

Part II

Cellular Compartments of Prokaryotes and Eukaryotes: Organization, Dynamics, and Functions

1 Cellular membrane systems

1.1 General Structure

When studying cellular membranes one must remember that membranes provide several critical functions to cells. Membranes provide protection, isolate functions of various chemicals/organelles, transduce signals, and provide a selective permeability barrier. The typical membrane consists of lipids and proteins which are held together by *noncovalent* interactions. The lipids form a bilayer which provides the basic fluid structure for the membrane. In animal cells, lipids make up approximately 50% of membranes, in a small animal cell this equates to approximately 10^9 lipids in the plasma membrane. It is also important to remember that all lipid molecules in cell membranes are **amphipathic**. The amphipathic nature of the lipid molecules allows for spontaneous formation of bilayers, as seen in the creation of artificial liposomes.

There are several classes of lipids, the most abundant being **phospholipids** (major phospholipids discussed below). These possess a polar head group with two hydrocarbon tails, which are typically fatty acids of differing length (14 – 24 carbons). Typically one tail is unsaturated (kinked), while the other is saturated and straight. Differences in the length and saturation of each tail are important influences on the ability of the lipids to pack together, thus affecting the fluidity of the membrane.

As discussed above, lipids are capable of spontaneously forming bilayers in aqueous environments. This reaction is due to the inability of lipids, with their hydrophobic regions, to interact favorably with water.

When hydrophobic molecules are placed into water they force water to rearrange into very ordered structures surrounding the molecules, creating an energetically unfavorable state, this is minimized through the clustering of hydrophobic molecules thereby affecting the smallest number of water molecules possible (Figure 1).

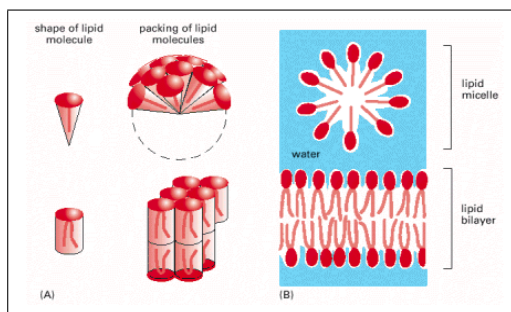


Figure 1: Spontaneous bilayer formation

1.1.1 Movement within a Bilayer

Within the bilayer lipids are capable of **lateral diffusion** within a monolayer as well as **”flip-flopping”** across a bilayer. In addition, individual lipid molecules have been shown to rotate along their long axis rapidly. These three types of movement are shown in figure 2.

While lateral diffusion occurs very rapidly approximately 10^7 times per second **”flip-flop”** across bilayers occurs less than once a month for any individual lipid. This presents an interesting problem for the production of new phospholipid molecules in the cell as they are primarily produced on the ER’s cytosolic membrane face. The overproduction on one side

would not produce a suitable bilayer without timely migration across to the noncytosolic monolayer. This problem is solved by a special class of membrane-bound enzymes known as **phospholipid translocators (flippases)**, which catalyze the "flip-flop" movement.

1.2 Membrane Composition

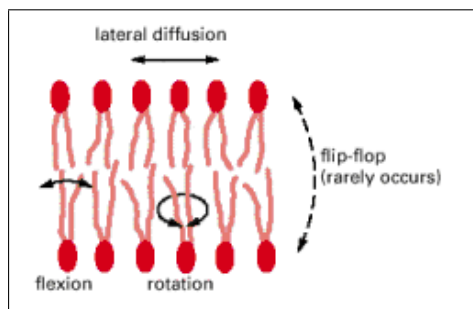


Figure 2: Spontaneous lipid movement

This brings us to another important point which is that cellular membranes are not comprised completely of phospholipids and the additional components of the membrane structure play critical roles in cellular function as well as membrane function. Other proteins which typically reside in membranes include but are not limited to, channels, carriers, signalling, and receptor proteins. In addition to phospholipids and proteins cellular membranes often contain **cholesterol** and **glycolipids**. Each of these serve important roles within the cell.

It is important to note that with the functional importance of many membrane components, their placement in one or both monolayers is equally important to the cell. Thus, the lipid, protein, and glycolipid compositions of each monolayer are asymmetrical. Take for example human red blood cells, where almost all of the choline containing lipids are in the outer monolayer, whereas all the terminal amine containing lipids are on the inner leaflet. This asymmetry is produced by the cell and ultimately critical to the survival of the cell. The distribution of specific head groups on the inner/outer leaflets is critical for protein interactions for a variety of functions. An example of this is **Protein Kinase C (PKC)** and its binding to the negatively charged **phosphatidylserine** located on the inner leaflet of membranes.

Sometimes phospholipids are also used in response to extracellular signals. The asymmetry allows for specific phospholipids to be cleaved with specific fragments remaining in the monolayer and others released as signalling molecules. An example of this is **Phospholipase C**, which cleaves an inositol phospholipid in the cytosolic monolayer which generates two fragments. One fragment remains in the monolayer activating PKC, while the other is released into the cytosol and stimulates the release of Ca^{2+} from the ER.

In addition to the discussed functions, the asymmetry of the membranes is utilized by animal cells to distinguish between live and dead cells. Animal cells which undergo apoptosis have phosphatidylserine rapidly translocated across to the extracellular monolayer to signal macrophages to phagocytose the dead cell. This translocation occurs through TWO mechanisms:

1. The phospholipid translocator that normally transports this lipid from the noncytosolic monolayer to the cytosolic monolayer is inactivated.
2. A "**Scramblase**" that transfers phospholipids nonspecifically in both directions is activated.

The most extreme asymmetry can be found in the distribution of glycolipids within cellular membranes. These sugar-containing lipids are found exclusively in the noncytosolic monolayer where they seem to partition specifically into **lipid rafts**. Lipid rafts are small microdomains of typically **sphingolipids** and cholesterol. They are held together into these domains through the van der Waals forces between the long saturated hydrocarbon

tails of sphingolipids. Due to the longer saturated hydrocarbon tails lipid rafts tend to be thicker than typical parts of the bilayer. Also, this increased thickness seems to better accommodate certain membrane proteins causing them to accumulate in lipid rafts. Lipid rafts are thought to help organize these proteins to allow them to function together, or concentrating them for transport. An important point to note about lipid rafts is that it appears that the two monolayers interact with each through their lipid tails.

1.2.1 Membrane Fluidity

Membranes are a critical part of cells and disruption of the membrane would most likely result in death of the cell. Thus an important aspect of the membrane is the maintenance of its fluidity. This occurs within eukaryotic cells on three levels. First, the length of the lipid tails within the bilayer alters the **phase transition temperature** (freezing temperature). The shorter the chain length the lower the tendency for hydrocarbon tail interactions thus lowering the phase transition temperature. Second, the cell can regulate membrane fluidity through the production of lipids with more/less kinks within the hydrocarbon tails. With more kinks it is more difficult for lipids to pack together resulting in greater fluidity at low temperature. The third level of control lies in the concentration of cholesterol within the membrane. Cholesterol molecules function as a buffer to temperature change within membranes as their rigid steroid rings provide a level of immobilization resulting in less hydrocarbon interaction which can lead to crystallization at low temperature. Note that cholesterol decreases the permeability of the bilayer to small water-soluble molecules as well making the membrane less fluid at normal temperatures.

1.2.2 Phospholipids

Four major phospholipids predominate the plasma membrane of mammalian cells: **phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and sphingomyelin** (structures in Figure 3). Remember that only phosphatidylserine carries a net negative charge, the importance of which will be clear later. These four make up the majority of lipid mass in membranes. Other phospholipids, such as the inositol phospholipids are present in smaller quantities but are functionally very important.

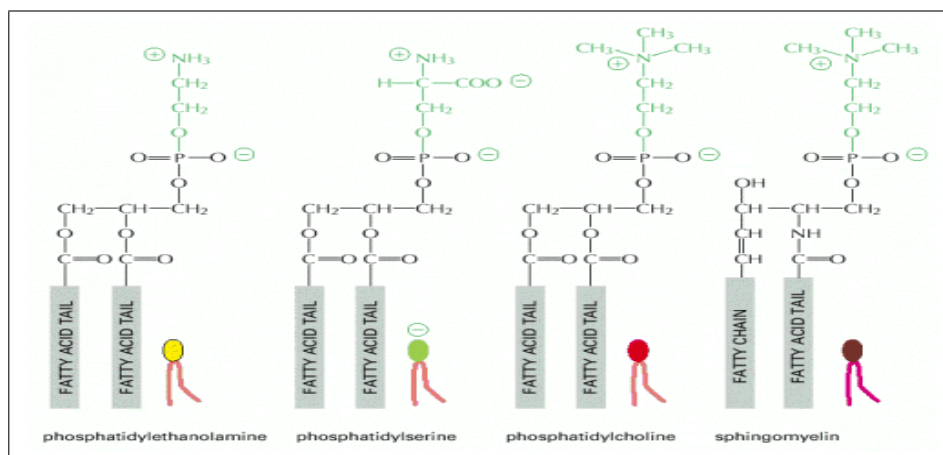


Figure 3: Various phospholipid structures

Remember Phosphatidylserine is negatively charged!

1.3 Transport

An important aspect of lipid bilayers is that they are selectively permeable. Given enough time, even with a protein-free lipid bilayer virtually any molecule will diffuse DOWN its concentration gradient across a lipid bilayer. The main limitations on this transit are the solubility of the molecule in oil. Small nonpolar molecules such as O_2 and CO_2 diffuse rapidly across membranes. Small uncharged polar molecules such as water and urea are also able to passively diffuse although at a slower rate. Charged molecules however no matter the size are practically unable to enter the hydrocarbon portion of the bilayer. Experimentally with synthetic bilayers, water is 10^9 times more permeable than small ions such as Na^+ and K^+ .

Membrane transport is conducted throughout the cell by two main classes of membrane transport proteins: **Carriers** and **Channels**. These proteins allow various polar molecules across the selectively permeable membrane. All transport proteins that have been studied have been found to be multipass transmembrane proteins. Carrier proteins (carriers, permeases, or transporters) bind a specific solute and undergo conformational changes resulting in movement of the solute across the membrane. In contrast, channel proteins form aqueous pores that allow specific solutes (through the use of a specificity filter) to pass through. Channels are much faster than carriers.

All channel proteins and many carriers proteins only allow downhill transfer of solutes across the membrane, or passive transport (facilitated diffusion). In the case of an uncharged molecule only its concentration gradient plays a role in the directionality of movement. However, in the case of a charged molecule not only does the concentration gradient influence the directionality but the electrical gradient across the membrane also plays a role, thus charged solutes move in a direction dependent on their electrochemical gradient. While its all well and good that a cell can have molecules move downhill across membranes, it is sometimes necessary for molecules to move against the electrochemical gradient, thus cells have the ability to *actively* transport molecules across the MB through the use of carriers (pumps). Active transport requires the coupling of the directional movement to a source of energy such as ATP hydrolysis or an ion gradient.

since Fat Alberts had a decently large section on this I'm going to toss this in as well

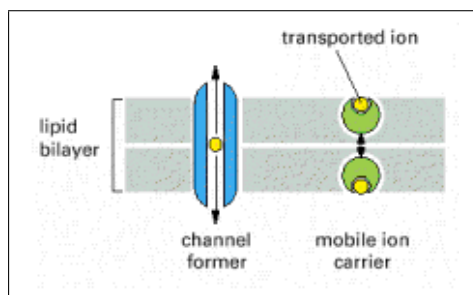


Figure 4: Channel former and Mobile ion carrier

Ionophores are small hydrophobic molecules which can dissolve in lipid bilayers and increase their permeability to specific ions. They are typically produced by microorganisms, but now are utilized by cell biologists in order to increase ion permeability for experiments. Two classes of ionophores exist, **mobile ion carriers** and **channel formers**. Both types shield the charge of the transported ion in order to allow it to pass through the bilayer. Ionophores are uncoupled and ONLY allow movement downhill. Figure 4 is a good representation of each of those ionophores.

Two examples of ionophores are *Valinomycin* and *Gramicidin A*. *Valinomycin* is a mobile ion carrier, which transports K^+ . *Gramicidin A* on the other hand is a channel forming ionophore, produced by certain bacteria which destroys other microorganisms by collapsing their H^+ , Na^+ , and K^+ gradients.

More detailed discussion of various MB transport methods will occur in later sections.

2 Nucleus(envelope and matrix):

2.1 Envelope

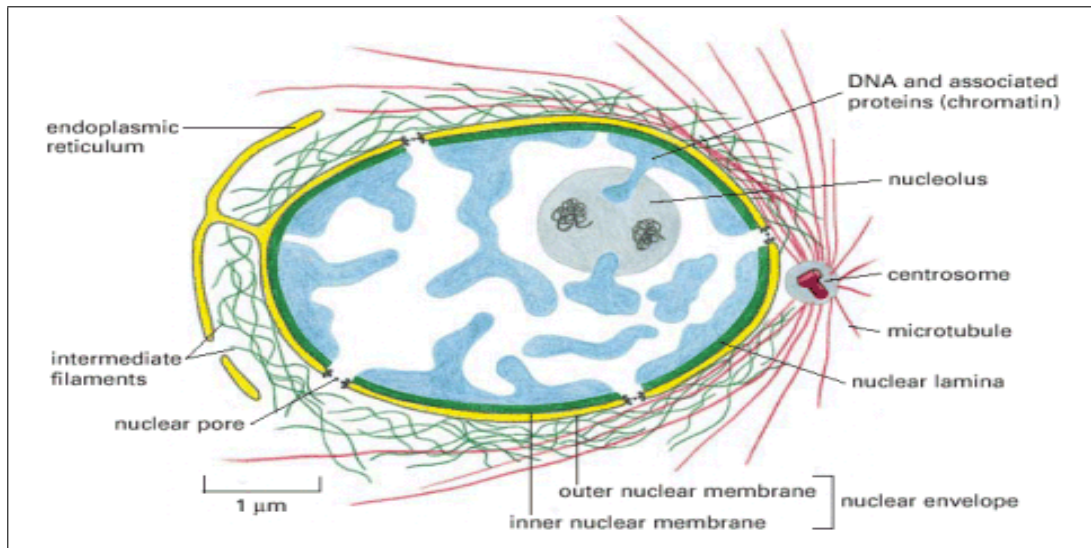


Figure 5: Cross-sectional view of typical cell nucleus

Figure 5 provides a clear representation of a typical animal cell nucleus. The nuclear envelope is a double membraned structure which defines the nuclear compartment. This dual-membrane structure is penetrated by nuclear pore complexes which allow for the transport of items into and out of the nucleus.

The double membrane consists of an inner and outer membrane. The inner membrane defines the nucleus itself and in many cells the outer membrane is continuous with the lumen of the rough endoplasmic reticulum. The outer membrane, like the ER, can be studded with ribosomes engaged in protein synthesis. These proteins are then transported into the space between the inner and outer membranes, which is continuous with the ER lumen allowing for packaging processes to continue.

The envelope is critical to the nucleus as bidirectional traffic is constantly occurring between the cytosol and the nucleus. The rapid exchange of proteins necessary for nucleus function as well as genetic material synthesized within the nucleus are important processes to be understood.

2.1.1 Nuclear Pores

quite frankly i'm not sure how much detail on pores we'll need to know but hopefully this covers the specifics well enough for the test...i wonder if nuclei can get acne from clogged pores as well =P

Looking at Figure 6a one can see that the inner AND outer membranes are crossed by nuclear pore complexes. Typically in animals, nuclear pore complexes are composed of more than 50 different proteins called **nucleoporins**. The subunits are broken up into four major structural groups: column subunits, which form the bulk of the pore wall; annular subunits, which extend into the pore like spokes; luminal subunits, that anchor the pore; and ring subunits which form the nuclear and cytosolic faces of the complex. In addition to

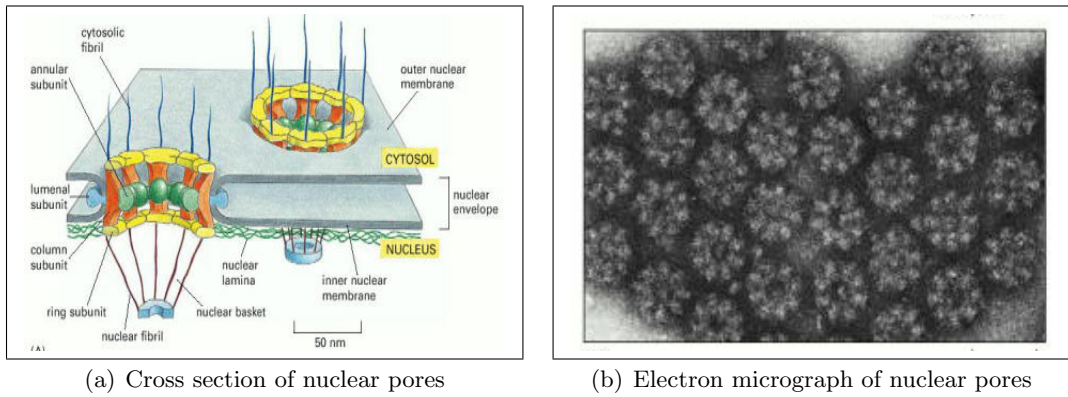


Figure 6: Various views of nuclear pores

these components fibrils protrude from each face (Figure 6a). Pore complexes are arranged in an octagonal symmetry seen quite clearly in Figure 6b.

A good rule of thumb when observing a nucleus for the level of transcriptional activity is the absence/presence of nuclear pore complexes. The greater the number the greater the rate. A typical mammalian cell contains 3000-4000 pore complexes. When a cell is synthesizing DNA it requires approximately 10^6 histones every three minutes. This translates to a rate of transport of approximately 100 histones per complex per minute. Considering that histones are not the only molecules traveling in and out of nuclear pores, it can be deduced that the rate is higher when looking at total transport.

The general structure of each pore is simple, it consists of one or more aqueous channels through which small water-soluble molecules can passively diffuse. Estimated diffusion levels for the pores are as follows:

| Table 1: Estimated diffusion levels | |
|-------------------------------------|--------------------|
| < 5000 Da | Freely permeable |
| 17,000 Da | 2 minutes |
| > 60,000 Da | Almost impermeable |

These data suggests that the pore diameter is approximately 9nm and 15nm long, which is a small percentage of the overall pore size. The size of the channel and the inability of large molecular weight proteins to pass through the pores makes sense as it excludes protein synthesis to the cytosol.

This data obviously precludes the transport of large proteins across the nuclear membrane, thus the question is how do DNA/RNA polymerases make their way into the nucleus post translation? The answer is through the use of receptor proteins which ferry molecules through nuclear pore complexes.

2.1.2 Nuclear Transport

Proteins which are destined for the nucleus after translation possess a **nuclear localization signal(NLS)**. This signal takes the form of one or two short sequences rich in positively charged amino acids K and R (Some nuclear proteins contain different signals, some of which have not been characterized). The signal can be located almost anywhere in the amino acid sequence of the protein, however they are thought to form loops or patches on the protein surface. Many proteins exhibit a functional NLS even when it is linked to K

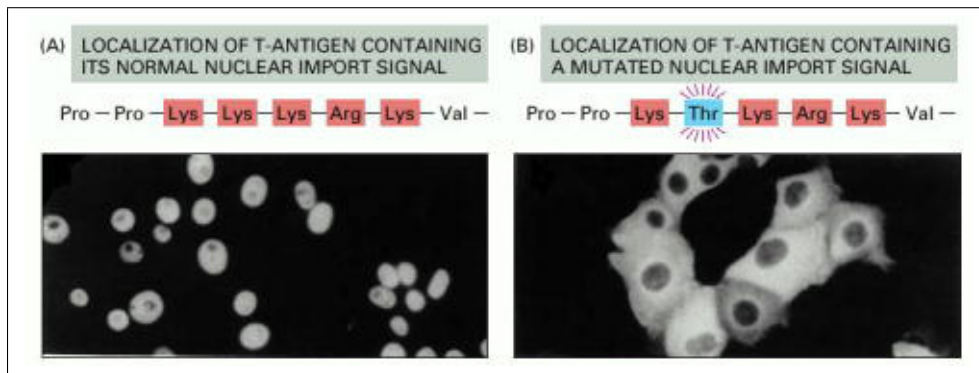


Figure 7: The NLS and it's effect on protein trafficking

side chains on the surface of cytosolic proteins, adding more support to the idea that signal location is unimportant in signal recognition. Experimental evidence points clearly to the K and R rich signal sequence as seen in the experiment shown in Figure 7. Immunofluorescence micrographs clearly show the location of SV40 virus T-antigen containing/lacking a complete NLS.

One important thing to remember with regards to nuclear transport is that it differs in comparison to other protein transfer mechanisms. This is true for two reasons. First, transfer occurs through a "large" aqueous pore rather than a transport protein. This then allows for the second difference which is that proteins can pass through in their fully folded functional conformation. In addition, newly formed ribosomes can be transported out as assembled particles. For very large complexes however it appears that some constriction occurs upon exit through the pore, so some restructuring can occur in certain situations.

Nuclear import is initiated through recognition of a NLS by **nuclear import receptors**, which are encoded by a family of genes. Each family member encodes a receptor protein which specializes in the transport of a specific group of nuclear proteins sharing similar NLSs. Import receptors not only bind to the NLS, but to nucleoporins as well. Fibrils from nuclear pore complexes as well as the nucleoporins themselves often contain numerous FG repeats which act as binding sites for the import receptors. The FG repeats are thought to line the path through the pore complex and allow the import protein to repeatedly bind, dissociate, and then rebind to adjacent repeat sequences. Once in the nucleus the import protein dissociates, returns to the cytosol, and begins the process anew. Sometimes import receptors do not bind cargo directly, additional adaptor proteins are sometimes used allowing for a broader repertoire of NLSs.

Nuclear export of large molecules also occurs through nuclear pore complexes and works just like nuclear import except in reverse. The system relies on a nuclear export signal (NES) and nuclear export receptors. Export and import receptors are not surprisingly encoded by the same gene family.

By importing specific proteins into the nucleus order is increased within the cell, therefore energy must be utilized to move proteins. The energy is thought to be provided by the hydrolysis of GTP by a monomeric GTPase known as **Ran**. Ran is found in both the cytosol and nucleus, playing a critical role in both nuclear import and export.

As a GTPase, Ran is a molecular switch that can exist in two states. Conversion between the two states is handled by two Ran-specific regulatory proteins. A cytosolic **GTPase-activating protein** (GAP) triggers GTP hydrolysis converting **Ran-GTP** \rightarrow **Ran-GDP** and a **nuclear Guanine exchange factor** (GEF) promotes exchange of GDP for GTP in the nucleus, converting Ran-GDP \rightarrow Ran-GTP. Because **Ran-GAP**

is located in the cytosol and **Ran-GEF** is located in the nucleus, the cytosol primarily contains Ran-GDP and the nucleus primarily contains Ran-GTP (Figure 8). This gradient drives nuclear transport in the appropriate direction.

Binding of the nuclear import receptors to FG-repeats on the cytosolic side of the nuclear pore complex occurs only when cargo is loaded to the receptor. The receptor with bound cargo then move along the tracks of FG-repeats into the nucleus where Ran-GTP binding causes the receptor to release its cargo. Having released its cargo the Ran-GTP and the empty receptor leave for the cytosol where Ran-GAP and **Ran Binding Protein** (RBP) collaborate to convert Ran-GTP \rightarrow Ran-GDP. Ran-GTP is first kicked off the import receptor by RBP, then Ran-GAP triggers GTP hydrolysis. Ran-GDP dissociates from the RBP and is reimported back into the nucleus completing the cycle.

Export occurs similarly with a few small differences. Ran-GTP in the nucleus promotes cargo binding. Once cargo is bound the receptor with Ran-GTP moves along the pore complex to the cytosol where once again Ran-GAP and RBP bind to Ran-GTP and hydrolyze its GTP. The export receptor then releases its cargo and returns to the nucleus.

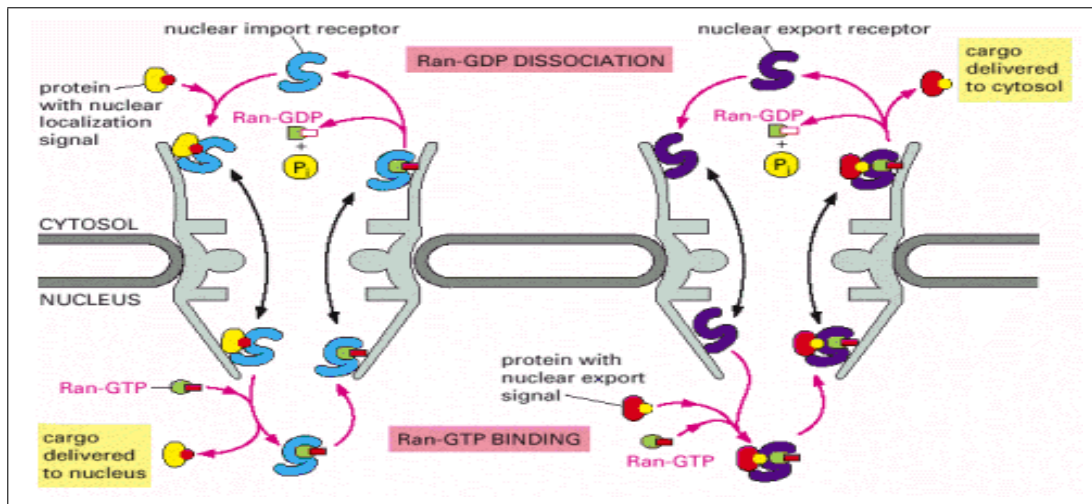


Figure 8: Ran gradient controlled nuclear transport

2.1.3 Nuclear Lamina and Matrix

Fibrous proteins called **lamins** form a 2D network along the inner surface of the inner nuclear MB providing it with shape as seen in Figure 3. Apparently lamins also bind DNA providing scaffolding for chromatin within the nucleus. Lamins breakdown early in the cell cycle, which is a critical step in cell division.

In addition to the fibrous lamins, the nucleus contains a complex network of protein and RNA fibrils known as the nuclear matrix. This matrix permeates the entire nucleus and includes regions which are defined as the chromosome scaffold. While the presence of this complex matrix may suggest a rigid internal structure, that is simply not the case as proteins can be seen traversing the nucleus. This allows for histone linking and delinking within the genome making chromatin a dynamic structure.

3 Mitochondria and Chloroplasts:

3.1 Evolution

Mitochondria and (in plants) chloroplasts generate a majority of the ATP used by cells to drive reactions. These two organelles serve very unique defined roles within the cells and their generation within the cell has been carefully studied. One evolutionary scheme known as the **endosymbiont hypothesis** is thought to have generated mitochondria/chloroplasts in animal and plant cells respectively. The hypothesis states that Mitochondria and plastids (chloroplasts are plastids) originated when a bacterium was engulfed by a larger pre-eukaryotic cell. They retained their autonomy and lived symbiotically with the larger cell over time becoming modern day organelles. This theory is supported by a few facts, first that plastids and mitochondria contain genomic material and proteins which closely resemble present-day bacteria. Second of all each organelle fits with the theory in that they contain a double membrane consistent with the idea of an endocytic origin as well as topologically equivalent portions (the inner membrane = original bacterial membrane, lumen evolved from bacterial cytosol). In addition, the organelles remain isolated from each other as well as the extensive vesicular trafficking network within the cell. This theory is outlined in figure 9 below.

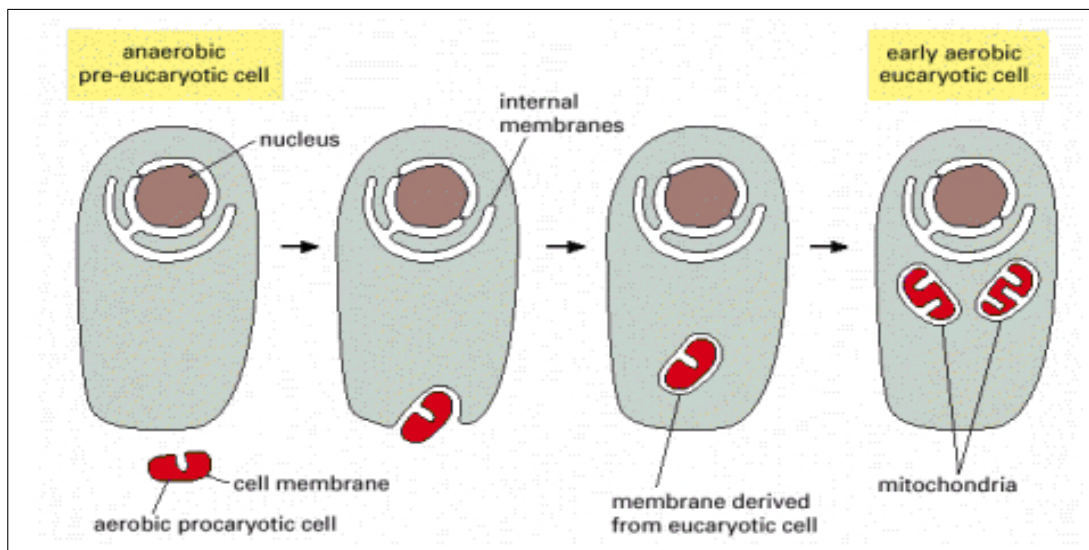


Figure 9: Evolutionary origin of mitochondria/plastids

For chloroplasts the endocytosis of the ancient plastid bacteria was not the final step. Thylakoid vesicles developed from proplastids in green leaves according to the needs of the cell. Membrane patches form and pinch off during differentiation resulting in the construction of a chloroplasts.

It is thought that mitochondria came from a photosynthetic bacteria which lost its photosynthetic abilities living it with only a respiratory pathway, however it is uncertain how many endocytic events resulted in the current population of mitochondria.

3.2 Biogenesis

All mitochondria and plastids contain their own genome, however each genome only encodes for a small percentage of the proteins necessary for their respective functions. This suggests that the majority of proteins must be encoded in the nucleus. Thus, many organelle

proteins are produced in the cytosol to be imported by the organelle for use. The organelle genome then encodes the remainder of the necessary proteins within the organelle. No processing of organelle produced proteins occurs within the cytosol, thus protein transit is unidirectional (One exception is mitochondria and their release intermembrane space proteins such as cytochrome C as a trigger for apoptosis). An interesting fact is that protein synthesis machinery within mitochondria and chloroplasts still resemble bacterial machinery. The resemblance is particularly close in chloroplasts. Thus the function of both are susceptible to antibiotics. The most important thing to remember about biogenesis however is that NO de novo organelles are produced. The organelles grow to a size sufficient for fission to produce two functional organelles.

Part III

Cell Surface and Communication

4 Extracellular Matrix:

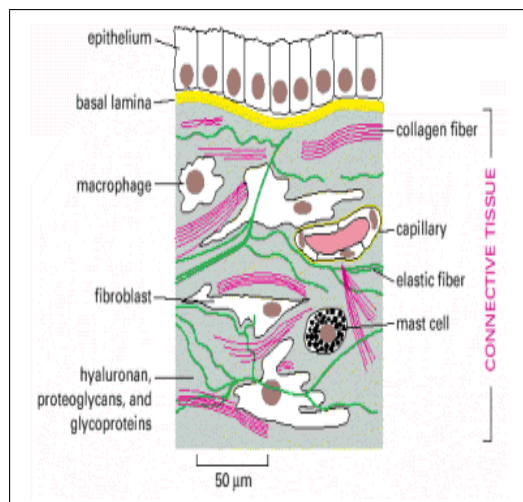


Figure 10: Example of Extracellular Matrix

A significant portion of tissue volume is made up of extracellular space, which is filled by an intricate network of molecules called the **extracellular matrix (ECM)**. The ECM is composed of a variety of polysaccharides and proteins which are secreted locally and then organized into a meshwork associated with the surface of the cell that produced it. While the Fat Alberts text focuses on the ECM of vertebrates, ECM can be found in most multicellular organisms including the cuticles of worms and insects, shells of mollusks, and the cell walls of plants. ECM is produced locally and the cells within it can control the orientation of the matrix outside. In most connective tissues the ECM is secreted by fibroblasts. Two main classes of macromolecules make up the ECM:

1. **Polysaccharide chains of glycosaminoglycans (GAGs)** which are typically found covalently linked to protein in the form of **proteoglycans** (Figure 11).
2. **Fibrous proteins** - e.g. Collagen, elastin, fibronectin, and laminin which all have structural and adhesive functions.

Overall these two classes of molecules serve a few general functions. The **proteoglycan** molecules form a highly hydrated gel-like "ground substance" in which the fibrous proteins are embedded. The polysaccharide gel resists compressive forces on the matrix while permitting the rapid diffusion of nutrients, metabolites, and hormones between the blood and the tissue cells. The collagen fibers both strengthen and help organize the matrix, and rubber like elastin fibers give it resilience. Finally many matrix proteins help cells attach in their appropriate locations.

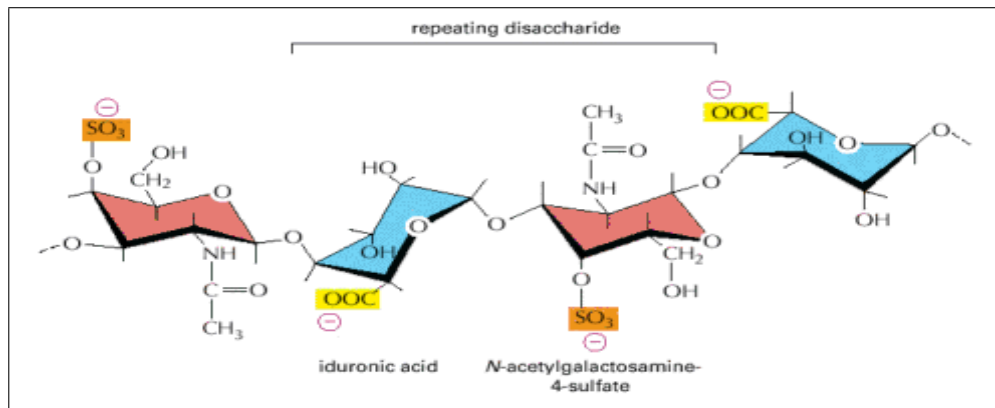


Figure 11: Repeating disaccharide sequence of GAGs

4.1 GAGs

There are four main groups of GAGs:

1. Hyaluronan - **does not covalently attach to protein.**
2. Chondroitin Sulfate and dermatan sulfate
3. Heparan sulfate
4. Keratan sulfate

GAGs are too stiff to fold up into compact globular structures like most polypeptide chains, thus they occupy a huge volume relative to their mass. Their highly negative charge pulls in a cloud of cations, most notably Na^+ , which results large amounts of water to be sucked into the matrix. This creates turgor pressure which provides the ECM with its gel like quality. Note that these four groups are prominent in animals, while plants contain different polysaccharides which dominate the ECM.

Proteoglycans are made in most animal cells, they are mainly assembled in the Golgi apparatus. They are easily distinguished from glycoproteins as at least one of the sugar side chains must be a GAG. In addition, proteoglycans can be enormous, with long unbranched GAG chains around 80 sugars long. There are no real classifications of proteoglycans as the heterogeneity of their populations are virtually limitless due to the number of possible modifications made to the core protein during their production. The great abundance and diversity of proteoglycans translates to a large functional repertoire. They can form gels of varying pore size and charge density (1), as well as playing a role in chemical signalling(2). In addition, they can bind and regulate the activities of secreted proteins (3). Examples of each of these functions follow:

1. The heparan sulfate proteoglycan (perlecan) produces a very fine gel, which serves as a filter in the kidney glomerulus for molecules passing into the urine from the bloodstream.
2. Heparan sulfate chains bind to fibroblast growth factors, which can stimulate a variety of cell types to proliferate (Figure 12). This interaction oligomerizes the growth factor molecules allowing them to cross-link their cell-surface receptors. While in most cases signal molecules bind the GAG portion of proteoglycans, it is possible for them to interact with the core protein as well. Some members of the transforming

growth factor β family bind to the core protein disabling their activity.

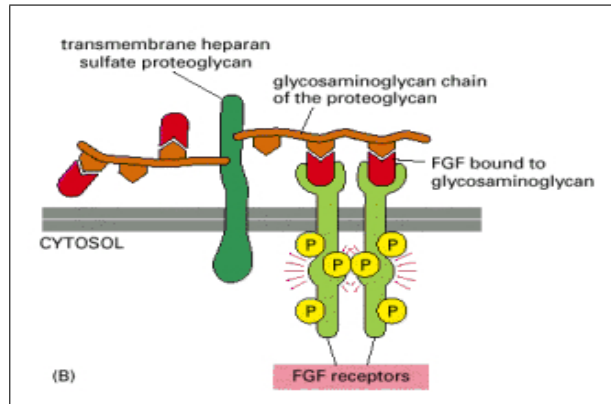


Figure 12: Proteoglycans aiding in signalling

3. Proteoglycans can act on secreted proteins in 5 ways.

- Immobilization of the protein in order to restrict its range of action.
- Sterically block the activity of the protein.
- Provide a reservoir of protein for delayed release.
- Protect the protein from proteolytic degradation, prolonging its action.
- Alter or concentrate the protein for more effective presentation to cell-surface receptors (e.g. immobilization of chemokines in blood vessels at inflammatory sites).

4.2 Fibrous Proteins

4.2.1 Collagens

Collagens are a family of fibrous proteins found in all multicellular animals. They are secreted by connective tissue cells, as well as by a variety of other cells. They are a major component of skin and bone, as well as the most abundant proteins in mammals (approximately 25% of total protein mass). The primary feature of collagen is its long stiff triple-stranded helical structure. Three collagen polypeptide chains are wound one around another in a ropelike superhelix (Figure 13). Collagens are extremely rich in proline and glycine. Proline stabilizes the helical conformation in each α chain with its ring structure. Glycine, being the smallest amino acid is spaced at every third residue allowing three α chains to pack tightly together.

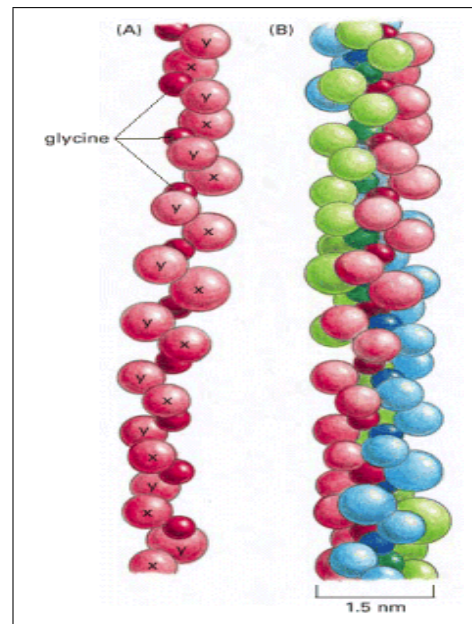


Figure 13: Triple helical structure of collagen

So far 25 distinct α chains have been identified, each coded by a separate gene, this translates to possibly more than 10,000 types of triple-stranded collagen. Only 20 superhelixes have been discovered. The main types of collagen found in connective tissue are types I, II, III, V, and XI (Table 19-5 has a larger list and is available in Fat Alberts).

Type I, II, III, V, and XI - the principle collagens of skin, bone, tendons, ligaments, etc, therefore the most common. Type I's are known as fibrillar collagens, they are secreted and form higher-order collagen fibrils. Collagen fibrils often aggregate into larger cablelike bundles which can be seen under light microscopes known as collage fibers.

Type IX and XII - called fibril-associated collagens. Decorate the surface of collagen fibrils, thought to link fibrils to one another and to other components in the ECM.

Type IV and VII - called network-forming collagens.

Type IV - assemble into feltlike sheets which are a major part of mature basal laminae.

Type VII - form dimers that assemble into specialized structures called anchoring fibrils, which help attach the basal lamina of multilayered epithelia to connective tissue. Very abundant in skin.

4.2.2 Collagen Fiber Production

Individual collagen polypeptide chains are synthesized on membrane bound ribosomes and translocated directly into the lumen of the ER as larger precursors called pro- α chains. These precursors not only have a short amino terminal ER localization signal peptide but additional propeptides as well at the N and C terminal ends. In the lumen of the ER certain prolines and lysines are hydroxylated to form hydroxyproline and hydroxylysine, some of the hydroxylysines are glycosylated. Each pro- α chain then combines with two others to form a triple-stranded molecule known as procollagen.

It is thought that hydroxylysines and hydroxyprolines help to stabilize the triple-stranded helix, evidence to support this lies in the disease Scurvy, which is a deficiency of ascorbic acid (vitamin C) necessary for proline hydroxylation. This results in a lack of stable procollagen, leading to a gradual loss of normal collagen in the ECM. Blood vessels and gums appear to be the most effected by this deficiency as they become very fragile in those suffering from Scurvy. Turnover for most collagen molecules however is thought to be very slow, in bone for example molecules can persist for 10 years before being degraded and replaced.

After secretion in procollagen form, the propeptides are removed by proteolytic enzymes outside the cell. The propeptides have at least two functions. First, they guide the intracellular formation of procollagen. Second, because they are only removed after secretion they prevent the formation of large collagen fibrils within the cell. After the propeptides are removed, the collagen molecules rapidly self-assemble close to the cell surface often in deep infoldings of the membrane. A large portion of the strength of collagen fibers is provided by the formation of covalent cross-links between K residues of the constituent collagen molecules. Collagen production is summarized in figure 14 below.

4.2.3 Fibril-associated Collagens

While GAGs resist compressive forces, **collagen fibrils** form structures which resist tensile forces. The varying tensile resistance is achieved through varying arrangements of collagen

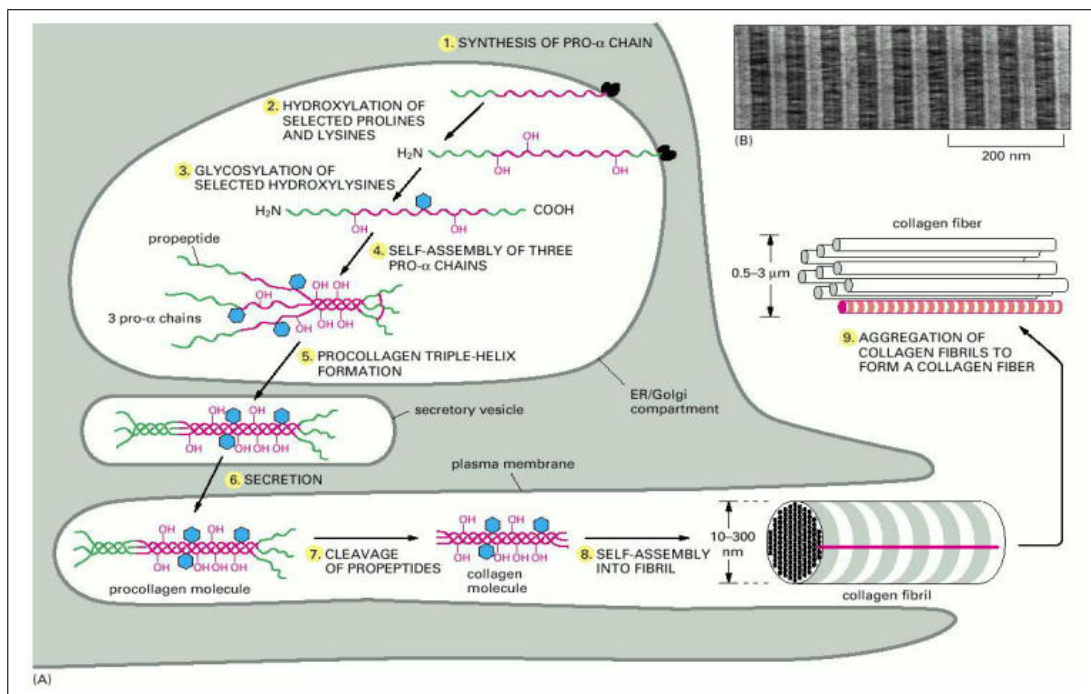


Figure 14: Formation of collagen fibers

fibrils. These arrangements are influenced by the secretion and presence of Fibril-associated collagens (Types IX and XII). Fibril-associated collagens differ from fibrillar collagens in several ways:

1. Their triple-stranded helical structure is interrupted by short nonhelical domains, which increases flexibility
2. They are NOT cleaved after secretion
3. They DO NOT aggregate to form fibrils, rather they bind periodically to the surface of fibrils to mediate the interactions of collagen fibrils.

4.2.4 Elastin

Tissues must be capable of surviving the recoil after a transient stretch, this is achieved through the presence of elastic fibers composed of a highly hydrophobic protein called **elastin** (approximately 750 aa long). Like collagen it is rich in proline and glycine, however it is not glycosylated and contains NO hydroxylysine. Soluble tropoelastin is secreted into the extracellular space and assembled into elastic fibers. After secretion the tropoelastin becomes highly cross-linked with other tropoelastin fibers. This cross-linking occurs through lysine residues as it did with collagen. Currently there is still controversy over the structure of elastin fibers and how the structure accounts for their rubberlike properties. One theory is a "random coil" conformation depicted below (Figure 15).

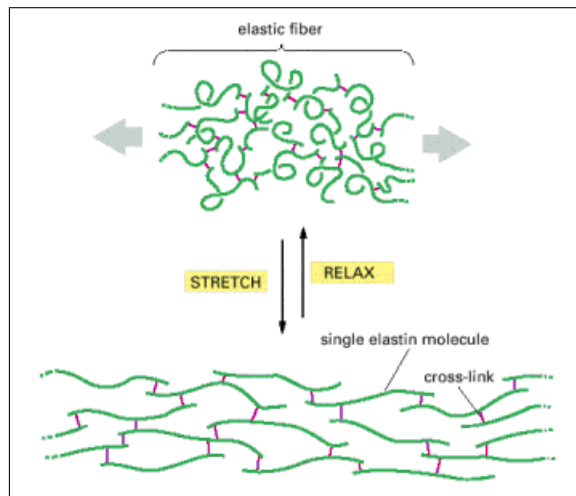


Figure 15: Random coil conformation of elastin fibers

4.2.5 Fibronectin

Fibronectin is a large glycoprotein found in all vertebrates which is important for cell adhesion to the ECM. It has multiple domains with multiple specific binding sites for other matrix macromolecules as well as for cell surface receptors. Fibronectin is a dimer composed of two large subunits joined by disulfide bonds (Figure 16).

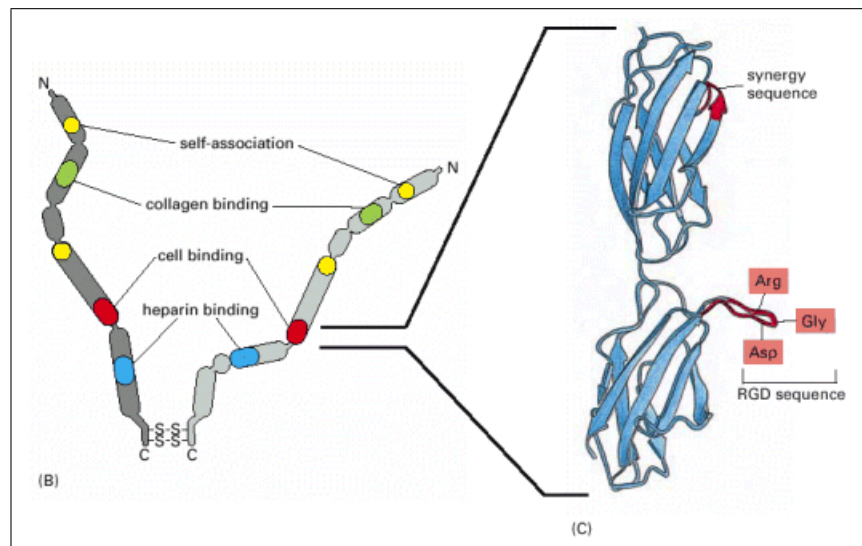


Figure 16: Structure of Fibronectin and its cell binding domain

Each domain seen in figure 16 consist of smaller modules which are serially repeated. The main type of module is the type III fibronectin repeat, which binds to integrins. It is the most common of all protein domains in vertebrates. A central feature of this type III repeat is the Arg-Gly-Asp (RGD) sequence within the binding site. This repeat allows for binding with integrins, however this is merely necessary for binding, but not sufficient for tight binding.

Fibronectin fibrils that form near the surface of fibroblasts are usually aligned with adjacent intracellular actin stress fibers. Actin filaments promote the assembly of secreted

fibronectin molecules into fibrils and influence their orientation. Interactions between intracellular actin and fibronectin are mediated by integrin transmembrane adhesion proteins.

Another function of fibronectin is believed to be guidance of cellular migration in vertebrate embryos. Other glycoproteins in the matrix are also capable of this function.

4.2.6 Cell Mobility Within the ECM

Matrix components are degraded by extracellular proteolytic enzymes, allowing cells to migrate through the vacated region. Most of these enzymes are metalloproteases, which depend on Ca^{2+} or Zn^{2+} . Serine proteases also play a role. Three mechanisms control the degradation of matrix components:

Local activation - proteases are secreted as inactive precursors that can be activated locally when needed (e.g. plasminogen, which is cleaved locally by other proteases to yield plasmin, which breaks blood clots).

Confinement by cell-surface receptors - cells can have receptors that bind proteases thereby confining the enzyme to the sites where it's needed (e.g. urokinase-type plasminogen activator (uPA) is bound to receptors on the growing tips of axons and the leading edge of some migrating cells, clearing a directed path for migration).

Secretion of inhibitors - protease action is confined by the secretion of inhibitors (e.g. Tissue inhibitors of metalloproteases (TIMPs) which are protease-specific and bind tightly blocking activity, possibly secreted at the margins of areas of active protease activity).

4.3 Cell Walls

The plant cell wall is an elaborate ECM that encloses each cell in a plant. Plant cell walls are generally thicker, stronger, and more rigid than the ECM produced by animal cells.

All cell walls in plants have their origins in dividing cells, typically these new cells are produced in special regions known as meristems. Their walls initially to accommodate growth are thin and known as primary cell walls. Later once growth stops, the primary wall is retained without alteration, however a rigid secondary cell wall is deposited inside the old cell wall. While all cell walls of higher plants differ in both composition and organization, they are constructed like animal ECM, using two components, one that provides tensile strength, the other resistance to compression.

In plant cell walls the tensile fibers are made from cellulose, tightly linked into a network with cross-linking glycans. In primary cell walls, the matrix in which the cellulose network is embedded is made up of pectin (a polysaccharide rich in galacturonic acid).

Cellulose provides tensile strength to the primary cell wall. Each molecule consists of at least 500 glucose residues forming a ribbonlike structure stabilized by hydrogen bonds. Intermolecular hydrogen bonds between adjacent cellulose molecules cause them to adhere strongly to one another in overlapping parallel arrays forming large bundles called cellulose microfibrils, which have a tensile strength comparable to steel. Microfibrils are arranged in layers or **lamellae**, connected by cross-linking glycans (Figure 17).

Cross-linking glycans functionally act the same as fibril-associated collagens discussed earlier. Along the network of microfibrils a network of **pectins** forms as seen in figure 17. Pectins are a heterogenous group of branched polysaccharides that contain numerous negatively charged galacturonic acid units. The negative charge produces the same effect that negatively charged GAGs produce creating a highly hydrated semirigid gel (pectin is

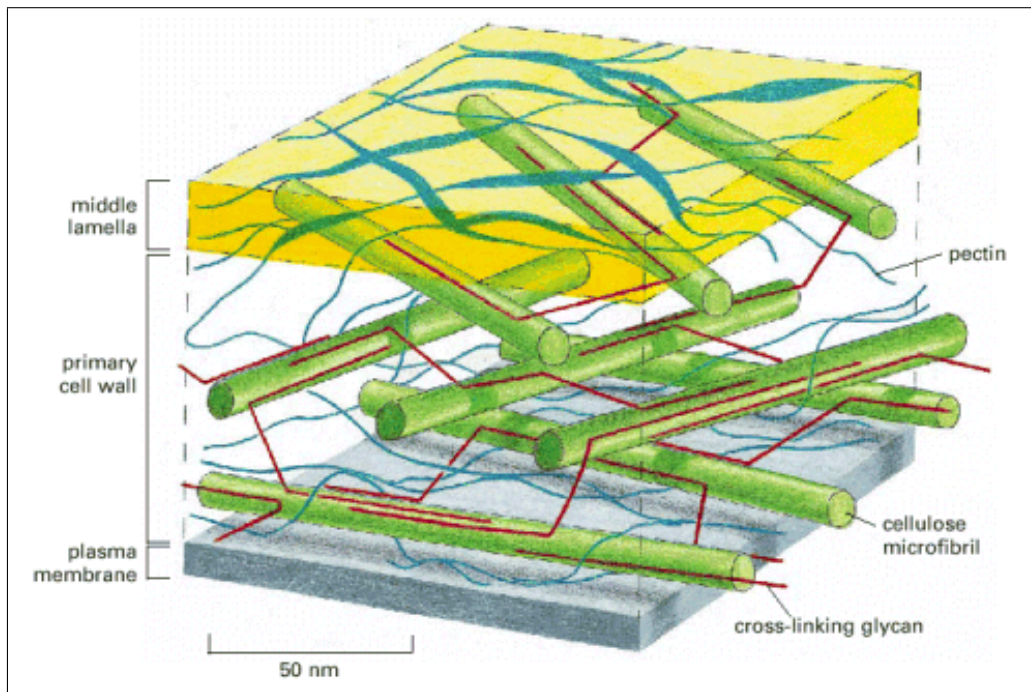


Figure 17: Plywood arrangement of cellulose microfibrils in primary cell wall.

added to fruit juice to make jelly). Some pectins are particularly abundant between the walls of adjacent cells.

In addition to the polysaccharide-based networks proteins can contribute up to about 5% of the cell wall's dry mass. Many of the proteins are enzymes responsible for wall turnover and remodelling. One class of wall proteins contains high levels of hydroxyproline as in collagen. These proteins are thought to strengthen the wall and are produced in increased amounts as a response to attack by pathogens.

5 Cell Adhesion and Junctions:

5.1 Cell Junctions

Cells in tissues are linked to one another and to the ECM through specialized contact sites known as cell junctions. Cell junctions fall into three functional classes:

Occluding Junctions - seal cells together in an epithelium in a way that prevents small molecules from leaking from one side to the other.

Anchoring Junctions - mechanically attach cells and their cytoskeletons to their neighbors/ECM.

Communicating Junctions - mediate the passage of chemical/electrical signals between cells.

5.1.1 Occluding Junctions

All epithelia serve at least one function in common which is to be a selective barrier separating fluids on either side that have different compositions. This requires that adjacent cells be sealed together by occluding junctions. **Tight junctions** have this role in vertebrates. An example of this role is shown in Figure 18.

While tight junctions provide a seal which allows a maintenance of directional transport within epithelial cells, tight junctions are not impenetrable. They are 10,000 times more permeable to inorganic ions such as Na^+ in certain situations. These differences are reflected in the proteins that make up the varying tight junctions throughout the body.

Each tight junction is made up of numerous sealing strands that completely circle the apical end of each cell in the epithelial sheet. Each sealing strand is composed of a long row of transmembrane adhesion proteins embedded in each of the two interacting plasma membranes. The extracellular domains of these proteins join directly to one another to occlude the intercellular space as seen in Figure 19. The major transmembrane proteins

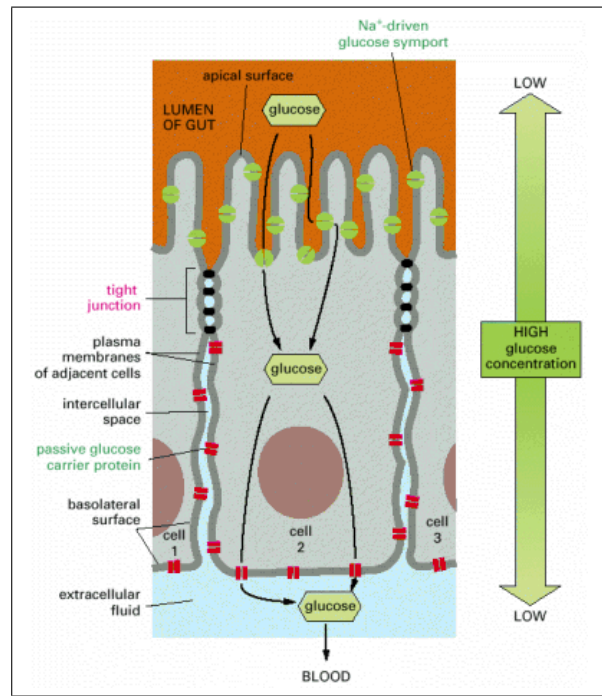


Figure 18: Epithelial cells and the role of tight junctions.

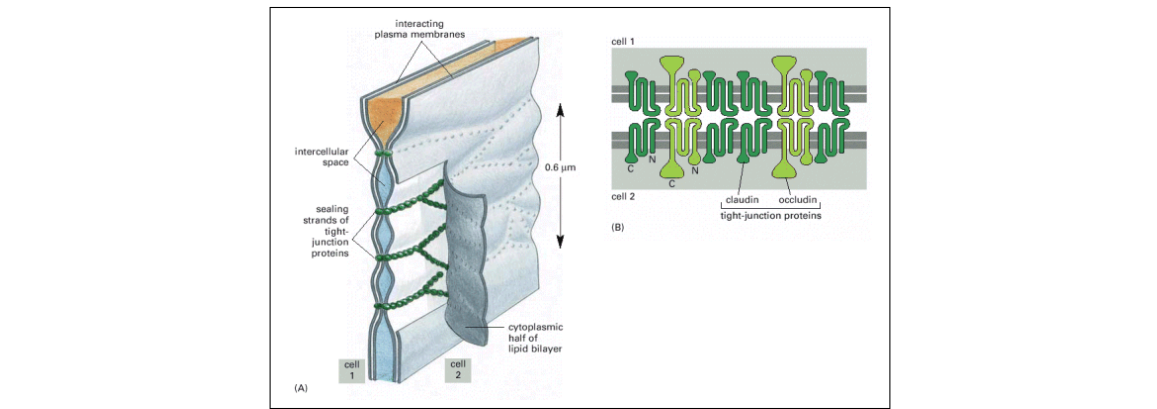


Figure 19: More detailed view of a tight junction.

in a tight junction are **claudins**. A second major transmembrane protein is **occludin**, whose function is still uncertain. Claudins and occludins associate with intracellular ZO proteins which anchor the strands to the actin cytoskeleton.

5.1.2 Anchoring Junctions

Ultimately the lipid bilayer is flimsy and unable to transmit large forces from cell to cell or from cell to ECM. Anchoring junctions solve this problem by forming a membrane spanning structure that is tethered inside the cell to the tension bearing filaments on the cytoskeleton. They are widely distributed in animal tissues and are most abundant in tissues subjected to severe mechanical stress (e.g. heart, muscle, and epidermis).

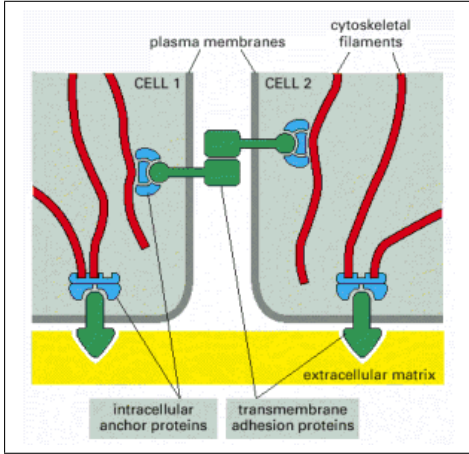


Figure 20: Construction of an anchoring junction.

Two main classes of protein make up anchoring junctions, intracellular anchor proteins and transmembrane adhesion proteins. Intracellular anchor proteins form a distinct plaque on the cytoplasmic face of the bilayer and connect the junction to either actin/intermediate filaments. Transmembrane adhesion proteins have a cytoplasmic tail which binds to one or more of the intracellular anchor proteins and an extracellular domain which interacts with either the ECM or the transmembrane adhesion proteins of another cell (Figure 20). In addition to these two classes, many anchoring junctions contain intracellular signaling proteins, enabling junctions to transmit signals.

The following table displays the different anchoring junctions, the transmembrane adhesion protein used, and the cytoskeletal component to which they bind.

Table 2: Adhesion Junctions their role and components.

| Type of Junction | Role | Protein | Cytoskeletal Anchor |
|--------------------|-------------|-----------|------------------------|
| Adherens junctions | | | Actin |
| Desmosomes | Cell-Cell | Cadherins | Intermediate Filaments |
| Focal adhesions | | | Actin |
| Hemidesmosomes | Cell-Matrix | Integrins | Intermediate Filaments |

5.1.3 Adherens Junctions

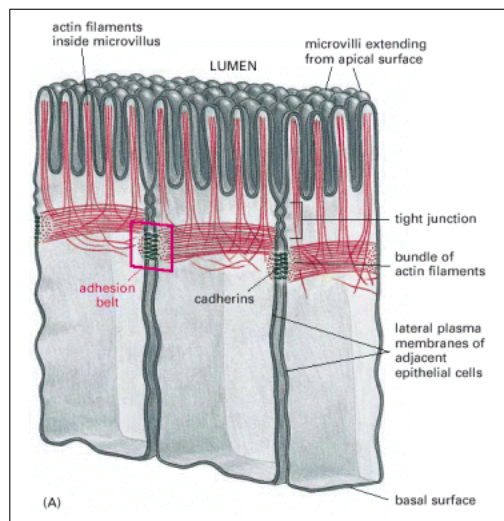


Figure 21: Adhesion belt in epithelia.

Adherens junctions occur in various forms however the typical example is in the

form of an **adhesion belt** seen in epithelia (Figure 21). These adhesion belts link actin bundles across an extensive transcellular network which can then contract with the help of myosin motor proteins. This mechanism is thought to be a fundamental process in animal morphogenesis. It also appears that formation of tight junctions can only occur after epithelial cells have formed adherens junctions. Anti-cadherin Ab that block the formation of adherens junctions block the formation of tight junctions as well.

5.1.4 Desmosomes

Desmosomes serve as rivets between cells and anchors for intermediate filaments within. Through desmosomes, intermediate filaments are able to form an extremely strong structural framework between adjacent cells of a tissue. Depending on the cell type different intermediate filaments will attach.

5.1.5 Communication Junctions

Most cells in animal tissues communicate through **gap junctions**. Each gap junction is a patch where the membranes of two adjacent cells are separated by a uniform gap of about 2-4 nm. The gap is spanned by channel forming proteins (**connexins**), which create channels allowing inorganic ions and other small water-soluble molecules to pass directly from the cytoplasm of one cell to the cytoplasm of the other. This exchange couples the cells electrically and metabolically. Experiments have suggested that the channel size is approximately 1.5 nm excluding macromolecules from exchange. Connexins are four-pass transmembrane proteins, six of which assemble to form a channel (**connexon**). When the connexons of two cells in contact are aligned, they form a continuous aqueous channel. Gap junctions in various tissues can have different properties such as permeability. This reflects the presence of numerous different connexins, each encoded by separate genes. Cells typically express more than one connexin, thus it becomes necessary for cluster numerous connexons at gap junctions in order to ensure proper channel formation.

5.2 Cell Adhesion

Cell adhesion occurs through the use of **cell adhesion molecules** (CAMs), which includes the previously discussed transmembrane adhesion proteins. CAMs can be cell-cell or cell-matrix. Some are Ca^{2+} dependent while others are Ca^{2+} independent.

Cadherins are the major CAMs responsible for Ca^{2+} dependent cell-cell adhesion in vertebrate tissues. They have been shown to be critical in cell-cell adhesion in embryonic development. While three mechanisms exist for how cadherins mediate binding (Figure 22), in animals cadherins usually link cells through the homophilic mechanism sorting out and aggregating cells both qualitatively and quantitatively. Thus it appears that both qualitative and quantitative differences in the expression of cadherins have a role in organizing tissues.

Another group of Ca^{2+} dependent adhesion molecules are **selectins**. Selectins are cell-surface carbohydrate-binding proteins that mediate a variety of interactions in the blood stream. Each selectin is a transmembrane protein with a highly conserved lectin domain that binds to a specific oligosaccharide on another cell.

Cadherins, selectins, and **integrins** are all Ca^{2+} dependent adhesion molecules (Mg^{2+} for some integrins; the molecules responsible for Ca^{2+} independent cell-cell adhesion belong mainly to the immunoglobulin (Ig) superfamily of proteins. One of the best studied examples is **neural cell adhesion molecule** (N-CAM). Like cadherins it is thought to bind through a homophilic mechanism. Some Ig-like cell-cell adhesion molecules however

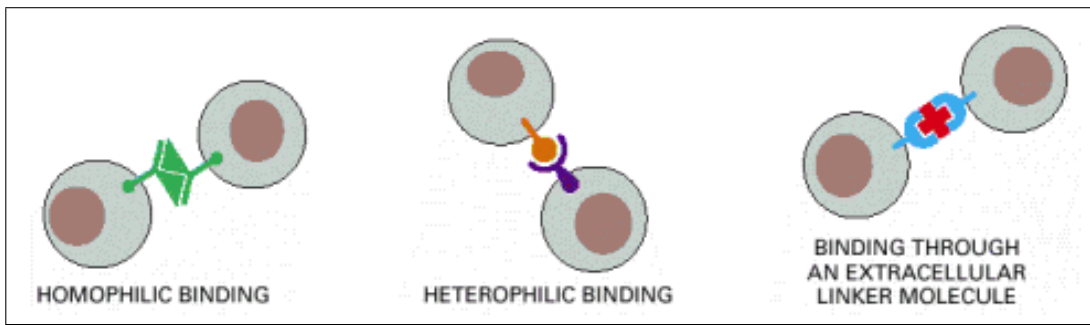


Figure 22: Three mechanisms for cadherin mediated cell-cell adhesion

use a heterophilic mechanism such as **Intercellular CAMs (ICAMs)**. While cadherin and Ig family molecules are expressed on the same cells, cadherin mediated adhesions are much stronger and are largely responsible for holding cells together. N-CAM and other members of the Ig family seem to fine tune interactions during development and regeneration.

Keep in mind that while it appears to be a pretty cut and dry system of molecule to molecule interactions a single cell type can utilize multiple mechanisms for adhering to other cells.

6 Signal Transduction and Receptor Function:

Signal transduction occurs after a water-soluble signal molecule binds to a specific receptor protein on the surface of target cells. These receptors must then "translate" the molecule's meaning to the cell, converting an extracellular ligand-binding event into signals that alter the behavior of the target cell.

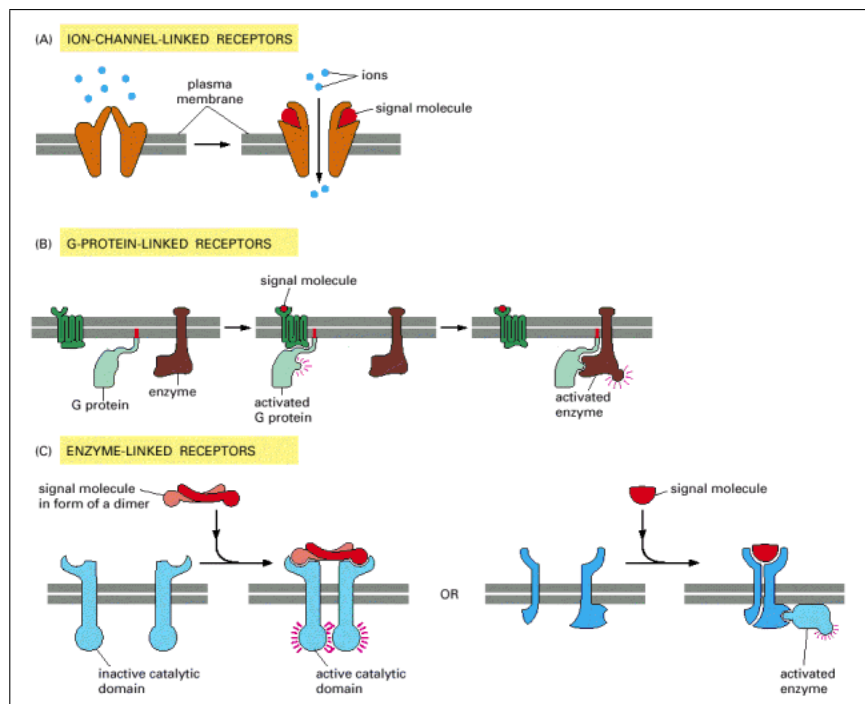


Figure 23: Three classes of cell-surface receptors.

The three largest classes of cell-surface receptor proteins are **Ion-channel-linked**, **G-**

protein-linked, and **Enzyme-linked receptors**. Receptors are placed in one of these three classes by the transduction mechanism they utilize (Figure 23). There are some cell-surface receptors that do not fit into any of the following classes.

6.1 Ion-Channel-Linked Receptors

These receptors are also known as transmitter-gated ion channels or ionotropic receptors. They are responsible for rapid synaptic signaling between electrically excitable cells. Signaling is mediated by a small number of neurotransmitters that transiently open or close an ion channel formed by the protein to which they bind (Figure 23a).

6.2 G-Protein-Linked Receptors

G-protein-linked receptors act indirectly to regulate the activity of a separate membrane bound target protein which can be an enzyme or an ion channel (Figure 23b). This interaction is mediated by a third protein, called a trimeric GTP-binding protein (**G-protein**). The activation of the target protein can result in various effects on the cell dependent on the nature of the activated protein. All G-protein-linked receptors belong to a large family of homologous seven-pass transmembrane proteins.

Due to their seven-pass structure G-protein-linked receptors are sometimes known as **serpentine receptors**. The G-proteins associated are made up of three subunits - α , β , and γ . In the unstimulated state the α subunit has GDP bound and is inactive. The general process by which G-protein-linked receptors work is as follows:

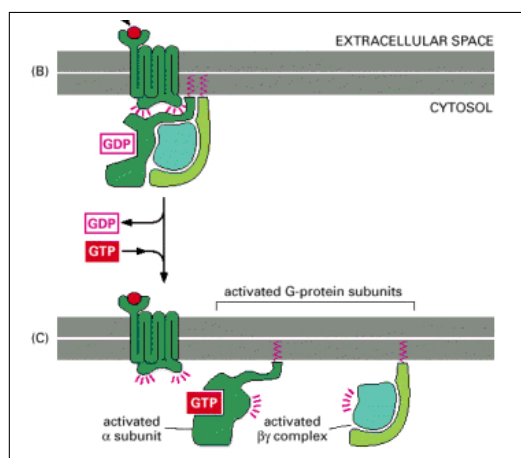


Figure 24: Activation and separation of the G-protein.

1. Ligand binding results in a conformational change that allows the receptor to activate the trimeric GTP-binding proteins associated with the receptor. While there are many G-proteins, they are all similar in structure and function the same way.
2. The activation of the G-protein occurs through a release of GDP from the α subunit and the binding of GTP in its place.
3. This exchange results in a separation of the G-protein into an α subunit and a $\beta\gamma$ complex (Figure 24).
4. The dissociation of the subunits causes the α subunit to adopt a new conformation allow it to interact with target proteins. This also reveals a face of the $\beta\gamma$ complex capable of interacting with a second set of proteins.
5. The α subunit is a GTPase and once it hydrolyzes its GTP it reassociates with the $\beta\gamma$ complex to inactivate the G protein. The time during this period is typically very short due to the interaction of the α subunit with

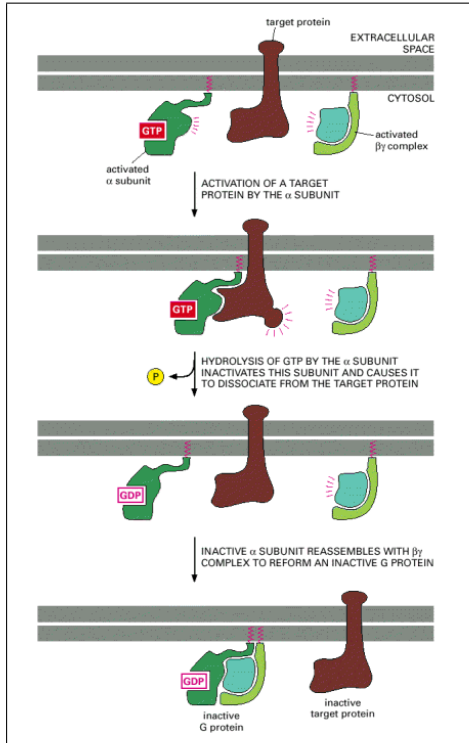


Figure 25: GTP hydrolysis and inactivation of the G-protein.

either its target protein (Figure 25) or regulator of G-protein signalling (RGS) proteins which act as GAPs.

Some G-protein pathways utilize cAMP as a signalling molecules. They do so by increasing the activity of adenylyl cylcase through a stimulator G protein (G_s). Typically the G_s is the α subunit of the G-protein, sometimes the $\beta\gamma$ complex has been known to increase or decrease the enzymes activity.

Another G-protein pathway example is the Inositol Phospholipid signaling pathway.

1. A G-protein known as G_q activates phospholipase C- β
2. Phospholipase C- β acts on phosphatidylinositol 4,5 biphosphate [PI(4,5)P₂] cleaving it into Inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol.
3. IP₃ rapidly difusses through the cytosol reaching the ER and binding to gated Ca²⁺ channels releasing Ca²⁺ into the cytosol.
4. Dephosphorylation of IP₃ to IP₂ and phosphorylation to IP₄ lead to termination of the Ca²⁺ response.
5. The diacylglycerol produced earlier activates protein kinase C (PKC), which is Ca²⁺ dependent, with the influx of Ca²⁺ from the IP₃ signal it translocates from the cytosol to the cytoplasmic face of the plasma MB. There it is activated by diacylglycerol, Ca²⁺, and phosphatidylserine.

G-protein-linked receptor desensitization depends on receptor phosphorylation. Desensitization occurs in three general ways:

1. Receptors can be altered so that they no longer interact with G-proteins (receptor inactivation).
2. Receptors can be temporarily moved to the interior of the cell so that they no longer have access to their ligand (receptor sequestration).
3. Lysosomes can destroy the receptors after internalization (receptor down-regulation).

In each of these three cases the process depends on phosphorylation of the receptor by PKA, PKC, or a member of the **G-protein-linked receptor kinases** (GRKs). GRKs specifically phosphorylate multiple serine and threonines on a receptor after it has bound ligand. Once the receptor has been modified in this manner it binds with high affinity to a member of the **arrestin** family of proteins. The bound arrestin can inactivate the receptor by preventing it from interacting with G-proteins and can also serve as an adaptor protein for **clathrin**-coated pits leading to receptor mediated endocytosis.

6.3 Enzyme-Linked Receptors

These receptors when activated function either directly as enzymes or are directly associated with enzymes that they activate (Figure 23c). They are formed by single-pass transmembrane proteins that have their ligand-binding site outside the cell and a catalytic domain inside. Enzyme-linked receptors are heterogenous in nature, however the vast majority are protein kinases or associated with protein kinases and ligand binding results in phosphorylation of specific sets of proteins in the target cell.

So far six classes of enzyme-linked receptors have thus far been identified:

1. Receptor tyrosine kinases phosphorylate specific tyrosines on a small set of intracellular signaling proteins.
2. Tyrosine-kinase-associated receptors associate with intracellular proteins that have tyrosine kinase activity.
3. Receptorlike tyrosine phosphatases remove phosphate groups from tyrosines of specific intracellular signaling proteins (called receptorlike because their ligands have not been identified).
4. Receptor serine/threonine kinases phosphorylate specific serines or threonines on associated latent gene regulatory proteins.
5. Receptor guanylyl cyclases directly catalyze the production of cyclic GMP in the cytosol.
6. Histidine-Kinase-associated receptors activate a "two-component" signaling pathway in which the kinases phosphorylates itself on histidine and then immediately transfers the phosphate to a second intracellular signaling protein.

6.3.1 Receptor Tyrosine Kinases

These receptors are one of the two most numerous receptors within cells. Typically ligand binding to two or more receptor tyrosine kinases induces the receptors to cross-phosphorylate their cytoplasmic domains on multiple tyrosines (Remember that oligomerization is necessary for enzyme-linked receptor activation). The autophosphorylation activates the kinases as well as producing a set of phosphorylated tyrosines that then serve as docking sites for a set of intracellular signaling proteins, which bind via their **SH2** (Src homology region) or **PTB** (phosphotyrosine-binding) domains. The receptors for insulin and IGF-1 act in a slightly different manner. While typically the receptor's phosphorylated tyrosines serve as docking sites, the receptors for insulin and IGF-1 phosphorylate a specialized docking protein called **insulin receptor substrate-1** (IRS-1) which creates more docking sites than would have been possible on the receptors alone.

A huge variety of proteins can dock to the phosphorylated tyrosines, some are enzymes such as **phospholipase C- γ** (PLC- γ) which activates the inositol phospholipid pathway discussed earlier resulting in higher Ca^{2+} levels. Often the proteins which bind are merely relay molecules which phosphorylate other molecules like Src which will be seen in an example. Some of the docked proteins serve as adaptors (e.g. SHC, Grb-2) to couple the receptors to the small GTPase, **Ras**, which in turn can activate cascades resulting in various results. Not all proteins which bind to the docking sites however result in a relay of the signal. Some decrease the signaling process, such as the c-Cbl protein which docks and catalyzes the receptor's conjugation with ubiquitin leading to internalization and degradation of the receptors (receptor down-regulation).

In summary the general receptor tyrosine kinase pathway is as follows:

1. Ligand binds to two or more receptors resulting in oligomerization of the receptors. There are various methods for this to occur such as multimer ligands as well as monomer ligands which pull receptors together.
2. Receptors autophosphorylate each other's tyrosines resulting in a further activation of their kinase activity as well as the production of numerous phosphotyrosine binding sites.
3. Either adaptor proteins can bind to the phosphorylated tyrosines through SH2 domains or relay proteins can bind, further relaying the signal after binding.
4. If adaptors bind such as Grb-2 and Shc, Sos can then bind.
5. Sos binding results in the exchange of GDP for GTP on Ras resulting in Ras activation.
6. Ras acts downstream on various signaling proteins can activate many pathways.

One of Ras' downstream pathways is a Serine/Threonine phosphorylation cascade that results in cell proliferation/differentiation. Ras' nature as a GTPase means that its overall signal life is short due to GAP induced GTP hydrolysis. The Ras signal is thus converted into the serine/threonine phosphorylation to prolong the signal and relay it downstream to the nucleus. While many phosphorylations occur in the cascade three constitute the core of the cascade. The last of the three is a MAP-kinase. The MAP-kinase requires phosphorylation of both a threonine and a tyrosine for activation, this is done by a MAP-kinase-kinase (known as **MEK** in mammalian cells). This dual phosphorylation adds a level of specificity for the activation of the MAP-kinase. The MAP-kinase-kinase is activated by a MAP-kinase-kinase-kinase (known as **Raf** in mammals).

The **Ras-MAP-kinase** pathway is extremely important and thus must be tightly regulated, one method to avoid cross-talk between the many different MAP-kinases present within cells is to utilize scaffolding proteins to couple certain MAP-kinases with their respective MEKs and Rafs.

Part IV

Cytoskeleton, Motility, and Shape

There are three types of cytoskeletal filaments which are common to eukaryotic cells:

Intermediate Filaments - provide mechanical strength and resistance to shear stress.

Microtubules - determine the positions of membrane-enclosed organelles and direct intracellular transport.

Actin filaments - determine the shape of the cell and provide the mechanism for whole-cell movement.

Each class of proteins serve specific functions in conjunction with accessory proteins (e.g. motor proteins).

All of these cytoskeletal filaments are constructed from protein subunits, often with the aid of accessory proteins that bind to the filaments to determine the sites of assembly, regulate disassembly/assembly, or various other functions. Accessory proteins thus bring cytoskeletal structure under the control of extra and intracellular signals.

Cytoskeletal polymers are built from protofilaments, which are long linear strings of subunits. They associate laterally, typically twisting around one another helically. This allows for the creation of many longitudinal bonds increasing the strength of the polymers.

The initial process for the assembly of filaments is called filament nucleation and is the rate limiting step. After this initial lag phase however addition of subunits is rapid until the system reaches a state of equilibrium where subunits added and subunits dissociating are equivalent. At this point the concentration of free subunits left in solution is called the critical concentration, C_c ($C_c = k_{off}/k_{on}$). This cycle is depicted below.

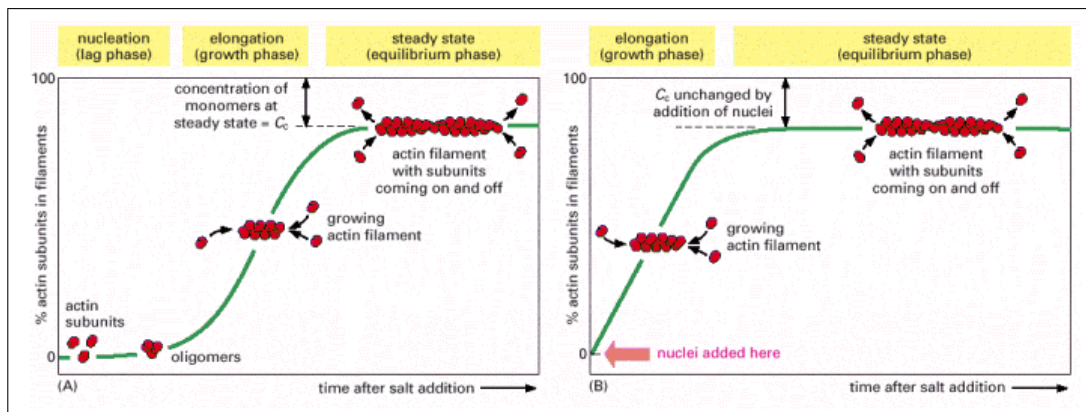


Figure 26: Basic process for filament construction.

Since the lag phase is the nucleation step, the cell can take advantage of this by utilizing special proteins in specific locations which catalyze the nucleation step. Thereby forcing filament production at specific sites. For a very comprehensive look at the polymerization of Actin and Tubulin see table 16-2 in Fat Alberts (Treadmilling and Dynamic Instability are discussed). Growth occurs preferentially on one end of actin and microtubule filaments. Where growth occurs fastest, depolymeration also occurs fastest. This more dynamic ends is called the plus end and the other is called the minus end.

7 Actin-based Systems

Actin subunits are single globular polypeptides with a binding site for ATP. Actin subunits assemble head to tail to generate protofilaments with a distinct polarity. These protofilaments then twist into a right-handed helix. Overall actin filaments are flexible in comparison to microtubules but can be very strong when cross-linked/bundled as they are in a cell.

Myosin was the first motor protein identified. It is an elongated protein formed from two heavy chains and two light chains (see figure below). Myosin is responsible for muscle contraction which will be discussed below.

Each myosin head binds and hydrolyses ATP, utilizing that energy to walk towards the plus end of an actin filament. There are numerous types of myosin, found in many organisms, however they all move towards the plus end of actin filaments except for myosin VI, which moves towards the minus end.

Myosin II which will be discussed with regards to muscle contraction, is involved with contractile activity in muscle and nonmuscle cells, thus the vast majority of myosin activity has yet to be elucidated.

7.1 Muscle Contraction

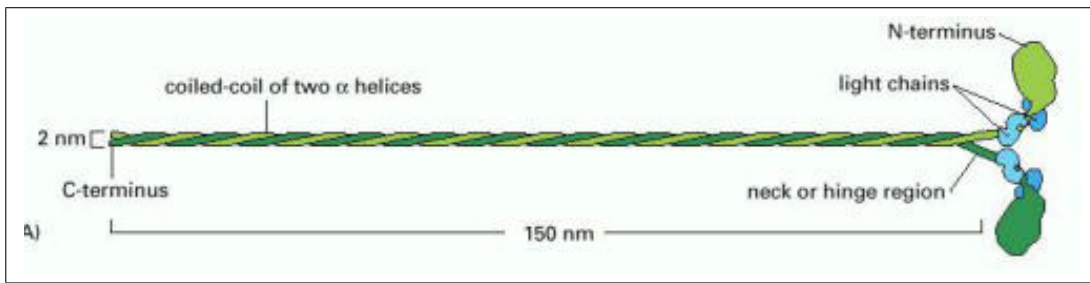
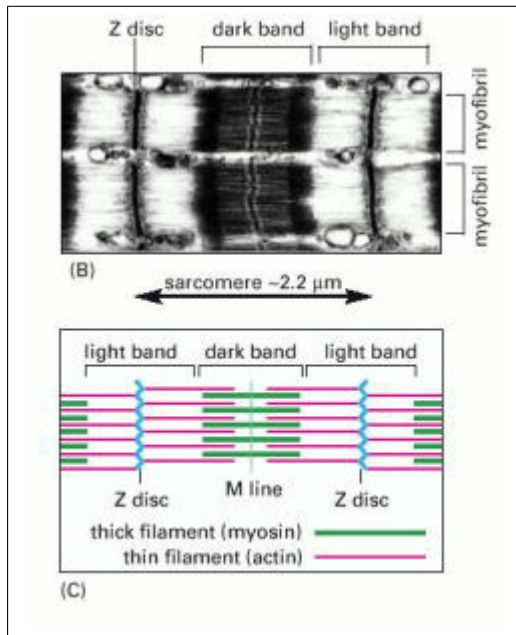


Figure 27: Structure of myosin.



This is by far the most familiar and well understood form of movement in animals. It is driven by the sliding of actin filaments against myosin II filaments through ATP hydrolysis. Figure 28 illustrates a typical myofibril which is the contractile unit of muscle cells. Each distance from Z disc to Z disc is called a sarcomere and each sarcomere is made up of arrays of parallel actin and myosin filaments. The actin filament's plus end is anchored in the Z disc which caps the filaments and prevents them from depolymerizing. Muscle contraction works in the following steps (Figure 29):

Figure 28: Myofibril composition.

1. The opening of Ca^{2+} channels in the sarcoplasmic reticulum in response to an action potential increases Ca^{2+} concentration resulting in the unbinding of an accessory protein, Troponin I. In the absence of Ca^{2+} Troponin I binds to actin keeping tropomyosin bound in a position which prevents proper myosin interaction with the actin filaments.
2. Tropomyosin moves off the myosin binding sites on the actin filament.
3. Myosin (Rigor state, no ATP bound) is bound to actin at an "original" position.
4. ATP binds to myosin, causing the myosin head to disassociate from the actin filament.
5. Myosin then hydrolyzes ATP, which pivots the head and binds it to a new actin subunit further towards the PLUS end of the filament.
6. P_i is then released from Myosin resulting in a "power stroke" which moves the filament and returns the myosin head to its Rigor position.
7. ADP is then released so that ATP can bind.

8. ATP binding to Myosin then allows the release of the actin filament resulting in a return to its original position. This is why rigor mortis occurs as the lack of new ATP to release myosin keeps the muscles in a contracted state.

8 Microtubule-based Systems

Microtubules are formed from subunits called **tubulin**. Each subunit is a heterodimer formed from α -tubulin and β -tubulin which are bound together noncovalently. Each subunit has a GTP binding site. The GTP bound to the α -tubulin monomer is trapped by the interface between the two subunits and is never hydrolyzed/exchanged. Once protofilaments are made there is a distinct polarity of exposed α -tubulin on one side and β -tubulin on the other.

Two types of microtubule motor proteins exist:

Kinesins - Motor proteins which carry membrane-enclosed organelles towards the plasma membrane. Similar in structure to myosin and walk toward the plus end of microtubules. Utilizes ATP for energy.

Dyneins - Minus-end directed microtubule motor proteins which are important in vesicle trafficking, localization of the Golgi near the center of the cell, and efficient movements of microtubules involved in the beating of cilia and flagella. Utilizes ATP.

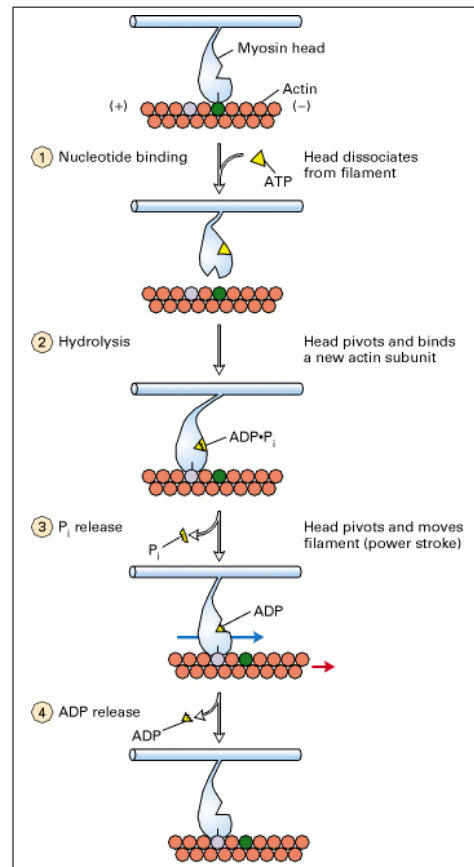


Figure 29: Muscle contraction.

9 Intermediate Filaments

Intermediate filaments are only found in some metazoans, vertebrates, nematodes, and molluscs.

Cytoplasmic intermediate filaments are closely related to their ancestors nuclear lamins. Intermediate filaments are elongated polypeptides with an extended central α -helical domain that forms a coiled coil with another monomer. A pair of dimers then associate in antiparallel fashion to form a tetramer, which is the soluble subunit for intermediate filaments. Lateral packing of these subunits and further organization creates a filament which can be seen in the figure below (Figure 30).

Fortunately little is understood about the assembly and disassembly dynamics of intermediate filaments, however they appear to be as dynamic as actin and tubulin.

Overall the function of intermediate filaments is to impart mechanical stability to cells. In other words, intermediate filaments serve as the ligaments of the cell, while microtubules and actin serve as the bones and muscles. There are many different filament families, which are expressed in various cell types. The most diverse family is that of the keratins. The diversity of keratins is useful in the diagnosis of epithelial cancers as the set of keratins expressed provides a fingerprint for the cell of origin. Another family is that of neurofilaments which are found along the axons of vertebrate neurons. ALS (Lou Gehrig's Disease)

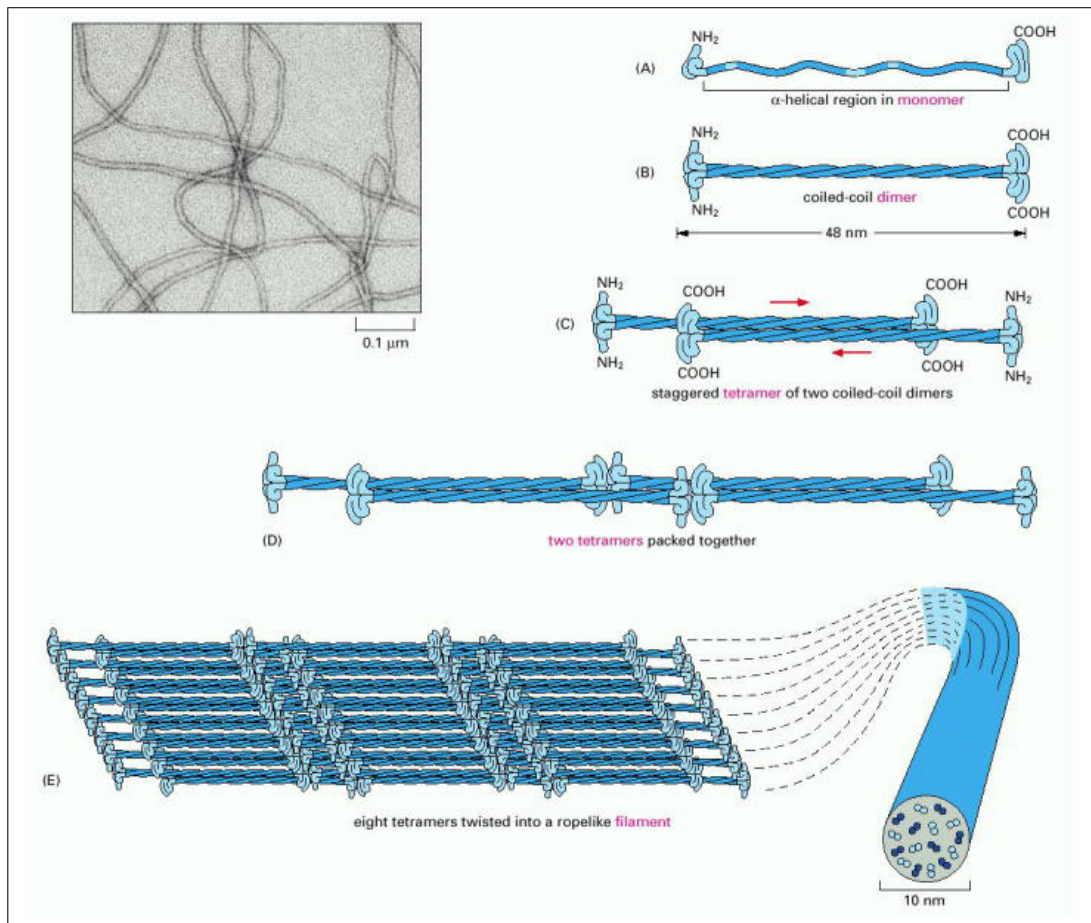


Figure 30: Intermediate filament construction.

is associated with abnormal assembly and accumulation of neurofilaments in motor neuron cell bodies.

10 Prokaryotic systems

Previously the cytoskeleton was thought to be a feature representative of eukaryotic cells, however recently homologues to all the major proteins of the eukaryotic cytoskeleton have been found in prokaryotes. These proteins are listed below:

FtsZ - Like tubulin, FtsZ forms filaments in the presence of GTP and is essential in the recruiting of other proteins for the production of a new cell wall between dividing cells.

MreB and ParM - Actin-like proteins which are involved in the maintenance of cell shape.

Crescentin - Intermediate filament-like protein.

Part V

Protein Synthesis and Processing

11 Regulation of translation

There are numerous methods of posttranscriptional regulation which provide control of translation. Several are discussed below.

11.1 Transcription Attenuation

Expression of certain genes can be inhibited by the premature termination of transcription. In prokaryotes the nascent RNA chain can adopt a conformation which interferes with the RNA polymerase forcing it to abort transcription. In eukaryotes an example is found in the HIV production cycle. After integration, RNA polymerase II will begin transcribing the integrated viral genome, but typically will halt itself (for unknown reasons at this time) thereby not producing the entirety of the viral genome. This prevents viral growth, however when the small transcripts which are produced encode for a protein called Tat, it binds to a specific stem-loop structure in the nascent RNA (binding site called Tar). Tat then recruits several cellular proteins which allow the RNA polymerase II to continue transcription. HIV has apparently evolved to coopt normal cellular machinery which allows RNA polymerase II to transcribe long introns efficiently without pausing.

11.2 Alternative Splicing

Cells are capable of producing numerous proteins from a single gene due to the process known as alternative splicing. This process can also be utilized to regulate the expression of proteins. The simplest example is that splicing is utilized to switch production from a nonfunctional protein to a functional one. In addition, splicing can be utilized to produce different versions of a protein in different cell type depending on the splicing pre-translation. A key example of this regulation is sex determination in *Drosophila*. Figure 31 illustrates the method.

11.3 Alteration of RNA Transcript Cleavage and Poly-A Addition

Control over the site of RNA cleavage can ultimately alter the C-terminus of the resultant protein. This is seen in the switch from MB bound Ab production to secreted Ab production in B lymphocytes. The switch in production is caused by an increased concentration of CStF, which is a protein that binds to G/U rich sequences and promotes RNA cleavage and poly-a addition. CStF is produced when Ab stimulation activates the B cell causing the RNA transcript to be cleaved at the shorter site producing RNA transcripts for secreted Ab.

11.4 RNA Editing

This is a remarkably complex system in which guide RNAs are utilized to alter certain genes after they have been transcribed. It is possible for gene editing to be so extensive that over half the nucleotides in the mature mRNA were inserted during the editing process. Overall not many specifics are known about the process, except for a few examples discussed in Fat Alberts.

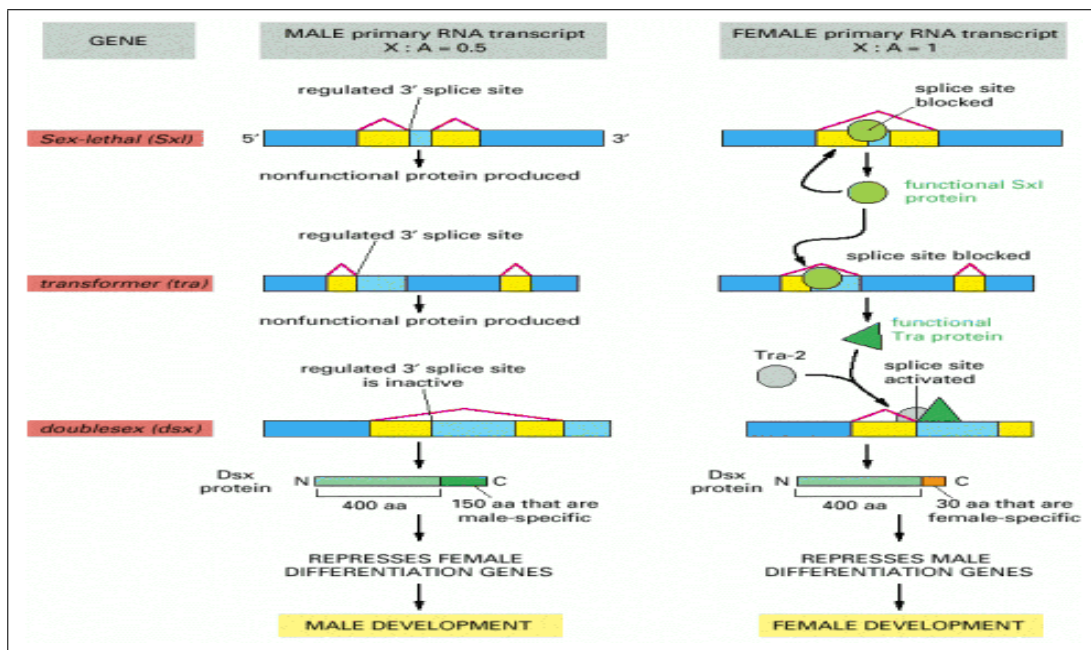


Figure 31: Sex determination in *Drosophila* via alternative splicing.

11.5 Regulation of RNA Transport

It is estimated that in mammals only about 1/20th of the total mass of RNA synthesized ever leaves the nucleus. This would be one of the simplest methods for regulating translation as no mRNA = no translation.

During typical gene expression RNA export is delayed until processing has been completed, therefore anything that prevents proper RNA splicing could block the exit of the RNA from the nucleus. Typically, unexported RNA are degraded by the exosome. The best example of regulated RNA transport is in the HIV life cycle.

During HIV's life cycle it must be able to produce a variety of transcripts to be spliced into many different mRNAs. In addition HIV must be able to produce full length RNA copies containing introns. This poses a problem as unspliced RNA is blocked for export by the host cell. HIV overcomes this through the production of a protein called Rev, which binds to the Rev Response Element (RRE) within viral introns. Rev then interacts with a nuclear export receptor and allows the viral RNA to leave the nucleus. Rev becomes a regulatory molecule for RNA transport with regards to viral replication. It is thought that Rev concentrations vary depending on the quality of conditions for viral replication.

11.6 Negative Translational Control

Negative translation control is seen in both prokaryotic and eukaryotic cells. In bacteria mRNAs contain a stretch of six nucleotides known as the Shine-Dalgarno sequence and is always found a few nucleotides upstream of an initiating AUG codon. This sequence pairs with the small ribosomal subunit and allows for correct positioning with regards to the AUG codon. Mechanisms in bacteria block this sequence providing a simple method for repressing translation of certain proteins.

Eukaryotic cells don't contain a similar sequence, however they utilize transcriptional repressors which bind to the 5' or 3'UTRs of mRNA inhibiting efficient translation. Iron regulation in eukaryotes is a good example. A translation repressor called aconitase binds to an iron-response element on the iron sequestering protein ferritin's RNA blocking trans-

lation. Aconitase is an iron-binding protein and exposure to iron allows it to dissociate and allow production of ferritin for iron storage.

11.7 Regulation of Translational Machinery

In addition to the above methods of posttranscriptional regulation the cell could easily regulate access/availability of translational machinery in order to regulate translation. An example of this is the phosphorylation of initiation factor eIF-2. Phosphorylation of eIF-2 prevents it from releasing eIF-2B (a GEF), preventing it from exchanging GDP for GTP. This slows down protein synthesis as little free eIF-2 is left for translation.

11.8 Internal Ribosome Entry Sites

While most proteins are translated from the first AUG downstream from the 5' cap, certain AUGs can be skipped. When this occurs, translation is initiated at an internal ribosome entry site (IRES). There are many different IRES, which bind different subsets of initiation factors, however they all bypass the need for a 5' cap structure and the presence of eIF-4E.

A perfect example of this regulation is when cells enter apoptosis. eIF-4G is cleaved resulting in a general decrease in translation due to the lack of the 5' cap initiation factor. Proteins critical for the control of cell death are encoded with IRES containing mRNAs ensuring that they are still translated despite the lack of the normal initiation factors.

12 Posttranslational Mod and Intracellular trafficking

After a protein is translated a host of other events can occur. Proteins can be translocated (during/after translation), packaged, and modified. Most posttranslational modifications occur in the ER and golgi apparatus. I'll begin by discussing translocation into the ER, some of the modifications that can occur there for different types of proteins and then finish up with trafficking of those proteins.

12.1 Translocation

Translocation of proteins can occur to a variety of organelles. Most of the time translocation occurs posttranslationally, such as mitochondrial translocation. For the ER however, most translocation occurs during translation. This explains why ribosomes cover the surface of the ER.

The Rough ER is peppered with ribosomes, which are actively translating mRNA. This translation translocates an ER signal sequence (variable, however each contains 8 or more nonpolar amino acids at its center) containing protein directly into the ER lumen. The steps for translocation are as follows:

1. Initially a protein is translated in the cytosol until the ER signal sequence has been produced.
2. This sequence is then bound by a Signal Recognition Particle (SRP), which pauses translation.
3. The SRP bound peptide and ribosome complex is then taken to the ER and binds to an SRP receptor on the ER membrane.
4. A protein translocator, Sec61 complex, lines up with the tunnel of the large ribosomal subunit and forms a tight seal.

5. The signal sequence on the growing polypeptide triggers the opening of the pore by binding to a site inside the pore after SRP releases.
6. On going peptide synthesis then pushes the peptide through the pore.
7. The ER signal sequence is then removed once it enters the ER lumen. The signal sequence is thought to remain bound until the C-terminus of the protein has passed through the membrane.

The above steps detail the translocation of a non-transmembrane protein. Single-pass transmembrane proteins on the other hand have a single internal hydrophobic region which stops the translocation process. The stop-transfer signal anchors the protein in the membrane and tells the translocator to release the cleaved ER signal sequence. The N-terminus is thus in the ER lumen while the C-terminus sits on the cytosolic side. The translocator opens up laterally and allows the transmembrane protein to slide into the ER MB. This process is similar for multi-pass transmembrane proteins, the difference being the presence of multiple start and stop transfer sequences within the polypeptide.

Once the proteins are translocated they are ultimately in transit to another destination. Some proteins stay in the ER and possess an ER retention signal of four amino acids at the C-terminus (KDEL). Two ER resident proteins which are important for posttranslational mods within the ER lumen are PDI and BiP. PDI catalyzes the formation of disulfide bonds and BiP pulls posttranslationally translocating proteins into the ER through the Sec61 and keeps the translocating protein from folding.

12.2 Protein Glycosylation

A major function of the ER is the covalent addition of sugars. Most MB bound proteins made in the ER are glycoproteins. Glycosylation begins with the transfer of a preformed precursor oligosaccharide composed of N-acetylglucosamine, mannose, and glucose and containing a total of 14 sugars. This transfer occurs with one enzymatic step, catalyzed by an enzyme called oligosaccharyl transferase, and typically occurs during protein synthesis. The precursor oligosaccharide is held in the ER membrane by a special lipid molecule known as dolichol, and is transferred by the oligosaccharyl transferase after the target Asparagine emerges into the ER lumen during translocation. All mature glycoproteins are derived from this initial molecule. This process is illustrated below (Figure 32).

While in the ER, three glucoses and one mannose are quickly trimmed from the oligosaccharides of most glycoproteins, the importance of this will be covered later, as more trimming occurs in the Golgi apparatus. In addition to N-linked oligosaccharides, less frequently oligosaccharides are linked to the hydroxyl group on the side chain of a serine, threonine, or hydroxylysine. These are known as O-linked oligosaccharides and are formed in the Golgi by pathways that are not yet fully understood.

While glycoproteins themselves exhibit functions after production the oligosaccharides serve another purpose within the ER. The oligosaccharides are used as tags to mark proteins as folded/unfolded. These tags tell calnexin and calreticulin (chaperones) to hold the incompletely folded proteins in the ER. Specifically calnexin and calreticulin recognize N-linked oligosaccharides that contain a single terminal glucose. This means that they only bind proteins that have had two of their initial three glucoses removed by ER glucosidases. When the third glucose is removed the protein dissociates from its chaperone and can leave the ER. The terminal glucose can be removed prior to complete folding, thus a **glucosyl transferase** within the ER lumen keeps adding a glucose to those proteins which are not completely folded. This prevents the glucose trimming reactions from sending a protein out of the ER too early.

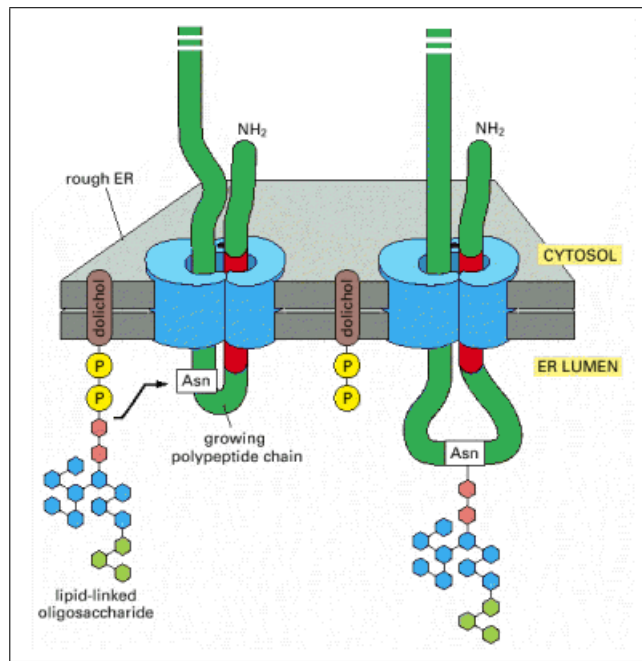


Figure 32: Protein glycosylation in the rough ER.

Despite these mechanisms attempting to produce properly folded proteins more than 80% for some proteins translocated into the ER fail to achieve their properly folded state. Such proteins are exported from the ER through a mechanism known as **dislocation** which utilizes the Sec61 complex to spit the protein back to the cytosol. It is not known which additional proteins and what signal sequence are used to initiate this mechanism.

12.3 Addition of Glycosylphosphatidylinositol (GPI)

Another modification that occurs in the ER is the addition of a GPI covalently to the C-terminus of some MB proteins destined for the plasma MB. This linkage is formed in the lumen of the ER while the transmembrane segment of the protein is cleaved off. GPI anchors are used to direct plasma MB proteins to lipid rafts.

12.4 Lipid Bilayer Assembly in the ER

In addition to protein folding and transfer the ER is responsible for synthesizing nearly all of the major classes of lipids. This makes sense as the vesicular trafficking from the ER itself requires a large number of lipids. The major phospholipid made is phosphatidylcholine, which is formed in three steps from choline, two fatty acids, and glycerol phosphate. Each step is catalyzed by enzymes in the ER membrane that face the cytosol, thus phospholipid synthesis occurs exclusively in the cytosolic leaflet of the ER. Phospholipid synthesis occurs in the following steps:

1. **Acyl transferases** successively add two fatty acids to glycerol phosphate to produce phosphatidic acid. This step enlarges the lipid bilayer.
2. Later steps are determined by the different head group to be attached to the phosphatidic acid. Phosphatidyl-ethanolamine, phosphatidylserine, and phosphatidylinositol are all synthesized in this manner.

Since this lipid production is asymmetrical and only enlarges one leaflet, a scramblase equilibrates phospholipids between the two leaflets. Flippases then specifically remove phospholipids containing free amino groups (phosphatidylserine and phosphatidylethanolamine) from the extracellular leaflet to the cytosolic side.

The ER also produces cholesterol and **ceramide**. Ceramide is made by condensing the amino acid serine with a fatty acid to form the amino alcohol sphingosine; a second fatty acid is then added to form ceramide. It is exported to the Golgi where it serves as a precursor for the synthesis of two types of lipids, glycosphingo-lipids, and sphingomyelin.

12.5 On to the Golgi

As discussed earlier, the translocation of proteins into the ER is merely the beginning of a protein's journey. The next stop is the Golgi apparatus where the proteins pass through successive compartments which can lead to successive modifications of the proteins. This pathway is a delicate balance of forward and retrieval vesicles within which proteins are shuttled to and from.

Proteins leave the ER in **COPII-coated** transport vesicles. They bud from ER exit sites when cargo proteins display exit signals that are recognized by the components of the COPII coat. Coat proteins are extremely important in vesicular transport as they allow for deformation of membranes. This allows budding of vesicles to occur. These regions are specialized areas of the ER known as **ER exit sites**, whose membrane lacks ribosomes. In most animal cells, ER exit sites seem to be randomly dispersed throughout the ER network.

Packaging of proteins into vesicles is a highly selective process. Some cargo proteins are actively recruited into transport vesicles. It is thought that these cargo proteins display transport (exit) signals on their surface which are recognized by complementary receptor proteins that become trapped in the budding vesicle. These receptors interact with the components of the COPII coat. At a much lower rate proteins without exit signals can also get packaged in vesicles, so that even proteins that normally function in the ER slowly leak out of the ER. Similarly, secretory proteins which are made in high concentrations may leave the ER without the help of sorting receptors. So far little is known about the exit signals that direct proteins out of the ER for transport to the Golgi and beyond. One exception is that the protein **ERGIC53** appears to serve as a receptor for packaging some secretory proteins into COPII-coated vesicles. It was discovered in individuals that have low serum levels of blood clotting factors V and VIII. It is thought that ERGIC53 recognizes mannose on Factor V and VIII thereby packaging it into transport vesicles.

Keep in mind that only proteins which have completed folding leave the ER in vesicles. Most likely BiP or calnexin may cover up exit signals in order to prevent them from moving on in the secretory pathway. This quality control mechanism however can be detrimental such as in cystic fibrosis, where individuals produce a slightly misfolded plasma membrane protein important for Cl^- transport. This misfolded protein does not lose function due to the mutation, but it is never allowed to get to the plasma membrane because of the quality control mechanisms within the ER. It ultimately is dislocated and degraded in the cytosol resulting in a complete lack of that protein.

12.6 Vesicular Tubular Clusters

After transport vesicles have budded from an ER exit site and shed their coat, they begin to fuse with one another. This fusion is known as homotypic and requires matching of SNARES (fusion proteins). In this case, v-SNAREs and t-SNAREs are contributed by both membranes.

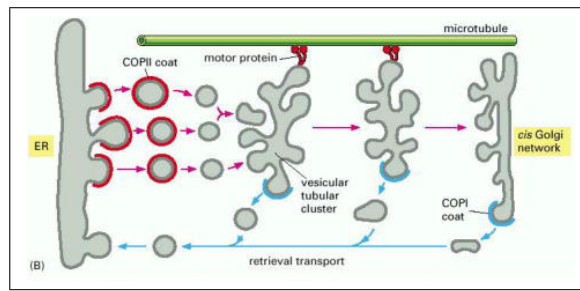


Figure 33: Schematic of vesicular transport between ER and Golgi apparatus

These structures that form are known as **vesicular tubular clusters**. These clusters constitute a new compartment separate from the ER and are short lived as they move quickly along microtubules to the Golgi, where they fuse and deliver their contents (Figure 33).

As soon as vesicular tubular clusters form, they begin budding off vesicles of their own. Unlike the COPII-coated vesicles from the ER, these vesicles are **COPI-coated**. They carry back to the ER, ER resident proteins which have escaped in the vesicles. In addition, they take back proteins that participated in the ER budding reaction. This **retrieval (retrograde) transport** continues as the vesicular tubular cluster moves to the Golgi apparatus.

12.7 Retrieval Pathway and Sorting Signals

The retrieval pathway depends on **ER retrieval signals**, which are present on resident ER proteins. This signal binds directly to COPI coats and allows the proteins to be packaged into COPI vesicles. The best characterized signal of this type is two lysines followed by any two amino acids (KKXX) at the extreme C-terminus of the protein. Soluble ER resident proteins, such as BiP, contain a short retrieval signal at their C-terminal end that consists of a Lysine, Asparagine, Glutamate, and Leucine (KDEL). Unlike the retrieval signals, the soluble proteins expressing a KDEL sequence must first bind to a **KDEL receptor**. The KDEL receptor then packages any protein displaying a KDEL sequence into COPI-coated vesicles.

While the retrieval pathway definitely provides a level of control for protein residency it is not the only mechanism through which the cell manages protein location. In cells where KDEL and ER retrieval signals have been mutated out of proteins, those mutant proteins still do not secrete at a rate comparable to nonresident proteins. It appears that secretion is slowed by some unknown mechanism as another layer of control in addition to retrieval mechanisms.

12.8 Golgi Apparatus

The Golgi apparatus consists of a collection of flattened, membrane-enclosed **cisternae**. Each of these Golgi stacks consists of four to six cisternae. The localization of these stacks is dependent on microtubules.

During their passage through the Golgi, transported molecules undergo an ordered series of covalent modifications. Each Golgi stack contains two distinct faces: a **cis** face (entry face) and a **trans** face (exit face). Both cis and trans faces are closely associated with special compartments composed of a network of interconnected tubular and cisternal structures. Thus, each side can be referred to as the **cis Golgi network (CGN)** and **the trans Golgi network (TGN)**. Proteins entering the CGN can either return

to the ER or move forward, similarly proteins exiting from the TGN can either move onward to lysosomes, secretory vesicles, or the cell surface, or be sorted back to an earlier compartment.

As discussed earlier, further modifications of oligosaccharide chains occurs in the Golgi. The outcome of these modifications is the creation of two broad classes of N-linked oligosaccharides, the **complex oligosaccharides** and the **high-mannose oligosaccharides**. Complex oligosaccharides are formed through the initial trimming in the ER and the addition of sugars in the Golgi. By contrast high mannose oligosaccharides have no new sugars added to them in the Golgi apparatus. They contain just two N-acetylglucosamines and many mannose residues often approaching the number originally present in the lipid-linked oligosaccharide. The location of the oligosaccharide on the protein determines whether it will remain high mannose or be processed to complex. The process by which oligosaccharides are processed to complex form is summed up in figure 34 below.

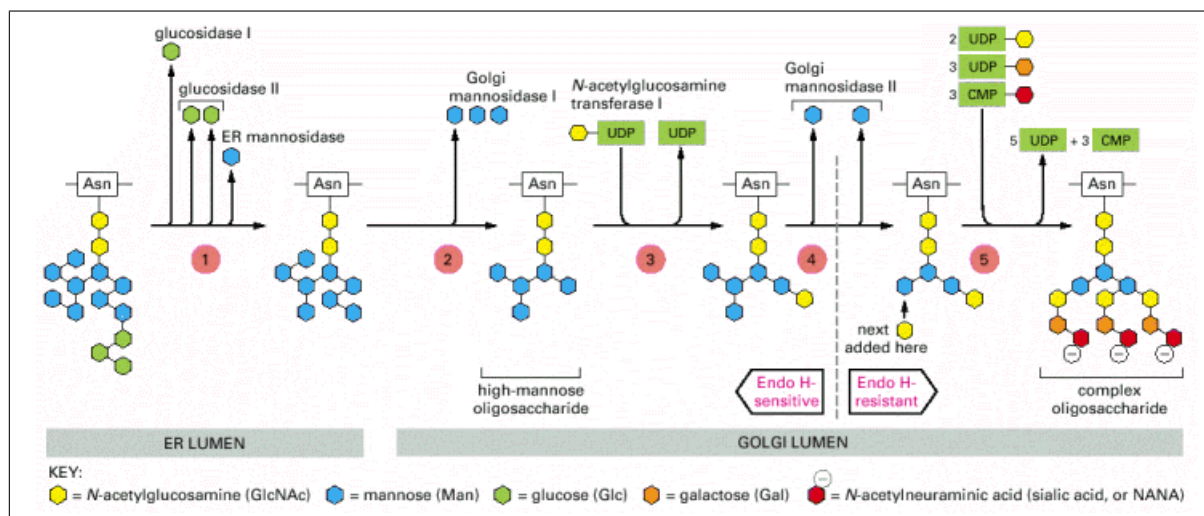


Figure 34: Processing pathway for oligosaccharides in the ER and Golgi.

12.9 Trans Golgi Network and Onward

All of the proteins that pass through the Golgi apparatus, except those that are retained there as permanent residents are sorted in the TGN according to their final destination.

12.9.1 Trans Golgi Network to Lysosomes

Lysosomes are specialized for the intracellular digestion of macromolecules. They contain unique membrane proteins and a wide variety of hydrolytic enzymes that operate best at a pH of 5. This low pH is maintained by an ATP driven H^+ pump in the lysosomal membrane. Newly synthesized lysosomal proteins are transferred into the lumen of the ER, transported through the Golgi, then carried from the TGN to late endosomes by means of clathrin-coated vesicles.

Lysosomal hydrolases contain N-linked oligosaccharides that are covalently modified in a unique way in the CGN so that their mannose residues are phosphorylated. This process requires two enzymes. The first is a **GlcNAc phosphotransferase** that binds the hydrolase and adds **GlcNAc-phosphate** to one or two of the mannose residues. A second enzyme (name unknown) cleaves the GlcNAc residue leaving behind the newly created M6P markers. These **mannose-6-phosphate (M6P)** groups are recognized by an **M6P**

receptor protein in the TGN that segregates the hydrolases and packages them into clathrin-coated budding transport vesicles that deliver the contents to **late endosomes** (the organelle that matures into lysosomes). The M6P receptors shuttle back and forth between the TGN and late endosomes. The low pH in the late endosome dissociates the receptor from the hydrolase making the transport unidirectional.

12.9.2 Trans Golgi Network to Cell Exterior: Exocytosis

Proteins can be secreted from cells by exocytosis in either a constitutive or a regulated fashion. In the regulated pathways, molecules are stored either in secretory vesicles or synaptic vesicles. They do not fuse with the plasma membrane until an appropriate signal is received. Secretory vesicles bud from the TGN and contain condensed proteins. As these vesicles mature (proteins condense within) they also undergo a retrieval pathway similar to that found in the ER mediated by clathrin-coated transport vesicles. This results in membrane recycling as well as returning Golgi resident proteins. Synaptic vesicles (confined to nerve cells and some endocrine cells) form from endocytic vesicles and from endosomes. They are responsible for regulated secretion of small-molecule neurotransmitters.

The difference between the regulated pathways and the constitutive pathway lies in the cells they operate in. Regulated pathways work in specialized secretory cells, while the constitutive secretory pathway operates in all eukaryotic cells, mediated by continual vesicular transport from the TGN to the plasma membrane.

Proteins from the TGN are delivered to the plasma membrane by the constitutive pathway unless they are diverted into other pathways or retained in the Golgi. In polarized cells, the transport pathways from the TGN operate selectively to ensure that different sets of membrane proteins, secreted proteins, and lipids are delivered to different membrane regions.

13 Endocytosis

Cells ingest many things through endocytosis, which is a process in which localized regions of the plasma membrane invaginate and pinch off to form **endocytic vesicles**. Endocytosis occurs both constitutively and as a triggered response to extracellular signals. Endocytosis is so extensive that in many cells a large fraction of the plasma membrane is internalized every hour, this requires a balance brought about by a continual exocytosis of plasma membrane components (proteins and lipids). This large scale endocytic-exocytic cycle is mediated largely by clathrin-coated pits and vesicles.

Many cell-surface receptors that bind extracellular molecules become localized in clathrin-coated pits. More than 25 different receptors are known to participate in receptor-mediated endocytosis. Examples of this are the **LDL receptor** and the **transferrin receptor**. Receptors contain a common endocytosis signal consisting of four amino acids Y-X-X-hydrophobic amino acid. This short peptide binds directly to **adaptins** in clathrin-coated pits beginning the budding process. As a result, they and their ligands are efficiently internalized in a process known as receptor-mediated endocytosis. The coated endocytic vesicles rapidly shed their clathrin coats and fuse with early endosomes.

Most of the ligands dissociate from their receptors in the acidic environment of the endosome and eventually end up in lysosomes, while most of the receptors are recycle via transport vesicles back to the cell surface for reuse. Both receptor-ligand complexes can follow other pathways from the endosomal compartment. In some cases, both the receptor and ligand end up being degraded in lysosomes, resulting in receptor down-regulation. In

other cases, both are transferred to a different plasma membrane domain, with the ligand exocytosed in a process known as **transcytosis**.

Part VI

Cell Division, Differentiation, and Development

14 Bacterial division

The cell cycle of prokaryotes is simple and fast. Replication of the genome begins at a particular DNA sequence, the replication origin, which is anchored to the cell membrane. Once replication is complete, assembly of new membrane and cell wall forms a septum, which eventually divides the cell in two. Because the origins of the two newly formed genomes are anchored to different membrane sites, each daughter cell receives one genome. In ideal growth conditions, the bacterial cell cycle is repeated every 30 minutes.

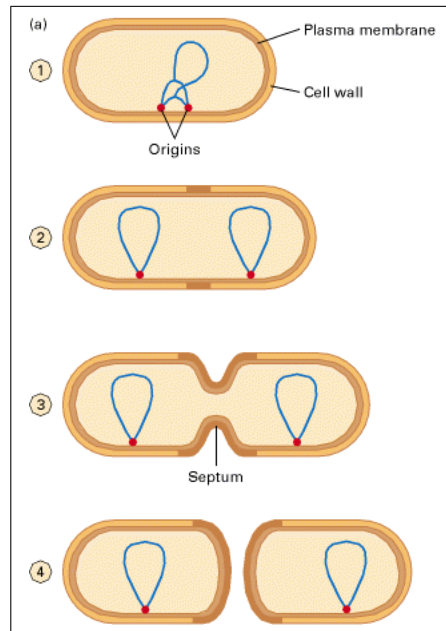


Figure 35: Bacterial cell division.

15 Eukaryotic Cell Cycles, Mitosis, and Cytokinesis

Most eukaryotic cells live according to an internal clock that proceeds through a sequence of phases known as the **cell cycle**. DNA is duplicated during the **synthesis (S) phase** (10-12 hours) and distributed to opposite ends of the cell during the **mitotic (M) phase** (less than an hour in a mammalian cell). Partly to allow more time for growth **gap phases** are inserted in most cell cycles - a G_1 phase between M phase and S phase and a G_2 phase between S phase and mitosis. In addition to serving as simple time delays the gap phases allow time for the cell to monitor internal and external conditions. Progress along the cycle is controlled at key checkpoints which monitor the status of the cell.

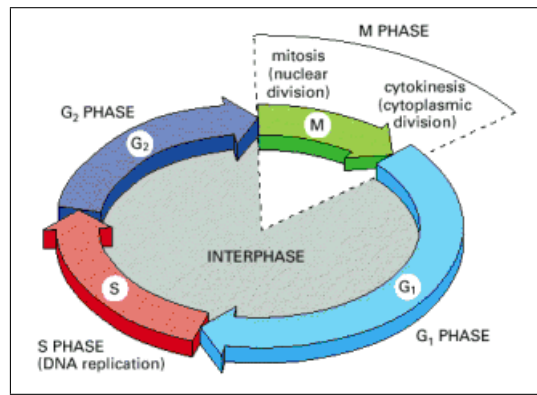


Figure 36: The phases of the Cell Cycle.

While this provides a general picture with which to view the eukaryotic cell cycle it is important to look at the cell cycle control system which regulates the major processes. As stated earlier, the cycle is controlled at key checkpoints. If any conditions are not met, the cell can arrest the cycle preventing entry into mitosis. An example of this would be if DNA replication is not complete or if chromosome separation is hindered by incorrect attachment of mitotic spindle. Most of the checkpoints within the cell cycle operate through negative intracellular signals. This makes sense when one considers the following example. If a cell is monitoring the attachment of chromosomes to the mitotic spindle a flurry of positive signals makes it difficult to detect the final "go" signals as it is a small change in the total intensity of the signal. On the other hand if a negative signal is sent from any unattached chromosome then the absence of signal will be easily detected as a "go" signal and any hint of stop will arrest the cell cycle.

The cell cycle control system is based on cyclically activated protein kinases. These kinases are a family known as **cyclin-dependent kinases (Cdks)**. The activity of these kinases rises and falls as the cell progresses through the cycle leading to direct cyclical changes in the phosphorylation of intracellular proteins that initiate or regulate major events.

Cyclical changes in Cdk activity are controlled by an array of enzymes and other proteins. The most important of these are proteins known as **cyclins**. Cdks, as their name implies are dependent on tight binding to a cyclin for their activity. No kinase activity occurs without cyclin binding. There are four classes of cyclins, each defined by the stage of the cell cycle at which they bind Cdks and function. Three of these classes are required in all eukaryotic cells.

1. G₁/S-cyclins bind Cdks at the end of G₁ and commit the cell to DNA replication.
2. S-cyclins bind Cdks during S phase and are required for initiation of DNA replication.
3. M-cyclins promote the events mitosis.
4. G₁-cyclins are the four class which help promote passage through the restriction point in late G₁. These are not required in all eukaryotic cells.

Cyclin-Cdk complexes drive cell cycle events because the cyclin not only activates Cdk activity, but directs the Cdk to specific target proteins. As a result, each cyclin-Cdk complex phosphorylates a different set of substrate proteins. Full activation of cyclin-Cdk complexes occur when a separate kinase, the **Cdk-activating kinase (CAK)**, phosphorylates

an amino acid near the entrance of the Cdk active site. This results in a conformational change that allows the Cdk to phosphorylate its target proteins effectively.

This added phosphorylation provides another level of control for the cell cycle. Thus, phosphorylation/dephosphorylation can regulate Cdk activity even in the presence of proper cyclins. In addition to phosphorylation control, the binding of **Cdk inhibitor proteins (CKIs)** can render Cdks inactive. These are mostly utilized in the control of G₁ and S phase.

15.1 Mitosis

Mitosis begins with **prophase**, which is marked by an increase in microtubule instability, triggered by M-Cdk. In animal cells, an unusually dynamic microtubule array (an **aster**) forms around each of the duplicated **centrosomes**, which separate to initiate the formation of two spindle poles. Interactions between the asters and a balance between minus-end-directed and plus-end-directed microtubule-dependent motor proteins result in the self-assembly of the bipolar spindle. In cells that lack centrosomes, a functional bipolar spindle self-assembles instead around the replicated chromosomes.

Prometaphase begins with the breakdown of the nuclear envelope, which is triggered when M-Cdk phosphorylates the nuclear lamina. The breakdown of the lamina and nuclear envelope allows the **kinetochores**, a complex protein machine that assembles onto the highly condensed DNA at the centromere, on the condensed chromosomes to capture and stabilize microtubules from each spindle pole. The end-on attachment to the kinetochore is on the plus end of the microtubule. The kinetochore microtubules from opposite spindle poles cross-link and then pull in opposite directions, creating a tension that helps bring the chromosomes to the equator of the cell as well as push the centrosomes apart to form the **metaphase plate**. The spindle microtubules at **metaphase** are highly dynamic and undergo a continuous poleward flux of tubulin subunits.

Anaphase begins with the sudden proteolytic cleavage of the cohesin linkage holding sister chromatids together. This cleavage is triggered by the activation of the **anaphase promoting complex (APC)**. Two crucial functions exist for the APC: (1) it cleaves and inactivates the M-phase cyclin, thereby inactivating M-Cdk and (2) it cleaves an inhibitory protein (**securin**), thereby activating a protease called **separase**. Separase cleaves a subunit in the cohesin complex to unglue the sister chromatids. This allows the chromosomes to pull to opposite poles. The chromosomes move by two independent and overlapping processes. The first, **anaphase A**, is the initial poleward movement of the chromosomes. It is accompanied by shortening of the kinetochore microtubules at their attachment to the chromosome and, to a lesser extent, by the depolymerization of spindle microtubules at the two spindle poles. The second process, **anaphase B**, is the separation of the poles themselves. Anaphase B depends on motor proteins at the poles that pull the poles apart, as well as motor proteins at the central spindle which push the poles apart. These two processes are sensitive to different drugs, which possibly reflects the different sensitivities of the motor proteins involved in each.

In **telophase**, the nuclear envelope reforms on the surface of each group of separated chromosomes. During this process, nuclear pore complexes are incorporated into the envelope and dephosphorylated lamins reassociate to form the nuclear lamina. Once the nuclear envelope is complete, pore complexes pump in nuclear proteins expanding the nucleus and the condensed chromosomes decondense into their interphase state.

15.2 Cytokinesis

The cell cycle culminates in the division of the cytoplasm by **cytokinesis**. In a typical cell, cytokinesis accompanies every mitosis, although some cells undergo mitosis without cytokinesis and become multinucleate. Cytokinesis begins in anaphase and ends in telophase, ending as the next interphase begins.

The first visible change of cytokinesis in an animal cell is the sudden appearance of a cleavage furrow on the cell surface. The furrow rapidly deepens until it completely divides the cell in two. In animal cells and many unicellular eukaryotes, the structure that accomplishes cytokinesis is the **contractile ring**, a dynamic assembly composed of actin filaments, myosin II filaments, and many structural and regulatory proteins. The ring assembles beneath the plasma membrane and splits the cell into two.

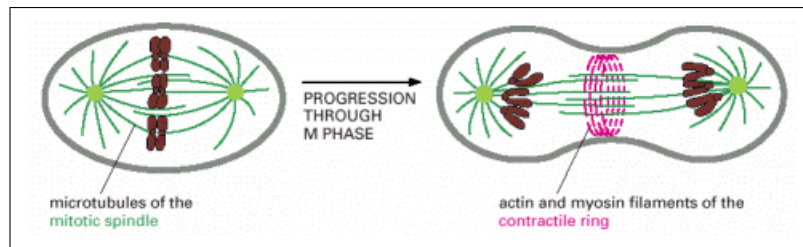


Figure 37: Formation of the contractile ring in anaphase.

At the same time, new plasma membrane is inserted by the fusion of intracellular vesicles. This compensates for the increase in surface area that accompanies cytoplasmic division.

Keep in mind that cytokinesis is different in plant cells. A large part of this is due to the presence of a cell wall. Rather than a contractile ring, the cytoplasm of the plant cell is partitioned from the inside out by the construction of a new cell wall, called the **cell plate**. The assembly of the cell plate begins in late anaphase and is guided by a structure called the phragmoplast, which contains the remaining overlap microtubules of the mitotic spindle. Small vesicles filled with polysaccharide and glycoproteins are transported along the microtubules to the equator of the phragmoplast, apparently by the action of microtubule motor proteins. The vesicles fuse to form a disc-like, membrane enclosed structure called the **early cell plate**. The plate expands outward by further vesicle fusion until it reaches the plasma membrane and the original cell wall. Later cellulose is laid down within the matrix of the cell plate to complete construction.

16 Meiosis and Gametogenesis

16.1 Meiosis

Meiosis is a special kind of nuclear division in which the chromosome complement is precisely halved. The essential events of meiosis were not fully established until the early 1930s. After a chromosome is duplicated by DNA replication, the twin copies of the fully replicated chromosome at first remain tightly linked along their length and are known as sister chromatids. While in mitosis sister chromatids line up at the equator of the cell in meiosis homologs recognize each other and become physically connected side-by-side along their entire length before they line up on the spindle. It is uncertain how this recognition occurs, but it is seemingly mediated by complementary DNA base-pair interactions at numerous sites along chromosomes.

The pairing of homologs produces a structure called a **bivalent**, which contains four chromatids. The pairing occurs during a long meiotic prophase, which can last for days or years. This pairing allows for genetic recombination. In metaphase I, all of the bivalents line up on the spindle. At anaphase I the two duplicated homologs separate from each other and move to opposite poles of the spindle and the cell divides. Another cell division follows to produce haploid gametes.

Because sister chromatids behave as a unit, each daughter cell from cell division I inherits two copies of one of the two homologs. The two copies are identical except where genetic recombination has occurred. The two daughter cells thus contain a haploid number of chromosomes, but a diploid amount of DNA. Formation of the actual gametes can now proceed through a second cell division, division II of meiosis.

Without further DNA replication the duplicated chromosomes align on a second spindle and the sister chromatids separate to produce cells with a haploid DNA content. Four haploid cells are therefore produced from each cell that enters meiosis.

Occasionally chromosomes fail to separate normally into the four haploid cells in a phenomenon known as **nondisjunction**. This can result in gametes forming abnormal embryos, most of which die. Some survive, such as in Down syndrome in humans. Trisomy 21 results from nondisjunction during meiotic division I or II. The vast majority of such segregation errors occur during meiosis in females, and the error rate increases with advancing maternal age. The frequency of missegregation in human oocytes is remarkably high, about 10%.

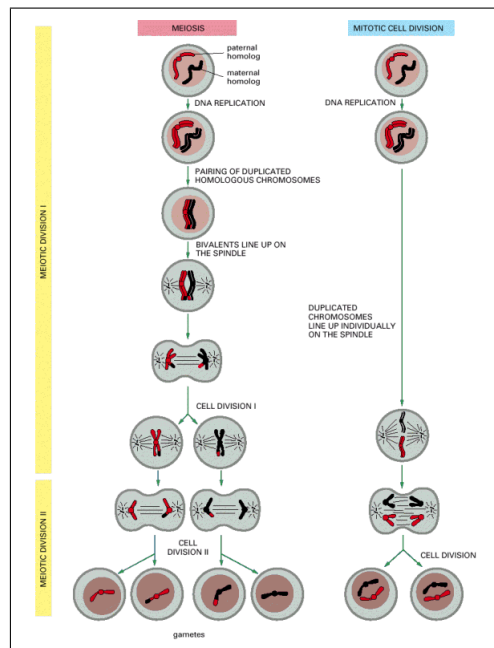


Figure 38: Comparison of Meiosis with Mitosis.

16.2 Gametogenesis in Animals

In most species there are just two types of gamete and they are radically different. The egg is amongst the largest cells in an organism, while the sperm is often the smallest. The egg and sperm are optimized in opposite ways for the propagation of the genes they carry.

Eggs develop in stages from primordial germ cells that migrate into the developing gonad early in development to become oogonia. After mitotic proliferation, oogonia be-

come primary oocytes, which begin meiotic division I and then arrest at prophase I for days to years, depending on the species. During this prophase-I arrest period, primary oocytes grow, synthesize a coat, and accumulate ribosomes, mRNAs, and proteins, often enlisting the help of other cells, including surrounding accessory cells. In the process of maturation, primary oocytes complete meiotic division I to form a small polar body and a large secondary oocyte, which proceeds into metaphase of meiotic division II. There, in many species, the oocyte is arrested until stimulated by fertilization to complete meiosis and begin embryonic development.

Typically a sperm is a small compact cell, highly specialized for fertilizing an egg. In human males, new germ cells enter meiosis continually from the time of sexual maturation, with each diploid primary spermatocyte giving rise to four haploid mature sperm. The process of sperm differentiation occurs after meiosis is complete. Maturing spermatogonia and spermatocytes fail to complete cytokinesis so that the progeny of a single spermatogonium develop as a large **syncytium**, mass of spermatids with cytoplasmic bridges. Sperm differentiation is therefore directed by the products from both parental chromosomes, even though each nucleus is haploid. Competition between sperm is fierce and the vast majority fail in their mission. Of the billions of sperm released during the reproductive life of a human male, only a few ever manage to fertilize an egg.

17 Fertilization and early embryonic development

Once released, egg and sperm alike are destined to die within minutes or hours unless they find each other and fuse in the process of **fertilization**. Of the 300,000,000 human sperm ejaculated during coitus, only about 200 reach the site of fertilization in the oviduct. Once it finds an egg, the sperm must first migrate through the layer of follicle cells and then bind to and cross the egg coat, the **zona pellucida**. Finally, the sperm must bind to and fuse with the egg plasma membrane. To become competent to accomplish these tasks, sperm must be modified in a process called **capacitation**. This occurs through the presence of bicarbonate ions (HCO_3^-) in the vagina, which enter sperm and activate a soluble adenylyl cyclase enzyme in the cytosol. This produces cAMP which helps to initiate changes associated with capacitation.

Two mechanisms operate to ensure that only one sperm fertilizes an egg. In many cases, a rapid depolarization of the egg plasma membrane prevents further sperm from fusing and thereby acts as a fast primary block to **polyspermy** (multiple fusions of different sperm with a single egg). This polarity returns to normal soon after fertilization so a second mechanism known as the egg **cortical reaction** kicks in to prevent polyspermy. When a sperm fuses with the egg plasma membrane it causes a local increase in cytosolic Ca^{2+} . This initial increase is followed by prolonged Ca^{2+} oscillations. The oscillations activate the egg to begin development, which initiates the cortical reaction. Cortical granules release their contents by exocytosis which change the structure of the zona pellucida. It becomes "hardened" so that sperm no longer bind to it.

Once fertilized the egg is called a **zygote**. Fertilization is not complete however until the two haploid nuclei have come together to form a single diploid nucleus. In fertilized mammalian eggs the two pronuclei do not fuse directly. They approach each other but remain distant until after the membrane of each pronucleus has broken down in preparation for the zygote's first mitotic division.

Part VII

Chromatin and Chromosomes

18 Karyotypes

Karyotype - a set of chromosomes viewed under the microscope, typically defined by chromosome number and other visible landmarks.

A karyotype can be used to quickly determine certain aspects of an organism's genetic profile (species, sex, and genetic abnormalities). The landmarks typically looked at when viewing a karyotype follow:

- Chromosome size allows for the identification of groups of chromosomes. For example in humans chromosomes can be placed into seven groups.
- Centromere position appears as a constriction at a specific position creating a ratio between the chromosome arms. Centromere positions can be categorized as telocentric (at one end), acrocentric (close to one end), or metacentric (in the middle). The shorter arm is delineated as p and longer arm q.
- Heterochromatin patterns are produced when chromosomes are treated with Feulgen reagent. Densely stained regions are called heterochromatin and reflect high compaction of the DNA. Lightly stained regions are euchromatin and indicate less densely packed regions. Active regions = euchromatin.
- Banding patterns are produced with various stains. Positions and sizes of bands are highly chromosome specific providing identifiable markers. Q bands are produced by quinacrine hydrochloride, G bands by Giemsa (binds to AT-rich regions), and R bands produced by reverse Giemsa staining.

One important thing to note about banding patterns is that a specialized banding occurs in organisms which can replicate their DNA many times without separating. This results in "giant" chromosomes with very visible banding known as polytene chromosomes. This can be seen in highly specialized cells of Malpighian tubules, rectum, gut, footpads, and salivary glands of insects such as houseflies, mosquitoes, and fruit flies.

19 Chromosome Rearrangements

Chromosome rearrangements encompass four different events: **deletions, duplications, inversions, and translocations**. Each of them can be caused by cleavage of DNA at two locations, followed by a rejoining of broken ends, forming a new gene order. Ionizing radiation can induce rearrangements artificially and results in numerous double-stranded breaks in DNA. Here are a few key points before discussing each class of rearrangement in detail:

1. Each one of these classes of rearrangement requires a double stranded break in DNA which initiates repair systems in the cell to attempt repair of the DNA.
2. Only DNA molecules with one centromere and two telomeres are recoverable rearrangements. **Acentric chromosomes** (lacks a centromere) are not inherited. **Dicentric chromosomes** (two centromeres) are dragged to opposite poles simultaneously and form **anaphase bridges**, which are not incorporated into either cell.

3. Large segment DNA loss/duplication can result in phenotypic abnormalities due to gene balance issues.

19.1 Balanced Rearrangements

Balanced rearrangements result in a change in gene order but no net gain or loss of genetic material. The following two rearrangements fall into this class. Note that for both of these rearrangements breaks can occur anywhere. Thus, breaks will often occur within genes resulting in disrupted gene function after rejoining.

19.1.1 Reciprocal Translocations

A reciprocal translocation is a rearrangement in which **acentric** fragments of two nonhomologous chromosomes trade places (Figure 39).

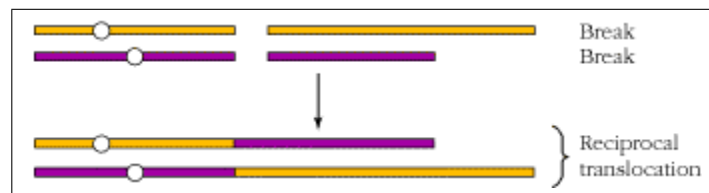


Figure 39: Reciprocal translocation

Meiosis in heterozygotes having two translocated chromosomes has important cytological effects. Pairing affinities of homologous chromosomes during meiosis results in a characteristic cross-shaped configuration in reciprocally translocated heterozygote chromosomes. Because independent meiotic assortment still plays a role in the chromosome separation two patterns occur. The segregation of each of the normal chromosomes with one translocated chromosome (T1-N2 and T2-N1) is called **adjacent-I segregation**. Often this results in inviable meiotic products. The other possible segregation is called **alternate segregation** and results in a N1-N2 and T1-T2 pairing (Figure 40). These products are viable as no genetic material is lost. Adjacent and alternate segregation occur equally frequently and thus half the overall population of gametes is nonfunctional. Due to these types of translocations pseudolinkages can be seen between genes. This is distinguished when genes on nonhomologous chromosomes appear linked providing a clue to the presence of a translocation.

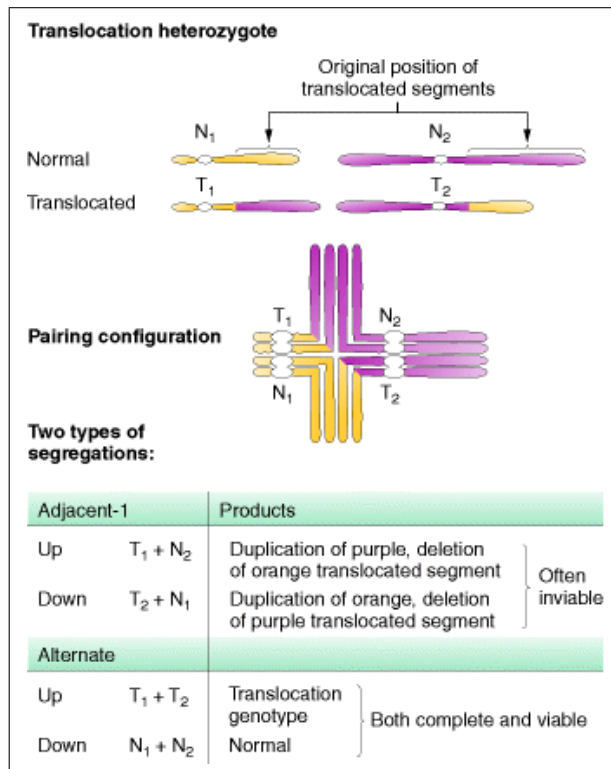


Figure 40: Alternate-I and Adjacent segregation due to reciprocal translocation.

19.1.2 Inversions

An inversion is a rearrangement in which an internal segment of a chromosome is broken twice then flipped 180 degrees and rejoined (Figure 41). There are two types of inversions

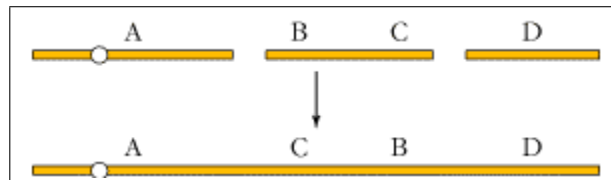


Figure 41: Inversion

which are defined by the location of the centromere. If the centromere is outside the inversion it is **paracentric**. Inversions spanning the centromere are **pericentric** (Figure 42). Pericentric inversions alter arm ratios on chromosomes, making this a good detection method. Paracentric inversions do not alter arm ratio, but can be detected with banding or other landmarks.

Typically inversions result in viable cells, which is evidenced by their breeding to homozygosity. However, some cases result in lethality when the breaks occur in a gene essential to survival. Cells that carry one normal chromosome plus one inverted are called inversion heterozygotes. Inversion heterozygotes which undergo meiosis can produce various chromosomal problems. When the chromosomes line up and pair, the normal chromosome remains untwisted, while the inverted one loops to pair gene segments with the normal chromosome. This produces an **inversion loop**. Crossing-over can then result in various new rearrangements or malformations of chromosomes. A paracentric inversion undergoing crossing-over is illustrated below (Figure 43a).

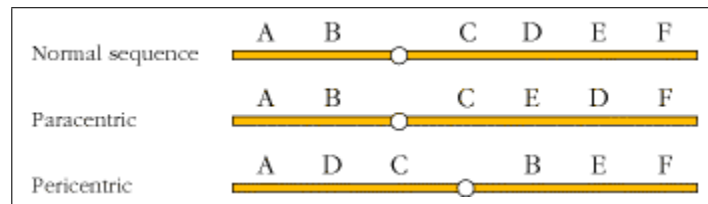


Figure 42: Types of inversion

The result of the paracentric inversion crossing-over is an acentric fragment as well as a dicentric bridge fragment. During anaphase I, the acentric fragment is lost while the dicentric bridge is eventually broken apart due to tension resulting in two chromosomes with terminal deletions. Genes within inversions have a **recombination frequency (RF)** of ZERO. RF for genes flanking an inversion is reduced in proportion to the relative size of the inversion (RF is VERY useful for gene mapping). In addition to gene mapping, one can diagnose an inversion as a genetic abnormality in a population of the progeny produced possess a much lower RF than expected. A pericentric inversion and its resulting meiotic products are illustrated (Figure 43b).

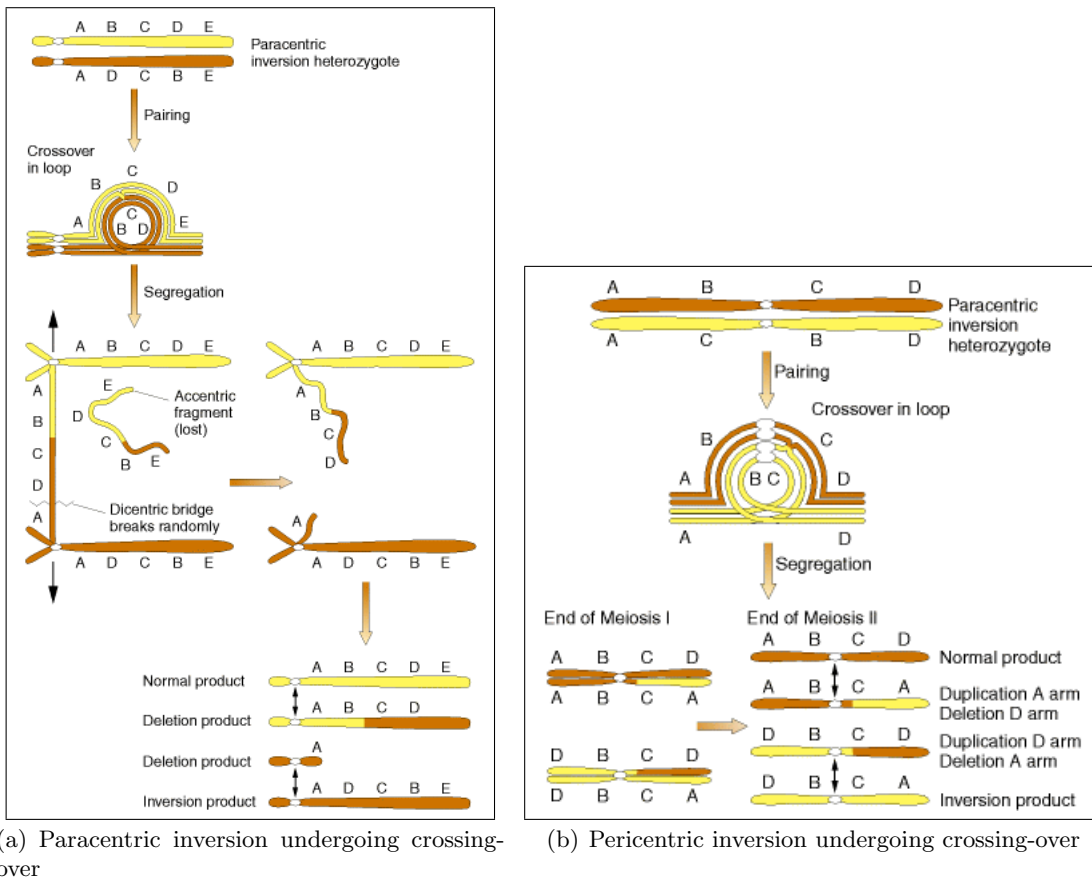


Figure 43: Inversion associated meiotic problems.

19.2 Imbalanced Rearrangements

These rearrangements result in a change in the gene dosage of affected chromosomes. The loss or gain of a copy of a segment of chromosome can disrupt gene balance.

19.2.1 Deletions

A deletion is a loss of a segment within one chromosome arm and the juxtaposition of the two segments on either side of the missing segment (Figure 44).

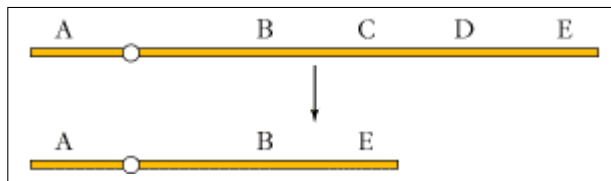


Figure 44: Deletion

A small deletion within a gene, an **intragenic deletion**, inactivates the gene and can result in a viable phenotype. These deletions are distinguished from point mutations because they are nonrevertible.

Deletions can be detected in various ways. One is to cross organisms and look for lethal variants. Typically **multigenic** deletions will result in lethal combinations simply due to

the disruption of gene balance or the uncovering of lethal recessive alleles.

Deletions can be utilized for gene mapping in a process known as **deletion mapping**. By deleting certain regions of DNA one can see where recessive alleles begin to assert dominance thus pinpointing the location of specific genes.

Most human deletions arise spontaneously in the germ line and are not witnessed in somatic chromosomes. Some deletion-bearing offspring can result from heterozygous translocations. Cri du chat syndrome is an example of this.

19.2.2 Duplications

A duplication is a repetition of a segment of a chromosome arm.

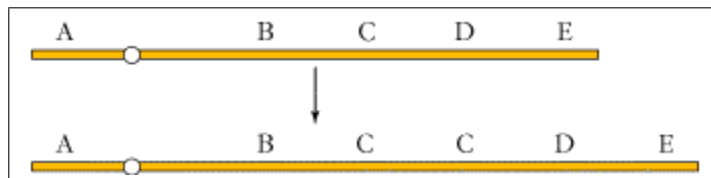


Figure 45: Duplication

Chromosome mutation sometimes results in the production of an extra copy of chromosome regions. In meiotic prophase, tandem duplication carrying heterozygotes display a loop representing the unpaired extra region of genetic material.

20 Changes in Chromosome Number

Alterations in chromosome number occur in two types: changes in whole chromosome sets (results in **aberrant euploidy**) and changes in parts of chromosome sets (results in **aneuploidy**).

20.1 Aberrant Euploidy

Organisms with multiples of the basic chromosome set are called **euploid**. Organisms with more or less than the normal set of chromosomes are called aberrant euploids. **Polyplloids** are individual organisms with more than two chromosome sets. This encompasses $3n$ (Triploid), $4n$ (Tetraploid), $5n$ (Pentaploid), and so on. If an organism is normally diploid, but suddenly only contains one set of chromosomes it is known as monoploid. An example of this is male bees, wasps, and ants.

Often a correlation between the number of chromosome set copies and the size of the organism exists. A tetraploid organism will be similar to a diploid, however it will be bigger as a whole. Amongst polyplloids two groups exist: **autopolyploids**, which are composed of multiple sets originating within one species, and **allo-polyploids** which are composed of sets from two or more different species.

20.1.1 Autopolyploids

Triploids are usually autopolyploids, arising in nature as well as constructed by geneticists through a cross of a tetraploid with a diploid. Triploids are characteristically sterile due to the inability for cells to segregate an odd number of chromosomes properly. In nature triploids are generated through a natural doubling of a $2n$ to a $4n$, then crossing with another $2n$. Plants are much more tolerant of polyploidy and are often larger when carrying a polyploid genome.

20.1.2 Allopolyploids

An allopolyploid is a plant which is a hybrid of two or more species. The prototypic allopolyploid was an allotetraploid synthesized by G. Karpechenko in 1928. He crossed two species (cabbage and radish) that had 18 chromosomes each. This hybrid was functionally sterile since the 9 chromosomes of each plant were different enough that they did not synapse for segregation. Eventually one part of the hybrid plant produced seeds and upon planting they produced fertile individuals with 36 chromosomes. Apparently the plants had derived from spontaneous chromosome doubling creating a genome which allowed for proper chromosome pairing and meiosis. One natural allopolyploid is bread wheat ($6n=42$) which appears to be composed of two sets of three ancestral genomes.

20.2 Aneuploidy

An aneuploid is an organism whose chromosome number differs from wild type by part of a chromosome set. Typically this variance is just one chromosome number greater or smaller than that of wt. Aneuploid nomenclature is based on the number of copies of the specific chromosome in the aneuploid state ($2n-1$, is monosomic).

The cause of most aneuploid conditions is nondisjunction in the course of meiosis or mitosis. In meiosis the failure is in either the first or second division.

21 Structure

A human cell contains about 2 meters of DNA. This enormous length of DNA must be efficiently packed into the nucleus. The material which comprises chromosomes is called chromatin, which is a complex of DNA and protein. At the simplest level, chromatin possesses a beaded necklace structure. The string of the necklace being the DNA and the beads nucleosomes. Nucleosomes are complexes of DNA and histones, a special chromosomal protein. Histones are remarkably conserved and each nucleosome contains an octamer of two units each of histones H2A, H2B, H3, and H4. DNA is wrapped twice around each octamer as seen in Figure 46.

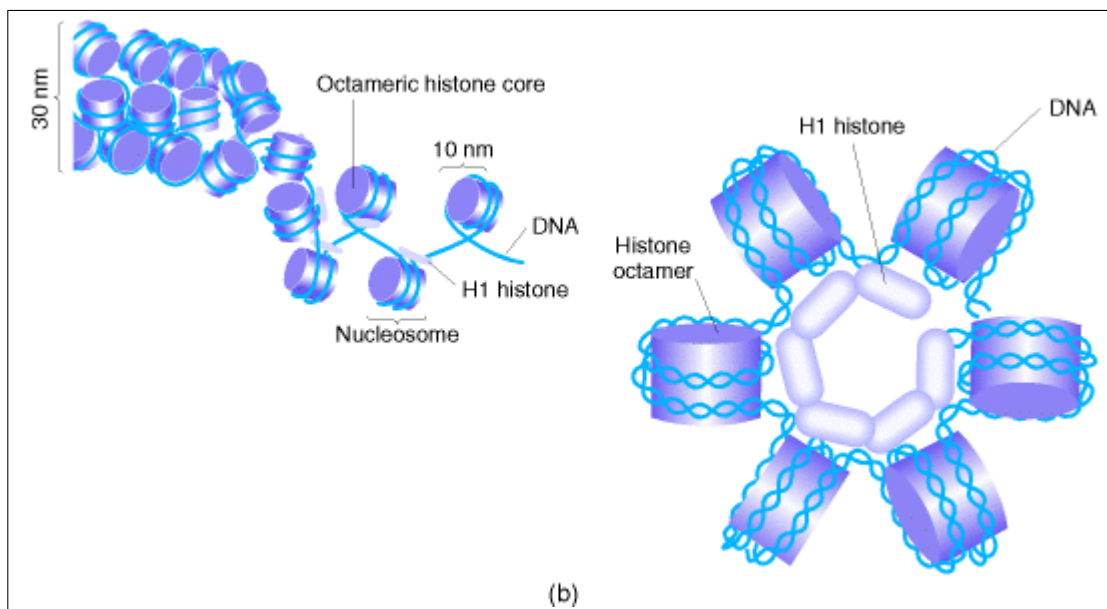


Figure 46: DNA wrapped around Histones and formation of a solenoid.

The next level of organization is that nucleosomes assume a coiled form known as a solenoid as seen above (Figure 46). The solenoid structure is stabilized by another histone known as Histone linker H1, which run down the center of the structure.

One more level of packaging exists before solenoids are converted into the final structure of a chromosome. This final level of organization is achieved through the use of a central nonhistone protein called the scaffold. DNA loops appear to project laterally off the scaffold. The solenoids appear to attach to the scaffold on special regions called scaffold attachment regions (SARs). The scaffold is mostly composed of the enzyme topoisomerase II, which can pass a strand of DNA through another strand. This final level of organization is seen in the figure below (Figure 47).

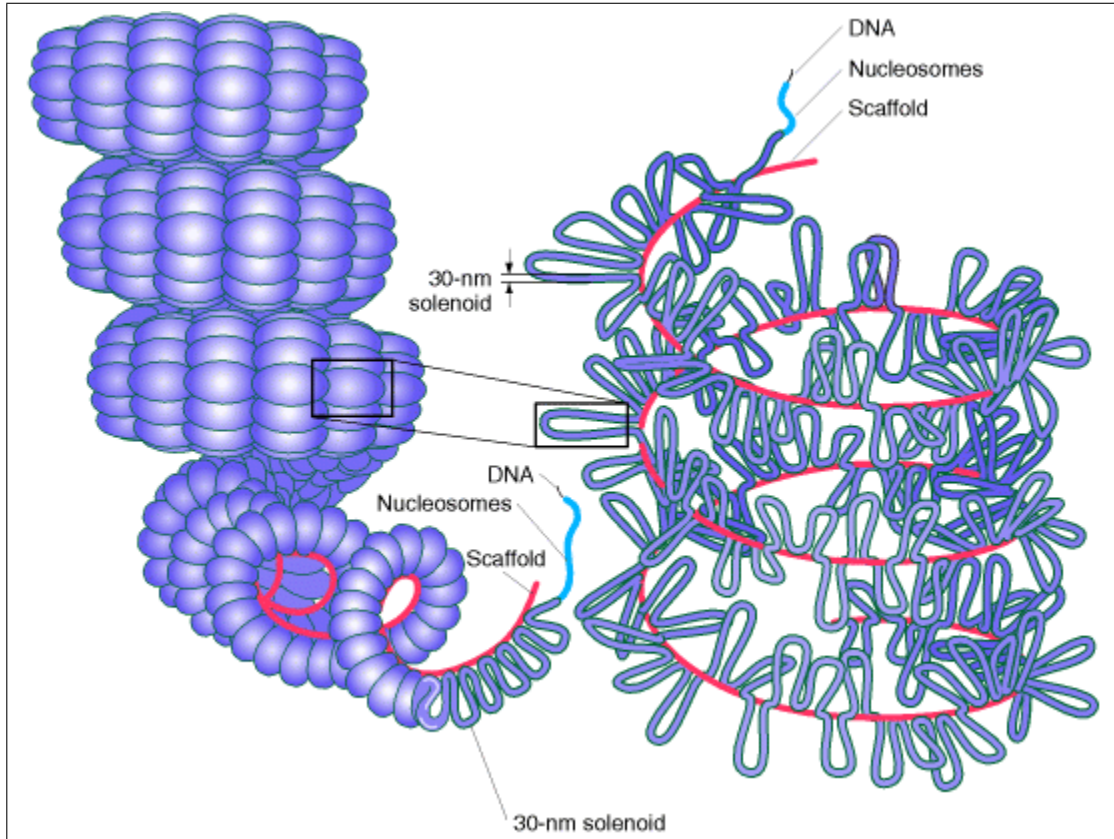


Figure 47: Solenoid coiling into chromatin.

Part VIII Genomics

22 Genome Structure

22.1 Structure of the Eukaryotic Genome

Eukaryotic organisms mostly have their genome split into two components: the nuclear genome and mitochondrial genome. The vast majority of genetic material is housed within the cell nucleus and a small portion in the mitochondria (chloroplasts in plants).

The nuclear genome is split into a set of linear DNA molecules which are organized into

chromosomes. No exceptions to the presence of chromosomes are known. All eukaryotes have at least two chromosomes and the DNA molecules are always linear. Variability within eukaryotes lies in the number of chromosomes. It is important to note that while chromosome number comparisons between organisms are interesting, the lack of correlation to any specific phenomena ultimately tells us nothing about each genome other than the simple fact that evolutionary events are non-uniform across species.

As discussed in the above section, DNA is packaged extremely efficiently into chromosomes through the use of histones. This results in a string and bead like conformation which is easily packed into dense structures that ultimately compose the chromosome. While knowing this structure can tell us a great deal about the hurdles which must be overcome for genetic information to be accessed and utilized, it does not tell us where genes are located.

It was noticed early on that chromosome staining produced highly variant banding within chromosomes. This suggested that gene distribution was uneven. A second line of evidence pointing to uneven gene distribution is derived from the **isochore** model of genome organization which contends that genomes of vertebrates and plants are composed of mosaics of segments of DNA. Each segment is at least 300 kb in length, with each segment having a uniform base composition that differs from that of the adjacent segments. Support for this model came from experiments involving DNA broken into 100 kb fragments and density gradient centrifugated. Human DNA produces five distinct fractions after this process, L1,L2 (AT-rich segments), H1,H2, and H3 (GC-rich segments). The H3 segment is the least abundant within the human genome however it contains over 25 percent of the genes. After the draft genome sequence was completed this uneven gene distribution was verified, however it suggests that the isochore theory over simplifies the complex pattern of variations within the human chromosome.

22.1.1 Genes Within the Eukaryotic Genome

There are many ways to categorize the genes in a eukaryotic genome. One way is to classify genes according to their function as shown in Figure 48.

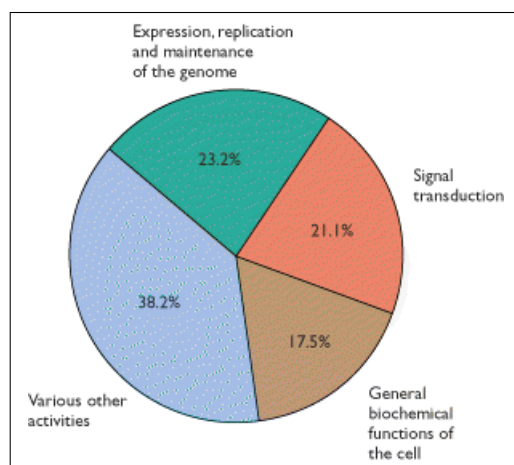


Figure 48: Functional classification of eukaryotic genes.

This method easily groups genes into broad hierarchies which can be further subdivided to produce more specific functional descriptions. However, this approach is useless for dealing with the many genes with which we have no idea the function. A more comprehensive approach is to utilize the structure of proteins encoded by genes. Analyzing protein

domains yields an effective method for categorizing gene sequences which do not have a known function. This method of classification has allowed for more accurate functional comparison between species and can be used to generate gene profiles and comparisons like the one seen below (Figure 49). Unfortunately, this data has not yet yielded any ground-breaking conclusions on genome structure or evolution.

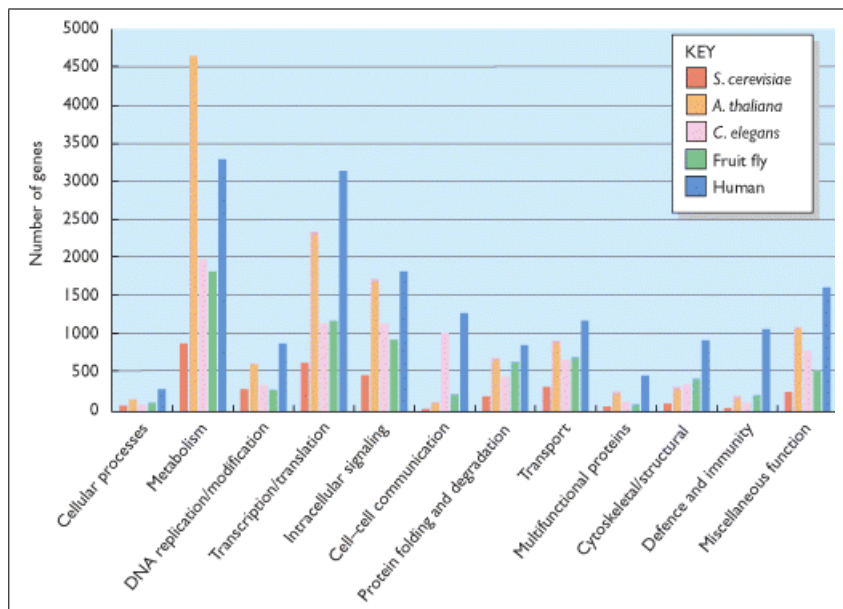


Figure 49: Species comparison for quantity of various genes categorized by function derived from protein domains.

22.2 Structure of the Prokaryotic Genome

Most prokaryotic genomes are less than 5 Mb in size, although there are a few which are substantially larger than this (e.g. *B. megaterium* with a genome of 30 Mb). The traditional view is that the prokaryotic genome is contained within a single circular DNA molecule localized within the **nucleoid**, the lightly staining region of an otherwise featureless prokaryotic cell. However, as time moves on it is becoming apparent that this traditional view may not be entirely accurate for all bacteria.

Most of what we know about the organization of DNA in the nucleoid comes from *E. coli*. The first thing noticed was that the *E. coli* genome is **supercoiled** (Positive supercoiling when turns are introduced, negative when turns are removed). This supercoiling is caused by the circular genome's inability to relieve torsional stress. Any rotation results in the genome further winding around itself to form a more compact structure. In *E. coli*, supercoiling is thought to be controlled by two enzymes, **DNA gyrase** and **DNA topoisomerase I**.

While the majority of bacterial and archaeal chromosomes are indeed circular, an increasing number have been found to be linear. This discovery coupled with the role of plasmids in the prokaryote genome complicate the typical prokaryotic genome model. *Borrelia burgdorferi* B31, contains a linear chromosomes of 911 kb, which is accompanied by 17 or 18 linear and circular plasmids which contributed another 533 kb of DNA. While the functions of most of these genes are unknown it has been noticed that the plasmids within *Borrelia* are essential components of the genome making prokaryotic genomes more multipartite than previously thought.

Another complication to the typical prokaryotic model is the presence of histone-like packaging molecules within archaeal cells. While we currently have no knowledge of archaeal nucleoid structure it is assumed that the histone-like proteins play a role in DNA packaging.

23 Repetitive DNA Content of Genomes

Repetitive DNA is found in all organisms and in some, including humans, it makes up a substantial fraction of the entire genome. While there are various types of repetitive DNA, we will look at them in two major categories: those clustered into tandem arrays and those that are dispersed around the genome.

23.1 Tandemly Repeated DNA

Tandemly repeated DNA is a common feature of eukaryotic genomes but is found much less frequently in prokaryotes. This type of repeat is also called **satellite DNA**, because the DNA fragments form satellite bands when genomic DNA is fractionated by density gradient centrifugation. Tandemly repeated DNA is thought to be generated either by replication "slippage" or by DNA recombination processes.

Alphoid DNA repeats found in the centromere regions of chromosomes are one type of satellite DNA. Some satellite DNA is scattered around the genome, most of it is located in the centromeres, where it may play a critical structural role. The repetitive DNA content of the centromere might also be a reflection of the fact that it is the last region of the chromosome to be replicated, the repetitions could ensure that there are no origins of replication present within the centromere.

Although not appearing in satellite bands on density gradients, **minisatellites** and **microsatellites** are classified as satellite DNA.

Minisatellites - an example of minisatellite DNA is Telomeric DNA (5'-TTAGGG-3').

In addition to telomeric minisatellites, some eukaryotic chromosomes contain various clusters of other minisatellite DNA near the ends of chromosomes. The functions of these minisatellites have not been identified (form clusters up to 20 kb in length, with repeat units up to 25 bp).

Microsatellites - are equally mysterious are typically consist of a 1,2,3, or 4 bp unit repeated 10-20 times as seen in the human β T-cell receptor locus. While each of these repeats is relatively short, there are many of them in the genome. While the function of microsatellites is still unknown they are useful to geneticists as they are variable within members of a species. Thus, no two humans alive today (except identical twins) contain the exact same combination of microsatellite length variants. (form clusters less than 150 bp, with repeats of 13 bp or less.

23.2 Interspersed Genome-wide Repeats

In contrast to tandemly repeated DNA, interspersed repeats arise through transposition. Most interspersed repeats contain an inherent transpositional activity. Two modes of transposition are utilized within the cell. One involves a RNA intermediate while the other does not.

23.2.1 Retrotransposition

The basic mechanism of **retrotransposition** follows:

1. An RNA copy of the transposon, a genetic element that can move from one position to another in a DNA molecule, is synthesized by normal transcription.
2. The RNA transcript is copied into DNA, which initially exists as an independent molecule outside the genome. Reverse Transcriptase is necessary for this step, and is often coded by a gene within the transposon itself.
3. The DNA copy integrates into the genome, sometimes into the same chromosomes, sometimes not. This results in two copies of the transposon.

So far three types of retroelement sequences have been elucidated.

1. LTR elements - transposable elements which contain Long Terminal Repeats at either end that play a role in the transposition process.

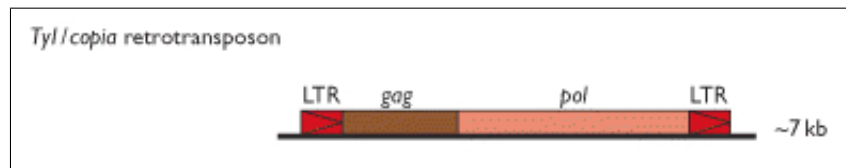


Figure 50: LTR retrotransposon.

2. LINEs (Long Interspersed Nuclear Elements) - contain a Reverse-transcriptase-like gene which is likely involved retrotransposition process.

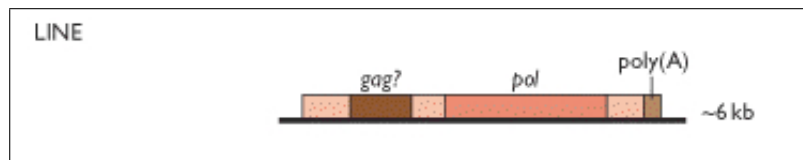


Figure 51: LINE.

3. SINEs (Short Interspersed Nuclear Elements) - do not contain a reverse transcriptase gene, but can still transpose. Most likely the RT is 'borrowed' from other retroelements.

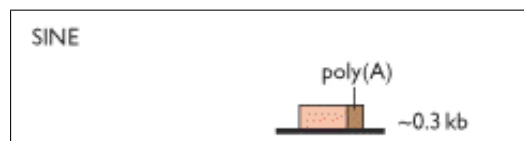


Figure 52: SINE.

23.2.2 DNA Transposons

Not all transposons require an RNA intermediate. Many are able to transpose without transcription to RNA. DNA transposons are less common than retrotransposons in eukaryotes, however they play a major role in prokaryotic genomes. Two distinct transposition mechanisms have been discovered involving direct DNA to DNA transposition.

The first method is known as **replicative transposition** and involves direct interaction between the donor transposon and the target site, resulting in copying of the donor element to its new location. The second method is **conservative transposition** and involves the excision of the element and re-integration at a new site. Both mechanisms require enzymes which are usually coded by genes within the transposon.

Both transposition mechanisms are detailed in Figure 53. DNA synthesis after ligation of the donor element with the target site ultimately determines if replicative transposition or conservative transposition is undertaken.

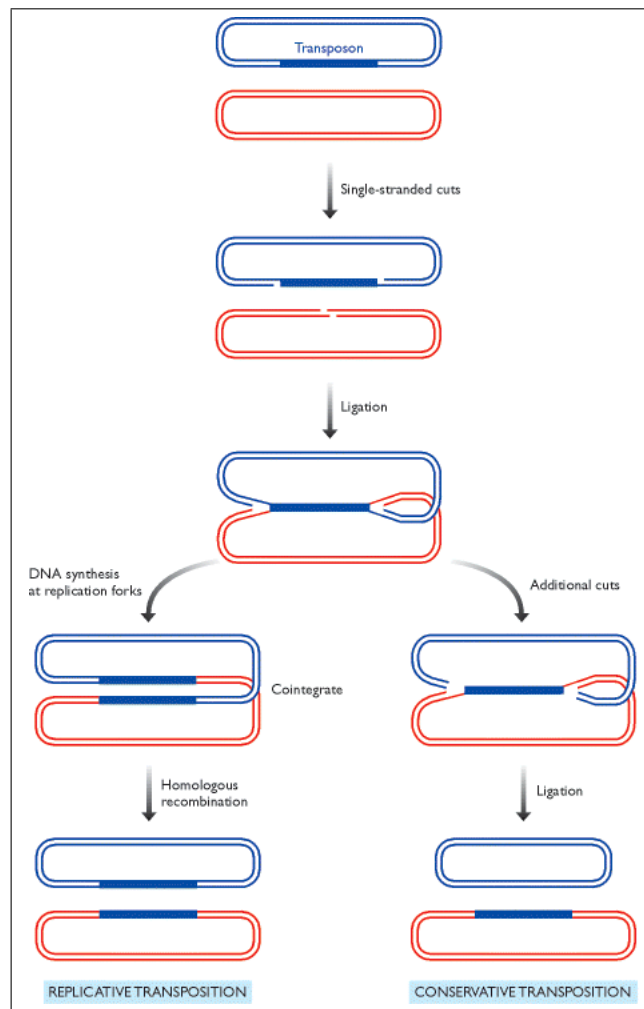


Figure 53: DNA transposition mechanisms.

24 Gene Mapping and Identification

24.1 Gene Mapping

For the purpose of mapping eukaryotic chromosomes a technique still used today is to look at linkage of genes in offspring and equating that to a spatial separation. This allows for the determination of sequences in the linear dimension of a chromosome. To understand this method lets look at the test cross used by Alfred Sturtevant to explain his method.

| | | | |
|---------------------|-----------|---|-------------|
| $pr\ vg/pr\ vg$ | 165 | } | parental |
| $pr^+\ vg^+/pr\ vg$ | 191 | | |
| $pr\ vg^+/pr\ vg$ | 23 | } | recombinant |
| $pr^+\ vg/pr\ vg$ | <u>21</u> | | |
| | 400 | | |

Figure 54: Sturtevant's test cross.

The progeny in this example represent 400 female gametes of which 44 (11 percent = .11/1.0) are recombinant. Sturtevant suggested that one could use this percentage (**recombination frequency**) as a quantitative index of the linear distance between two genes on a genetic map, otherwise known as a **linkage map**.

The basic idea is simple. Imagine two genes positioned a specific distance apart. Now with random crossing-over along the paired homologs over a large number of recombinations there should some recombinants between the two genes produced, as well as offspring with no recombination. Sturtevant postulated that there was a rough proportionality between the distance and the rate of cross over. He contended that the greater the distance between the linked genes, the greater the chance that nonsister chromatids would cross over in the region between the genes resulting in a higher **recombination frequency**. We then can define one **genetic map unit (m.u.)** as that distance between genes for which one product of meiosis in 100 is recombinant. Another way to look at this is a recombination frequency of .01 (1 percent) is defined as 1 m.u. (Sometimes a map unit is referred to as a centimorgan (cM) in honor of Thomas Hunt Morgan.)

A direct consequence of this analysis allowed Sturtevant to begin mapping genes in a linear order, by utilizing relative distances between genes with one another. An example of this is if 5 m.u. separate genes A and B, while 3 m.u. separate genes A and C, then B and C should be 8 or 2 m.u. apart. The place on the map/chromosome where a gene is located is called the **gene locus** and in the test cross posed by Sturtevant the locus for eye color and the locus for wing length are 11 m.u. apart.

The reciprocal of this analysis then is given a genetic distance in map units one can predict the frequencies of progeny in different classes. There is a strong implication that the "distance" on a linkage map is a physical distance along a chromosome, however a linkage map can be derived without even knowing that chromosomes exist. Thus, it is possible that linkage does not accurately map genes to their correct chromosome.

24.2 Gene Identification

As understanding of genetics and science have advanced methods for identifying genes have evolved from the baseline which was observing mutants and working backwards from the phenotypic differences observed. This classical approach was most easily performed in organisms that reproduced rapidly and are amenable to genetic manipulations. While

large populations could generate sufficient mutants for analysis often times mutagens were utilized to quickly create particular defects of interest. Unfortunately this method was not conducive for study of human genetics. Fortunately, with the highly conserved nature of some genes, we have been able to apply many of the lessons learned from the study of less complex model organisms to corresponding human genes.

One issue that can occur when working on large-scale genetic screens is that different mutants can show the same phenotype. These defects might lie in different genes which play a role in a single process. This problem can be quickly solved through the use of a **complementation test**. In the simplest complementation test an individual that is homozygous for one mutation is mated with an individual that is homozygous for another mutation that results in the same phenotype. If the two mutations are in the same gene then all the offspring will show the mutant phenotype, because there are no possible normal copies of the gene. However, if the mutations are in different genes then the resulting offspring would show a normal phenotype as they would possess one normal copy and one mutant copy of each gene. The mutations thereby **complement** each other and restore the normal phenotype.

This type of study can only begin to elucidate gene function as multiple genes can be involved with many biological pathways. Therefore it can be necessary for a scientist to carefully analyze numerous individual mutations to begin to piece together the order of gene function in biochemical pathways and slowly piece together the puzzle.

Gene mapping discussed above can then be utilized to map the genes to specific areas on chromosomes or relative to other genes.

Part IX

Genome maintenance

25 DNA Replication

DNA replication is **semiconservative** as each new duplex of DNA produced consists of an old template strand as well as a newly synthesized strand. This semiconservative method allows for several possible mechanisms of DNA strand growth. Three possible methods are shown in the following figure. In the first method one new strand derives from one origin and the other new strand derives from another separate origin. The second possibility entails one origin and one "growing fork" which moves along the DNA in one direction with both strands of DNA being copied. Some bacterial plasmids are replicated in this manner. The third possibility is that synthesis begins at one single origin and proceeds in both directions so that there are two replication forks. The available evidence suggests that this method is the most common amongst prokaryotic and eukaryotic cells. Thus, DNA is replicated semiconservatively through a bidirectional mechanism that emerges from a single origin site.

DNA replication ultimately takes place through three general steps:

1. Initiation - This step occurs when several key factors are recruited to an origin of replication. The origin is unwound forming a "replication bubble" with one replication fork at either end. The factors which bind at this step are collectively called the **pre-replication complex** and consist of the following enzymes:

topoisomerase - introduces negative supercoils into the DNA in order to minimize the torsional strain induced by the unwinding of the DNA by helicase.

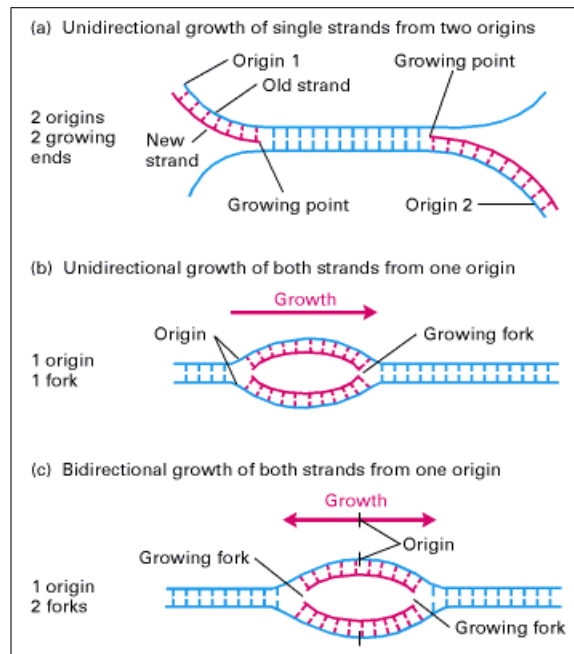


Figure 55: Different possible semiconservative DNA replication mechanisms.

helicase - unwinds and splits the DNA ahead of the fork.

single-strand binding proteins (SSB) - swiftly bind to the separated DNA preventing strands from reannealing.

primase - generates RNA primer to be used in DNA replication

DNA polymerase complex - complex consisting of DNA polymerase I and III which generate the new DNA strands.

2. Elongation - DNA polymerase III binds to the RNA primer produced by primase and synthesizes DNA from 5' to 3'. This strand is known as the **leading strand**. The other strand which replicates in pieces known as Okazaki fragments from 3' to 5' is known as the **lagging strand**. Lagging strand synthesis is also completed by DNA polymerase III. RNaseH removes RNA primers and DNA polymerase I replaces them. DNA Ligase ligates together gaps in between Okazaki fragments as well as the nucleotides placed by DNA polymerase I after removal of RNA primers. Keep in mind that in both prokaryotes and eukaryotes the DNA polymerases have proof-reading and 3' exonuclease activities resulting in increased fidelity.
3. Termination - this occurs when DNA replication forks meet one another or run to the end of a linear DNA molecule. Replication forks can also be deliberately stopped by a **replication terminator protein**. Keep in mind that the end of lagging strand synthesis results in a problem as there is no DNA template for RNA primer placement. This problem is solved by the end of chromosomes consisting of noncoding DNA. The end of these chromosomes are called **telomeres**. Cells endure the shortening of the chromosome until it reaches coding regions. **Telomerase** is an enzyme which adds repeat units to the end of the chromosome so the ends do not become too short after multiple rounds of DNA replication.

One thing to add about the above process is that replication origins between prokaryotes and eukaryotes have very different nucleotide sequences however, they share a few key

properties. First, replication origins are unique DNA segments with multiple short repeated sequences. Second, the short repeat units are recognized by multimeric origin-binding proteins. These proteins play a key role in assembling DNA polymerases and the other replication machinery at the origins of replication. In addition, the origin sequences usually contain an AT-rich stretch. This facilitates unwinding of duplex DNA as AT base pairs require less energy to melt than GC base pairs.

26 DNA Damage and Repair

In addition to the proofreading activity of DNA polymerases that can correct miscopied bases during replication cells have evolved mechanisms for repairing DNA damaged by chemicals or radiation. Unfortunately, repair mechanisms are relatively inefficient and not all genetic damage is repaired in a timely manner which ultimately lead to cancer.

Most DNA-repair mechanisms have been studied most extensively in *E. coli* using a combination of genetic and biochemical approaches. These studies have found three broad categories of repair mechanisms:

- Mismatch repair, which occurs immediately after DNA synthesis, uses parental strand as a template to correct an incorrect nucleotide incorporated into the newly synthesized strand.
- Excision repair, which entails removal of a damaged region by specialized nuclease systems and then DNA synthesis to fill the gap.
- Repair of double stranded DNA breaks in multicellular organisms occurs primarily by an end-joining process.

26.1 Mismatch Repair of Single-Base Mispairs

Many spontaneous mutations are point mutations, which involve a change in a single base pair in the DNA sequence. They arise from errors in replication, during recombination, and particularly by **base deamination**, where a C residue is converted into a U residue.

Bacterial and eukaryotic cells have a mismatch-repair system that recognizes and repairs all single-base mispairs except C-C, as well as insertions and deletions. The *E. coli* methyl-directed mismatch-repair system often referred to as the **MutHLS** system is the one we will look at.

In *E. coli* DNA, adenine residues in a GATC sequence are methylated at the 6 position. Since DNA polymerases incorporate adenine not methyl-adenine, methylated adenine is only present on the parental strand. An *E. coli* protein called MutH, binds specifically to hemimethylated sequences and distinguishes the methylated parental strand from the unmethylated daughter strand. If an error occurs during DNA replication, another protein MutS binds to this abnormally paired segment triggering the binding of MutL, a linking protein that connects MutS with a nearby MutH. This cross-linking activates a latent endonuclease activity of MutH which cleaves specifically the unmethylated daughter strand. After the initial incision, the segment of daughter strand containing the mistake is excised and replaced with the correct DNA sequence.

26.2 Excision Repair

Cells utilize excision repair to fix DNA regions containing chemically modified bases, often called **chemical adducts**, that distort the normal shape of DNA locally. The key to

this type of repair is the ability of certain proteins to slide along the surface of a dsDNA molecule looking for bulges or other irregularities in shape of the double helix. Thymine-Thymine dimers are repaired through this mechanism and are the most common type of damage caused by UV light. These dimers interfere with both replication and transcription of DNA (Figure 56). Excision repair can also repair DNA regions containing bases altered by the attachment of large chemical groups.

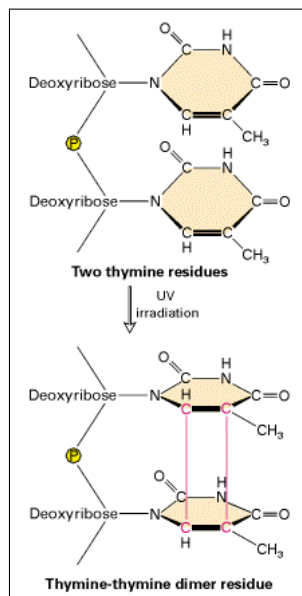


Figure 56: Thymine-thymine dimers.

The best understood excision repair example is the UvrABC system from *E. coli*. Initially a complex comprising two molecules of UvrA and one molecule of UvrB forms and binds to DNA. Both the formation and binding of the complex requires ATP. It seems likely that the UvrA-B complex binds to an undamaged segment and translocates along the helix into a distortion is detected. Translocation along the helix also requires ATP. An ATP-dependent conformational change in the damaged DNA region bound to the UvrA-B complex then produces a bend, in the DNA backbone. After the UvrA dimer dissociates, the UvrC protein, which has endonuclease activity, binds to the damaged site. It is thought that UvrC opens up space in the DNA allowing cleavage of the DNA. In the case of thymine-thymine dimers UvrC cleaves eight nucleotides 5' to the dimer and four to five nucleotides 3'. After cleavage the fragment is removed by a helicase and degraded with the gap repaired by DNA polymerase I and DNA ligase (Figure 57).

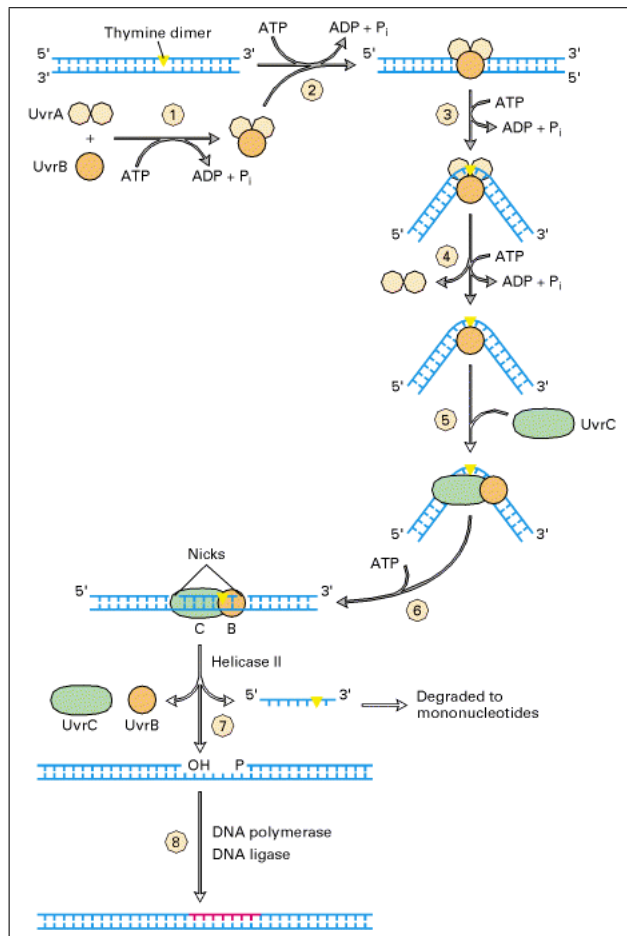


Figure 57: UvrABC excision repair system.

26.3 End-Joining Repair of Nonhomologous DNA

Double-strand breaks are particularly damaging for cells as they can result in improper joining of broken ends from different chromosomes. Such translocations may trigger abnormal cell growth by placing proto-oncogenes under inappropriate control of a promoter for another gene. In multicellular organisms the predominant mechanism for repairing double-stranded breaks involves rejoining the ends of two DNA molecules through the loss of several base pairs at the joining point. Formation of these mutagenic deletions can introduce mutations (Figure 58).

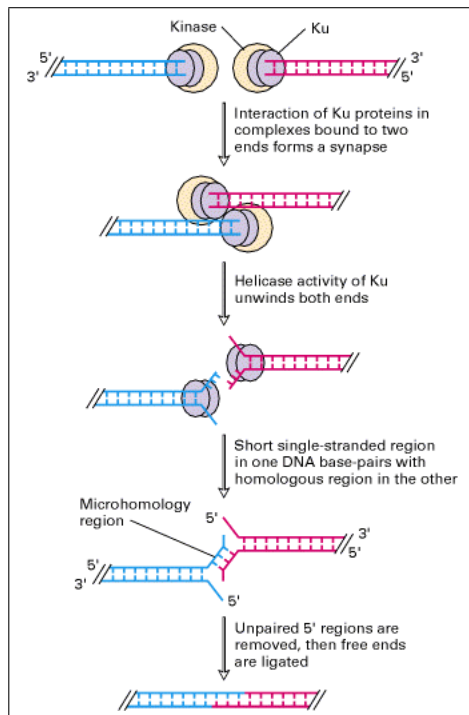


Figure 58: Repair of double-strand break.

27 DNA Recombination and Gene Conversion

In organisms genetic diversity is extremely important as it allows for response to changing environments. The DNA rearrangements that cause this variance are known as **genetic recombination**. Two broad classes of recombination are recognized: **general recombination** and **site-specific recombination**.

27.1 General Recombination

In general recombination (also known as homologous recombination), genetic exchange takes place between a pair of homologous DNA sequences. These are usually located on two copies of the same chromosome, although other DNA molecules that share the same nucleotide sequence can participate. The general recombination reaction is necessary for every proliferating cell, because accidents occur during every round of DNA replication that interrupt the replication fork and require general recombination mechanisms to repair. The details of the use of recombination machinery in this role is still incompletely understood however it appears that variations of the homologous end-joining reaction are utilized to restart replication forks that run into a break in the parental DNA template.

The central features of the general recombination mechanism seem to be the same in all organisms. Most of what we know was originally derived from studies of bacteria, especially *E. coli* and its viruses. General recombination has the following characteristics:

1. Two homologous DNA molecules that were originally part of different chromosomes break their double helices and cross over to reform two intact double helices.
2. The site of exchange can occur anywhere in the homologous nucleotide sequence of the two participating DNA molecules.

3. At the site of exchange, a strand of one DNA molecule has become base-paired to a strand of the second DNA molecule to create a **heteroduplex joint** that links the two double helices.
4. No nucleotide sequences are altered at the site of exchange; some DNA replication usually takes place, however cleavage and rejoining events occur so precisely that not a single nucleotide is lost or gained. Despite this precision however, the heteroduplex joints can tolerate a small number of mismatched base pairs and the sequences are usually not exactly the same on either side of the joint. This allows for novel sequences to be created.

Keep in mind that heteroduplex joints can be thousands of bp long. This ensures that general recombination can only occur at regions with an extensive sequence similarity.

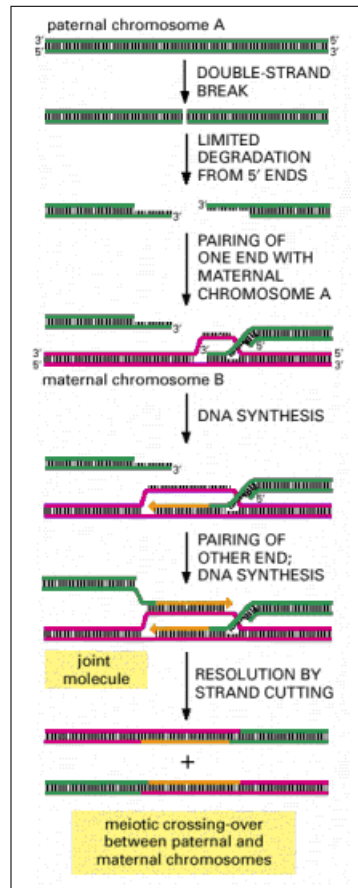


Figure 59: General process of general recombination in meiosis.

Recombination occurs after a process called **DNA synapsis**, in which base pairs form between complementary strands from the two DNA molecules, takes place. The DNA synapsis triggers recognition of the event and prompts the general recombination process to take place. Keep in mind that DNA synapsis can only occur after a special endonuclease simultaneously cuts BOTH strands of the double helix, chews back the 5' ends leaving a sticky 3' end. This free strand can then search for a homologous DNA helix for synapsis. The partner that a DNA single-strand needs to find in the synapsis step of general recombination is a DNA double helix, NOT another single strand. A single-strand DNA-binding-like protein known as **RecA** is responsible for allowing a single-stranded DNA to pair with a homologous region of double helix DNA. Each RecA monomer possesses more

than one DNA-binding site and can hold a single strand and a double helix together. This allows it to catalyze a multistep DNA synapsis reaction. The region of homology is identified before the duplex DNA target is opened up, through a three-stranded intermediate in which the DNA single strand forms transient base pairs with bases that flip out from the helix in the major groove of the double-stranded DNA molecule.

Once DNA synapsis has occurred, the short heteroduplex region where the strands from two different DNA molecules have begun to pair is enlarged through a process called **branch migration**. Branch migration can take place at any point where two single DNA strands with the same sequence are attempting to pair with the same complementary strand. In this case, an unpaired region of one of the single strands displaces a paired region of the other single strand, moving the branch point without changing the total number of DNA base pairs. Branch migration can occur spontaneously thus the RecA protein catalyzes unidirectional branch migration, readily producing a region of heteroduplex DNA that is thousands of bp long. This catalysis depends on the fact that RecA is a DNA-dependent ATPase. RecA thus "treadmills" unidirectionally along a DNA strand driving branch migration.

The synapsis that exchanges the first single strand between two different DNA double helices is presumed to be the limiting step in a general recombination event. After this step, extending the region of pairing and establishing further strand exchanges between the two DNA helices is thought to proceed rapidly. In most cases, a key recombination intermediate, the **Holliday junction (cross-strand exchange)** forms as a result. In a Holliday junction, the two homologous DNA helices that have initially paired are held together by the reciprocal exchange of two of the four strands present, one originating from each of the helices. This structure can isomerize by undergoing a series of rotations, catalyzed by specializing proteins to form a more open structure in which both pairs of strands occupy equivalent positions as well as further rotate to produce a conformation closely resembling the initial junction except that crossing strands have been converted into noncrossing strands (Figure).

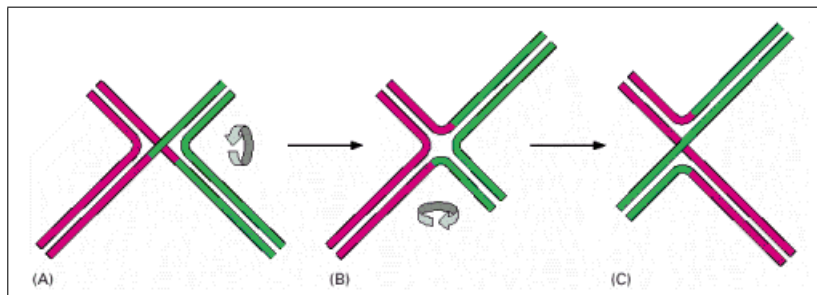


Figure 60: Holliday junction isomerization.

To regenerate two separate DNA helices the strands must eventually be cut in a process referred to as **resolution**. There are two ways in which a Holliday junction can be resolved. In one, the original pair of crossing strands is cut (the invading, or inside, strands in Figure a). In this case, the two original DNA helices separate from each other nearly unaltered, exchanging the single-stranded DNA that formed the heteroduplex. In the second possibility, the original noncrossing strands are cut (the outer strands in Figure a). This generates a much more profound alteration of the chromosomes. Major segments of double-stranded DNA are exchanged after this resolution.

General recombination thus can result in meiotic products which violate the standard rules of genetics. Occasionally, for example, meiosis yields three copies of the maternal version of the gene and only one copy of the paternal allele. This phenomenon is known as

gene conversion. It is believed to be a straightforward consequence of the mechanisms of general recombination and DNA repair. Genetic studies show that only small sections of DNA typically undergo gene conversion and in many cases only a part of a gene is changed.

27.2 Site-specific Recombination

Site-specific recombination can alter gene order and also add new information to the genome. Site-specific recombination moves specialized nucleotide sequences, called **mobile genetic elements**, between nonhomologous sites within a genome. The movement can occur between two different positions in a single chromosome, as well as between two chromosomes.

Mobile genetic elements range in size from a few hundred to tens of thousands of nucleotide pairs. Some of these elements are viruses in which site-specific recombination is utilized to move their genomes into and out of the chromosomes of their host cell. The abundant repeated DNA sequences found in many vertebrate chromosomes are mostly derived from mobile genetic elements. Over time, the nucleotide sequences of these elements have been altered by random mutation rendering them incapable of movement.

In addition to moving themselves, mobile genetic elements occasionally move or rearrange neighboring DNA sequences of the host cell genome. These movements can cause deletions of adjacent nucleotide sequences producing genetic variants.

Two different site-specific recombination mechanisms are known.

1. Transpositional site-specific recombination usually involves breakage reactions at the ends of mobile DNA segments embedded in chromosomes and the attachment of those ends at one of many different nonhomologous target DNA sites. It does NOT involve the formation of heteroduplex DNA.
2. Conservative site-specific recombination involves the production of a very short heteroduplex joint and therefore requires a short DNA sequence that is the same on both donor and recipient DNA molecules.

27.2.1 Transpositional Site-specific Recombination

Transposons are an example of transpositional site-specific recombination. An enzyme, usually encoded by the transposon, called **transposase** acts on a specific DNA sequence at each end of the transposon. Based on their structure and transposition mechanisms transposons can be grouped into three large classes.

1. DNA-only transposons, which exist as DNA throughout its life cycle. They are cut out of the donor DNA directly and joined to the target site by a transposase. Replicative transposition can also occur, which produces a copy of the donor DNA prior to transfer and insertion.
2. Retroviral-like transposons on the other hand are first transcribed to RNA, reverse transcribed back to DNA, then reinserted into the genome.
3. Nonretroviral retrotransposons are different from the second class, in that the RNA intermediate is directly involved in the transposition reaction, rather than just a messenger.

27.2.2 Conservative Site-specific Recombination

Conservative site-specific recombination mediates the rearrangements of other types of mobile DNA elements. In this pathway, breakage and joining occur at two special sites, one on each participating DNA molecule. Depending on the orientation of the recombination sites, DNA integration, excision, or inversion can occur.

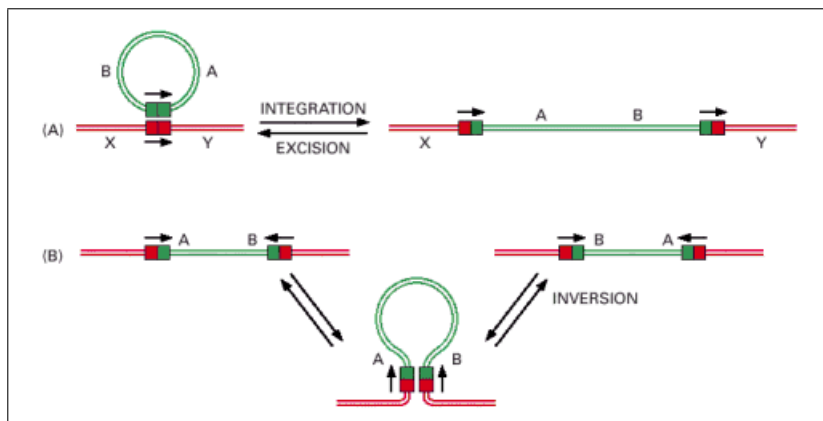


Figure 61: Possible outcomes of conservative site-specific recombination.

Interestingly site-specific recombination enzymes utilized for breaking and rejoining DNA double helices at specific sequences can do this in a reversible manner. This can allow for the restoration of the two original DNA molecules, thus the name conservative site specific recombination. *Bacteriophage lambda* was the first mobile DNA element to be understood to possess this form of recombination.

Part X

Biomolecules: Structure, Assembly, Organization, and Dynamics

28 DNA

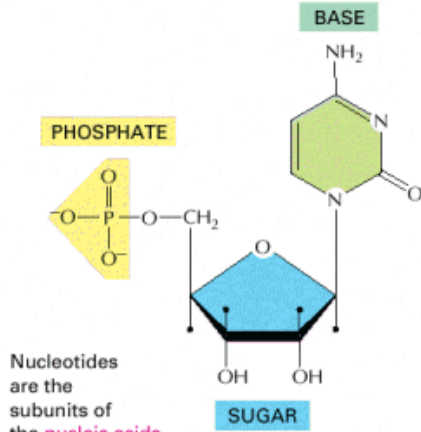
Biologists in the 1940s had difficulty accepting DNA as the genetic material because of its simple chemistry. DNA was known to be a long polymer composed of only four types of subunits, which resemble each other chemically. The discovery that DNA's structure was double-stranded and helical was a major breakthrough and provided the information necessary for understanding DNA's replicative cycle and role in information encoding.

DNA consists of two long polynucleotide chains composed of four types of nucleotide subunits. Each of these chains is known as a **DNA strand**. Hydrogen bonds between the base portions of the nucleotides hold the two chains together. Nucleotides are composed of a five carbon sugar to which are attached one or more phosphate groups and a nitrogen-containing base. In the case of the nucleotides in DNA, the sugar is deoxyribose attached to a single phosphate group and the base maybe either **Adenine (A)**, **Cytosine (C)**, **Guanine (G)**, or **Thymine (T)**(See the following panels for more detailed analysis of nucleotides).

Nucleotides are covalently linked together in a chain through the sugars and phosphates, which form the backbone of DNA. The way in which the nucleotide subunits are lined

NUCLEOTIDES

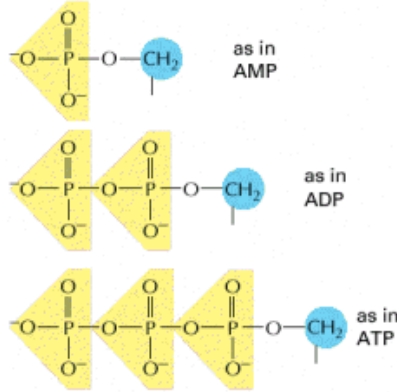
A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.



Nucleotides are the subunits of the **nucleic acids**.

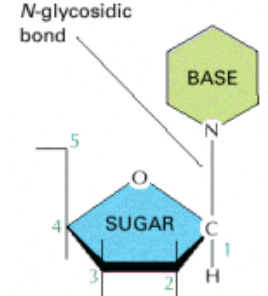
PHOSPHATES

The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar (designated 5'). Mono-, di-, and triphosphates are common.



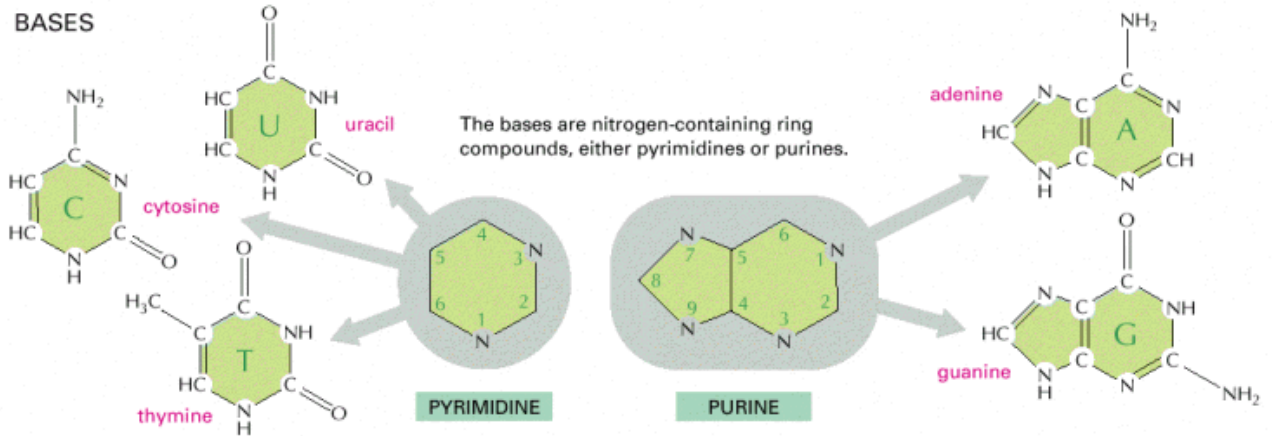
The phosphate makes a nucleotide negatively charged.

BASIC SUGAR LINKAGE



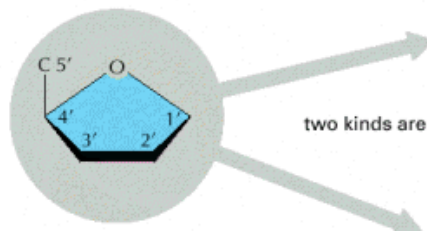
The base is linked to the same carbon (C1) used in sugar-sugar bonds.

BASES

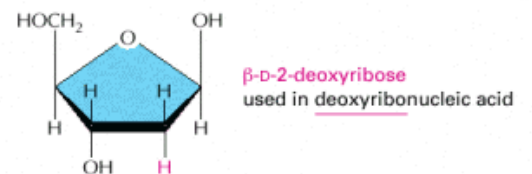
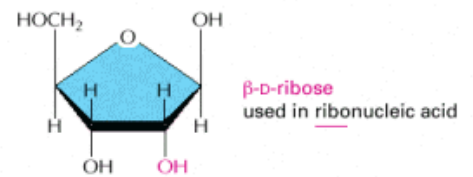


SUGARS

PENTOSE
a five-carbon sugar



two kinds are used



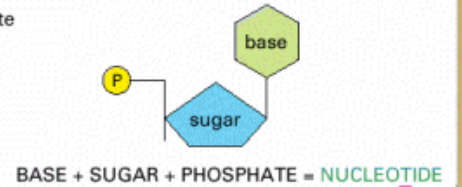
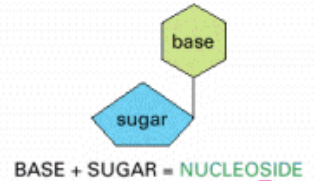
Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of the "5-prime carbon," etc.

NOMENCLATURE The names can be confusing, but the abbreviations are clear.

| BASE | NUCLEOSIDE | ABBR. |
|----------|------------|----------|
| adenine | adenosine | A |
| guanine | guanosine | G |
| cytosine | cytidine | C |
| uracil | uridine | U |
| thymine | thymidine | T |

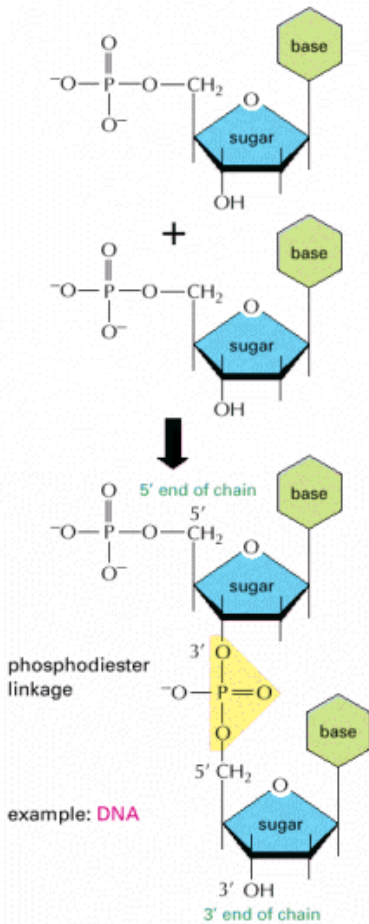
Nucleotides are abbreviated by three capital letters. Some examples follow:

AMP = adenosine monophosphate
 dAMP = deoxyadenosine monophosphate
 UDP = uridine diphosphate
 ATP = adenosine triphosphate



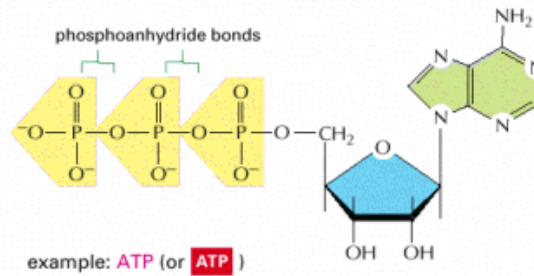
NUCLEIC ACIDS

Nucleotides are joined together by a **phosphodiester linkage** between 5' and 3' carbon atoms to form nucleic acids. The linear sequence of nucleotides in a nucleic acid chain is commonly abbreviated by a one-letter code, A-G-C-T-T-A-C-A, with the 5' end of the chain at the left.

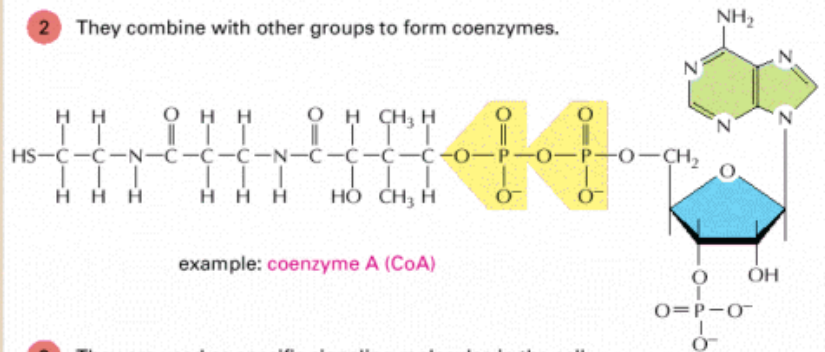


NUCLEOTIDES HAVE MANY OTHER FUNCTIONS

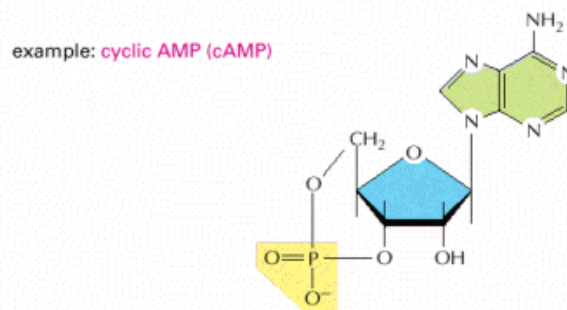
- 1 They carry chemical energy in their easily hydrolyzed phosphoanhydride bonds.



- 2 They combine with other groups to form coenzymes.



- 3 They are used as specific signaling molecules in the cell.



together provides DNA strands with a chemical polarity. The 5' phosphate signifies one pole, while the 3' hydroxyl signifies the other. These two sides could be viewed as a knob and a hole respectively in which linkage can occur.

The three dimensional structure of DNA arises from the chemical and structural features of its two polynucleotide chains. Base pairing always occurs with a bulkier two-ring base (a **purine**) paired with a single-ring base (a **pyrimidine**). Complementary base-pairing of A to T and G to C enables the base pairs to pack in the most energetically favorable arrangement in the interior of the double helix, with the sugars and phosphates on the outside. To maximize the efficiency of base-pair packing, the two sugar phosphate backbones wind around each other to form a double helix, with one complete turn every ten base pairs. The strands of each double helix run **antiparallel**, that is the polarity of one strand runs opposite to the other. Keep in mind that the pyrimidine and purine pairing creates two different grooves due to the width of the molecule during turns. A major groove (the larger) and a minor groove (the smaller) can be seen when viewing 3D models of a double helix. These grooves serve as important reference areas for proteins involved in various DNA functions.

29 RNA

Construction of RNA is pretty much identical to DNA except that it does not typically take on a double stranded conformation within the cell. In addition, RNA nucleotides contain the sugar **ribose**, which contains an O on the 2nd carbon that is not present in deoxyribose (hence the name). Also, RNA utilizes a nucleotide called **Uracil** (U) rather than Thymine. Uracil is structurally identical to Thymine minus a methyl group (Figure).

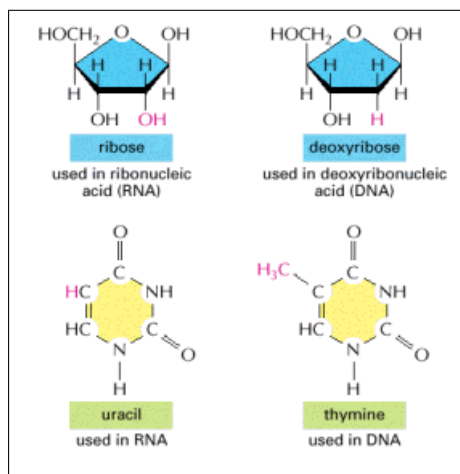


Figure 62: Differences between RNA and DNA

30 Polysaccharides

As the name implies, polysaccharides (sugar chains) are made up of monosaccharides. Monosaccharides can be described as polyhydroxy aldehydes and have the general empirical formula $(\text{CH}_2\text{O})_n$, where then n is an integer number ranging from 3 to 9. Regardless of the number of carbons, all monosaccharides can be grouped into one of two general classes: aldoses or ketoses. Aldoses contain a functional aldehyde, whereas ketoses contain

a functional ketone. Subclasses are then distinguished based on the number of carbon atoms according to the following terms: aldotriose, ketotriose, ketotetrose, etc.

When a glycoside linkage is formed between a hemiacetal (ROH) and an alcohol (R'OH) a full acetal (ROR') results. This full acetal product is a disaccharide. The glycosidic linkage represents the covalent bond of all monosaccharide-monosaccharide interactions. The glycosidic linkage involves the anomeric hydroxyl group, in α or β configuration (this differentiation is important in amylose vs. cellulose). Examples of these linkages can be seen in the figure below.

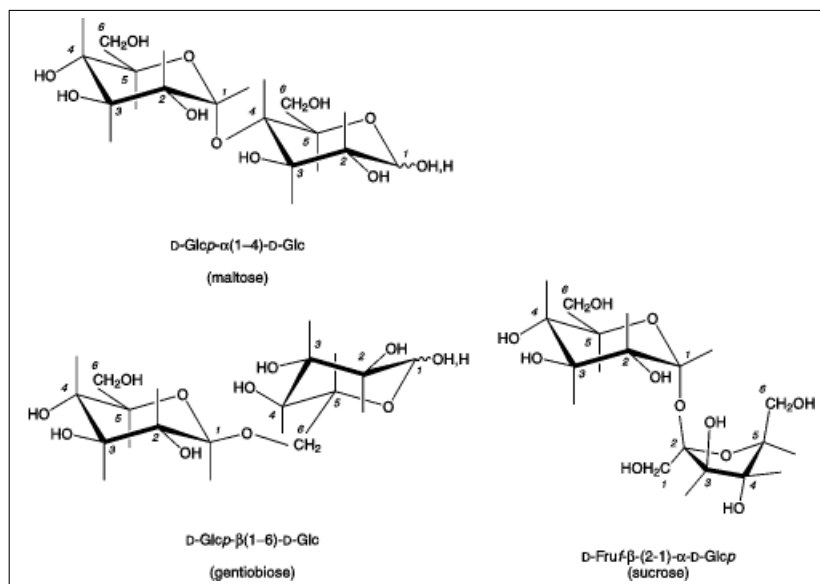


Figure 63: Disaccharide examples.

Two disaccharides that differ only in the position or configuration of their linkage can have significantly different conformations, which account for distinctive physical properties (e.g. maltose and gentiobiose in the above figure). This trend continues in the higher order oligo- and polysaccharides as well. In addition, the glycosidic linkage is the potentially most flexible part of a disaccharide structure.

Depending on the number of monomers linked to each other via glycosidic bonds, an oligosaccharide is termed a disaccharide, trisaccharide, or so on, with an upper limit of 10 which is the arbitrary cut off for polysaccharide status. The term homopolysaccharide is used to indicate a carbohydrate polymer that is composed of identical monosaccharide residues, such as cellulose and amylose (made up of glucose).

31 Proteins

From a chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known. A protein molecule is made from along chain of amino acids; each linked to its neighbor through a covalent **peptide bond** (bond formed when the carbon atom from the carboxyl group of one amino acid shares electrons with the nitrogen atom from the amino group of a second amino acid. A molecule of water is lost in this condensation reaction). Proteins are therefore known as **polypeptides**. Each type of protein possesses a unique sequence of amino acids, exactly the same from one molecule to the next.

The repeating sequence of atoms along the core (amino acid structure without the R

group) of a polypeptide chain are referred to as the **polypeptide backbone**. The side chains that protrude from this backbone vary depending on which amino acid is in each position. The amino acid side chains all possess different properties, however important groupings are polar, nonpolar, negatively charged, and positively charged. A brief list of the 20 different amino acids and their grouping can be found below (Figure).

| AMINO ACID | | SIDE CHAIN | | AMINO ACID | | SIDE CHAIN | |
|---------------|-----|------------|-----------------|---------------|-----|------------|----------|
| Aspartic acid | Asp | D | negative | Alanine | Ala | A | nonpolar |
| Glutamic acid | Glu | E | negative | Glycine | Gly | G | nonpolar |
| Arginine | Arg | R | positive | Valine | Val | V | nonpolar |
| Lysine | Lys | K | positive | Leucine | Leu | L | nonpolar |
| Histidine | His | H | positive | Isoleucine | Ile | I | nonpolar |
| Asparagine | Asn | N | uncharged polar | Proline | Pro | P | nonpolar |
| Glutamine | Gln | Q | uncharged polar | Phenylalanine | Phe | F | nonpolar |
| Serine | Ser | S | uncharged polar | Methionine | Met | M | nonpolar |
| Threonine | Thr | T | uncharged polar | Tryptophan | Trp | W | nonpolar |
| Tyrosine | Tyr | Y | uncharged polar | Cysteine | Cys | C | nonpolar |

┌ POLAR AMINO ACIDS ───────────┐
┌ NONPOLAR AMINO ACIDS ───────────┐

Figure 64: The amino acids and their properties.

The constraint that the side chains impose due to their varying sizes and charges of each amino acid in a polypeptide chain define the folding pattern of the polypeptide. Regardless of this however, any polypeptide can fold in an enormous number of ways. Keep in mind that not all of the conformations are functional or the lowest energy conformation.

The folding of a protein chain is however further constrained by many different sets of weak noncovalent bonds that form between one part of the chain and another. These interactions involve atoms in the backbone as well as amino acid side chains. These weak bonds are hydrogen bonds, ionic bonds, and van der Waals attractions. Many weak bonds can act in parallel to hold two regions of a polypeptide chain tightly together.

Finally, the hydrophobicity of the amino acid side chains play a major role in guiding protein folding. The nonpolar side chains in a protein tend to cluster in the interior of the molecule thus allowing them to avoid contact with the water that surrounds them inside a cell. In contrast polar amino acids are typically on the outside of proteins, however they can be found buried in the core of proteins, usually bonded to other polar amino acids or to the polypeptide backbone.

Ultimately, proteins fold into a particular three-dimensional structure that provides the lowest free energy. Protein folding occurs spontaneously and all the information needed for specifying the three-dimensional shape of a protein is contained in its amino acid sequence.

While folding can occur without outside interference, within a cell protein folding is often assisted by **molecular chaperones**. These proteins bind to partly folded polypeptide chains and help them progress along the most energetically favorable folding pathway. Chaperones are vital in the crowded conditions of the cytoplasm since they prevent the temporarily exposed hydrophobic residues from aggregating incorrectly.

Despite the apparent complexity of protein folding, study of protein structure has found that most proteins follow certain folding patterns. The first pattern discovered is called the α helix, which is generated when a single polypeptide chain twists around on itself to form a rigid cylinder. A hydrogen bond is made between every fourth peptide bond linking the C=O of one peptide bond to the N-H of another (Figure). This gives rise to a regular helix with a complete turn every 3.6 amino acids. Short regions of α helix are especially abundant in proteins located in cell membranes, such as transport proteins and receptors. In some proteins α helices wrap around each other to form a particularly stable structure

called a **coiled-coil** which was seen earlier with α -keratin as well as myosin.

The second pattern is called a β sheet, which can form either from neighboring polypeptide chains that run in parallel or antiparallel. Both types of β sheet produce a very rigid structure, held together by hydrogen bonds that connect the peptide bonds in neighboring chains (Figure).

Biologists ultimately distinguish four levels of organization in the structure of a protein. The amino acid sequence is **primary**, stretches of polypeptide forming α helices and β sheets are **secondary**. The full three-dimensional structure of a chain is referred to as **tertiary** and in multiple subunit proteins that structure is termed **quaternary**.

Protein structure can be analyzed based on globular regions of 40-350 amino acids known as **domains**. Small proteins can consist of a single domain, while large proteins are formed from several domains linked together. Thus far, we know of about 1000 different ways of folding up a domain.

Keep in mind that while most complex structures involving proteins within a cell can assemble spontaneously, some steps are irreversible and thus when disrupted, may not reassemble without outside help.

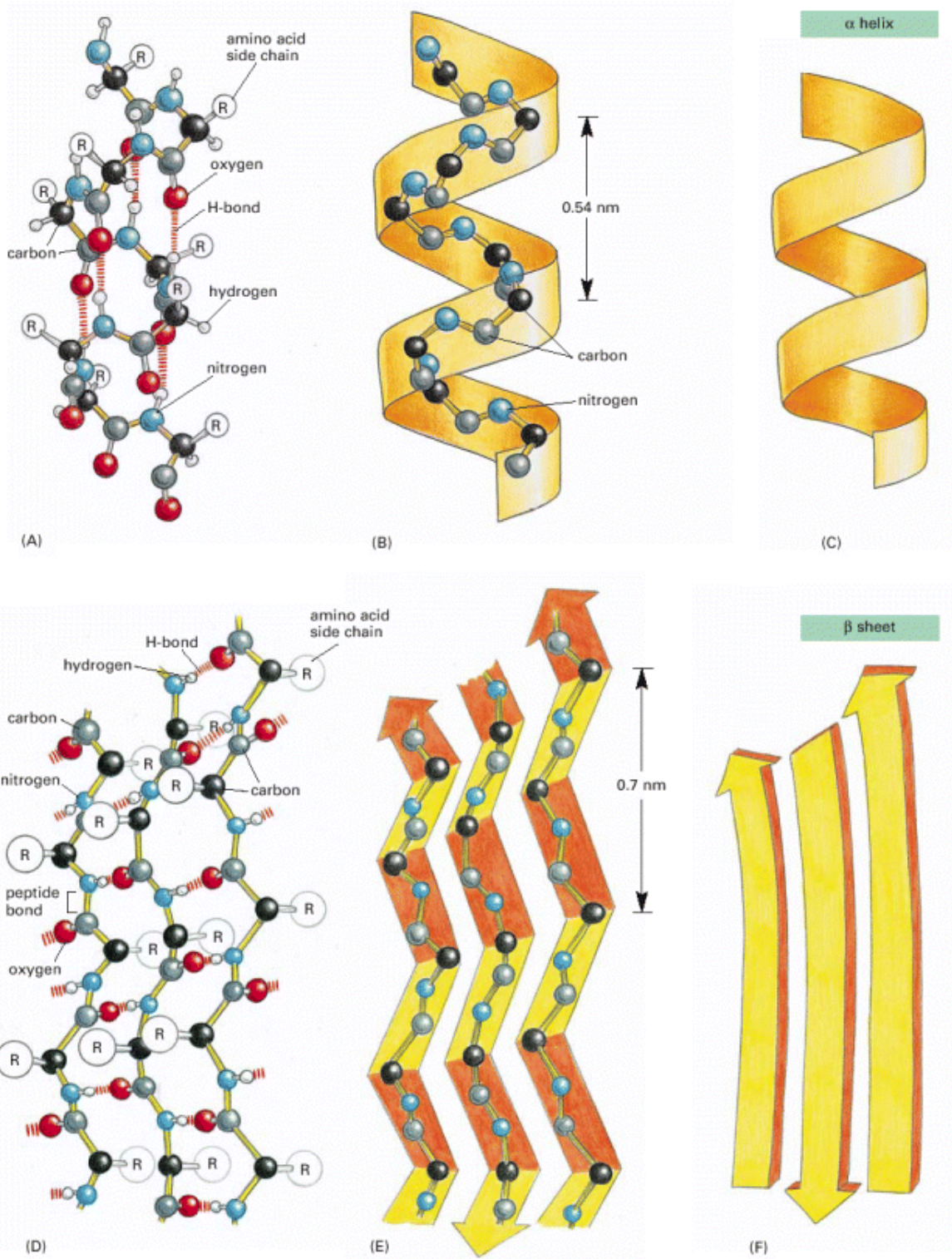


Figure 65: Examples of α helix and β sheet conformations.