Lecture 4
Hierarchical Organization of the Genome
by John R. Finnerty

Genome Architecture
Structural Subdivisions

1. Nucleotide: monomer — building block of DNA
2. DNA: polymer — string of nucleotides
3. Gene: stretch of DNA encoding a functional RNA or protein; typically 100’s-1000’s of nucleotides
4. Chromosome: unit of inheritance; DNA + protein packaging. Independent assortment of maternal and paternal chromosomes is responsible for “Mendelian inheritance.”
5. Genome: all the organism’s DNA packaged into 1 or more chromosomes
Nucleotide

Nucleotides are monomers — building blocks of DNA
The DNA alphabet uses four different nucleotides: G A T C

Each nucleotide consists of 3 parts.
1. one or more phosphates [PO₄]
2. a 5-carbon sugar called deoxyribose
3. a nitrogen containing “base”

The 5 carbons of deoxyribose are numbered.
The 1’ carbon is bound to the “base.”
The 3’ carbon is bound to a hydroxyl group (OH)
The 5’ carbon is bound to the phosphate (PO₄)

deoxyribose
Nucleotides

Four different nucleotides are present in DNA — 4 letters of the DNA alphabet: A, C, G, T

DNA

Each strand of DNA is a linear polymer of nucleotide monomers held together by covalent bonds (G-A-T-C-C-G). The DNA double helix comprises two intertwined linear strands of DNA that are held together by hydrogen bonds between nucleotides from opposing strands. The structure of DNA fits a general pattern with biological macromolecules:

- a chain of monomers is bound together using covalent bonds
- diverse 3 dimensional shapes are achieved via hydrogen bonds among non-adjacent members of the chain
Biological Polymers

A polymer consists of many similar units bound together. The individual units are called monomers. Polysaccharides (carbs), proteins, and nucleic acids (DNA, RNA) are polymers.

Building the polymer releases water [Dehydration]
Breaking apart the polymer requires water [Hydrolysis]

Covalent & Hydrogen Bonds in Biological Macromolecules

In polysaccharide chains, amino-acid chains (polypeptides or proteins), and nucleotide chains (DNA or RNA), the bond between adjacent monomers is a covalent bond.

Bonds between non-adjacent members of the chain help give the chain a more complex 3D shape (such as a donut shape, a double helix, or a pleated sheet). These bonds are generally hydrogen bonds.

These hydrogen bonds are more flexible because they are more easily dissolved and re-formed under different conditions.
Monosaccharides / Polysaccharides

Cellulose is a polymer of glucose.

Amino acids / Protein

A protein is a polymer of amino acid monomers.
Protein “secondary structure”

Hydrogen bonds stabilize protein structure.

DNA

Each strand of DNA is a linear polymer of nucleotide monomers held together by covalent bonds (G-A-T-C-C-G). The DNA double helix comprises two intertwined linear strands of DNA that are held together by hydrogen bonds between nucleotides from opposing strands.
DNA Polymerization

Building the polymer:
(1) Covalent bonds between nucleotides create a DNA strand.
(2) Directionality: the polymerizing strand grows from the 5' to the 3' direction.

5' : Phosphate (— PO₄)
3' : Hydroxyl (—OH)

Imposes directionality on the monomer. [like Legos]

DNA Polymerization

Favored connection

Not favored
DNA Polymerization

**Polymerization reaction involves removal of water (dehydration reaction)**

How does the DNA polymerization process “know” the proper sequence of nucleotides?

- **“The Double Helix”** — DNA exists as a double-stranded molecule. Bases on one DNA strand engage in hydrogen bonds with bases on the other DNA strand.

- Each strand can act as a TEMPLATE for the synthesis of a new second strand.

- The rules of base pairing dictate the sequence of bases in each newly synthesized strand.
  - the base C pairs (H-bonds) with the base G.
  - the base T pairs (H-bonds) with the base A.
Watson & Crick (1953)

The "Dark Lady of DNA" who generated the data that inspired the double-helix model of Watson and Crick.

Rosalind Franklin (1952)

The "Dark Lady of DNA" who generated the data that inspired the double-helix model of Watson and Crick.

X-ray diffraction image of DNA
“Until now, however, no evidence has been presented to show how it (DNA) might carry out the essential operation required of genetic material, that of exact self-duplication.”

“We have recently proposed a structure for the salt of deoxyribonucleic acid which, if correct, immediately suggests a mechanism for its self duplication.”

“The first feature of our structure which is of biological interest is that it consists not of one chain, but of two. These two chains are both coiled around a common fibre axis....”
Watson & Crick (1953)

“The other biologically important feature is the manner in which the two chains are held together. This is done by hydrogen bonds between the bases.”

“The important point is that only certain pairs of bases will fit into the structure. One member of a pair must be purine (A/G) and the other a pyrimidine (C/T) in order to bridge the two chains. If a pair consisted of two purines, for example, there would be no room for it.”

Watson & Crick (1953)

“Previous discussions of self-duplication have usually involved the concept of a template, or mould. Either the template was supposed to copy itself directly, or it was to produce a ‘negative,’ which in its turn was to act as a template and produce the original ‘positive’ once again.”

“Now our model for deoxyribonucleic acid is, in effect, a pair of templates, each of which is complementary to the other.”
Double Helix
Base Pairing
H-bonding between bases on opposing strands

Covalent bonds
“sugar-phosphate backbone”

DNA Replication

Hydrogen bonds
Base pairing
DNA Replication

Step 1. Separate the two strands (the protein Helicase unwinds the helix).

DNA Replication

Step 2. The protein DNA polymerase adds nucleotides to a new strand.
DNA Replication

Gene

A stretch of DNA that “encodes” a functional RNA or protein. It provides enough information to produce the RNA or protein molecule in the right context. In a particular subcellular location, in a particular cell type at a particular time depending on an organism’s developmental state or physiological condition.
### ATP-synthase
Membrane channel & enzyme that catalyzes production of ATP.

Protein must be produced in all the cells of the body.

### Insulin
Hormone that circulates in the blood and regulates sugar metabolism.

Protein produced only in the islet cells of the pancreas.

---

## Two kinds of information contained in a gene.

<table>
<thead>
<tr>
<th>ATP-synthase regulatory region</th>
<th>ATP-synthase protein coding region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sythesize ATP-synthase protein in all cells</strong></td>
<td><strong>To make ATP-synthase protein, assemble this particular sequence of amino acids</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insulin regulatory region</th>
<th>Insulin protein coding region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sythesize insulin in particular pancreas cells</strong></td>
<td><strong>To make insulin protein, assemble this particular sequence of amino acids</strong></td>
</tr>
</tbody>
</table>

**ATP-SYNTHASE gene**

**INSULIN gene**
Interrupted Genes


Alternate Splicing

- An important source of complexity in the transcriptome of eukaryotes.
- “Many different studies predict that from 40 to 60% of human genes are alternatively spliced (Croft et al., 2000; Hide et al., 2001; Kan et al., 2001; Modrek et al., 2001).”
Conserved Intron Locations

- Evolutionarily related genes often have conserved gene structure (conserved positions of introns)
- Three mammalian dihydrofolate reductase (DHFR) genes
  (source: Lewin, Genes X)

Drosophila Development

- (A) 2 hours
- (B) 5-8 hours
- (C) 10 hours
- (D) 2 hours
- (E) 5-8 hours
- (F) 10 hours

0.5 mm
Even-skipped

- Even-skipped is a homeodomain protein, closely related to Hox genes.
- In Drosophila, it functions as a pair-rule gene.
- It is expressed in every other parasegment (includes the posterior part of a future segment plus the anterior part of the next most posterior segment).
- A mutation in even-skipped or another “pair-rule” gene (fushi-tarazu) causes a deletion of part of every other segment.

The even-skipped promoter
Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo

Stephen Small, Adrienne Blair and Michael Levine

Biology Department, Center for Molecular Genetics, Baner Hall, UCSD, La Jolla, CA 92037-0322, USA

Communicated by E. Hafen

In an effort to determine how crude gradients of transcriptional activators and repressors specify sharp stripes of gene expression in the early embryo, we have conducted a detailed study of even-skipped (eve) stripe 2. A combination of promoter fusions and P-transformation assays were used to show that a 480 bp region of the eve promoter is both necessary and sufficient to direct a stripe of LacZ expression within the limits of the endogenous eve stripe 2. The maternal morphogen bicoid (bcd) and the gap proteins hunchback (hb), Kruppel (Kr) and giant (gt) all bind with high affinity to closely linked sites within this small promoter element. Activation appears to depend on cooperative interactions among bcd and hb proteins, since disrupting single binding sites cause catastrophic reductions in expression. gt is directly involved in the formation of the anterior border, although additional repressors may participate in this process. Forming the posterior border of the stripe involves a delicate balance between limiting amounts of the bcd activator and the Kr repressor. We propose that the clustering of activator and repressor binding sites in the stripe 2 element is required to bring these weakly interacting regulatory factors into close apposition so that they can function both cooperatively and synergistically to control transcription.

---

**Chromosome**

A unit of inheritance—genes are passed from parent to offspring on chromosomes.

A single long DNA molecule + associated packing proteins.

May be a few thousand nucleotides long or hundreds of millions of nucleotides long.

May comprise many thousands of individual genes.
Composition of Chromosomes

- Chromosomes consist of DNA in association with proteins that facilitate its packaging and un-packaging.

  • When a cell is dividing, DNA needs to be tightly bound in a stable form so that it can be dragged into a new daughter cell efficiently and safely (without being ripped to pieces).

  Cooper (1997) *The Cell. A Molecular Approach*

Composition of Chromosomes

- Hierarchy of Organization:
  - DNA double helix
  - Wrapped around the nucleosomes, spherical protein structures consisting of histones.
  - DNA in association with these proteins is called chromatin
  - Chromatin is condensed (tightly folded) and compacted into discrete chromosomes during cell division.

  Cooper (1997) *The Cell. A Molecular Approach*
Histone Packing of DNA

10 nm fiber

30 nm fiber
## Chromosome Length

<table>
<thead>
<tr>
<th>Chromosome Length</th>
<th>Length of DNA molecule (nucleotides)</th>
<th>Length of DNA molecule (cm)</th>
<th>Length of condensed chromosome (m)</th>
<th>Shortened by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortest Human Chromosome (#22)</td>
<td>$33.4 \times 10^6$</td>
<td>$1.1 \times 10^{-2}$ m</td>
<td>$1 \times 10^{-6}$ m</td>
<td>about 10,000 X</td>
</tr>
<tr>
<td>Longest Human Chromosome (#1)</td>
<td>$250 \times 10^6$</td>
<td>$8.5 \times 10^{-2}$ m</td>
<td>$9 \times 10^{-6}$ m</td>
<td>about 10,000 X</td>
</tr>
</tbody>
</table>

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Chromosomes.html