Chapter 1: DNA: Its Discovery, Structure, and Mechanisms

The Discovery of DNA and its Structure

At the base of every lifeform on earth lies the basic unit of heredity known as genes. More importantly, genes are segments of Deoxyribonucleic acid also known as DNA. DNA is a polymer by which all life on earth's genetic information is contained. Before we explore the composition and functions of DNA, we will dive into the history around its discovery and the major players involved in the research surrounding the establishment of DNA's structure and composition.

Discovering DNA

Swiss-born scientist Friedrich Miescher began his formal education at the age of 17 at the University of Basel in Germany in 1862 and eventually graduated and passed his board exams in 1868. After a brief career in the medical field, Miescher pursued a new career path in the realm of research in the laboratory of Felix Hoppe-Seyler, a renowned physiologist and chemist known for establishing many principles of biochemistry. In his initial experiments, Miescher attempted to study lymphocytes but due to the logistics and technology of the time, couldn't efficiently obtain samples. With guidance from his mentor, he began to collect samples of drainage from surgical bandages at the local hospital and isolating the cells within the pus. During his testing, Miescher determined the ingredients comprising the cytoplasm of the cells, protein and lipids. In the process of performing his experiments, Miescher observed the presence of a mysterious byproduct after the addition of acid to his solution of cells and adversely, its disappearance after the addition of a base. This precipitate was a rudimentary form of DNA. Due to the properties of this material, Miescher attributed its origination to the nucleus of the cells. Subsequently, Miescher developed a novel method for isolating the DNA for study involving a series of salt washes and solvent additions. To begin, cells from the surgical bandages were cleaned and separated from contaminants. Next, the cells were washed with a sodium sulfate solution. The cells were then filtered allowing for any other materials to be left behind. In order to split the cell nuclei from the cytoplasm, Miescher then used a series of washings with Hydrochloric acid solution over a number of weeks. These nuclei were then agitated in an ether solution to break down the lipids within the cytoplasm still tarnished by cytoplasm while leaving the clean, cytoplasm-free nuclei alone.

The Structure of DNA

After the discoveries of Miescher, a Russian biochemist by the name of Phoebus Levene would continue to lead the charge towards understanding the structure and components of DNA. Levene pioneered the discovery of the carbohydrate-based components of both DNA and RNA, these being deoxyribose and ribose respectively. However, some of his most significant work was the discovery of the structure of a lone

nucleotide giving us the model that we know now. This model, presented in 1919, was known as the Polynucleotide model. Levene described the makeup of nucleic acids as a chain of nucleotides that when broken down were composed of one of the four nitrogenous bases (A,T,C,G), a sugar molecule, and an attached phosphate group.

Another important contribution to the realm of biochemistry and the establishment of the true structure of DNA was the work of Erwin Chargaff. In Chargaff's experiments performed in the late 1940's, he found that no organisms possessed the same order of nucleotides in their DNA and more importantly that the same exact DNA existed a similar amount of base pairs when comparing adenine (A) to thymine (T) and guanine (G) and cytosine (C). This was the first evidence of an equal distribution of purines (A + G) and pyrimidines (C +T) and the links between the two.

In addition to Chargaff's contributions to the field, other major contributors included figures such as Maurice Wilkins and Rosalind Franklin with discoveries in the field of X-ray crystallography. Wilkins and Franklin most notably are attributed with the first image of DNA in an image revered as "Photo 51" depicting its helical structure. Despite this breakthrough, only Wilkins would gain recognition for this contribution with the help of two other pioneers in the field.

Over 80 years after Miescher's contributions to biochemistry and genetics, two major players discovered another revolutionary advancement in the field. These figures were James Watson and Francis Crick. Watson and Crick immortalized themselves in the field of biochemistry through their work that substantiated the true structure of DNA. With the help of the scientists before them, eventually, through a combination of x-ray diffraction, a process in which x-rays are shot at atoms and are scattered by their negatively charged particles to produce an image, and troubleshooting by building physical models of DNA, and eventually in 1953 they were able to present to the world the double helix we associate with DNA today. This double helix consisted of two strands of complementary base pairs bonded by hydrogen and attached to a sugar-phosphate backbone. Watson and Crick also noted the anti-parallel layout of DNA, which is depicted as the 5' end of a strand being attached to the 3' end of the strand complementary to it. Although in recent years there have been small changes in the interpretations of results, four constants about DNA based on the Watson and Crick model still hold:

1. The structure of DNA is a double-stranded helix bonded by hydrogen in which base C's are always paired with G's and A's always with T's following Chargaff's discoveries.
2. Most of the helices of DNA are right-handed twists, the only that is left-handed is a type of DNA known as Z-DNA.
3. The structure of the double-helix is antiparallel, meaning that the 5' end of one strand is complementary or in conjunction with the 3' end of the pair it is attached to.
4. The outsides of the nitrogenous bases remain open for bonding with proteins and other molecules essential to the production of DNA.

Process of DNA Replication

One of the most important processes about heredity is DNA replication. DNA replication is the method by which our DNA is copied into new strands during the process of cell division. It is the basis by which our bodies can produce healthy cells

whether it's due to injury or the process of growing. The process of DNA replication is semiconservative meaning one of the two strands in the DNA helix is conserved and used as a framework for the building of a new strand. Replication begins with an initiator protein and the subsequent attachment of a protein known as the DNA helicase. The DNA helicase cuts the hydrogen bonds between the nitrogenous bases in the DNA strand causing them to separate into single strands. This breaking apart happens at a spot known as the origin. After the helicase splits the strands, another protein called DNA primase creates and attaches an RNA primer which is a short segment of RNA used to signal a specialized enzyme known as DNA polymerase to come to the site. DNA polymerase then interacts with the single strands and uses free-floating nucleotides to assemble the new strand of DNA by choosing nucleotides that pair with each other (A+T, G+C) which is known as complementary base pairing. Complementary base pairing allows for the new copy to be the opposite or anti-sequence of the template strand so that when the new copy itself is replicated later on the original template sequence will once again be created. This process allows for the original DNA strand to be preserved.

The Central Dogma of Molecular Biology

The Central Dogma of Molecular Biology is the basis of all we know when it comes to what we currently know about genetics. It states that it is only possible for genetic information to flow in the sequence of DNA to RNA to Protein. This flow is made up by the steps of transcription in which DNA is transcribed into RNA and translation in which the information with RNA is used to create proteins.

Transcription

The transcription of DNA was first seen in the 1970's under an electron microscope. Scientists noticed a structure resembling a tree branch in that there was a central structure with various smaller structures growing out from it. The researchers added DNase and RNase, enzymes that degrade DNA and RNA, and noticed with the addition of DNase that the base structure disappeared and with the addition of RNase the branching structures were destroyed. This experiment provided first-hand visual evidence for DNA being used to create RNA through transcription. As stated previously transcription is the copying of genetic information from DNA into RNA for some later use such as the creation of a protein. The major steps of transcription are: initiation, elongation, and termination. Intertwined with these three steps is an enzyme called RNA polymerase which produces single-stranded RNA in contrast with the double-stranded DNA produced by DNA polymerase. The transcription process is as follows:

1. Initiation: Similar to replication, this step begins with the unwinding of DNA at a location known as the promoter. Here an enzyme known as a holoenzyme begins to separate the two DNA strands, the template and the coding strands, allowing for them to be used further.
2. Elongation: In this step of the process, one of the separated strands begins to be copied, this strand is called the template strand. RNA polymerase builds the RNA strand out of base pairs complementary to the template strand resulting in an order identical to the strand that wasn't used, also called the coding strand, with the exception that all Thymines (T) are replaced with Uracil (U). These nucleotides are added starting at the 3' end of the strand and move towards the 5' end.

3. Termination: In this final step, sequences of nucleotides called terminators are recognized by RNA polymerase. After the transcription of a terminator, RNA polymerase is signaled to release the strand it is coding, forming a U-shaped fold called a hairpin loop.

After transcription, the genetic information is known as a messenger RNA or mRNA due to the nature of its role which is delivering data from the DNA in the nucleus to a ribosome to begin translation. The mRNA's usability immediately after transcription differs in prokaryotes and eukaryotes. In Prokaryotes the mRNA can be used directly after translation, however, in eukaryotes a few more steps are required. After eukaryotic transcription instead of mRNA, pre-mRNA is made and components are added on to make it mRNA. The structures added are a 5' cap consisting of a single guanine (G) nucleotide attached by a triphosphate, similar to that in the structure of DNA, and a poly-A tail composed of a chain of adenine (A) nucleotides. These additions assist the now mature mRNA by protecting it from degradation allowing it to safely be transported to the cytoplasm for usage by a ribosome.

Translation

Once transcription is completed and the mature mMRA sequence is transported from the nucleus into the cytoplasm, translation may begin. In translation, the information within the mRNA is interpreted and used to create a chain of amino acids, called a polypeptide, resulting in a protein or component of a protein. Within the genetic information encoded in mRNA, the directions come in the form of codons, 61 to be precise, which are sets of 3 nucleotides that each encode for one of the 20 specific amino acids typically found in proteins. An important codon is AUG also known as the start codon because it signals the building of new proteins. Other noteworthy codons include UAA, UAG, and UGA which are known as stop codons. These codons indicate the end of a polypeptide signaling translation to end but do not code for amino acids. The two major components involved in translation are a form of RNA called a Transfer RNA, or tRNA for short, and ribosomes. Each tRNA reads and attaches to specific codons with a component called an anticodon all while bringing along the matching amino acid to be assembled. Ribosomes are a packet of RNA and associated proteins, comprising two parts called subunits. These subunits, one small and one large, surround a piece of mRNA and provide tRNA with docking points known as A (amino acid),P (polypeptide), and E (exit) sites in which they can bring the amino acids. The ribosome also links the amino acids brought by the tRNAs eventually creating a protein. Just as in transcription, the three steps in translation are: initiation, elongation, and termination. Below is the sequence in which translation is performed:

1. Initiation: In the first step, the ribosome engulfs the mRNA segment and the primary codon (AUG) is attached by the tRNA, this combination of ingredients is known as the initiation complex and begins every instance of translation.
2. Elongation: The step is where the chain of amino acids begins to be built upon and extended. It begins with a tRNA and its attached amino acid linking to the mRNA in the A site on the large subunit via its anticodon. Next in the P site the amino acid transferred from the tRNA and added to the chain. The tRNA then exits through the E site to find another amino acid to attach itself to and repeat

the cycle. Meanwhile, the ribosome is cementing the amino acids together and moving along the mRNA in the 5' to 3' direction until the protein is complete.

3. Termination: The final step occurs with the incorporation of one of the stop codons (UAA, UAG, UGA)which cannot be detected by tRNA, instead when these codons enter the ribosome they induce the production of something called a release factor that causes the translation to end.

The Sequencing of DNA

DNA Sequencing

To begin, it is important to know the steps and purpose of Polymerase Chain Reaction or PCR as it will provide a basis for the sequencing method that will be discussed later on. At its core PCR consists of 3 major steps. The first is known as denaturation in which the strand of DNA being used is heated to a temperature in which the two strands of DNA that make up the helical structure can no longer hold together and break apart. Next is the annealing step. In annealing, small fragments of DNA known as primers are bound to the single strands of DNA that have been separated, giving the free nucleotides that are being added a place to attach. The third and final step of the cycle is the extension phase. In this phase, DNA polymerase is used to extend the end of the split strands to allow for more nucleotides to be added. This 3 step cycle is repeated over two dozen times in a machine called a thermocycler to produce the desired amount of copies.

The method of DNA sequencing we're going to discuss is Sanger sequencing. This method of sequencing was developed by Frederick Sanger in 1977 and is also called chain-termination sequencing. In Sanger sequencing, as a special method of Polymerase Chain Reaction, a technique used to produce millions of copies of a particular DNA fragment to be studied, is the primary mechanism by which the end product is achieved. In this technique PCR is performed but where as typically it would only use a structures called deoxynucleoside triphosphates or dNTPs which contain the building blocks of DNA that are used in the process of PCR to create copies of strands, the process also uses dideoxynucleoside triphosphates or ddNTPs that are marked fluorescently. These ddNTPs lack a part called the 3' hydroxyl group that is found on dNTPs and necessary for the formation of phosphodiester bonds, an important component for the extension of DNA sequences. This causes the chain of DNA to stop extending. This process continues thousands of times allowing the researcher to distinguish all of the genes in the sequence. This method of sequencing is extremely accurate and is still used today in many basic tasks. After these terminated chains called oligonucleotides are separated by gel electrophoresis. This is able to be performed because of DNA's negative charge. The gel is polarized pulling the oligonucleotides through the gel towards the positive pole. This separates the oligonucleotides by size since the smaller strands experience less friction and are able to move further into the gel.

References

National Institutes of Health. (n.d.). *The discovery of The double helix, 1951-1953* | *Francis Crick - profiles in Science*. U.S. National Library of Medicine. Retrieved March 9, 2023, from https://profiles.nlm.nih.gov/spotlight/sc/feature/doublehelix

Pray, L. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* 1(1):100 https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/

***PCR Basics*. Thermo Fisher Scientific - US. (n.d.). Retrieved March 9, 2023, from https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-basics.html**

Clancy, S. (2008) DNA transcription. *Nature Education* 1(1):41

Clancy, S. & Brown, W. (2008) Translation: DNA to mRNA to Protein. *Nature Education* 1(1):101

***Intro to gene expression (central dogma) - khan academy*. (n.d.). Retrieved March 31, 2023, from https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/translation/a/intro-to-gene-expression-central-dogma**

Nature Publishing Group. (n.d.). Nature news. Retrieved March 30, 2023, from https://www.nature.com/scitable/topicpage/cells-can-replicate-their-dna-precisely-6524830/#

Nature Publishing Group. (n.d.). Nature news. Retrieved March 30, 2023, from https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393/

Sanger Sequencing Steps & Method. (n.d.). Retrieved March 30, 2023, from https://www.sigmaaldrich.com/US/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing