An Engineers Guide to Biology and Biochemistry

- Chemical reactions
- Macromolecules
- Genomes
- Cells and their structure
- Genetic circuits
- Metabolic networks
- Protein networks
- Cell replication
- Cell differentiation
- Viruses
- Bioinformatics and systems biology

Chemical Reactions

- All material is created or destroyed via chemical reactions.
- Chemical reactions combine atoms to form molecules and combine simpler molecules to form more complex ones.
- They also work in reverse.
- Atoms are the basic building block for all matter.
- About 98 percent of any living organism consists of: hydrogen (H), carbon (C), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S).
- Atoms form molecules via covalent, ionic, and hydrogen bonds.

Chemical Reaction Example

\[ 2H_2 + O_2 \xrightarrow{k} 2H_2O \]

- \( H_2 \), \( O_2 \), and \( H_2O \) are chemical (or molecular) species.
- Subscripts indicate \( H \) and \( O \) are present in dimer form.
- The molecules \( H_2 \) and \( O_2 \) are known as the reactants.
- The water molecule is known as the product.
- The 2's indicate 2 \( H_2 \) molecules are used to produce 2 water molecules.
- These numbers are known as the stoichiometry of the reaction.
- Since matter is conserved, atom counts on each side must equal.
- Some reactions in this course may not have this property.

Rate Constants

- The \( k \) above the arrow is known as the rate constant.
- It indicates the probability or speed of this reaction.
- Used in many of the modeling techniques in this course.
- Often difficult to determine for bio-chemical reactions.

Law of Mass Action

- Rate of a chemical reaction is governed by rate constant and concentrations of reactants raised to power of stoichiometry.
- This is known as the law of mass action.
- The rate of water formation is:
  \[ \frac{d[H_2O]}{dt} = 2k[H_2]^2[O_2] \]
  where \([H_2O]\), \([H_2]\), and \([O_2]\) represent the concentration of water, hydrogen dimers, and oxygen dimers.
- 2 in front of \( k \) is due to this reaction producing two water molecules.
Laws of Thermodynamics

- Chemical reactions must obey the laws of thermodynamics.
- First law is that energy can be neither created nor destroyed.
- Second law is entropy (disorder in the universe) must increase.
- These two laws can be combined into a single equation:

\[ \Delta H = \Delta G + T \Delta S \]

where \( \Delta H \) is change in bond energy, \( \Delta G \) is change in free energy, \( T \) is the absolute temperature, and \( \Delta S \) is change in entropy.

Gibb’s Free Energy

- \( \Delta G \) is also known as the Gibb’s free energy after J. Willard Gibbs who introduced this concept in 1878.
- Consider a reversible reaction of the form:

\[ A + B \rightleftharpoons C + D \]

where \( K_{eq} = k_{f}/k_{r} \) is the equilibrium constant.
- The Gibb’s free energy for the forward reaction is:

\[ \Delta G = \Delta G^\circ + RT \ln \left( \frac{[C][D]}{[A][B]} \right) \]

where \( R = 1.987 \text{ cal/mol} \text{ K} \) is the gas constant and \( T \) is the temperature.
- When negative, forward reaction can occur spontaneously.
- When positive, reverse reaction can occur spontaneously.
- When zero, the reaction is in a steady state.

Hydrolysis of ATP

- How do chemical reactions with positive free energy occur?
- Free energies of chemical reactions are additive.
- Coupling with other reactions allows them to occur.
- Hydrolysis of ATP releases energy:

\[ \text{ATP} + \text{H}_2\text{O} \rightarrow \text{HPO}_4^{2-} + \text{ADP} \]

- These types of ATP reactions occur in all living organisms.
- ATP is the universal energy currency of living organisms.

Enzymes

- Activation energy barrier must be overcome.
- An enzyme, or catalyst, can accelerate a reaction without being consumed by the reaction.
- Modifier is a species that is not consumed by a reaction.
- Often enzyme amount much smaller than other reactants.
- Enzymes do not effect free energy of the reaction, but only help the reaction overcome its activation energy barrier.

Michaelis-Menten Equation

- A basic enzymatic reaction:

\[ E + S \rightleftharpoons E_S \rightleftharpoons E_P \]

where enzyme, \( E \), used to change substrate, \( S \), into product, \( P \).
- When \( |S| >> |E| \), rate is often approximated using the Michaelis-Menten equation:

\[ \frac{d[P]}{dt} = V_{max} \frac{|S|}{|S| + K_M} \]

where \( V_{max} = k_3|E| \) and \( K_M = (k_2 + k_3)/k_3 \).
Michaelis-Menten Derivation

Derived using steady state assumption that \([ES]\) reaches final concentration quickly (i.e., \(d[ES]/dt \approx 0\)).

Using the law of mass action:
\[
\frac{d[P]}{dt} = k_3[ES]
\]
\[
\frac{d[ES]}{dt} = k_1[E][S] - k_2[ES] - k_3[ES]
\]
Using total concentration of enzyme, \([E_t]\):
\[
[E] = [E_t] - [ES]
\]

Michaelis-Menten Derivation (cont)

Using \(d[ES]/dt = 0\) and above equations:
\[
[ES] = \left[\frac{k_1[E][S]}{k_2 + k_3}\right] \frac{[S]}{[S] + K_M}
\]

After substitution:
\[
\frac{d[P]}{dt} = \frac{k_3[E][S]}{K_M + [S]} = V_{max} \frac{[S]}{[S] + K_M}
\]

This is the quasi-steady state assumption.
If \(k_3 \ll k_2\), can be simplified even further:
\[
\frac{d[P]}{dt} = \frac{k_3[E][S]}{K_M + [S]}
\]

This is the rapid equilibrium assumption.

Macromolecules

Nearly 70 percent of all living organisms are made up of water.
Remainder largely macromolecules of 1000s of atoms.
There are four types:
- Carbohydrates
- Lipids
- Nucleic acids
- Proteins

Carbohydrates

Made up of carbon and water \((C_n(H_2O)_m)\) where \(m \approx n\).
Often called sugars.
An example is glucose.
Important source of chemical energy.
Powers nearly all processes of a cell.
Also part of the backbone for DNA and RNA.

Lipids

Made up mostly of carbon and hydrogen atoms.
Primary use is to form membranes.
Membranes separate cells from one another and create compartments within cells as well as having other functions.
Examples include fats, oils, and waxes.

Nucleic Acids

(Courtesy: National Human Genome Research Institute)
Deoxyribonucleic acid (DNA)

- Stores information within living organisms.
- Sequence of nucleotides encode the instructions to construct proteins.
- Most organisms use DNA, but a few viruses use RNA.
- A strand of DNA is always synthesized in the 5' to 3' direction.
- DNA is a double-stranded with each strand running in opposite directions.
- A-T and G-C base pairs are complementary.
- Chemical makeup of this base pairing creates a force that twists the DNA into its coiled double helix structure.
- DNA is readily copied since one strand of DNA can act as a template to direct the synthesis of a complementary strand.

Ribonucleic acid (RNA)

- A single-stranded chain of nucleotides with the same 5' to 3' direction.
- Uses a different sugar and uracil replaces the thymine nucleotide.
- All genes that code for proteins are first made into an RNA strand called a messenger RNA (mRNA).
- mRNA carries the information encoded in DNA to the protein assembly machinery, or ribosome.
- The ribosome complex uses mRNA as a template to synthesize the exact protein coded for by the gene.
- DNA also codes for ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs).

Proteins

- Basic building blocks of nearly all the machinery of a cell.
- Each cell contains thousands of different proteins.
- Long chains with as many as 20 kinds of amino acids.
- Genetic code carried by DNA specifies order and number of amino acids and, therefore, shape and function of the protein.
- Code from DNA is transferred to RNA through transcription.
- mRNA is translated by a ribosome into protein.
- mRNA decoded in blocks of three bases, or codons.
- Protein built one amino acid at a time, with order determined by the order of the codons in the mRNA.

The Genetic Code

<table>
<thead>
<tr>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UCU Serine</td>
<td>UAU Tyrosine</td>
<td>UGU Cysteine</td>
</tr>
<tr>
<td>U</td>
<td>UCC Serine</td>
<td>UAC Serine</td>
<td>UGA Stop</td>
</tr>
<tr>
<td>U</td>
<td>UCA Serine</td>
<td>UAG Stop</td>
<td>UGG Tryptophan</td>
</tr>
<tr>
<td>C</td>
<td>CCU Leucine</td>
<td>CUA Leucine</td>
<td>CUG Leucine</td>
</tr>
<tr>
<td>C</td>
<td>CCC Proline</td>
<td>CCA Proline</td>
<td>CGA Arginine</td>
</tr>
<tr>
<td>C</td>
<td>CCA Proline</td>
<td>CCG Proline</td>
<td>CAG Arginine</td>
</tr>
<tr>
<td>A</td>
<td>AUC Leucine</td>
<td>ACC Leucine</td>
<td>AUA Isoleucine</td>
</tr>
<tr>
<td>A</td>
<td>ACC Threonine</td>
<td>ACA Threonine</td>
<td>AUC Asparagine</td>
</tr>
<tr>
<td>A</td>
<td>ACA Threonine</td>
<td>AAC Asparagine</td>
<td>AGA Arginine</td>
</tr>
<tr>
<td>L</td>
<td>GUU Valine</td>
<td>GCU Alanine</td>
<td>AGU Asparagine</td>
</tr>
<tr>
<td>G</td>
<td>GCC Valine</td>
<td>GCA Alanine</td>
<td>AGC Asparagine</td>
</tr>
<tr>
<td>G</td>
<td>GCA Alanine</td>
<td>GGU Glutamine</td>
<td>AGG Arginine</td>
</tr>
<tr>
<td>G</td>
<td>GGU Glutamine</td>
<td>GGC Glycine</td>
<td>AGU Asparagine</td>
</tr>
</tbody>
</table>

Proteins (cont)

- In 1961, Nirenberg and Matthaei correlated the first codon (UUU) with the amino acid phenylalanine.
- A given amino acid can have more than one codon.
- These redundant codons usually differ at the third position.
- Serine is encoded by UCU, UCC, UCA, and/or UCG.
- Redundancy is key to accommodating mutations that occur naturally as DNA is replicated and new cells are produced.
- Some codons do not code for an amino acid at all but instruct the ribosome when to stop adding new amino acids.

Protein Structure

- Protein folds into a specific 3-dimensional configuration.
- Shape and position of the amino acids in this folded state determines the function of the protein.
- Understanding and predicting protein folding is an important area of research.
- The structure of a protein is described in four levels.
  - Primary structure - sequence of amino acids.
  - Secondary structure - patterns formed by amino acids that are close (e.g., α-helices and β-pleated sheets).
  - Tertiary structure - arrangement of far apart amino acids.
  - Quaternary structure - arrangement of proteins that are composed of multiple amino acid chains.
What is a Genome?

- All of the 30 million types of organisms use the same basic materials and mechanisms to produce building blocks necessary for life.
- Information encoded in the DNA within its genome is used to produce RNA which produces proteins.
- A genome is divided into genes where each gene encodes the information necessary for constructing a protein.
- Some also control the production of proteins by other genes.

What are Genes?

- In 1909, Danish botanist Wilhelm Johanssen coined the word gene for the hereditary unit found on a chromosome.
- 50 years earlier, Gregor Mendel characterized hereditary units as factors—differences passed from parent to offspring.
- Mendel was an Austrian monk who experimented with his pea plants in the monastery gardens.
- Normally they self-fertilize, but he manipulated their parentage and thus their traits using a pair of clippers.
- Discovery went largely ignored for nearly 50 years until three researchers essentially duplicated his results.

Where are Genes?

- Until 1953, it was not known for certain that genes are made of DNA.
- In 1953, Watson and Crick, with support from x-ray data from Franklin and Wilkens, discovered the double helix structure of DNA.
- This discovery showed that DNA is composed of two strands composed of complementary bases.
- This base pairing idea shed light on how DNA could encode genetic information and be readily duplicated during cell division.
- Between 1953 and 1965, work by Crick and others showed how the DNA codes for amino acids and thus proteins.

How Many Genes Do Humans Have?

- In February 2001, two largely independent draft versions of the human genome were published in Nature.
- Both estimated between 30,000 to 40,000 genes in the human genome (today’s estimate is between 20,000 and 25,000).
- How do scientists estimate the number of genes in a genome?
  - Open reading frames, a 100 bases without a stop codon;
  - Start codons such as ATG;
  - Specific sequences found at splice junctions; and
  - Gene regulatory sequences.
- When complete mRNA sequences known, software can align start and end sequences with the DNA sequence.

What is Contained in Our Genome?

- Sequences that code for proteins are called structural genes.
- Regulatory sequences are start/end of genes, sites for initiating replication/recombination, or sites to turn genes on/off.
- Over 98% of our genome has unknown function (“junk” DNA).
- “Repetitive DNA”, short sequences repeated 100s of times, make up 40 to 45 percent of our genome.
- Although have no role in the coding of proteins, they are an excellent “marker” by which individuals can be identified.
- “Pseudogenes” are believed to be a remnant of a real gene that has suffered mutations and is no longer functional.
- Believed to carry a record of our evolutionary history.
Introns and Exons

- Genes make up about 1% of the total DNA in our genome.
- A eukaryotic gene is not found in a continuous stretch.
- The coding portions of a gene, called exons, are interrupted by intervening sequences, called introns.
- Both exons and introns are transcribed into mRNA, but before being transported to the ribosome, the mRNA transcript is edited.
- Removes introns, joins exons together, and adds unique features to end of transcript to make a “mature” mRNA.
- It is still unclear what all the functions of introns are, but may serve as the site for recombination.

One Gene–One Protein?

- About 40 percent of the expressed genome is alternatively spliced to produce multiple proteins from a single gene.
- This process may have evolved to limit effects of mutations.
- Genetic mutations occur randomly, and the effect of a small number of mutations on a single gene may be minimal.
- However, an individual having many genes each with small changes could weaken the individual, and thus the species.
- If single mutation affects several alternate transcripts, it is likely that the individual will not survive.

What is a Cell?

- The structural and functional unit of all living organisms.
- Some organisms, such as bacteria, are unicellular.
- Other organisms, such as humans, are multicellular.
- Humans have an estimated 100,000,000,000,000 cells!
- Each cell can take in nutrients, convert these into energy, carry out specialized functions, and reproduce as necessary.
- Each cell stores its own set of instructions in its genome for carrying out each of these activities.

Prokaryotic Organisms

- Life arose on earth about 3.5 billion years ago.
- The first types of cells were prokaryotic cells.
- They are unicellular organisms that lack a nuclear membrane.
- They do not develop or differentiate into multicellular forms.
- Bacteria are the best known and most studied form.
- Some bacteria grow in masses, but each cell is independent.
- They are capable of inhabiting almost every place on the earth.
- They lack intracellular organelles and structures.
- Most functions of organelles are taken over by the plasma membrane.
Eukaryotic Organisms

- Eukaryotes appear in the fossil record about 1.5 billion years ago.
- They include fungi, mammals, birds, fish, invertebrates, mushrooms, plants, and complex single-celled organisms.
- Eukaryotic cells are about 10 times the size of a prokaryote and can be as much as 1000 times greater in volume.
- Use same genetic code and metabolic processes as prokaryotes, but higher level organizational complexity permits multicellular organisms.
- Have membrane-bounded compartments called organelles.
- Most important is the nucleus that houses the cell's DNA.

Eukaryotic Features

(Courtesy: National Human Genome Research Institute)

The Plasma Membrane (A Cell’s Protective Coat)

- Outer lining of a eukaryotic cell.
- Separates and protects a cell from its environment.
- Made mostly of lipids, proteins, and carbohydrates.
- Embedded within are a variety of molecules that act as channels and pumps, moving molecules into and out of the cell.
- In prokaryotes, usually referred to as the cell membrane.

The Cytoskeleton (A Cell’s Scaffold)

- Acts to organize and maintain the cell’s shape.
- Anchors organelles in place.
- Helps during endocytosis, the uptake of external materials.
- Moves parts of the cell in processes of growth and motility.
- Involves many proteins each controlling a cell’s structure by directing, bundling, and aligning filaments.

The Cytoplasm (A Cell’s Inner Space)

- The large fluid-filled space inside the cell.
- In prokaryotes, this space is relatively free of compartments.
- In eukaryotes, “soup” within which organelles reside.
- Home of the cytoskeleton.
- Contains dissolved nutrients, helps break down waste products, and moves material around the cell.
- Contains many salts and is an excellent conductor of electricity, creating the perfect environment for the mechanics of the cell.
- The function of the cytoplasm, and the organelles which reside in it, are critical for a cell’s survival.

Organelles

- They are a set of “little organs” that are adapted and/or specialized for carrying out one or more vital functions.
- Organelles are found only in eukaryotes and are always surrounded by a protective membrane.
The Nucleus: A Cell’s Center

- A spheroid membrane-bound region that contains genetic information in long strands of DNA called chromosomes.
- Most conspicuous organelle found in a eukaryotic cell.
- Separated from the cytoplasm by nuclear envelope which isolates and protects DNA from molecules that could damage its structure or interfere with its processing.
- Where almost all DNA replication and RNA synthesis occurs.
- During processing, DNA is transcribed into mRNA.
- mRNA is transported out of the nucleus, where it is translated into a specific protein molecule.
- In prokaryotes, DNA processing takes place in the cytoplasm.

The Ribosome: The Protein Production Machine

- Found in both prokaryotes and eukaryotes.
- It is a large complex composed of RNAs and proteins.
- They process genetic instructions carried by mRNA.
- Translation is the process of converting a mRNA’s genetic code into the exact sequence of amino acids that make up a protein.
- Protein synthesis is extremely important, so there are a large number of ribosomes (100s or 1000s) in a cell.
- They float freely in the cytoplasm or sometimes bind to another organelle called the endoplasmic reticulum.

The Endoplasmic Reticulum and the Golgi Apparatus: Macromolecule Managers

- The endoplasmic reticulum (ER) is the transport network for molecules targeted for modifications and specific destinations.
- The ER has two forms: the rough ER and the smooth ER.
- The rough ER has ribosomes adhering to its outer surface.
- Translation of the mRNA for proteins that either stay in the ER or are exported out of the cell occurs at these ribosomes.
- The smooth ER serves as the recipient for those proteins synthesized in the rough ER.
- Proteins to be exported are passed to the Golgi apparatus for further processing, packaging, and transport to other locations.

Mitochondria and Chloroplasts

- Mitochondria are self-replicating organelles that occur in various numbers, shapes, and sizes.
- Have two functionally distinct membrane systems:
  - The outer membrane, which surrounds the organelle; and
  - The inner membrane, which has folds called cristae that project inwards to increase its surface area.
- Plays a critical role in generating energy in the eukaryotic cell.
- Chloroplasts are similar but are found only in plants.
- Both surrounded by double membrane and involved in energy metabolism.
- Chloroplasts convert sun’s light energy into ATP using photosynthesis.

Organelle Genome

- Plants and animals have genome within mitochondria and chloroplasts.
- Mitochondrial DNA is only inherited from our mother.
- Independent aerobic function may have evolved from bacteria living inside other organisms in a symbiotic relationship.
- These organisms evolved to become incorporated into the cell.
- Many diseases caused by mutations in mitochondrial DNA.
- Mitochondrial Theory of Aging suggests accumulation of mutations in mitochondria drives aging process.

Lyosomes and Peroxisomes

- Often referred to as the garbage disposal system of a cell.
- Spherical, bound by membrane, and rich in digestive enzymes.
- Peroxisomes often resemble a lysosome, but are self-replicating, whereas lysosomes are formed in the Golgi complex.
- Lyososomes contain three dozen enzymes for degrading proteins, nucleic acids, and certain sugars called polysaccharides.
- These enzymes work best at low pH, reducing risk that they will digest their own cell if they escape from the lysosome.
- The cell could not house such destructive enzymes if they were not contained in a membrane-bound system.
- Lyosome can digest foreign bacteria that invade a cell.
They also recycle receptor proteins and other membrane components and degrade worn-out organelles.

They can even help repair damage to the plasma membrane by serving as a membrane patch, sealing the wound.

Peroxisomes rid the body of toxic substances, such as hydrogen peroxide, and contain enzymes for oxygen utilization.

High numbers of peroxisomes can be found in the liver.

All enzymes in them are imported from cytosol and have a special sequence, called a PTS or peroxisomal targeting signal.

They also have membrane proteins for importing proteins into their interiors and to replicate.