Eukaryotic Cell
The Lysosome

- involved in the degradation of molecules
- found in animal cells
- membrane limited
- varies in shape and size
- contains acid hydrolases (phosphatase, nucleases, proteases, etc.), enzymes that work only at acid pH levels
- acid pH level maintained through H+ pumps
- if contents spill into the cytosol no degradation takes place (pH in [7.0-7.3])
- it degrades membranes and organelles that have outlived their usefulness
- it plays role in the degradation of extracellular macromolecules (endocytosis: exterior -> cell)

Tay-Sachs disease: inherited, recessive neurodegenerative disease, leads to death by age 5. Defective lysosome is missing an enzyme (β-N-hexosaminidase-A) needed to degrade continuously accumulating constituent of plasma membrane of mammalian cells (particularly nerve cells)
Molecules In A Cell: DNA/RNA
Molecules In A Cell: Carbohydrates

- **general formula:**
  \[
  \left[ \text{CH(OH)} \right]_n
  \]

- **disaccharides** = two mono-s linked together
  - e.g. lactose = glucose + galactose
  - sucrose = glucose + fructose

\[
\begin{align*}
  &\text{O} \\
  | & | \\
  C_{nH_{2n+1}} - C - H & \text{aldehyde} \\
  & \text{or} \\
  &\text{O} \\
  | & | \\
  C_{nH_{2n+1}} - C - C_{mH_{2m+1}} & \text{ketone}
\end{align*}
\]

and two or more hydroxyl groups
Molecules In A Cell: Lipids

- Biomembranes separate a cell from its surroundings
- Lipids are the structural elements of biomembranes
- Fatty acids are the principal components of lipids
- General formulas for acids:

  \[
  \begin{align*}
  &\text{saturated} \\
  &C_{n+1}H_{2n+1} - C - OH \\
  &\text{unsat. / single double bond} \\
  &C_{n-1}H_{2n} - C - OH \\
  &\text{unsat. / 2 double-1 triple bond} \\
  &C_{n-3}H_{2n-3} - C - OH
  \end{align*}
  \]
Molecules In A Cell: Proteins

Figure 2-2 The structures of the 20 common amino acids. In each structure, a central carbon atom (the α carbon) is bonded to an amino group (or to an imino group in proline), a carboxyl group, a hydrogen atom, and an R group. The R groups are in red.
Molecules In A Cell: Proteins (cont.)

H O
\[ \begin{array}{c}
| \ \ \ |
\end{array} \]
NH₂ – C_a – C – OH
\[ \begin{array}{c}
| \\
R
\end{array} \]

POLAR:
uncharged (=S, T, Q, N),
+ charged (=K, R, H),
- charged (=D, E)

HYDROPHOBIC: I, L, M, V, F, A, Y, W

OTHER: C (=SH allows for disulphide bond)
G (=small) P (=rigid)
Polymerization

\[
\text{NH}_2 - \text{C}_a - \text{C} - \text{OH} + \text{HNH} - \text{C}_a - \text{C} - \text{OH} \quad \rightarrow \quad \text{peptide bond}
\]
And in 3 dimensions...

(a) Because the carbon-nitrogen peptide bond has a partial double-bond character, the peptide group is planar.
(b) However, there is considerable flexibility in the geometry of polypeptides: rotation is possible about the two covalent single bonds that connect each α carbon to the two adjacent planar peptide units. But some restrictions do apply to the values of ψ and φ. For example, if the pictured adjacent peptide groups were coplanar, then certain oxygen and hydrogen atoms would be separated by less than their van der Waals radii and would repel one another.
Enzymes

- Proteins that catalyze chemical reaction
- Mediators of dynamic events.
- Enzymes increase the rates of reactions.

Naming convention:

XXX - ase is the name of the enzyme acting on molecule XXX

Examples:

- XXX = protein -> protease
- XXX = rna -> ribonuclease
- XXX = dna -> deoxyribonuclease
Antibodies

- produced after invasion by infectious agent(s)
- the recognition site of an antibody can bind tightly to very specific sites (generally on surface proteins or carbohydrates of the infectious agent)
- experimentally, animals will produce antibodies for any foreign injected polymer
- they act as signals for the elimination of infectious agents
- they have exquisite specificity
- composed of 2 heavy and 2 light chains; the N-termini of heavy/light chains are highly variable
- also useful in isolating proteins from a mixture
Composition Of Cells

Figure from: http://www.accessexcellence.org/AB/GG/macroMols.html
http://www.essentialcellbiology.com Published by Garland Publishing,
a member of the Taylor & Francis Group.
The Central Dogma
Genes On DNA: Prokaryotes

template strand
Genes On DNA: Eukaryotes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length</th>
<th>#introns</th>
<th>introns as % length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin</td>
<td>1.4 Kb</td>
<td>2</td>
<td>67</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>18 Kb</td>
<td>13</td>
<td>88</td>
</tr>
<tr>
<td>Cystic Fibrosis</td>
<td>250 Kb</td>
<td>26</td>
<td>98</td>
</tr>
<tr>
<td>Dystrophin</td>
<td>2.3 Mb</td>
<td>&gt;100</td>
<td>99</td>
</tr>
</tbody>
</table>
Genes On DNA (cont.)

Figure from: http://www.accessexcellence.org/AB/GG/exon2.html Artist: Darryl Leja
## Codons

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td><strong>Stop (och)</strong></td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td><strong>Stop (amb)</strong></td>
</tr>
</tbody>
</table>

| **C**                   |                 |                        |
| Leu                     | Pro             | His                    | Arg         | U          |
| Leu                     | Pro             | His                    | Arg         | C          |
| Leu                     | Pro             | Gln                    | Arg         | A          |
| Leu                     | Pro             | Gln                    | Arg         | G          |

| **A**                   |                 |                        |
| Ile                     | Thr             | Asn                    | Ser         | U          |
| Ile                     | Thr             | Asn                    | Ser         | C          |
| Ile                     | Thr             | Lys                    | Arg         | A          |
| **Met (start)**         | Thr             | Lys                    | Arg         | G          |

| **G**                   |                 |                        |
| Val                     | Ala             | Asp                    | Gly         | U          |
| Val                     | Ala             | Asp                    | Gly         | C          |
| Val                     | Ala             | Glu                    | Gly         | A          |
| Val **(Met)**           | Ala             | Glu                    | Gly         | G          |
Genes On DNA (cont.)

- gene clusters
  - same gene / different genes
- gene discontinuity
  - introns / exons
Genes On DNA (cont.)

How about the following?

other?
"promoters" signal initiation site

-35  -10  +1  Gene

Transcription In Prokaryotes
Transcription In Eukaryotes

- 3 types of promoters
- 3 types of polymerases (I, II, III)
- 3 types of genes
Transcript vs. Gene

Which of the following is true?

- prokaryotes: complementary palindromes
- eukaryotes: not exactly known
Modifications of Eukaryotic mRNA

- All eukaryotic mRNAs are 'capped'
  - methylated G at 5' end

- Most eukaryotic mRNAs are 'polyadenylated'
  - First cleavage downstream from *poly-A signal*, then attachment of poly-A by 'poly-A polymerase'

- Intron splicing

- RNA editing

- Not exactly known why/how
Ribosomes

- composed of ribosomal RNA and proteins
- different in prokaryotes/eukaryotes

**Prokaryotes**
- 50s: 5s, 23s, 34 polypeptides
- 30s: 21 polypeptides

**Eukaryotes**
- 60s: 5s, 5.8s, 28s, 49 polypeptides
- 40s: 33 polypeptides

- the need for multi-genes / transcription
Translation In Prokaryotes

DNA → mRNA → RNA polymerase → Protein
Translation In Eukaryotes

NUCLEUS

RNA polymerase

DNA

mRNA

CYTOPLASM

Protein
Gene Finding

- generally, a difficult problem

- even more difficult for eukaryotes
  - split genes (Phil Sharp / Richard Roberts 1977)
  - limited knowledge of reality

- varying claims
Gene Finding: The Schools

▶ statistics based
  ◦ use consensus accumulated from knowledge base to draw conclusions
    • Genmark, Glimmer, ...

▶ similarity-based
  ◦ determine similarities (if any) with known coding sequences
    • Critica, BLASTX, ...

▶ dictionary-based (…in 3 lectures)
  ◦ use recurring elements from known sequences to decompose a candidate
Gene Finding: Statistics

The Basic Genmark Approach:

- k-th order Markov model
- characterized by $P(b/W)$ and $P_o(W)$
  - where $|W| = k$ and $b$ is one of {A, C, G, T}

\[
P(b/W) = \frac{n(Wb)}{(n(WA)+n(WC)+n(WG)+n(WT))}
\]
\[
P_o(W) = \frac{n(W)}{\sum_{all\ w\ of\ length\ k} n(w)}
\]

- distinguishes between 'coding' and 'non-coding'
- distinguishes 'positional phases'
- a total of 7 categories
- determines $P(\text{category } i / S)$
Gene Finding: Statistics (cont.)

\[ P(\text{category } i \mid S) = \frac{P(S \mid \text{category } i) \ast P(\text{category } i)}{\sum_{\text{all } i} P(S \mid \text{category } i) \ast P(\text{category } i)} \]

\[ \sum_{\text{all } i} P(S \mid \text{category } i) \ast P(\text{category } i) \]

- if for some \( i \):

\[ P(\text{category } i \mid S) > T \]

then \( S \) is declared an instance of category \( i \).
Gene Finding: Statistics - Concerns

- do not work well if no appropriate training set is available
- organism-dependent
- do not work well on genes that are the results of horizontal transfer
- cannot predict short (60-80 aa) genes
- overlapping genes are a potential difficulty
Consensus Or Not?

- MELON
- MANGO
- HONEY
- SWEET
- COOKY

__________

MONEY (??)

By Ed. Trifonov
Gene Finding: Similarity

The Basic Critica Approach:

- given a piece of DNA, run BLASTN to determine any hits
- look at each alignment of DNA fragments in turn
- for each of 6 reading frames do
  - a) compare the codons and score
  - b) compare the codon translations and score
  - c) compute total score
- identify those regions with score > threshold
- extend right until STOP codon or EOS
- extend left using alternative START codons
Gene Finding: Similarity - Concerns

- do not work well if examined sequence is a "pioneer"
- do not work well if existing similarity is low
- search-scheme dependent
- threshold-dependent
Considerations

- positives: A
- true positives: $A \cap B$
- negatives: $\Omega - A$

sensitivity: $|A \cap B| / |A|$

- false positives: $B \setminus A \cap B$
- false negatives: $A \setminus A \cap B$

specificity: $|A \cap B| / |B|$