DNA
The discovery of the structure and function of the genetic substance
The Discovery of DNA

- DNA was discovered by Friedrich Miescher in 1869 in Tubingen, Germany.
- He used bandages from wounded soldiers to obtain cell *nuclei* (discovered in 1831 by Robert Hooke).
- From these he obtained a gelatinous substance he called *nucleic acid*.
- This was the first substance found to contain nitrogen and phosphorus.
Proteins as the genetic substance

• Throughout most of the 19th and into the 20th century proteins were considered to be the basis of heredity

• It was known that proteins were made up of some 20 amino acids joined together and the large number of different proteins that had been discovered were considered to be a basis for the variation needed in the genetic substance, although no actual mechanism for heredity was known

• For most of this time nucleic acid was considered an obscure substance of no particular interest
Chemistry of DNA

• The chemistry of nucleic acid remained obscure.
• Its chemical structure was elucidated by Phoebus Aaron Levene working at the Rockefeller Inst. in NYC.
• Levene was one of the few Jews who obtained an MD at the Imperial Medical Inst. in St. Petersburg.

P.A. Levene (1869-1940)
Nucleotide

- Levene showed by mild hydrolysis in 1909 that the three components of nucleic acid: phosphate, sugar and base were arranged in that order.
- This basic unit of nucleic acid was termed a nucleotide.
- At that time it was thought that animals and plants had different nucleic acids with distinct sugar components.
Nucleic Acid Sugars

- Levene showed in 1928 that the sugar component of “animal” nucleic acid (DNA) was the previously unknown 2-deoxy-ribose
- Since ribose had been found in yeast nucleic acid, at this point two nucleic acids were thought to exist: ribonucleic acid (RNA) in plants and deoxyribonucleic acid (DNA) in animals
Tetranucleotide Hypothesis

- Four bases (A,T,G,C) had been discovered in animal nucleic acid
- This led to the tetranucleotide hypothesis
- However, the specific structure of the hypothetical tetranucleotide was unknown
The chemical structure of DNA

The actual chemical structure of DNA including the nature of the inter-nucleotide bond was proven by synthesis by Dan Brown and Lord Todd in Cambridge in the early 1950s.

Lord Todd (1907-97) Nobel Prize 1957
Physical Studies on DNA

- Experiments showed that DNA was in fact a very large molecule (it was very viscous) which argued against the tetranucleotide hypothesis.
- **William Astbury**, working in Leeds, pioneered X-ray diffraction studies of fibers in the 1930s, and was the first to apply this method to DNA.
- He was also the first to use the term “Molecular Biology” in 1950.
X-ray Diffraction of DNA

• In 1938 Astbury passed X-rays thru a dried sample of DNA and obtained a diffuse pattern that showed the presence of periodicity within the molecule

• He calculated that the main structural elements were 3.3 Å apart
How X-ray Diffraction works

• X-rays are reflected off the periodic layers of atoms in crystals or fibers
• By analyzing the pattern obtained the structure of the molecules can be elucidated
Linear Structure

- Assuming that the nucleotides were flat and co-planar, Astbury in 1938 proposed a single-stranded structure that he likened to a "pile of pennies" joined together by phosphate groups.

- He estimated that there must be ca. 300 such nucleotides in one molecule.
In 1944 Oswald Avery and his group at Rockefeller Inst. published results of studies of the pathogenicity of pneumococcus bacteria. They proved that it was the DNA component of the bacteria that conferred pathogenicity and not the protein component. This was the breakthrough that was needed to spur interest in the study of the 3D structure of DNA.
Nucleotide structure

- After WWII Sven Furberg went from Norway to work with J.D. Bernal in London
- He solved the first crystal structure of a nucleoside in 1949
- It had the base perpendicular to the sugar, so that Astbury’s model for DNA was incorrect
- Bernal showed that water molecules bound together through weak hydrogen bonds

The 3d structure of Cytidine

Sven Furberg (1920-83)

J.D. Bernal (1901-71)
Helices in Proteins

- Linus Pauling was the first in 1950 to show the presence of a helix in a protein, called the $\alpha$-helix.
- He realized that the peptide bond was planar, but the inter-peptide links could twist.
- It became clear that this was a common element in most protein structures.
DNA as a Helix

- Based on his structure of cytidine with the base perpendicular to the sugar (the “standard configuration”) Furberg speculated on the structure of DNA
- In his thesis (1949) he proposed two possible structures, one stepwise and the other helical
- This was a single stranded helix, but it was the first helical structure proposed for DNA
Titration of DNA

- J.M. Gulland found in 1947 that if he titrated DNA carefully that there was a hysteresis (a lagging effect) at low and high pH
- The reason for this phenomenon was unknown at that time
Base Ratios of DNA

- Erwin Chargaff at Columbia University showed in 1950 that the amount of purine and pyrimidine bases were equal in DNA such that $G=C$ and $A=T$.
- But he had no rational explanation for this result.

Erwin Chargaff (1905-2002)
DNA Fibers

- **Maurice Wilkens** working at King’s College in London discovered in 1950 that he could pull fibers out of a DNA gel.
- When dried, a bundle of these fibers gave a fairly defined X-ray diffraction pattern.

Maurice Wilkens (1916-2004) Nobel Prize 1953
B-form DNA

- Wilkens recruited Rosalind Franklin to solve the structure of DNA. But, they soon had a falling out.

- Franklin realized she must control the humidity and discovered that there were in fact two forms of DNA.

- At low humidity she called it the A-form and at high humidity the B-form.

- She decided to try to solve the structure of the A-form.
Watson & Crick

• James Watson and Francis Crick began working on the structure of DNA in 1950
• Watson saw Franklin’s B-form photo and described it to Crick
• Crick who was a crystallographer realized that it had the characteristic X-pattern of a helix
Base Pairing

- Jerry Donohue introduced Watson to the concept of base pairing thru hydrogen bonds.
- Putting all this information together, Watson and Crick proposed a double-helical model for DNA.
- Pauling proposed a triple helical model that was wrong.
The Double-Helical Structure of DNA

Watson and Crick’s proposed structure when published in 1953 caused a sensation and was perhaps the greatest discovery in biology in the 20th century.
Watson and Crick published a note in 1953 that stated that they realized that their structure had genetic implications.

Meselson & Stahl in a classic expt. in 1958 showed that DNA replication was semi-conservative, i.e. one of the original strands remained but a new strand was synthesized.

The unchanged strand acts as a template for the synthesis of the new strand.
Evidence for the W-C structure

- It is consistent with the chemical structure
- It explains the hysteresis of DNA titration
- It explains Chargaff’s base ratios
- It is consistent with semi-conservative replication
- The Hershey-Chase expt. in 1952 proved that bacteriophage infect bacteria by injecting their DNA, not their protein
- A crystal structure of an oligomer of DNA was published by Dickerson in 1968 that has a double helical structure
- It explains genetics thru the base sequences
The crystal structure of an oligomer of DNA shows that it has the Watson-Crick double helical structure.
The Genetic Code

• The sequence of the 4 bases (A, T, G, C) in DNA determine the genetic function, but how?
• In 1961 Marshall Nirenberg at NIH developed a cell-free system to study the translation of DNA into proteins
• He elucidated the genetic code, showing that three bases were required to code for one amino acid in a protein
• A combination of three bases was called a codon

Marshall Nirenberg (1927-2010)
Nobel Prize 1968
Base sequencing

• DNA is the genetic substance and the base sequence of DNA determines the genetic characteristics of an organism (genotype determines phenotype)
• Therefore determining the DNA base sequence can provide genetic information
• Frederick Sanger developed the first automated chemical method to determine the sequence of the four bases A, T, G, C in DNA in Cambridge in 1977
• This process was made much more practical with the introduction of the Polymerase Chain Reaction (PCR) by Kary Mullis in 1983 that amplifies the amount to DNA to be analyzed
Applications of DNA Sequencing

- The Human Genome Project published the total sequence of the human genome in 2003, consisting of ca. 20,000 genes and 6.6 billion bases
- Sequencing can determine parentage and heredity (e.g. Anastasia)
- In forensics to determine guilt or innocence from samples of blood or tissue left at a crime scene (e.g. the Innocence Project)
- To determine the relationships of ethnic groups (e.g. Polynesians)
- To elucidate the evolutionary tree connecting species (e.g. cytochrome c)
- In genetic diseases such as cancer to define the precise genetic pattern of the disease (e.g. lung cancer)
- To produce superior genetically modified (GM) forms of food crops, such as wheat, etc.
- To transfer desirable characteristics from one organism to another
- To use gene therapy to eliminate genetic diseases (e.g. Cystic Fibrosis)
Chronology of DNA

- 1869 – Discovery of DNA by Miescher
- 1909 – Chemical structure of nucleotide shown by Levene
- 1928 – Deoxyribose discovered by Levene
- 1939 – First X-ray diffraction of DNA by Astbury
- 1944 – Genetic function of DNA proven by Avery
- 1949 – X-ray structure of nucleoside by Furberg
- 1950 – \( \alpha \)-helix shown in proteins by Pauling
- 1950 – Base ratios discovered by Chargaff
- 1950 – DNA fibers prepared by Wilkens
- 1951 – B-form DNA observed by X-ray diffraction by Franklin
- 1952 – Inter-nucleotide bond proven by Brown & Todd
- 1953 – Double-helical structure of DNA proposed by Watson & Crick
- 1958 – Semi-conservative replication proven by Meselson & Stahl
- 1961 – Genetic code elucidated by Nirenberg
- 1977 – First automated base sequencing by Sanger
- 1983 – PCR invented by Mullis
- 2003 – Human genome published