A Crash Course in Genetics

General Overview:
- DNA Structure
- RNA
- DNA Replication
- Encoding Proteins
- Protein Folding
- Types of DNA
- Manipulating DNA
- PCR

DNA Is Structured Hierarchically

Levels of Structure
- Double Helix
- Histones / Nucleosomes
- Solenoid Supercoil
- Chromatin / Minibands
- Chromosomes

DNA Compacted to Conserve Space

There are several levels at which DNA is compacted:
1. The double helix — the DNA in a single cell contains 2.9 x 10^9 base pairs and would be a meter long.
2. Nucleosome — DNA is wound around a histone protein core to form a nucleosome. This gives a 5- to 9-fold reduction in length.
3. Solenoids — Nucleosomes (beads on a string) supercoil and form solenoid structures. 4-6-fold reduction in length.
4. Minibands — Solenoid turns loop around a protein-RNA scaffold to form Minibands. 18-fold reduction in length.
5. Chromosomes — Minibands further condense to form Chromosomes, the form of DNA as seen during cell division and genetics studies.

Putting the Puzzle Pieces Together

In 1953, James Watson and Francis Crick discovered the structure of the DNA double helix

What DNA is Made Of

DNA = deoxyribonucleic acid
- deoxyribose sugar with the 2’OH (hydroxyl) group missing
- Phosphate group(s) (not shown here, attach to 3’OH)
- Nitrogenous base — Adenine, Guanine, Thymine, Cytosine
- Together these components make up a nucleotide
DNA Replication — Making Copies

As cells divide, identical copies of the DNA must be made. The following sequence of events occurs:
- The weak hydrogen bonds between the strands break, leaving exposed single nucleotides.
- The unpaired base will attract a free nucleotide that has the appropriate complementary base.
- Several different enzymes are involved (unwinding helix, holding strands apart, gluing pieces back together, etc).
- DNA Polymerase, a key replication enzyme, travels along the single DNA strand adding free nucleotides to the 3’ end of the new strand (directionality of 5’ to 3’).
- DNA Polymerase also proofreads the newly built strand in progress, checking that the newly added nucleotide is in fact complementary (avoidance of mutations).
- This continues until a complementary strand is built (semi-conservative model).

More About DNA Replication

The rate of DNA replication is relatively slow, about 40-50 nucleotides per second. Recalling the length of DNA, it would take 2 months to replicate from one end to the other. Nature overcomes this by having many replication start points: replication origins.
The “Central Dogma”

The central dogma relates to the flow of ‘genetic’ information in biological systems.

DNA
\[\text{transcription}\]
mRNA
\[\text{translation}\]
Protein

DNA's Purpose in Nature: Encoding Proteins

Before proteins can be assembled, DNA must undergo two processes:
1) Transcription
2) Translation

Step 1: DNA Transcription

- Process involves formation of messenger RNA sequence from DNA template.
- Although DNA is the same in all tissues, there are different promoters which are activated in different tissues, resulting in different protein products being formed.
- Gene splicing (removing introns) further modifies the sequences that are left to code, ultimately producing different protein products from the same gene.
- RNA polymerase enzymes bind to promoter site on DNA, pull local DNA strands apart.
- Promoter sequence orientates RNA polymerase in specific direction, as RNA has to be synthesized in the 5' to 3' direction (same linking pattern as DNA).
- One DNA strand is used preferentially as template strand, although either could be used.
- Post-transcriptional modifications (5' methyl cap and poly-A-tail protect mRNA from degradation).

Transcription Example

DNA double-strand sequence:
5' CAG AAG AAA ATT AAC ATG TAA 3'
3' GTC TTC TTT TAA TTG TAC ATT 5'
mRNA sequence:
5' CAG AAG AAA A U U AAC AUG UAA 3'

NOTE: same as template strand of DNA, except T->U

A Crash Course in Genetics

General Overview:
- DNA Structure
- RNA
- DNA Replication
- Encoding Proteins
- Protein Folding
- Types of DNA
- Manipulating DNA
- PCR

Step 2: Translation & The Genetic Code

Proteins are made of polypeptides, which are in turn composed of amino acid sequences. The body contains 20 different amino acids, but DNA is made up of 4 different bases. Thus we need combinations of bases to denote different amino acids.

Amino Acids are specified by triplets of bases (codons):

<table>
<thead>
<tr>
<th>DNA (bases)</th>
<th>RNA (bases)</th>
<th>Protein (amino acid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td>U</td>
</tr>
<tr>
<td>G</td>
<td>U</td>
<td>C</td>
</tr>
</tbody>
</table>

Protein Folding

Tertiary Structure

Quaternary Structure
For a given sequence there are three possible reading frames.

One-Letter Amino Acid Naming Scheme

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>One-Letter Name</th>
<th>One-Letter Name</th>
<th>One-Letter Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>ala</td>
<td>A</td>
<td>U</td>
</tr>
<tr>
<td>Arginine</td>
<td>arg</td>
<td>R</td>
<td>C</td>
</tr>
<tr>
<td>Asparagine</td>
<td>asn</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>Aspartic</td>
<td>asp</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>Cysteine</td>
<td>cys</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Glutamine</td>
<td>gln</td>
<td>Q</td>
<td>G</td>
</tr>
<tr>
<td>Glutamic</td>
<td>glu</td>
<td>E</td>
<td>G</td>
</tr>
<tr>
<td>Glycine</td>
<td>gly</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>his</td>
<td>H</td>
<td>M</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>ile</td>
<td>I</td>
<td>F</td>
</tr>
<tr>
<td>Leucine</td>
<td>leu</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>lys</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>met</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>phe</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>pro</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>ser</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>thr</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>trp</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>tyr</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>val</td>
<td>V</td>
<td>V</td>
</tr>
</tbody>
</table>

A Crash Course in Genetics

Example continued:

mRNA sequence:
5’ CAG AAG AAA AUU AAC AUG UAA 3’

amino acid sequence (using Genetic Code):
Gln-Lys-Ile-Asn-Met-STOP
The final folded three dimensional (tertiary) structure is an intrinsic property of the primary structure.

In general, proteins are unstable outside of the cell and very sensitive for solvent conditions.

Proteins do follow rules in folding, but which rules they apply are unpredictable. Rules include:
- Interior is densely packed.
- Minimal exposure of nonpolar groups
- Backbone of polar groups are buried
- Folding with minimal conformational strains preferred
- Elements of secondary structure that are adjacent in sequence tend to be adjacent in tertiary structure

Although protein structure can be determined relatively easily using various crystallography and spectroscopy methods, as of until now, it is impossible to predict protein folding based on the primary amino acid sequence.

A Crash Course in Genetics

Predictability of Protein Folding

One Model of Protein Folding

• Many (10^{16}) different unfolded states (U) quickly equilibrate to a small number of partially folded, marginally stable intermediates (I).
• Kinetic restraints under refolding conditions cause (U) to converge to a common folding pathway.
• Intermediates have a preference for partially folded conformations.
• Last transition from I4 to F is a slow equilibrium with a nearly folded transition state.

General Overview:
- DNA Structure
- RNA
- DNA Replication
- Encoding Proteins
- Protein Folding
- Types of DNA
- Manipulating DNA
- PCR
Types of DNA

Fewer than 10% of the three billion nucleotide pairs in the human genome actually encodes proteins.

There are several categories of DNA:

3. **Single copy DNA** - seen only once in a cell, makes up about 75% of the genome, includes protein-coding genes. Most of this DNA is found in introns or in sequences that lie between genes.

4. **Dispersed Repetitive DNA** - as name suggests, this repetitive DNA is scattered singly throughout the genome.

5. **Satellite DNA** - repetitive DNA found in clusters around certain chromosome locations. Called so because they can be easily separated by centrifugation.
   - Makes up about 10% of genome.
   - Highly variable, source of differentiation between people.

Manipulating DNA: Laboratory Uses

Measuring length of DNA molecules:

Knowing the atomic weight of a nucleotide, and markers, **gel electrophoresis** separates pieces of DNA by weight, with the heavier (longer) segments moving slower and the lighter (shorter) segments moving faster through the gel.

These bands are compared with “markers” (pieces of DNA with known molecular weights and lengths), which are run simultaneously with the segments to be measured.

Gel Electrophoresis

Gel electrophoresis allows us to separate pieces of DNA. The DNA can be stained with ethidium bromide and then detected using ultraviolet light. The bright “bands” on this gel are DNA fragments produced by restriction enzyme digestion. The bands to the far sides are DNA size markers.

Nature’s Secret: Denaturation & Renaturation

Recall that the two strands of DNA are held together by weak hydrogen bonds. Thus, heating the double-strand DNA increases the kinetic energy, breaking these bonds.

A-T-rich regions separate first (recall, two H-bonds between A and T as opposed to three bonds between G and C).

This property allows researchers to estimate the relative AT vs GC content in a segment of DNA, according to how quickly the DNA denatures.

If the temperature is lowered again slowly, the DNA can renature.

Process must be done slowly so that correct base pairing can occur.

Consequently, DNA denaturation is REVERSIBLE and is useful in the laboratory!
Other Operations on DNA

- **DNA can be lengthened**, through polymerases, providing there is a primer (existing sequence partially bonded to a template) and a free 3’ end to which bases can be added.

- **DNA can be shortened**, via DNA nucleases:
  - **Exonucleases** (like Pacman, eating from one side to the other) cut from the ends, removing one nucleotide at a time.
  - **Endonucleases** (like a pair of scissors taken to the middle of a strip of paper) cut from the inside, leaving either “sticky ends” or blunt ends.
  - **Restriction endonucleases** are most useful in genetics research.
    - They cut at specific sites, and only cut ds DNA.

This will become very useful for DNA Computing!

A Crash Course in Genetics

**General Overview:**
- DNA Structure
- RNA
- DNA Replication
- Encoding Proteins
- Protein Folding
- Types of DNA
- Manipulating DNA
- PCR

Multiplying DNA — What is PCR?

PCR = Polymerase Chain Reaction

**Problem:**
To be able to use DNA segments in the laboratory, one often needs multiple copies of the segment. Nature’s solution (DNA replication) is too slow, and requires in vivo conditions.

**Purpose:**
Some potential uses for many copies of DNA include:
- Forensics - identifying the guilty party through genetic analysis.
- Genetically Inherited Diseases - some diseases are inherited through a mutation of a single gene. The presence of that gene could be detected using PCR to exaggerate its presence, allowing detection.

The Scientist’s Solution: PCR

PCR (polymerase chain reaction) is a laboratory-based method of immitating nature’s DNA replication.

**We need:**
3. Two primers, each 15-20 bases long (oligonucleotides), corresponding to the DNA sequences on either side of the sequence of interest
4. DNA polymerase, a thermally stable form (thermophilic bacterium origin) to mimic DNA replication
5. A large collection of free DNA nucleotides
6. A template strand (Genomic DNA from an individual)

Schematic Representation of PCR

PCR — How It Works

1. Heat the genomic DNA to denature, resulting in a single stranded template.
2. Expose the DNA to the primers, allowing them to hybridize (under cooler conditions) to the appropriate locations on either end of the sequence of choice.
3. Reheat the DNA to an intermediate temperature and expose the mixture to free DNA bases, allowing a new DNA strand to be synthesized by DNA polymerase.
   This results in a double stranded sequence of DNA.
4. Heat the double-stranded DNA to a high temperature, causing it to denature.
5. Repeat steps 2-4 to multiply sequence of choice.
The Pros and Cons of PCR

Benefits and Advantages of PCR
- The heating-cooling cycle takes minutes, allowing amplification of sequence to occur quickly.
- PCR can be used with extremely small quantities of DNA (blood stain, single hair, saliva on postage stamp).
- DNA produced is very pure, thus do not need radioactive probes to detect certain sequences or mutations.

Downfalls of PCR
- Primer synthesis requires knowledge of the DNA sequence around the DNA segment of interest.
- PCR is extremely sensitive, therefore prone to contamination in laboratory.
- Limited to short sequences (1000 bases) only.

References
- Julie Stromer, A Crash Course in Genetics, CPSC 601.73 (W2002) presentation.
- Mike Surette, CPSC 605 (F2003).