Syllabus for sessions 122, 123, 124, 212, 213, 222

Session 122: Building blocks of proteins and nucleic acids

SLO 1. Describe the key properties of amino acids (charge, polarity/hydrophobicity, aromatic character, reactivity) and the structure of the peptide bond.

SLO 2. Explain the “Central Dogma” of molecular biology.

SLO 3. Understand the chemical structure of nucleotides: sugar, phosphate, and base; explain the polarity of DNA and RNA chains.

SLO 4. Understand the structure of the DNA double helix and RNA secondary structure. Explain the forces that stabilize the double helix, in particular the role of water.

SLO 5. Explain how DNA and RNA polymers are synthesized. Explain the pivotal role of phosphotransfer reactions in DNA and RNA biology.

Session 123: Information pathways

SLO 1. Explain how cells overcome three major challenges to replicate their genomes.

SLO 2. Explain why different genes need to be expressed at different rates, in different cells, at different times.

SLO 3. Understand the functions of the three mammalian RNA polymerases.

SLO 4. Explain the structure of a mammalian gene, including regulatory and coding elements. Explain the general role of transcription factors.

SLO 5. Understand how RNA polymerases are directed to gene promoters. Compare general transcription factors.

SLO 6. Explain how transcription factors control which proteins are made in different cell types.

Session 124: RNA and protein synthesis

SLO 1. Outline the processing reactions that precede export of a mRNA transcript from nucleus to cytoplasm.

SLO 2. Outline the major steps of protein synthesis. Understand the roles of mRNA, tRNA, tRNA synthetases, and the ribosome.

SLO 3. Describe the basic ways in which micro RNA molecules control gene expression.

Session 212: Protein Structure and Function

SLO 1. Know the elements of protein secondary, tertiary, and quaternary structure.

SLO 2. Explain the roles of hydrophilic vs. hydrophobic aminoacyl residues in protein folding.

SLO 3. Explain the importance of correct protein folding, chaperone proteins, and how misfolding can lead to pathology.

SLO 4. Understand common post-translational modifications of proteins (phosphorylation; disulfide bond formation; glycosylation) and know why specific modifications occur predominantly on proteins within the cytoplasm or in extra-cytoplasmic environments.

SLO 5. Know that different proteins are targeted to specific locations inside and outside of cells.
Session 213: Hemoglobin; prosthetic groups; enzymes

SLO 1. Explain the diversity of protein-protein and protein-ligand interactions.

SLO 2. Understand how the quaternary structure of hemoglobin and explain the function of the heme prosthetic group.

SLO 3. For a general ligand-receptor pair, be able to explain and calculate the relationship between $k_{on}$, $k_{off}$, and $K_0$.

SLO 4. Explain how allosteric cooperativity in hemoglobin enhances $O_2$ delivery to peripheral tissues, especially during exercise.

SLO 4. Describe the mechanistic bases of hemoglobinopathies.

SLO 5. Explain the key properties of enzymes. Explain why many enzymes contain bound prosthetic groups. Describe enzymatic co-factors and consider how deficiency of these factors could lead to disease.

SLO 6. Explain and calculate the relationships between $k_{cat}$, $K_m$, and $V_{max}$.

Session 222: Secretory and endocytic pathways;

SLO 1. Describe the major organelles of the secretory pathway.

SLO 2. Understand the difference between targeting of integral membrane proteins and secretory proteins.

SLO 3. Outline how proteins are folded and modified in the endoplasmic reticulum and Golgi organelles.

SLO 4. Explain how proteins and lipids are packaged into carrier vesicles, and how carrier vesicles fuse with target membranes including the plasma membrane.

SLO 5. Explain key mechanisms that underlie endocytosis, receptor recycling, and traffic to the lysosome.

SLO 6. Outline the synaptic vesicle cycle, and explain how clostridial neurotoxins selectively block neurotransmission.
Session 122: Building blocks of proteins and nucleic acids

SLO1. Describe the key properties of amino acid side chains (charge, polarity/hydrophobicity, aromatic character, and reactivity) and the structure of the amino acid backbone and peptide bond.

Proteins

Proteins are the major components (by mass) of the body. Proteins have a multitude of functions: they provide structure (tensile strength; elasticity); they do mechanical work (moving chromosomes; contracting muscle); they sense the internal and external environment; they process and transmit signals in response to this information; and they carry out most of the enzymatic functions required for metabolism, for DNA replication and repair, and for gene expression.

To understand the functions of proteins, we must think about their synthesis and structure.

Amino Acids & Polypeptides

A protein is made from one or more linear polypeptide chains — strings of covalently linked amino acids. The general structure of an amino acid is shown in Fig. 1. There are twenty major amino acids in eukaryotes, including humans. They differ by the side chain [R]. (Proline is cyclized; technically, it’s an imino acid).

For the purposes of this course you do not need to memorize the structures of the amino acids. However, we will need to consider their chemical and physical properties.

1. Some amino acids we can make ourselves from other chemical precursors. Others must be obtained through dietary intake. We’ll consider amino acid metabolism in some detail later in the course.
2. **Amino acid side chains** are chemically diverse. Consequently, a polypeptide’s properties result from its linear sequence of amino acid residues.
   a. Amino acids can be hydrophilic (polar or charged), or hydrophobic (apolar; “greasy”).
   b. In a folded protein, hydrophilic side chains are usually exposed to the aqueous solvent, while hydrophobic side chains tend to be buried so that they are shielded from the aqueous solvent.
   c. Amino acid side chains have diverse chemical reactivities. For example, serine, threonine, and tyrosine all have terminal hydroxyl groups that can form ester bonds. Arginine and lysine contain positively charged amine groups. Cysteine contains a redox-active sulfhydryl group.
3. In a polypeptide, amino acids are linked in a linear chain, head-to-tail. Linkages between amino acids are called peptide bonds (Fig. 2).

4. A polypeptide backbone has a polarity:
   a. At one end, there is primary amine group. This is the amino- or N-terminus (Figs 1,2).
   b. At the other end of the backbone is a carboxylic acid group. This is the carboxy- or C-terminus (Figs 1,2).
   c. A peptide bond is formed by condensation of the amine on one residue with the carboxyl group on another. In the overall reaction a water molecule is lost.

5. Almost all polypeptides synthesized by cells are synthesized one amino acid residue at a time. New residues always added to the C-terminal end of the growing chain. For this reason, we write down amino acid sequences from N-to-C.

6. The N-to-C primary sequence of a polypeptide, along with any additional covalent modifications to the polypeptide, controls how the polypeptide folds into a three-dimensional structure. This is a key point: sequence controls structure, and hence function.

SLO 2. Explain the “Central Dogma” of molecular biology.

Sequence Information and the “Central Dogma” of Molecular Biology

The linear sequence of almost all polypeptides (there are a few exceptions) is stored, in encoded form, in the DNA of our genome. The flow of sequence information occurs whenever DNA, RNA, or protein polymers are synthesized:

\[
\text{DNA} \xrightarrow{\text{replication}} \text{DNA} \xrightarrow{\text{transcription}} \text{RNA} \xrightarrow{\text{translation}} \text{polypeptide}
\]
A critical point: each of these processes entails a series of chemical reactions. Each is catalyzed by specific enzymes and is controlled by the laws of statistical thermodynamics. Consequently, these information transfer processes are never error-free. They cannot be. Consequently, cells spend enormous energy, materials, and time to reduce and cope with errors in nucleic acid and protein synthesis. When these tactics fail, the consequences can be devastating. A significant portion of this course will focus on errors in biological information transfer.

On the flip side, sequence changes (through DNA replication errors and other mutational processes) are responsible for all the richness and splendor of human genetic diversity. Understanding this diversity is essential to understanding and treating human disease and will only become more important as we are deluged with human DNA sequence data.

Fig. 3 shows that the cost of DNA sequencing has dropped faster than the price of electronic integrated circuits dropped from the 1970s to the present (Moore’s Law). In 2016 it costs about $1000 to sequence a human genome.

At present, you won’t see much DNA sequence data used in most clinical settings. But by the time you complete your medical training, the cost will have fallen to only a fraction of the current $1000 per genome — comparable to many standard lab tests. Thus, it is essential that you should obtain a working understanding of human genetic variation and its consequences for health and disease.

Fig. 3. Cost of DNA sequencing and synthesis, per base. The cost scale (y-axis) is logarithmic. Source: Rob Carlson www.synthesis.cc
SLO 3. Understand the general structure of nucleotides: sugar, phosphate, and base; explain the polarity of DNA and RNA chains.

The protomer of a DNA or RNA polymer is the nucleotide (Fig. 5). A nucleotide contains a base, a 5-carbon (pentose) sugar, and one or more phosphate groups. If the sugar doesn’t have a phosphate group on it, the pentose-base unit is called a nucleoside.

In RNA the pentose is ribose. In DNA the pentose is 2’-deoxyribose — ribose lacking a hydroxyl at its 2’ position.

To each sugar is attached a base (Fig. 6). The base is invariably attached at the 1’ position of the pentose through a glycosidic bond. In DNA the bases are: adenine (A), thymine (T) guanine (G), and cytosine (C). In RNA, thymine (T) is replaced by uracil (U).

The nucleotides (base-sugar-phosphate) are called: adenosine (A), guanosine (G), thymidine (T), cytidine (C), and uracil (U). Often the names are written to indicate the phosphorylation state: Adenosine diphosphate (ADP), etc.
DNA and RNA chains are strings of linked nucleotides. Each chain (Fig. 4) consists of a **backbone** made out of alternating sugar and phosphate groups. Put a bit differently, the pentose sugars are linked by **phosphodiester bonds**.

As with proteins, DNA and RNA chains have **polarity**. This is defined by the orientation of the pentose sugar: one linking phosphate is attached at the 3’ position on the pentose, and one is attached at the 5’ position (see Fig. 5 and compare to the backbone in fig. 4).

**In biological polymerization reactions, nucleotides are always added at the 3’ end of an elongating chain.** That is, the chain is polymerized 5’-to-3’. This leads to a convention: **we write down DNA and RNA sequences 5’-to-3’** — unless we explicitly specify otherwise.
SLO 4. Understand the structure of the DNA double helix and RNA secondary structure. Explain the forces that stabilize the double helix, in particular the role of water.

Strands of DNA or RNA can hybridize (anneal) to form double-stranded structures like the familiar DNA double helix.

1. Specificity in hybridization is provided by base-pairing (Fig. 7). A pairs with T (or U), G with C. Accuracy in pairing is promoted by favorable, non-covalent hydrogen bonds and by shape complementarity.
2. The pentose-phosphate backbone is at the exterior of the double helix. This makes sense: the sugars are very polar and every phosphate group carries a negative charge. Both groups are hydrophilic — they favorably interact with water.
3. The two sugar-phosphate backbones run antiparallel, like the traffic on a two-way street: one strand runs 5’-3’, top-to-bottom, and the other runs 3’-5’, top-to-bottom.
4. The base-pairs stack like plates at the center of the helix, so close together that water is excluded. This also makes sense: the bases are flat and relatively hydrophobic. Their flat surfaces favorably interact (stack) with one another, and are shielded from aqueous solvent. This also protects the bases from certain kinds of chemical attacks.
5. To summarize: the stability of a double helix is controlled by base pairing and also by other forces: separation of the negatively-charged phosphates, favorable stacking interactions between the flat bases, and the resulting shielding of the hydrophobic bases from the aqueous solvent.
6. The double helix has a minor groove and a major groove (Fig. 8). The major groove is critical: it allows proteins to touch the bases and “read” the DNA sequence, as if by braille. Thus, regulatory proteins can identify and bind to specific short DNA sequences without pulling the two strands apart.

Fig. 7. Base pairing. Circled groups show complementary hydrogen bonds.

Fig. 8. DNA double helix.
7. RNA can hybridize with RNA, or with DNA. In RNA biology, hybridization between complementary sequences on a single strand is of special importance (Fig. 9). Hybridization allows the formation of hairpin structures with short regions of double helix. These secondary structure elements can combine to generate complex tertiary structures including tRNAs and ribosomes, which are critical in protein synthesis.

Fig. 9. Hybridization within and between RNA strands. Note that the backbones in hybridized regions are antiparallel.

SLO 5. Explain how DNA and RNA polymers are synthesized. Explain the pivotal role of phosphotransfer reactions in DNA and RNA biology, and in the cellular energy economy.

DNA replication and RNA transcription are chemical reactions that involve information transfer.

1. To provide a template for DNA replication or RNA transcription, the DNA double helix must be locally pulled apart. Separation of the strands is also called melting or denaturation.
2. A highly specialized DNA or RNA polymerase enzyme moves along a template strand. The polymerase elongates the nascent chain by testing the base-pairing of incoming nucleotides with the template.
3. If the base-pairing is correct, the enzyme triggers the chemistry: the incoming nucleotide is added to the 3’ end of the nascent chain.
4. To power polymerization, incoming nucleotides are **NTPs** (nucleotide triphosphates, for RNA) or **dNTPs** (deoxynucleotide triphosphates, for DNA):

\[ \ldots pNpN-3` + NTP \rightarrow \text{polymerase} \rightarrow \ldots pNpNpN-3` + PP_i \]

Each “p” in the above scheme is a phosphodiester bond. The product of the reaction is a nascent chain with one additional nucleotide residue. A molecule of inorganic pyrophosphate \((PP_i = P_2O_7^{4-})\) is evolved.

The polymerization reaction is potentially reversible. To make the reaction irreversible, the enzyme **inorganic pyrophosphatase** destroys the evolved pyrophosphate in a highly favorable, effectively irreversible, reaction:

\[ PP_i + H_2O \rightarrow \text{pyrophosphatase} \rightarrow 2 P_i + \text{heat} \]

In later sessions, we will see the destruction of pyrophosphate used to make additional metabolic reactions, such as protein and lipid synthesis, irreversible.

5. In most cases the polymerase enzyme remains tightly bound to the template and nascent strands, and the elongation cycle begins again. The ability of an enzyme to catalyze many polymerization cycles without falling off (dissociating) from a template is called **processivity**. Some DNA and RNA polymerases have processivities of a million bases or more.

The main polymerases we’ll think about in this course are the enzymes that catalyze DNA replication RNA transcription.

However, other DNA and RNA polymerase enzymes exist (Fig. 11, dashed lines): RNA polymerases that use RNA as a template, and DNA polymerases that use RNA as a template (“**reverse transcriptases**”). Our cells use these alternative polymerases for specialized housekeeping functions such as telomere maintenance. RNA viruses and retroviruses use these classes of enzymes in their infection and replication cycles, as you’ll see in the Invaders and Defenders course.

![Fig. 11. Flow of sequence information in biological systems.](image-url)
**Session 123: Information pathways**

**Common Themes of Information Transfer**

DNA → *replication* → DNA → *transcription* → RNA → *translation* → polypeptide

There are common themes in these reactions.

1. Biological polymerization reactions always have an intrinsic directionality (polarity).
   - In DNA replication and RNA transcription, the **template strand is always read 3´-to-5´**, and the **nascent chain is always synthesized 5´-to-3´**.
   - In protein synthesis (translation), the **mRNA template is always read 5´-to-3´**, and the **nascent polypeptide is always synthesized N-to-C**.

2. The polymerase must be accurately positioned at a start site on the template. In each case this process is called **initiation**, and in each case it entails several regulated steps.

3. The polymerase has an **elongation** cycle, which is what it sounds like. This is the major biosynthetic stage.

4. A signal on the template signals **termination** of polymerization. Termination involves disassembly of the elongation machinery and the release of templates and products.

This is a general framework. We shall not focus on every stage for each process, but rather on key stages that illustrate important concepts.

**SLO 1. Understand three concepts that explain how cells overcome challenges to replicate their genome.**

To replicate its genome, the cell must overcome several challenges:

- There are two strands to be replicated.
- The two strands run in opposite directions (they’re antiparallel).
- Replication must be accurate.
- Enormous amounts of DNA be replicated: 6 Gbp/cell (= 6x10⁹ base pairs per cell).
- One and only one copy of each of the 46 chromosomes must be segregated into each daughter cell.

How is this done? Here we provide a cursory outline of DNA replication. Later in the course we will look more closely at replication, mitosis, and meiosis, in the contexts of the cell division cycle, genetic inheritance, and cancer.
For now, there are three core concepts about DNA replication to remember:

1. **DNA replication is semiconservative** (Fig. 1). This means that when a cell divides, the DNA duplexes in each daughter cell contain one of the parent cell’s original DNA strands (which is used as a template for polymerization, and one newly synthesized, or nascent strand).

2. **DNA replication is discontinuous** (Fig. 2). This means that the nascent strand associated with one of the two strands is synthesized in short segments called Okazaki fragments that are then knitted together.

Discontinuous replication is necessary because the DNA polymerase can only add nucleotides to the 3’ end of a nascent chain, but the template strands are antiparallel (Fig. 1).

An important difference between DNA and RNA polymerases is that DNA polymerases can only extend pre-existing nascent chains, while RNA polymerases can begin new chains from a single nucleotide. Thus, DNA polymerases invariably require a short RNA primer. The primer is made by a special RNA polymerase called primase (Fig. 2).

Each replication fork has a leading strand which is synthesized continuously, and a lagging strand which is synthesized discontinuously as a series of Okazaki Fragments.

As each Okazaki fragment is completed, the preceding RNA primer is removed, and the fragments are linked, or ligated, by an enzyme called DNA ligase.
3. DNA replication is bidirectional (Fig. 3). This means that at the DNA replication origin — the site where polymerization is initiated — two replication forks diverge from the origin. The replication machinery at each fork synthesizes one leading strand and one lagging strand (which is assembled from Okazaki fragments).

![Fig. 3. Bidirectional replication of mammalian chromosomes.](image)

Because human chromosomes can be hundreds of millions of base pairs in length, replication is done in parallel starting at many replication origins on each chromosome.

Each chromosome begins as one piece of double-stranded DNA. Replicating the very ends of the chromosomes, the telomeres, presents special problems on the lagging strand. Telomere DNA is therefore maintained by a special enzyme called telomerase.

**SLO 2. Explain why different genes need to be transcribed at different rates, in different cells, at different times.**

Most of the mammalian genome consists of non-coding DNA. A minority (~2%) of the human genome actually serves to encode mRNAs that are used as templates for protein synthesis. A similarly small fraction of the genome encodes other biologically important RNA species.

The fundamental question about gene expression is: each of us has many different cell types, but only one genome. The different cell types are different because they make different proteins: muscle cells make contractile proteins; nerve cells have the enzymes needed to manufacture neurotransmitters; osteoblasts have the protein machinery needed to manufacture bone, and so on.
Moreover, even a single cell type needs to make different proteins at different times: we synthesize and secrete insulin (a peptide hormone) when we eat. We remodel entire tissues and organs throughout growth, in response to injury, and during pregnancy.

How does this happen?

\[\text{DNA} \rightarrow \text{transcription} \rightarrow \text{RNA} \rightarrow \text{translation} \rightarrow \text{polypeptide}\]

The level of any given protein is a function of competing processes: synthesis and destruction.

Protein synthesis requires an mRNA template, and the abundance of the mRNA encoding any given protein is also regulated by a balance of synthesis and destruction.

**SLO 3. Understand the functions of the three mammalian RNA polymerases.**

**RNA Transcription**

The cell makes RNA molecules for various reasons:

- **mRNA** is the template for protein synthesis (translation).
- **tRNA** and rRNA are core parts of the protein synthesis machinery.
- Diverse RNA molecules are involved in regulating gene expression and other processes. Examples include micro RNA (miRNA) and long non-coding (lncRNA). Many other examples are emerging.

The various RNAs are made by three RNA polymerase (RNAP) enzymes:

1. **RNAP I** makes most of the ribosomal rRNA — the most important part of the ribosome, the enzyme that synthesizes polypeptides.
2. **RNAP II** makes all of the messenger mRNA — the templates for polypeptide synthesis.
3. **RNAP III** makes transfer tRNA — the carrier of activated amino acids for polypeptide synthesis.

We will focus on transcription by RNAPII, because its activity controls the levels of mRNA templates for protein synthesis. The underlying principles by which RNAP I and III operate are similar.
SLO 4. Explain the structure of a mammalian gene, including regulatory and coding elements. Explain the general role of transcription factors.

A gene contains two kinds of sequences:

1. The **transcription unit** is the DNA sequence used as a **template** to synthesize RNA.
2. **Regulatory sequences** tell RNAP where to initiate and terminate transcription. They allow cells to control which genes are actively transcribed (“expressed”), and which are silent. These sequences can be further sub-divided:
   a. The **promoter** sequence directs RNAP II and associated general transcription factors to the transcriptional **start site**.
   b. **Enhancer** sequences, usually 10-30 bp in length, bind **transcription factors**, or **activator proteins**, that instruct RNA polymerase to become active at the promoter. Each gene is controlled by different enhancer elements.
   c. Different cells contain specific sets of activators. This is the main basis for cell-type-specific gene regulation! Enhancers can sit right next to the promoter, or tens of thousands of base pairs distant. Enhancers are usually upstream of the promoter but they can also be embedded within the transcription unit or even downstream of it.
   d. The **terminator** tells RNAP that it has reached the end of the transcription unit.
   e. There are other regulatory elements in the genome as well. For example, **Silencer** sequences bind factors that, as the name suggests, curtail transcription. **Insulator elements** ensure that different genes are subject to independent regulation.

![Structure of a typical protein-coding gene](image)

**Fig. 2. Structure of a typical protein-coding gene.**

**Top:** Arrangement of a typical transcription unit and its regulatory DNA sequences (not drawn to scale).

**Bottom Left:** Activating transcription factor proteins bound to an enhancer sequence. Note that the transcription factors are “reading” the sequence by probing the **major groove** of the DNA double helix.

**Middle:** RNAP II (“Pol”) and the GTFs bound to a promoter sequence.

**Right:** Diagram of a transcription bubble with a nascent RNA chain. (In this case, the bubble would be moving from right-to-left. You should be able to identify the 5’ and 3’ ends of each DNA and RNA strand.)

**Sources:** *Nature* 392:42; Kornberg, 2006 Nobel Lecture; Calladine, *Understanding DNA.*
SLO 5. Understand how RNA polymerases are directed to gene promoters. Compare general transcription factors.

Sequence of Events in Transcription:

1. Genomic DNA is elaborately packaged. Transcription is regulated in part by how densely packaged a given gene is, and hence, how accessible its regulatory sequence elements are. Later, we will discuss how DNA packaging is controlled.

2. To initiate transcription of a gene, RNAP II must be directed to the promoter. This is done by the General Transcription Factors (Figs. 2 and 3). The GTFs recognize and bind to promoter sequences. They place RNAP II at the start site. GTFs then locally melt the DNA at the promoter, separating the two DNA strands to form a transcription bubble.

3. The GTFs are “general” transcription factors because they are always needed for initiation of transcription. However, the GTFs do not have the ability to regulate when RNAP actually initiates RNA synthesis.

4. In other words, the GTFs are necessary but not sufficient for initiation of transcription.

5. Transcription factors that bind to regulatory enhancer sequences are needed to activate transcription by RNAP II and the GTFs (Fig. 2).

6. Activating transcription factors “talk” to RNAP II and the GTFs by binding to coactivators that in turn touch RNAP II and the GTFs (Fig. 3). Together, these events cause the pre-initiation complex — RNAP II and the GTFs — to initiate RNA polymerization.

7. The chemistry of RNA polymerization is similar to the chemistry of DNA polymerization. The most important differences are:
   a. No primer is needed for RNA synthesis.
   b. NTPs are used for RNA synthesis, not 2-deoxy dNTPs.

8. As RNAP II elongates the nascent mRNA chain, it moves along the template strand of the transcription unit (Fig. 2). The replication bubble moves as RNAP II “crawls” along the DNA template strand. In other words the DNA double helix melts in front of RNAP II, and re-hybridizes (anneals) behind it.

9. When RNAP II reaches a terminator sequence (Fig. 2), the newly-synthesized RNA chain is released, RNAP is removed from the template strand, and the transcription bubble collapses.
SLO 6. Explain how transcription factors control which proteins are made in different cell types.

The reason we care so much about the mechanics of RNAP II transcription is that this process controls which mRNA transcripts are produced, and in what abundance. This in turn controls the specific repertoire of proteins that can be made by each cell.

Fig. 4 shows the regulatory sequences of genes that encode some key proteins made only in specific kinds of cells: skeletal muscle cells, heart muscle cells, and cells in the lens of the eye.

The colored shapes show different enhancer sequences. Each type of enhancer element is recognized and bound by specific activating transcription factors. Humans have about 2,000 different transcription factors.

By producing specific combinations of transcription factors, each cell specifies which subsets of genes are actively transcribed, and in what quantities.

Fig. 4. Regulatory enhancer elements of protein-coding genes that are selectively transcribed in three different cell types. The arrow at the right end of each diagram identifies the promoter and the first base (+1) of the transcription unit. Source: Davidson, The Regulatory Genome.
This concept is so important that it bears repeating:

The specific array of transcription factors present in a given cell shapes that cell's pattern of gene expression — and thus, that cell's overall protein complement, its identity (muscle, fibroblast, neuron, etc.), and thus the cell's functional characteristics.

Another important point is that transcription factors are themselves proteins. Consequently, the genes that encode transcription factors are themselves subject to transcriptional regulation. By transcribing and translating specific transcription factors the cell can execute temporary or stable programs of gene expression in response to developmental cues and other signals.

Fig. 5. Positive feedback loops maintain gene expression programs.

Here, an “initiator” transcription factor binds an enhancer on the gene encoding a “terminal selector” transcription factor. When this gene is transcribed and the resulting mRNA is translated, the resulting protein binds to another enhancer in its own gene, and ensures that the terminal selector gene continues to be transcribed. The terminal selector also stimulates transcription of other genes needed for specific functions.

Source: PNAS 110:7101
Session 124: RNA and protein synthesis

SLO 1. Outline the processing reactions that precede export of a mRNA transcript from nucleus to cytoplasm. Think about how errors in mRNA processing might result in pathology.

mRNA Processing and Export

DNA replication and RNA transcription both occur in the cell’s nucleus. However, proteins are synthesized in the cytoplasm. To serve as templates for protein synthesis, mRNA molecules must be exported from the nucleus to the cytoplasm. This occurs at a special portal in the nuclear membrane, the **nuclear pore** (Fig. 1). The nuclear pore is an immense molecular assemblage that precisely controls the passage of mRNA and other macromolecules into, and out of, the nucleus. This is another theme that we will encounter again and again: biosynthetic products made in one cellular organelle are shuttled to another location — as with stations on an assembly line.

Fig. 1. **Nuclear pores control the movement of macromolecules in and out of the nucleus.** On the left, an entire nucleus is shown, with pores spanning the nucleus’s double membrane. On the right, a partial structure of a single nuclear pore is depicted. The nuclear membrane is shown in gray. Protein components of the pore are shown in color, with each protein shown in a different hue.

Sources: Fawcett, *The Cell*; *Science*, DOI: 10.1126/science.aaf1015
In the nucleus, the initial “raw” RNA transcript made by RNAP II must be processed, in three ways (Fig. 2). Only then is the mature mRNA exported from the nucleus to the cytoplasm, where it will be used as a template for protein synthesis.

1. At the 5´ end of the mRNA (the “beginning” of the transcript), a special nucleotide, 7-methylguanosine, is attached through a covalent bond. This is called the 5´ cap.

2. The 5´ cap is a signal. In the cytoplasm, the cap will tell the protein synthesis machinery that the RNA bearing the cap is in fact a messenger mRNA — a template for protein synthesis — and not some other type of RNA.

3. At the 3´ end of the transcript, a string of adenosine (A) nucleotides is added. This is the poly-A tail of the mRNA. The poly-A tail is constructed by a special enzyme (poly-A polymerase) through a non-templated polymerization process.

4. The poly-A tail signals export of the mRNA from nucleus to cytoplasm. It also controls the stability (the half-life) of the mRNA once it’s in the cytoplasm.

5. The mRNA is spliced to remove introns and ligate (join) exons together. This step is somewhat involved, and extremely important, so we’ll examine it in a bit more detail.

![Diagram](https://via.placeholder.com/150)

**Fig. 2.** Processing of primary transcripts in the nucleus yields mature mRNA molecules. Mature mRNAs are then exported to the cytoplasm to serve as templates in translation. The poly-A tail is usually 150-250 nt long.

The RNA processing machinery (the capping and poly-A addition enzymes, and the splicing machinery) are grabbed and positioned onto the growing transcript by a flexible “tail” on the RNAP II enzyme itself. In other words, mRNA elongation by RNA polymerase II, and mRNA processing, are coupled reactions, all of which occur in the nucleus.
**Differential mRNA splicing**

The maturation of a large mRNA molecule may entail *dozens* of splicing reactions. In different cell types, these splicing reactions may be regulated so that not every exon ends up in each final, mature mRNA molecule. Consequently, a **single transcription unit may encode more than one mRNA variant**, with each derived from a different combination of exons (Fig. 3).

Differential splicing allows the ~20,000 protein-coding *genes* in the human genome to encode substantially more than 20,000 distinct mRNA templates and, thus, a much greater diversity of proteins.

![Diagram](image)

**Fig. 3.** Differential splicing of pre-mRNA’s transcribed from a single gene can yield several different mature mRNA transcripts, each encoding a different polypeptide variant. The exons are shown as boxes, and the introns that are spliced out, by connecting lines. Note that the splicing pattern is different in different cell types, and that one cell type may produce more than one mRNA splice variant. Introns in the pre-mRNA are shown in gray. UT indicates un-translated regions at the 5’ and 3’ ends of the transcript. The 5’ cap and 3’ poly-A tail are not depicted in this diagram, but would be present on each mature mRNA.

To summarize:

- Transcription initiation controls how many raw transcripts get made.
- Different cells have different activating transcription factors.
- Different genes have different enhancers that bind different transcription factors.
- Cap addition signals that an mRNA will be a template for protein synthesis.
- Differential splicing controls which exons are in the mature mRNA template, and thus the sequence of the resulting polypeptides.
- The poly-A tail (along with other features of the mRNA) controls nuclear export and the stability of the mRNA — how long it persists in the cytoplasm.
SLO 2. Outline the major steps of protein synthesis. Understand the roles of mRNA, tRNA, tRNA synthetases, and the ribosome.

Here we summarize how polypeptide chains are synthesized and how they fold into their correct three-dimensional configurations.

The notes for this section begin with two charts: first, the assignments of RNA codons to amino acids (the genetic code); second, the chemical and physical properties of amino acids.

You do NOT need to memorize these tables! You do need to be able to apply their content.

**Table 1**

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>ψ Phe</td>
<td>UAU</td>
</tr>
<tr>
<td>UUC</td>
<td>UCU</td>
<td>UAU</td>
</tr>
<tr>
<td>UUA</td>
<td>ψ Leu</td>
<td>UAA</td>
</tr>
<tr>
<td>UUG</td>
<td>UUC</td>
<td>UAG</td>
</tr>
<tr>
<td>CUU</td>
<td>CCC</td>
<td>CAU</td>
</tr>
<tr>
<td>CUC</td>
<td>CCA</td>
<td>CAC</td>
</tr>
<tr>
<td>CUA</td>
<td>CCG</td>
<td>CAG</td>
</tr>
<tr>
<td>CUG</td>
<td></td>
<td>pro</td>
</tr>
<tr>
<td>AUU</td>
<td>ψ Ile</td>
<td>AAU</td>
</tr>
<tr>
<td>AUC</td>
<td>ACC</td>
<td>AAC</td>
</tr>
<tr>
<td>AUA</td>
<td>ACA</td>
<td>AAA</td>
</tr>
<tr>
<td>AUG</td>
<td>ψ Met</td>
<td>AAG</td>
</tr>
<tr>
<td>GUU</td>
<td>GCU</td>
<td>GAU</td>
</tr>
<tr>
<td>GUC</td>
<td>GCC</td>
<td>GAC</td>
</tr>
<tr>
<td>GUA</td>
<td>GCA</td>
<td>GAA</td>
</tr>
<tr>
<td>GUG</td>
<td>GCG</td>
<td>GAG</td>
</tr>
</tbody>
</table>

ψ = relatively hydrophobic residues

U                  C               A                 G

ψ = relatively hydrophobic residues
Table 2

**Twenty-One Amino Acids**

<table>
<thead>
<tr>
<th>A. Amino Acids with Electrically Charged Side Chains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
</tr>
<tr>
<td>Arginine (Arg)</td>
</tr>
<tr>
<td>Histidine (His)</td>
</tr>
<tr>
<td>Lysine (Lys)</td>
</tr>
<tr>
<td>Aspartic Acid (Asp)</td>
</tr>
<tr>
<td>Glutamic Acid (Glu)</td>
</tr>
<tr>
<td><img src="image1.png" alt="Chemical structures" /></td>
</tr>
<tr>
<td>$pK_a$ values:</td>
</tr>
<tr>
<td>Arginine: $pK_a = 2.63$</td>
</tr>
<tr>
<td>Histidine: $pK_a = 1.70$</td>
</tr>
<tr>
<td>Lysine: $pK_a = 2.15$</td>
</tr>
<tr>
<td>Aspartic Acid: $pK_a = 1.95$</td>
</tr>
<tr>
<td>Glutamic Acid: $pK_a = 2.16$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Amino Acids with Polar Uncharged Side Chains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serine (Ser)</td>
</tr>
<tr>
<td>Threonine (Thr)</td>
</tr>
<tr>
<td>Asparagine (Asn)</td>
</tr>
<tr>
<td>Glutamine (Gln)</td>
</tr>
<tr>
<td>Cysteine (Cys)</td>
</tr>
<tr>
<td>Selenocysteine (Sec)</td>
</tr>
<tr>
<td>Glycine (Gly)</td>
</tr>
<tr>
<td>Proline (Pro)</td>
</tr>
<tr>
<td><img src="image2.png" alt="Chemical structures" /></td>
</tr>
<tr>
<td>$pK_a$ values:</td>
</tr>
<tr>
<td>Serine: $pK_a = 1.73$</td>
</tr>
<tr>
<td>Threonine: $pK_a = 2.00$</td>
</tr>
<tr>
<td>Asparagine: $pK_a = 2.16$</td>
</tr>
<tr>
<td>Glutamine: $pK_a = 2.16$</td>
</tr>
<tr>
<td>Cysteine: $pK_a = 1.91$</td>
</tr>
<tr>
<td>Selenocysteine: $pK_a = 2.34$</td>
</tr>
<tr>
<td>Glycine: $pK_a = 1.55$</td>
</tr>
<tr>
<td>Proline: $pK_a = 10.47$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Special Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cysteine (Cys)</td>
</tr>
<tr>
<td>Selenocysteine (Sec)</td>
</tr>
<tr>
<td>Glycine (Gly)</td>
</tr>
<tr>
<td>Proline (Pro)</td>
</tr>
<tr>
<td><img src="image3.png" alt="Chemical structures" /></td>
</tr>
<tr>
<td>$pK_a$ values:</td>
</tr>
<tr>
<td>Cysteine: $pK_a = 1.91$</td>
</tr>
<tr>
<td>Selenocysteine: $pK_a = 10.28$</td>
</tr>
<tr>
<td>Glycine: $pK_a = 10$</td>
</tr>
<tr>
<td>Proline: $pK_a = 10.47$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. Amino Acids with Hydrophobic Side Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine (Ala)</td>
</tr>
<tr>
<td>Valine (Val)</td>
</tr>
<tr>
<td>Isoleucine (Ile)</td>
</tr>
<tr>
<td>Leucine (Leu)</td>
</tr>
<tr>
<td>Methionine (Met)</td>
</tr>
<tr>
<td>Phenylalanine (Phe)</td>
</tr>
<tr>
<td>Tyrosine (Tyr)</td>
</tr>
<tr>
<td>Tryptophan (Trp)</td>
</tr>
<tr>
<td><img src="image4.png" alt="Chemical structures" /></td>
</tr>
<tr>
<td>$pK_a$ values:</td>
</tr>
<tr>
<td>Alanine: $pK_a = 2.33$</td>
</tr>
<tr>
<td>Valine: $pK_a = 2.27$</td>
</tr>
<tr>
<td>Isoleucine: $pK_a = 2.26$</td>
</tr>
<tr>
<td>Leucine: $pK_a = 2.18$</td>
</tr>
<tr>
<td>Methionine: $pK_a = 2.34$</td>
</tr>
<tr>
<td>Phenylalanine: $pK_a = 9.34$</td>
</tr>
<tr>
<td>Tyrosine: $pK_a = 9.34$</td>
</tr>
<tr>
<td>Tryptophan: $pK_a = 10.10$</td>
</tr>
</tbody>
</table>

Source: Wikimedia
SLO 1. Outline the major steps of protein synthesis. Understand the roles of mRNA, tRNA, tRNA synthetases, and the ribosome.

Protein Synthesis: the Genetic Code

Like nucleic acid synthesis, protein synthesis is a template-directed process. However, in protein synthesis, the process is a bit more complicated because the template does not directly interact with the polypeptide product. How this works has never been explained more plainly than by Francis Crick, in an astonishing proposal made in 1955:

...Each amino acid would combine chemically, at a special enzyme, with a small molecule which, having a specific hydrogen-bonding surface, could combine specifically with the nucleic acid template. This combination would also supply the energy necessary for polymerization... there would be 20 different kinds of adaptor molecule, one for each amino acid, and 20 different enzymes to join the amino acid to their adaptors...

We now know that the “adapter” is tRNA. Crick’s proposal was correct in every detail save one: there are 61 possible combinations of 3-base mRNA codons (see Table 1), so there are more than 20 tRNA “adapters”. This system provides the biochemical basis of the genetic code — the rules through which each 3-base codon in an mRNA template specifies one specific amino acid.
tRNA and aminoacyl tRNA synthetase enzymes

1. Each tRNA has 2 “business ends”:
   a. The **anticodon** pairs with the 3-base **codon** on the mRNA template. Look closely at the diagram (Fig. 5). Note that as in other nucleic acid hybrids, the two strands are **antiparallel**.
   b. The **aminoacyl acceptor site** is a terminal adenosine (A) nucleotide, where the carboxyl group of the amino acid is esterified to the 3'-OH of the adenosine. **Note that this is a relatively unstable, high-energy bond.** It will make polypeptide elongation thermodynamically downhill, and hence favorable — exactly as predicted by Crick.

2. Any given tRNA can be esterified to *any* amino acid. However, the **fidelity of translation depends absolutely on the accurate matching of a tRNA bearing a specific anticodon to the corresponding amino acid**.

3. The job of coupling each amino acid to its corresponding tRNAs done by 20 different enzymes: the **aminoacyl tRNA synthetases** (also called aaRS enzymes). This is an absolutely key point: **the specificity of the genetic code is controlled by the tRNA synthetases**.

4. The tRNA synthetases produce aminoacylated tRNA molecules in a two-step reaction (Fig. 6).
   a. First, an ATP is coupled to the amino acid to form an **amino adenylate**. That is, the amino acid is coupled to AMP. This is a high-energy **activated intermediate**.
   b. In the first step, pyrophosphate (PP_i) is also released. As we saw in DNA and RNA polymerization, pyrophosphate is immediately destroyed by pyrophosphatase, making the first sub-reaction an irreversible **committed step**.
   c. Second, the aminoacyl group is transferred to the tRNA. The products are an aminoacyl tRNA, and AMP. Because we start with ATP, and end up with AMP and two inorganic phosphates, coupling of an amino acid to tRNA has an **energetic cost of 2 ATP equivalents**.
The Ribosome

The “polypeptide polymerase” is the ribosome, an enormous ribonucleoprotein complex. The ribosome has two subunits. Each subunit contains both RNA and many different polypeptides.

1. The small subunit (it is not small, just not as big as the large subunit!) is the “decoding center.” The small subunit’s job is to match each codon on the template to a corresponding aminoacyl-tRNA. This is no easy task. There are 61 codons that specify 20 amino acids (look at Table 1), so a large majority of the aminoacyl-tRNA molecules that enter the ribosome must be rejected.
2. When a correct codon-anticodon interaction is detected by the small subunit, the large subunit catalyzes the peptidyltransfer reaction — the chemistry of polypeptide elongation.

Remarkably, although each ribosome subunit contains both peptides and rRNA, both the decoding center within the small subunit, and the peptidyltransfer center in the large subunit, are made of ribosomal rRNA. The enzymatic core of the ribosome is a “ribozyme”. This is probably a relic of the ancient origin of ribosomes at the dawn of life, in the so-called RNA world.

Polypeptide synthesis

As in nucleic acid polymerization, there are three phases of polypeptide synthesis.

1. **Initiation**: the polymerase — the ribosome — must be placed precisely over the start codon on the mRNA template.
2. **Elongation**: this is where template-mediated polymerization of the polypeptide occurs.
3. **Termination**: A stop codon is identified, triggering release of the polypeptide and removal of the ribosome from the mRNA template.

Initiation of protein synthesis

As with DNA replication and transcription, initiation of protein synthesis is pretty complicated. And since the frequency of initiation controls the rate of protein synthesis, this step is also highly regulated.

Regulation occurs at both a global level (the cell asks how much protein synthesis it can support overall, given the available energy and resources), and for specific mRNA transcripts, which support initiation with different efficiencies.
Steps in initiation (Fig. 7)

1. The ribosome small subunit — the decoding center — assembles with other initiation factors.
   a. Together, they bind to the 5’ cap of the mRNA.
   b. The small subunit and initiation factors move along the mRNA from 5’-to-3’, scanning the mRNA for a start codon.
   c. In most cases the start codon is AUG. If you look at Table 1, you’ll see that AUG encodes the amino acid methionine (Met, M). Thus, the first amino acid in a polypeptide is usually Met.

2. Once the decoding center is accurately placed over the AUG start codon, the large subunit (in Fig. 7, the larger oval) docks onto the small subunit and mRNA, and the initiation factors fall off (dissociate). Now the elongation cycle can begin.

Two additional points about translation initiation must be emphasized. First, as mentioned above, this step is highly regulated. Initiation factors are largely responsible for this regulation.

Second, the accuracy with which the ribosome’s small subunit is placed over the start codon is critical: a positional error of ± 1 or 2 nucleotides will put the mRNA transcript out-of-frame, and result in a totally different, and incorrect, polypeptide sequence.

Similarly, if a mutation in the genome inserts or deletes one or two DNA bases in the coding region of a gene, the resulting mRNAs will contain frameshift errors. When this happens, every codon following the frameshift will be incorrectly decoded during translation.
**Note:** There are two major, important differences in mRNA structure and translation initiation between eukaryotes (humans included) and bacteria.

1. **In bacteria, mRNA molecules do not have a 5´ cap or 3´ poly-A tail** (Fig. 8). In addition, bacterial mRNA molecules often have a series of translation start sites, and a series of coding regions. Such an mRNA is referred to as “polycistronic.”

   In eukaryotes, a mature mRNA template usually has only a single start site and encodes only a single polypeptide.

   ![Fig. 8. mRNA structure in prokaryotes (bacteria) versus eukaryotes (including humans).](image)

   **Source:** Alberts, Molecular Biology of the Cell

2. **In bacteria, the first amino acid** is usually a Met derivative, fMet (formylmethionine). Peptides beginning with fMet are recognized by our innate immune system as a danger signal, because they can indicate an active bacterial infection. You will learn more about bacteria, danger signals, and innate immunity in the Invaders & Defenders Block.
Polypeptide elongation

The assembled ribosome (large and small subunits) has three sites that can accommodate tRNA molecules: the A, P, and E sites. These names are shorthand for aminoacyl-tRNA, peptidyl-tRNA, and exit sites. It will become clearer in a moment why these names are used. The three sites are used in sequence. Here’s how it works (Fig. 9).

1. At the A site, the ribosome samples incoming aminoacyl-tRNA (aa-tRNA) molecules. The ribosome is looking for an aa-tRNA with an anticodon correctly pairs with the mRNA codon positioned under the A site. Dozens of incorrect aa-tRNAs are rejected for each correct match.

2. When a correct codon-anticodon match is identified, the ribosome initiates the peptidyltransfer reaction — the chemistry.
   a. The growing polypeptide chain, still esterified at its C (carboxyl)-terminus to a tRNA, resides in the P site.
   b. In the peptidyltransfer reaction, the nasent polypeptide is transferred from the peptidyl-tRNA sitting in the P site, to the aminoacyl residue sitting in the A site. This is counterintuitive. Study fig. 6 to see how it works.

3. Now the ribosome undergoes a ratchet-like twisting motion. This motion causes translocation of the mRNA, the peptidyl-tRNA, and the deacylated (discharged) tRNA. Now the peptidyl-tRNA is one residue longer, and in the P site. The deacylated tRNA is in the E site, where it is ejected from the ribosome. And the A site is unoccupied, ready to begin the cycle again for the next codon on the mRNA template.

A final point about energy. Two ATP equivalents used to charge each aa-tRNA. This powers the peptidyltransfer reaction. However, additional ATP equivalents are consumed during the tRNA selection and translocation portions of the elongation cycle. In terms of both energy and materials, protein synthesis is very expensive. Many cell types use most of their energy on protein synthesis.
SLO3. Describe the basic ways in which micro RNA molecules control gene expression.

1. **MicroRNAs (miRNAs)** are a class of post-transcriptional/translational regulators of gene expression. More than 2,000 miRNAs have been annotated in human genome. 60% of all human genes are estimated to be regulated by one or more miRNA. Therefore, miRNAs are crucial regulators of gene expression.

2. miRNAs are **short noncoding RNAs**, usually about 22 nucleotides long.

3. miRNA molecules are formed through **two major pathways**.
   a. Many miRNA precursors are coded as stand-alone genes, which can be transcribed by RNA polymerase II. Note that in the figure above, the miRNA is derived from a Pol II transcript with a 5´cap and 3´poly-A tail.
   b. As they are not coding sequences, miRNA precursors can also be derived from intron sequences that are embedded within other mRNA precursor transcripts. In these cases, splicing excises the intron and its miRNA from the coding exons of the mature mRNA.

4. miRNA does not function alone. miRNAs bind to the Argonaute family of proteins in the cytoplasm.

5. miRNA-Argonaute complexes bind to specific mRNA transcripts via complementary hybridization between the miRNA seed region, and target sequences in the 3'UTR of the targeted mRNA transcript.

6. In most cases, miRNAs function to **repress** (decrease) the production of specific sets of proteins.

7. miRNA-Argonaute-mRNA complex can repress protein expression in several ways:
   a. destabilization of the mRNA via shortening poly (A) tail;
   b. inhibition of translation initiation;
   c. cleavage and degradation of the target mRNA.

8. Note that all of these mechanisms are post-transcriptional, meaning that they operate on mature mRNA molecules in the cytoplasm. In contrast, that transcription factors act in the nucleus to control the rate at which different mRNA molecules are synthesized (transcribed) by RNA polymerase II.

9. The first miRNA was discovered in a genetic model nematode *C. elegans* in 1993. Studies in humans have revealed that mutations in miRNAs can cause or contribute to various human diseases. For instance, mutation in miRNA-96 is linked to hereditary progressive hearing loss, and deletion of the miR-17~92 cluster causes skeletal abnormality and growth defects. Dysregulation of miRNA function is also implicated as a causative factor in several cancers.
Session 212: Protein Structure and Function

SLO 1. Know the elements of protein secondary, tertiary, and quaternary structure.

*If you want to understand function, study structure.* — F. Crick

As a nascent polypeptide emerges from the ribosome, it must **fold** into a specific, functional, three-dimensional structure. **The functional “native” fold of a protein is determined by the linear sequence of amino acids in the polypeptide.**

We think about folding as a hierarchical process (Fig. 8, next page):

- **A polypeptide’s primary structure** its **linear sequence of amino acids.** This sequence, as you have just seen, is specified by the sequence of codons in an mRNA template.
- **Secondary structure** elements are “folding motifs” that form through local interactions between residues within the polypeptide chain (H-bonding, salt bridges, van der Waals interactions, etc.). The most common and important secondary structure motifs are the **α-helix** (Fig. 7) and the **β-sheet** (Fig. 8).
- Many proteins contain mainly α-helix or β-sheet motifs, while others use both kinds of motifs. An example is shown in Fig. 9.

![Fig. 7. α-helix motif.](image)

The polypeptide backbone forms a right-handed helix. The helix is stabilized by hydrogen bonds formed between backbone amino and carbonyl groups on successive turns of the helix. The amino acid side chains (blue) project outward from the helix.
**Tertiary structure** describes the overall spatial geometry of a polypeptide’s secondary structure elements. This is the overall 3-dimensional fold of the polypeptide.

Many, many proteins operate as larger **complexes**. The assembly of more than one polypeptide into a protein complex is the **quaternary structure**. This can mean as few as two small polypeptides, or an assemblage as big as the nuclear pore or — even bigger — silk or human hair.

**Fig. 8. β-sheet motif.**
Hydrogen bonds between backbone amine and carbonyl groups connect adjacent segments of the polypeptide. The side chains project above and below the sheet. The strands can run parallel (then N- and C- termini are on the same side) or anti-parallel.

Source: Pauling & Corey, 1951

**Fig. 9. Influenza virus HA protein.** This protein is a **homotrimer**, meaning that the quaternary structure is a complex containing three identical copies of a single type of polypeptide.

In this rendering, each of the three chains has both α-helix (red corkscrews) and β-sheet (blue arrows) motifs, connected by “loop” segments (purple). The polypeptides are post-translationally **glycosylated** with carbohydrate molecules (green).
SLO 2. Explain the roles of hydrophilic vs. hydrophobic aminoacyl residues in protein folding.

The biophysical principles that control protein folding are the exactly same ones that we have already seen with RNA and DNA:

1. Water molecules form many hydrogen bonds with one another, and they have high entropy (they can diffuse freely, translate, and rotate).
2. Hydrophobic (greasy) amino acid side chains are surfaces where water cannot hydrogen bond (this is an enthalpic penalty). Near these surfaces the water has reduced entropy, as well. Because water “hates” hydrophobic side chains, these chains “want” to be shielded from the aqueous solvent. Thus, hydrophobic amino acid residues tend to be buried within folded portions of the protein.
3. Hydrophilic (polar or charged) amino acid side chains can form energetically favorable hydrogen bonds with water. They are often exposed to the aqueous solvent. If they cannot interact with the solvent (if they are buried), they generally interact with other portions of the polypeptide through hydrogen bonds or salt bridges.
4. Additional inter-chain interactions that contribute to protein stability include van der Waals contacts, aromatic stacking interactions (analogous to the base stacking that we saw with DNA and RNA), and electrostatic repulsion between similarly charged (-/- or +/+ groups on the polypeptide.
Primary Structure

Primary structure is the linear sequence of amino acids as encoded by the DNA. This sequence defines how the protein will fold and therefore also defines how it will function. A single change in the amino acid sequence of hemoglobin can cause the proteins to clump together, resulting in the disease sickle cell anemia.

Secondary Structure

Hydrogen bonds between amino acids form two particularly stable structural elements in proteins: alpha helices and beta sheets. Alpha helices (shown in blue) are the basic structural elements found in hemoglobin, but many other proteins also include beta sheets. The inset highlights the pattern of hydrogen bonds (shown in green) that stabilizes alpha helices.

Tertiary Structure

Many functional proteins fold into a compact globular shape, with many carbon-rich amino acids sheltered inside away from the surrounding water. The folded structure of hemoglobin includes a pocket to hold heme, which is the molecule that carries oxygen as it is transported throughout the body.

Quaternary Structure

Two or more polypeptide chains can come together to form one functional molecule with several subunits. The four subunits of hemoglobin cooperate so that the complex picks up and delivers more oxygen than is possible with single subunits.

Fig. 10. Summary of the hierarchy of protein folding. Notice that, in contrast to the flu virus protein in Fig. 9, hemoglobin is an exclusively α-helical protein.
SLO 3. Explain the importance of correct protein folding, chaperone proteins, and how misfolding can lead to pathology.

Protein folding, chaperone proteins, and how misfolding can lead to pathology

Mutations that cause protein sequence changes, or errors in transcription or translation, can change the balance of forces that we have just describes, causing misfolding of a protein loss of its function. Other mutations may still allow a protein to fold more or less correctly, but change the protein’s activity. For example, some mutations result in ion channels that open more easily than they otherwise would, leading to neurological disorders.

Both within our cells, and in extracellular spaces (cartilage, blood, cerebrospinal fluid, etc.), very many different proteins are present at extremely high overall concentrations. This dense proximity means that proteins will touch other proteins during and after folding, with enormous potential for inappropriate interactions that can lead to non-specific aggregation.

You’re probably familiar with one protein aggregation process: making Jell-O™. We start with a clear aqueous solution of soluble proteins at high concentration. We then heat the solution so that the proteins unfold. That is, they are denatured. As the unfolded proteins cool, they aggregate into a single disordered gel.

In cells, protein aggregates are major sources of cytotoxicity, and — as we will see — they contribute to pathologies ranging from Alzheimer’s disease to type II diabetes.

Mutations can cause proteins to misfold at elevated rates, but even non-mutant proteins sometimes misfold, especially in the presence of stresses such as heat or oxidation. To mitigate inappropriate contact between un-folded or partially-folded proteins, cells use special proteins called chaperones. There are many different chaperones, and they shield proteins from inappropriate contacts, or actually use energy to pull apart proteins that have formed inappropriate contacts, giving the proteins a “second chance” to fold correctly.

When a protein cannot fold correctly even with the assistance of chaperones, the cell may recognize the misfolded polypeptide as hopeless, and mark it for destruction. This cellular surveillance process operates in almost every cell and is called protein quality control. The quality control system has at least two branches: the ubiquitin-proteasome system, and the autophagy-lysosome system. We’ll discuss these systems later in the block.
SLO 4. Understand common post-translational modifications of proteins (phosphorylation; disulfide bond formation; glycosylation) and know why specific modifications occur predominantly on proteins within the cytoplasm or in extra-cytoplasmic environments.

Once synthesized, most polypeptides undergo **covalent post-translational modifications**. These fall into several different categories. The following list is not comprehensive! It’s illustrative, showing some important examples.

1. **Proteolysis.** Many proteins are precisely clipped before they are fully functional. For example, many digestive enzymes are made as inactive **proenzymes** — a form safe for transport through sensitive cellular compartments. Upon secretion into the digestive tract an inhibitory portion of the polypeptide is clipped off, and the enzyme is activated.

2. **Disulfide bonding.** The terminal sulfhydryl group on the amino acid **cysteine** can be oxidized to form a **cysteine-cysteine disulfide bond.**
   a. Disulfide bonds are most often used to form mechanically stabilizing cross-links within a polypeptide chain or to cross-link two chains together in a protein complex.
   b. In general, the cytoplasm and nucleus of a cell have a chemically reducing potential, while the extracellular environment has a relatively oxidizing potential. What this means: we very seldom see proteins with disulfide bonds in the cytoplasm or nucleus, but **lots** of secreted proteins such as antibodies (Fig. 11), and many cell-surface proteins, have disulfide bonds.

![Fig. 11. An antibody (IgG) molecule.](image)

Each IgG is a **heterotetramer** containing two identical light chains and two identical heavy chains. This is an all β-sheet protein. The quaternary structure of the complex is stabilized by non-covalent interactions between the chains and also by disulfide bonds that covalently cross-link the two heavy chains together.
3. **Glycosylation.** Sugars are attached to most, but not all, **secreted and cell surface proteins**. Usually these are short-chain, branched carbohydrates. These sugars are used in cell-cell recognition and in cell signaling processes, and they can stabilize and protect proteins that are exposed to harsh extracellular environments. On the down side, viruses such as HIV use glycosylation to shield their surface proteins from recognition and attack by our immune systems. Very few cytoplasmic or nuclear proteins are glycosylated (though the ones that are may be of great importance).

![Image of glycosylation](image)

**Fig. 12. The glycan “shield” of a Coronavirus spike protein.**

The spike protein is used by Coronaviruses, such as SARS, to gain entry into host cells during infection. The spike consists of a protein homotrimer, anchored to the viral envelope (membrane) at the base.

Attached to each monomer are over twenty complex carbohydrate molecules (blue). Each sphere represents a hexose. The structure of the carbohydrates, ascertained using mass spectrometry, is shown in schematic form on the right.

The carbohydrates both stabilize the spike protein and shield it from proteases, antibodies, and other host defenses.


4. **Phosphorylation.** The covalent transfer of phosphate groups from ATP to polypeptides (protein phosphorylation) is a critical regulatory mechanism that controls almost every aspect of cell physiology.
   a. The phosphotransfer reaction is mediated by **protein kinase enzymes**. The recipient is always an amino acid residue with a **terminal hydroxyl group** on its side chain: serine, threonine, or tyrosine. The product is a **phosphoester**.
   b. **Phosphorylation is reversible.** Hydrolysis of the phosphoester removes the phosphophoryl group from the protein. **Dephosphorylation** is catalyzed by **protein phosphatase** enzymes.
   c. Several hundred kinases and phosphatases are encoded in the human genome.
   d. Protein kinases and phosphatases were discovered here at the UW School of Medicine, by Professors Ed Krebs and Eddie Fischer. For their seminal discoveries, these lifelong friends and collaborators shared the Nobel Prize.
SLO 5. Know that proteins are targeted to specific locations inside and outside of cells.

Different proteins have different functions and the proteins must carry out these functions in different locations. Inside a cell, for example, some proteins operate in the cytoplasm, some in the nucleus, some within organelles such as mitochondria, and some within the plane of the plasma membrane. Many proteins are secreted from cells. Examples include the collagen that holds our tissues together, antibodies and other proteins in blood and serum, digestive enzymes in the gut, and polypeptide hormones such as insulin.

Each protein must have a way to get to its site of action. **Protein targeting** typically involves a specific amino acid sequence within the polypeptide that serves as a “molecular zip code” used to direct the protein to its destination.

For example, the RNA polymerase II complex consists of several polypeptides. It is synthesized, folded, and assembled into a complex in the cytoplasm, but it has a “nuclear localization sequence” that directs the folded complex through the nuclear pore and into the nucleus, where it will transcribe mRNA molecules. As we’ll see, mutations in some proteins can cause mis-targeting, resulting in disease.

**Fig. 12. Extracellular matrix.**

Many big and small proteins including collagen (long white fibers) assemble into complex structural webs that link cells together within tissues. Each of these proteins carries a signal sequence that directs its secretion after it is synthesized inside the cell. Many secreted proteins are heavily post-translationally modified.

Source: David Goodsell
Session 213: Hemoglobin; prosthetic groups; enzymes

SLO 1. Explain the diversity of protein-protein and protein-ligand interactions.

In this session, we begin to think seriously about how proteins do their jobs. Proteins come in a dazzling variety of shapes, sizes, abundances, and tissue distributions.

Nearly without exception, all proteins are functionally similar: what proteins do is recognize specific chemical entities, and bind — stick — to them.

That’s it. That’s almost (though not quite) the whole deal. **Proteins stick to specific things.**

Antibodies stick to antigens to mediate immune responses during infection. Cell adhesion molecules allow cells to stick to each other and to extracellular matrix proteins. Extracellular matrix proteins stick to each other, and to cells. Transcription factors recognize and stick to specific enhancer sequences in our genes. Odorant receptors stick to specific volatile molecules (from perfume to putrescine). Hormone receptors stick to insulin, estrogen, and other hormones.

Enzymes are also most simply understood through their ability to stick to things. They stick to their substrates, and they stick to transition states more tightly, favoring the formation of those states and accelerating reaction rates. Enzymes often bind products less tightly, allowing their dissociation (release) from the enzyme.

Within membranes, ion channels, transporters, and pumps are again sticking to substrates and using a series of binding steps to move things from one side of a membrane to the other.

Molecular motors like myosin — same thing again. Myosin sticks and un-sticks to the actin thin filament. The order of myosin-actin sticking-unsticking is coupled to myosin sticking (binding) to ATP, to the ATP hydrolysis transition state, and finally to the release of ADP and Pi. Here, coupling of two binding cycles allows energy derived from ATP hydrolysis to make myosin’s binding and un-binding to actin directional — and our muscles contract.

With a general quantitative description of a protein’s binding characteristics we can understand an enormous amount about what a protein does, how it does it, how it can fail in its functions, leading to pathology. The same concepts, as we will see in the next session, allow us to think about how drugs interact with their molecular targets. After all, most drugs are just introduced chemical entities that particular proteins recognize, and stick to.
SLO 2. Understand the structure of hemoglobin and explain the function of heme in hemoglobin.

We begin this exploration with a protein we’ve already seen. Hemoglobin (Hb) is a protein present at enormously high concentration in the cytoplasm of red blood cells. The most important (but not only) thing hemoglobin sticks to is molecular oxygen (O₂).

**Fig. 1. Heterotetramer of α₂β₂ Hb.**
Note that there are four chains, each with a bound heme group.

Rendered from PDB 2DHB dataset (M. Perutz, 1970).

**Structure of Hb**

1. Adult Hb (HbA) is a complex of four polypeptides: two α chains and two β chains (α₂β₂).
2. Fetuses and infants have fetal HbF containing two γ (gamma) chains instead of β chains (α₂γ₂). The α, β, and γ chains are not identical, but they have very similar primary sequences and a nearly identical, all α-helical, tertiary fold.
3. **Terminology note:** α-helices and β-sheets are **not the same as the α and β chains of Hb**.
4. Separate genes encode mRNA templates for the various Hb chains.
5. Each of the four chains cradles one heme molecule (Figs 1 and 2): a porphyrin ring that coordinates an ion of iron (Fe²⁺) at its center. The bound heme is an example of a **prosthetic group**. As we’ll soon see, vitamins often serve as prosthetic groups in metabolic enzymes.
6. Within each subunit (or chain), amino acid side-chains and the bound heme operate together to bind one O₂ molecule (Fig. 2). Thus, the binding capacity of a hemoglobin heterotetramer is 4 O₂.
7. It follows that a one Hb tetramer can be ¼, ½, ¾, or entirely saturated with O₂.
8. With huge numbers of Hb molecules, O₂ saturation of the population can be anywhere from 0 to nearly 100%.
**Function: what Hb needs to do**

As erythrocytes pass through our lungs, their Hb binds O\textsubscript{2}. The erythrocytes flow with the blood to our peripheral tissues where they dump the O\textsubscript{2}. The tricky bit is that Hb needs to hold on to its precious cargo of O\textsubscript{2} until it is in a part of the body that’s most in need of O\textsubscript{2}.

In other words, Hb needs to bias its O\textsubscript{2} binding characteristics to accelerate dissociation in relatively hypoxic environments, rather than dumping O\textsubscript{2} randomly. This, along with control of vasoconstriction, allows us to maintain a relatively shallow pO\textsubscript{2} gradient from our lungs to our fingers, toes, and brain in the periphery.
SLO 3. For a general ligand-receptor pair, be able to explain and calculate the relationship between $k_{on}$, $k_{off}$, and $K_D$.

We will say that $O_2$ is a ligand and HbA is a receptor. As the partial pressure of oxygen, $pO_2$, increases, the fractional saturation of HbA increases (Fig. 3).

The first thing to notice is this is an ensemble measurement of many molecules of HbA.

We note that there is a concentration of $O_2$ where 50% the receptor (HbA) is 50% saturated by its ligand ($O_2$).

1. This concentration is the dissociation constant ($K_D$), and it indicates how tightly a ligand binds its receptor. For this reason, $K_D$ is also called the affinity constant.
2. $K_D$ has units of concentration. For a gas, we may use partial pressure (units: mm Hg, atmospheres, Pa, p.s.i, etc.). For solutes in liquid we will more often use concentration per volume (M, g/L, etc.).
3. We say that an interaction with a small $K_D$ (say, $10^{-9}$ M) is high-affinity. We say that an interaction with a large $K_D$ (say, $10^{-3}$ M) is low-affinity.
4. For a simple ligand-receptor pair, $L + R \rightleftharpoons LR$,

$$
K_D = ([L][R]) / [LR]
$$

where $[L]$, $[R]$, and $[LR]$ are the concentrations of the ligand, the receptor, and the complex, at equilibrium (when the concentrations of free and receptor-bound ligand are not changing).
5. $K_D$ is an equilibrium constant, which reflects two rates: the rate at which a particular ligand sticks to the receptor (association), and the rate at which that ligand falls off (dissociation). For the simple case, $L + R \rightleftharpoons LR$,

$$
K_D = k_{off}/k_{on}
$$

where $k_{on}$ is the rate constant for association and $k_{off}$ is the rate constant for dissociation. $k_{on}$ has units of inverse concentration and time (mol$^{-1}$ s$^{-1}$), and $K_D$ has units of inverse time (s$^{-1}$).
6. Take the time to satisfy yourself that if the association rate constant increases, the receptor-ligand affinity increases. Conversely, if dissociation rate constant increases, the affinity decreases.

Now take another look at the shape of the saturation curve for HbA. What does the shape of the curve tell us? We will be thinking through this question in class.
SLO 4. Describe the mechanistic bases of hemoglobinopathies.

Diseases caused by mutations that alter hemoglobin function are probably the most prevalent and best-understood of all Mendelian genetic disorders. At least 1 in 15 people carry genetic variants that contribute to hemoglobin-related disorders.

To understand the hemoglobinopathies we need to start by looking at the genes that encode the various Hb chains, and their expression patterns (Fig. 4).

1. First, note that during gestation, Hb is initially produced in the yolk sac and then in the liver. Toward the end of gestation, Hb production gradually switches to the bone marrow.
2. Concomitant with the changes in the locations of Hb production, the chain types produced are also changing. The major form of fetal and infant Hb, HbF (α2γ2), is gradually supplanted by adult HbA (α2β2).
3. These changes are called Hb switching. They represent a classical example of developmental regulation of gene transcription. The mechanisms through Hb switching occurs include both the types of mechanisms we have already seen, and “epigenetic” changes in DNA packaging (chromatin regulation) that we’ll see in more detail later.

Fig. 4. Expression patterns of Hb α-like and β-like chains. On top is a map of the β-like globin loci on chromosome 11. The Locus Control Region (LCR) contains a series of enhancer sites required for expression of all β-like globin loci. Additional enhancer and repressor elements then control the expression of each specific globin gene.

On the bottom is a map of the α-like globin loci on chromosome 16. Note that there are two identical copies of the α-globin gene.


4. When we look at the arrangement of the genes that encode Hb chains, we see that they are arranged in clusters: the β-like loci sit together on chromosome 11, and the α-like loci sit together on chromosome 16.
5. Remarkably — and unusually — each set of genes is arranged in its temporal order of expression.
Sickle cell disease (hemoglobin S disease)

Sickle Cell Anemia is the most prevalent single hemoglobinopathy. It falls within the broader class of hemolytic anemias.

**Note — Cyanosis vs. Pallor.** You cannot manifest red (or blue) color without blood. Pallor (being pale) reflects anemia. Cyanosis (being blue) indicates that blood is poorly oxygenated. As mentioned above, HbS carries O2 normally, but leads to clogging of capillaries and hemolysis, the latter resulting in anemia.

1. Sickle disease results from a single nucleotide **missense substitution**, $\beta^S$, that changes a codon for glutamate to a codon for valine at amino acid position 6 in the Hb $\beta$-chain ($\beta$-globin).
2. The $\beta^S$ mutation has **no effect on the ability of Hb to carry O2**.
3. Heterozygote carriers of the sickle allele, referred to as $\beta^S$, generally present few or no symptoms. Extreme physical exertion can lead to rhabdomyolysis.
4. Heterozygote carriers exhibit partial resistance to malaria. The $\beta^S$ allele appears to have appeared *de novo* multiple times. Its prevalence is elevated in populations in the Mediterranean, Africa, and southern Asia, all areas with endemic malaria.
5. Heterozygous carriers of the sickle trait produce a mix of Hb tetramers: $\alpha_2^A\beta_2^A$ (normal HbA), $\alpha_2^A\beta_2^S$ (sickle HbS), and $\alpha_2^A\beta^A\beta^S$.
6. Homozygotes with sickle cell disease produce mainly $\alpha_2^A\beta_2^S$ (sickle) Hb tetramers.
7. The O2-free form of HbS, deoxyHbS (deoxy-HbS), is five times less soluble than deoxyHbA.
8. At the high concentrations of HbS that are present in homozygotes, **deoxy-HbS tetramers assemble into long, higher-order filaments**. These deoxy-HbS filaments distort the normal rounded shape of erythrocytes, leading to clogging of capillaries. The fibers can also puncture the cell’s membrane, causing erythrocyte lysis (hemolysis).

This is one of many examples you’ll see of protein folding diseases.
Other hemoglobinopathies

In addition to S disease there are a large number of mutations in the various Hb chains that can lead to disease. These fall into three categories:

1. **Structural hemoglobinopathies.** These disorders are generally caused by **missense mutations** that alter the primary structure of Hb chains. Sickle (HbS) disease is an example.

2. **Thalassemias.** These disorders are caused by **imbalances** in the amounts of α- and β-chain synthesis, degradation, or Hb tetramer assembly, leading to excess production of unassembled globin chains.

3. **Hereditary persistence of fetal hemoglobin (HSPS).** These are regulatory disorders in which **the switch from HbF to HbA fails to occur** in early childhood. HSPS by itself does not lead to major pathology but it can be a strong genetic modifier of other hemoglobinopathies and thalassemias.

An absolutely key point is that **there are two identical copies of the Hb α-chain gene** on chromosome 16 (See fig. 4). This means that most people have **four copies** of the α-chain gene. Consequently, recessive disorders of Hb β-chains are far more frequent, and tend to be more severe, than disorders of the Hb α-chains.

Moreover, because of the **temporal order** of Hb chain expression (Fig. 3), disorders of Hb β-chains tend to manifest only in childhood (recall that HbA is αβ2), while disorders of Hb α-chains can begin manifesting prenatally (HbF is α2γ2).

**Structural hemoglobinopathies**

For the Hb β-chain (β-globin) alone, mutations leading to the synthesis of well over 700 different structural variants of the protein have been identified. Additional variants in other globins can also cause disease.

The Hb structural variants were originally named with letters (HbS, HbE, etc., and later, they were named by the location where the carriers of the mutations were identified (Hb Hammersmith, etc.). Structural mutations can lead to changes in the O2 saturation curve (and cyanosis, as in Hb Hammersmith), changes in Hb solubility, and hemolysis (as in HbS), to thalassemias, or combinations of these defects.

In addition, the iron in the heme center can be oxidized by bound O2, from ferrous (Fe2+) to ferric (Fe3+). Hb containing an oxidized heme center **cannot bind O2**, and is called methemoglobin. An enzyme, **methemoglobin reductase** converts the heme iron to the ferrous (Fe2+) state, and restores its ability to carry O2.

Mutations that impair methemoglobin reductase cause gradual conversion of Hb to methemoglobin, resulting in cyanosis. Moreover, some structural variants of Hb (such as Hb Hyde Park) cannot productively engage with methemoglobin reductase, and these also gradually convert into methemoglobin, (also leading to cyanosis).
Thalasseimias

A vast variety and number of mechanisms and mutations can cause imbalances in globin chain abundance. Consequently, thalassemias occur with a huge variety of severities and presentations.

1. α-thalassemia. Recall that there are two α-globin genes on chromosome 16 (Fig. 4). α-thalassemias most commonly arise through deletion of entire α1 or α2 genes.
   a. The silent carrier genotype is −α/αα, resulting in 75% of normal α-globin production.
   b. The α-thalassemia trait occurs in two forms: −/−αα, and −α/−α. These genotypes result in 50% of normal α-globin production.
   c. Simple α-thalassemia disease is associated with the −α−/−α genotype. Only 25% of the normal α-globin amount is produced, and an excess of Hb β4 tetramers are produced. This variant is also called HbH or Hb Bart’s.
   d. The −α−/−α − genotype results in assembly of Hb γ4 tetramers. Because this genotype is embryonic lethal, the switch from γ to β expression never occurs.

2. β-thalassemia. A wide variety of mutations can cause reduced production of β globin chains, and β-thalassemia. We review these because they illuminate and summarize many of the ways that mutations can change protein production.
   a. The locus control region (LCR; Fig. 4) is needed for transcription of the entire set of β-like globin chains. Deletions within the LCR can decrease or eliminate transcription of all of the β-like genes.
   b. Similarly, mutations in the β-chain promoter can impair transcription of the β-chain mRNA alone, but leave transcription of the other β-chain genes intact.
   c. Mutations in the β-chain transcription unit can prevent formation of a functional mRNA template:
      d. Defects in 5’cap addition;
      e. mRNA splicing defects;
      f. Failure of poly-A tail addition;
      g. Point mutations can eliminate the start codon;
      h. Insertions or deletions can cause frameshifts;
      i. Introduction of premature stop codons (nonsense mutations).
      j. Missense mutations (like those in the structural hemoglobinopathies) can also cause instability β-chain protein and its destruction by the protein quality control system.

3. As with sickle cell trait (HbS), heterozygotes with the α- and β-thalassemia traits are partially protected from malaria, and these traits (and hence, the associated diseases) occur at elevated levels in populations affected by endemic malaria.

Because most disease alleles are in carriers compared to affecteds for recessive disorders (2pq>>q²), a fitness advantage to carriers often outweighs the loss of fitness among those with disease. (As used here, fitness refers to survival to reproduction.)
SLO 4. Explain the key properties of enzymes. Explain why many enzymes contain bound prosthetic groups. Describe enzymatic co-factors and how deficiency of these factors leads to disease.

Enzymes

Enzymes are highly selective catalysts.

1. Enzymes bring together specific substrate molecules in specific geometries, and accelerate chemical reactions that convert the substrates into specific products.
2. As catalysts, enzymes cannot change the thermodynamics — the overall free energy balance — of a reaction (Fig. 3).
3. Instead, enzymes change the kinetics of the reaction. They do this by reducing activation energy barriers that would otherwise prevent the reaction from occurring at a biologically relevant rate.

![Fig. 3. Principle of enzyme-mediated catalysis. E, enzyme. S, substrate(s), P, Product(s). The enzyme does not alter the overall free energy balance of the reaction pathway. Rather, the enzyme lowers the activation energy for formation of the transition intermediate. Source: Wikimedia](image)

4. By bringing specific substrates together in the appropriate geometry, and by shielding intermediates from reactive non-substrate molecules, enzymes suppress the formation of off-pathway products.
5. The vast majority of enzymes are proteins. (But, as we saw with the ribosome, a small critical subset of enzymes have catalytic centers made out of RNA.)
6. Many enzymes use covalently or non-covalently bound prosthetic groups to control the electronic environment within their active sites, promoting catalysis. Prosthetic groups are sometimes, but not always, vitamins. Thus, vitamin deficiency often leads to compromised function in specific enzymes, leading to metabolic or structural pathologies.
7. In summary, we can think of enzymes as receptors that bind ligands (substrates) in highly specific ways, to promote specific chemical reactions among those substrates.
SLO 5. Explain and calculate the relationships between \( k_{\text{cat}} \), \( K_M \), and \( V_{\text{max}} \).

**Michaelis-Menten enzyme kinetics**

As with ligand-receptor interactions, a small number of parameters provides a vivid description of an enzyme’s critical properties.

1. \( V \) is the reaction rate at which substrate is converted to product, under some specified set of conditions (substrate concentration, temperature, pH, etc.).
2. The \( V_{\text{max}} \) is the *maximum* rate at which the enzyme can convert substrate to product. \( V_{\text{max}} \) occurs when the substrate is at a sufficiently high concentration that its availability is not rate-limiting.
3. The \( K_M \), or **Michaelis-Menten constant**, is the concentration of substrate at which the rate of product formation per molecule of enzyme is half-maximal. Thus, *at a substrate concentration \([S]\)*, where \( V = V_{\text{max}}/2 \),

   \[
   K_M = [S]
   \]
4. \( K_M \) is closely analogous to \( K_D \), because each refers to the concentration of ligand or substrate where 50% of the receptor or enzyme is occupied.
5. \( k_{\text{cat}} \) is the rate constant for the enzyme-catalyzed conversion of the enzyme-bound substrate to product: For the scheme in Fig. 3, \( k_{\text{cat}} \) is the rate for \( ES \rightarrow E+P \). \( k_{\text{cat}} \) is also called the enzyme’s *turnover rate*.
6. The **catalytic specificity** of an enzyme for any given substrate can be defined as a ratio,

   \[
   k_{\text{cat}} / K_M
   \]
7. By comparing \( k_{\text{cat}} / K_M \) ratios for one enzyme and various substrates, we can learn how **selective** an enzyme is. For this reason \( k_{\text{cat}} / K_M \) is often called the **specificity constant**.
8. *For example*: the active site pockets of DNA polymerase enzymes have high affinity (small \( K_M \)) for dNTP (DNA) nucleotides, but extremely low affinity (very large \( K_M \)) for NTP (RNA) nucleotides. This is because DNA polymerase active sites are usually shaped so that the 2’-OH group of dNTPs does not fit within the pocket.
9. Thus for DNA polymerases \( k_{\text{cat}} / K_M \) for dNTPs is relatively large, while \( k_{\text{cat}} / K_M \) for NTPs is very small. *We therefore say that DNA polymerases are selective for dNTP substrates versus NTP substrates.*
10. As we’ll see in following sessions on Pharmacology, **competitive enzyme inhibitors** have high affinities for enzyme active sites (low \( K_M \)), but small — or zero — \( k_{\text{cat}} \). Competitive inhibitors “clog” an enzyme’s active site, preventing legitimate substrates from binding.
11. We explore how this works in Fig. 4. Here, we plot a surface of constant \( k_{\text{cat}} / K_M \) for a family of enzymes and substrates. As \( K_M \) decreases, and \( k_{\text{cat}} \) decreases proportionately, the \( V_{\text{max}} \) decreases. If \( k_{\text{cat}} \) falls faster than \( K_M \), the reaction rates would be below the plotted surface: this is the space where competitive inhibitors operate.
Fig. 4. Relationship between [S], $K_M$, and reaction rate ($V$), at constant $k_{cat}/K_M = 1$.

For E:S pairs with small $K_m$, the reaction saturates at very low [S], but the Vmax is correspondingly small. This is because the “high affinity” E:S pairs bind tightly to the enzyme.

To increase the $V_{max}$, affinity must be lowered ($K_m$ must increase), or the turnover must be faster (the value of $k_{cat}$ must be increase). Units are arbitrary.

Session 222: Secretory and endocytic pathways

nearly all proteins have distinctive subcellular (or, for secreted proteins, extracellular) localizations. Many proteins are directed to specific locations by “molecular zip codes” — structures within the protein that contain targeting information. Membrane and secreted proteins are encoded by 30-40% of our genes.

SLO 1. Describe the major organelles of the secretory pathway.

1. Secreted proteins, and most integral membrane proteins, are synthesized by ribosomes at the rough endoplasmic reticulum (ER), co-translationally translocated into (membrane proteins) or across (secreted proteins) the ER membrane.

2. At the ER, proteins are folded with the aid of chaperone proteins. In many but not all cases secreted and membrane proteins are glycosylated: complex carbohydrates are covalently attached to the proteins by ER-resident enzymes.

3. Following folding, the proteins are packaged into carrier vesicles and transported to the Golgi apparatus. In the Golgi further post-translational modifications are performed and additional quality control checks are made. The proteins enter the Golgi complex at the cis Golgi, traverse the medial Golgi, and end up in the trans Golgi network (TGN).

4. The TGN is the Grand Central Station of the secretory pathway: here, proteins are sorted into different carrier vesicles that can go to many different locations in the cell, including endosomes, lysosomes, and secretory vesicles.

5. There are two broad categories of secretory vesicles.
   a. Constitutive secretory vesicles fuse with the plasma membrane “automatically,” by default. This is a “housekeeping” pathway.
   b. Regulated secretory vesicles fuse with the plasma membrane only in response to a signal. There are many different types of regulated secretory vesicles in different cell types. They may contain hormones (e.g., insulin), neurotransmitters, blood clotting factors, etc. The commonality is that regulated secretory vesicles are held in reserve in the cytoplasm, until a signal indicates that their contents are required at the cell surface.

6. Fusion of a secretory vesicle at the plasma membrane is called exocytosis.
SLO 2. Understand the difference between targeting of integral membrane proteins and secretory proteins.

1. Integral membrane proteins, including receptors and ion channels, have one or more membrane-spanning domains that traverse the lipid bilayer.

2. Secreted proteins generally do not have transmembrane domains, and are soluble in the aqueous lumens of the ER, Golgi, and secretory vesicles.

3. Both integral membrane proteins, and secreted proteins, are usually translocated into the ER as they are synthesized. That is to say, translocation is co-translational.

4. A common protein machinery is used to target both secreted and integral membrane proteins at the ER: the Preprotein Translocase. This complex is also called the Sec61 complex.

5. Recall that the N-terminus is the first part of a protein to emerge from the ribosome. Secreted and integral membrane proteins emerging from the ribosome are marked by a “zip code” — an N-terminal signal sequence.

6. The signal sequence is identified by the Signal Recognition Particle (SRP). The SRP directs the mRNA, ribosome, and nascent polypeptide to the Preprotein Translocase.

7. As the polypeptide elongates, it is threaded through the preprotein translocase.

8. For secreted proteins, the signal sequence is removed by a site-specific protease. This allows the protein to diffuse in the aqueous phase.

Fig. 2. Targeting of secreted and membrane proteins to the endoplasmic reticulum.

Fig. 3. SRP recognizes N-terminal signal sequences to target nascent polypeptides to the ER.
SLO 3. Outline how proteins are folded and modified in the endoplasmic reticulum and Golgi organelles.

1. Chaperones in the ER lumen, such as BiP, intercept the nascent protein and attempt to help it fold correctly. In many cases, folding requires the assembly of multi-protein complexes (oligomerization).

2. Transmembrane proteins are assembled by allowing trans-membrane segments to escape from the preprotein translocase laterally, into the membrane bilayer (Fig. 4). The transmembrane segments are hydrophobic. Thus integral membrane proteins are soluble (that is, they can diffuse) within the two-dimensional plane of the membrane bilayer, but they are not freely soluble in the aqueous phase.

3. 20% of the time or more, the folding of membrane proteins fails! The protein mis-folds: it must then be identified as a “dud” and degraded. The polypeptide is extracted back into the cytoplasm, and degraded by the proteasome (which you will learn about later in the course).
   a. Mutations, or errors in transcription or translation, can cause mis-folding.
   b. Chemical and physical stresses (oxidation, elevated temperature, etc.) can cause mis-folding.
   c. Some amount of mis-folding happens through stochastic processes, even without unusual mutations or stresses.
   d. Note that quality control occurs both in the cytoplasm and in the secretory system.

---

**Fig. 3.** Transmembrane domains are inserted into the ER membrane by lateral escape from the preprotein translocase.

**Fig. 4.** Overview of protein targeting and protein quality control mechanisms.
4. The protein folding and quality control machinery can be overwhelmed. This causes **ER stress**. The **ER stress response** is a **signaling pathway** that does three major things:
   a. It senses the amount of unfolded protein in the ER lumen;
   b. it slows down translation (protein synthesis); and
   c. it causes transcription of genes encoding ER chaperones.

Thus the ER stress response is a **homeostatic mechanism** that allows the efficiency of ER folding to respond appropriately to demand.

5. The ER stress pathway can be overwhelmed. At this point, the cell may choose self-destruction (apoptosis).

6. Proteins exported to the Golgi are subjected to additional quality control checks for proper folding and post-translational modification. If they fail these checks, they are often **retrogradely** transported back to the ER. At the ER they get an additional chance to fold or to be degraded.
SLO 4. Explain how proteins and lipids are packaged into carrier vesicles, and how carrier vesicles fuse with target membranes including the plasma membrane.

Transport vesicles are membrane-delimited containers that carry lipids, integral membrane proteins, and solutes (including proteins) between organelles, and between organelles and the plasma membrane. We can outline in general how transport vesicles operate. This is a general scheme showing common themes for different pathways.

1. Cargo lipids and proteins are identified by cytoplasmic coat complexes. Note that soluble secretory proteins must be identified by transmembrane receptors that traverse the membrane to engage with the coat complex. **Note: different coat complexes assemble transport vesicles with different contents.**

2. The coat complex forms a shell that both captures cargo (lipids, proteins), and deforms the membrane (budding).

3. The coated transport vesicle detaches from the donor membrane.

4. The coat dissociates.

5. Proteins and lipids on the transport vesicle are recognized by tethering factors that attach the transport vesicle to an acceptor membrane. We can think of these as molecular “Velcro” – but this “Velcro” has specificity: only appropriate vesicles will tether to a given acceptor membrane. **Note: Different vesicles and acceptor membranes use different tethering factors.**

6. Tethering leads to a tighter association: docking. During docking, the fusion complex is assembled. The fusion complex contains SNARE proteins.

7. Zippering of SNARE proteins is energetically favorable, and leads to fusion of the vesicle and acceptor membranes. **Different combinations of SNARE proteins can assemble. Only some of these actually trigger fusion — an additional layer of specificity.**
Here is an example of two specific pathways, operating in concert. If you examine this diagram closely, you’ll see specific examples of all the general mechanisms described on the previous page.

a. The forward (anterograde) pathway uses the **COPII** coat to make vesicles that carry secretory cargo molecules from the ER to the cis-Golgi.

b. The backward (retrograde) pathway uses the **COPI** coat to retrieve SNARE proteins, cargo receptors (e.g., H/HDEL) back to the ER. The retrograde pathway also retrieves mis-folded proteins that have failed quality control checks in the Golgi.

![Fig. 6. Distinct coats control anterograde and retrograde traffic between the ER and Golgi.](image)

**SLO 5.** Explain key mechanisms that underlie endocytosis, receptor recycling, and traffic to the lysosome.

If you imagine how the secretory pathway operates in isolation, there is a problem: membrane is continuously deposited on the plasma membrane, but not retrieved. The plasma membrane grows and grows. Thus, there is a second system that brings membrane back into the cell: the **endocytic pathway**.

The endocytic system also has a **degradative** role, bringing membrane lipids, proteins, and surface-adsorbed particles into the cell, where they can be routed back into the secretory pathway or, alternatively, sent to the hydrolytic **lysosome** for destruction and recycling. The endocytic system also has a **recycling pathway**, as we’ll see below.
While the secretory pathway begins at the ER and ends with **exocytosis** (exit or ejection from the cell), the endocytic pathway begins at the plasma membrane, initiating with **endocytosis** (taking into the cell).

Fig. 7. Overview of secretory and endocytic systems, and coat complexes that mediate transport within them.

Note: **ERGIC** is the ER-Golgi intermediate compartment. You don’t need to know about it.

In endocytosis, the coat consists of an inner shell that recognizes cargo and an outer shell that organizes membrane deformation. The outer shell complex, **clathrin**, also operates at other steps of intracellular transport. The inner shell adaptor system is diverse, with molecules specialized for capturing different cargo at different locations in the cell.

Fig. 8. The clathrin vesicle coat system.

Note: You **do not** need to know the names of the different clathrin adapter proteins.
SLO 6. Outline the synaptic vesicle cycle, and explain how clostridial neurotoxins selectively block neurotransmission.

At chemical synapses, vesicles loaded with neurotransmitter fuse with the plasma membrane of the presynaptic cell. This is a regulated exocytosis event. This followed by endocytosis.

Together, these events constitute the synaptic vesicle cycle. This is an example of an endocytic recycling pathway: the donor membrane from which the vesicle is derived, and the acceptor membrane with which the vesicle fuses, is the same membrane — the plasma membrane.

![Synaptic Vesicle Cycle Diagram](image)

**Fig. 9. The synaptic vesicle cycle.**

Clostridial neurotoxins are secreted by bacteria of the genus *Clostridium*, including the species that cause botulism and tetanus.

The clostridial neurotoxins bind to receptors on the plasma membrane, are endocytosed, and then the active toxin “payload” is translocated across the membrane into the cytoplasm. The active toxin is a protease enzyme that requires a bound Zn$^{2+}$ metal ion for activity (a Zn$^{2+}$ protease).

These proteases have exquisite selectivity: they recognize only SNARE proteins involved in exocytosis, and cleave them at specific sites. This potently blocks membrane fusion, and therefore prevents neurotransmission.