., Liu, H., Yao, M., Zhang, C., Chen, J., Marina, R. J., Lee, K., Jones, K. L., Huang, M. Y., Li, A., Smith-Geater, C., Thompson, L. M., Duan, W., & Yeo, G. W. (2023). An RNA-targeting CRISPR-Cas13d system alleviates disease-related phenotypes in Huntington's disease models. *Nature neuroscience,* 26(1), 27–38.

Morgan, T. H., Sturtevant A. H., Muller H. J., & Bridges C. B. (1915). The mechanism of mendelian heredity. New York: *Henry Holt*.

Mullis K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American,* 262(4), 56–65.

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature,* 463(7280), 457–463.

Palade G. (1975). Intracellular aspects of the process of protein synthesis. *Science* (New York, N.Y.), 189(4200), 347–358.

Pauling, L., & Corey, R. B. (1953). A Proposed Structure For The Nucleic Acids. *Proceedings of the National Academy of Sciences of the United States of America,* 39(2), 84–97.

Peterson E. A., & Sober H. A. (1956). Chromatography of Proteins. I. Cellulose Ion-exchange Adsorbents. *Journal of the American Chemical Society,* 78(4), 751–755.

Petersen, C. P., Bordeleau, M. E., Pelletier, J., & Sharp, P. A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Molecular cell,* 21(4), 533–542.

Philips T. (2008). Genetically modified organisms (GMOs): Transgenic crops and recombinant DNA technology. *Nature Education,* 1, 213.

Roberts, T. C., Langer, R., & Wood, M. J. (2020). Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery,* 19(10), 673–694.

Roy, S., & Caruthers, M. (2013). Synthesis of DNA/RNA and Their Analogs via Phosphoramidite and H-Phosphonate Chemistries. *Molecules,* 18(11), 14268–14284.

Saad R. (2005). Discovery, development, and current applications of DNA identity testing. *Proceedings (Baylor University. Medical Center),* 18(2), 130–133.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America,* 74(12), 5463–5467.

Scheuerbrandt, G., Duffield, A. M., & Nussbaum, A. L. (1963). Stepwise synthesis of deoxyribo-oligonucleotides. *Biochemical and biophysical research communications,* 11, 152–155.

Simkin, A. J. (2019). Genetic Engineering for Global Food Security: Photosynthesis and Biofortification. *Plants,* 8(12), 586.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature,* 321(6071), 674–679.

Smith, H. O., Hutchison, C. A., 3rd, Pfannkoch, C., & Venter, J. C. (2003). Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America,* 100(26), 15440–15445.

Smith M. (2017). DNA Sequence Analysis in Clinical Medicine, Proceeding Cautiously. *Frontiers in molecular biosciences,* 4, 24.

Sørensen, M. A., Kurland, C. G., & Pedersen, S. (1989). Codon usage determines translation rate in Escherichia coli. Journal of molecular biology, 207(2), 365–377. Statello, L., Guo C-J., Chen L-L., Huarte, M., (2021), Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology,* 22, 96-118

Statello, L., Guo, C., Chen, L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology,* 22(2), 96–118.

Strauss B. S. (2018). Why Is DNA Double Stranded? The Discovery of DNA Excision Repair Mechanisms. *Genetics, 209*(2), 357–366.

Sutton, W. S. (1903). The chromosomes in heredity. *The Biological Bulletin.* 4, 231–251.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., ... Zhu, X. (2001). *The sequence of the human genome. Science* (New York, N.Y.), 291(5507), 1304–1351.

Vogelstein, B., & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences of the United States of America,* 76(2), 615–619.

Watson, J. D., & Crick, F. H. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature,* 171(4356), 737-738.

Watzinger, F., Ebner, K., & Lion, T. (2006). Detection and monitoring of virus infections by real-time PCR. *Molecular aspects of medicine,* 27(2-3), 254–298.

Williamson J. R. (2009). The ribosome at atomic resolution. *Cell,* 139(6), 1041–1043.

Wolf G. (2003). Friedrich Miescher: The man who discovered DNA. *Chemical Heritage,* 21, 10-11.

Yuen, M., Heo, J., Jang, J., Yoon, J., Kweon, Y., Park, S., Tami, Y., You, S., Yates, P., Tao, Y., Cremer, J., Campbell, F., Elston, R., Theodore, D., Paff, M., Bennett, C. F., & Kwoh, T. J. (2021). Safety, tolerability and antiviral activity of the antisense oligonucleotide bepirovirsen in patients with chronic hepatitis B: A phase 2 randomized controlled trial. *Nature Medicine,* 27(10), 1725-1734.

Zhou, L. Y., He, Z. Y., Xu, T., & Wei, Y. Q. (2018). Current Advances in Small Activating RNAs for Gene Therapy: Principles, Applications and Challenges. *Current gene therapy,* 18(3), 134–142.

Zhu, H., Li, C., & Gao, C. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. Nature reviews. *Molecular cell biology,* 21(11), 661–677.

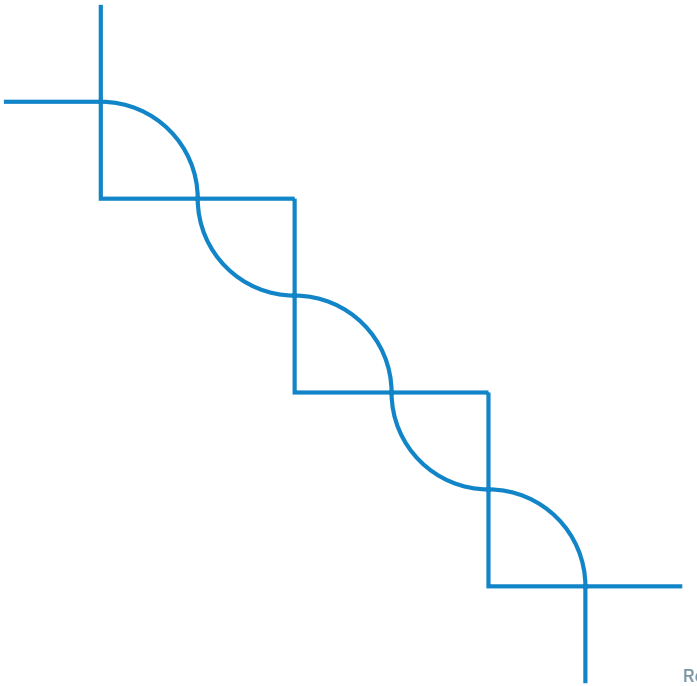
Zwart H. (2015). Human Genome Project: history and assessment. *International Encyclopedia of Social & Behavioral Sciences. 2nd ed. Oxford: Elsevier,* 311–317.

General Interest

Cooper G. and Adams K. (2022), The Cell, A Molecular Approach, 9th edition, Oxford University Press

Venter C. (2007). A Life Decoded: My Genome: My Life. Viking/the Penguin Group, New York.

Watson, J.D. (1968). The Double Helix. Weidenfeld and Nicholson.



BIOSYNTH®

Global locations

Discover all our locations at www.biosynth.com/worldwide-locations

Switzerland	India
United Kingdom	South Korea
Ireland	Japan
Slovakia	Austria
United States	Netherlands
China	Germany

APEX CHEMICALS

Apex Chemicals.,Ltd. (Head Office)

2034/87, Italthai Tower 19th Floor New Petchburi Road, Bangkapi, Huaykwang, Bangkok 10310, Thailand

Tel. 02-038-9999 Fax. 02-038-0099 Email : info@apexchemicals.co.th