

Notes for BMB/Micrb 251 (lec 1-21)

Molecular & Cellular Biology I Fall 2005

Textbook: Molecular Biology of the Cell, 4th Edition

Author: Alberts et al.

Instructor: Dr. B. Franklin Pugh

Meeting time: 12:20-1:10 MWF

Meeting place: 26 Hosler

CHAPTER 1 CELLS AND GENOMES

THE UNIVERSAL FEATURES OF CELLS ON EARTH

The basic unit of life is the **cell**. Fig. 1-1
Most forms of life are just a single cell.
Humans have 10^{13} cells, all derived from a single cell!
Cells must consume energy in order to grow and multiply.

All Cells Store Their Hereditary Information in the Same Linear Chemical Code (DNA)

DNA: **A, T, C, G** (each is called a **nucleotide**)
computer: 0, 1

All Cells Replicate Their Hereditary Information by Templated Polymerization

Each nucleotide is composed of three parts: **phosphate, sugar**, and a **base**. Fig. 1-2
Only the base is different between A, T, C, G.
A polymer of nucleotides is called a **polynucleotide**.
The unique arrangement of nucleotides forms the **genetic code**.
The genetic code provides all the information necessary to make an organism.
When cells grow and multiply, they must duplicate the genetic code.
The polynucleotide provides a template for its own **replication**.
A only pairs with T
C only pairs with G

All Cells Transcribe Portions of Their Hereditary Information into the Same Intermediary Form (RNA)

In order to use the information in the genetic code, the DNA must be 'read'. Fig. 1-4
Transcription is the process by which parts of the DNA are 'read'.
Transcription is similar to DNA replication, except that an **RNA** polynucleotide is made.
DNA: A, T, C, G
RNA: A, U, C, G (a 'U' is used instead of a 'T').
Same pairing rules apply.
(RNA also has an extra -OH group on each sugar.)
RNA and DNA are two different kinds of polynucleotides.
Different RNAs have different sequences of nucleotides.
Some RNAs direct chemical reactions (more on this later). Fig. 1-6
Messenger RNA (mRNA) codes for the production of **proteins**.
Translation is the process by which mRNA is 'read' into protein. Fig. 1-4

All Cells Use Proteins as Catalysts

Proteins do almost all the work in the cell.

Proteins make each cell different.

Proteins are made up of a linear polymer of **amino acids**.

Proteins range in size (~100-1000 amino acids).

There are 20 different kinds of amino acids.

The arrangement of the amino acids is dictated by the DNA → RNA nucleotide sequence.

A protein is also called a **polypeptide**.

Polypeptides fold up into very precise 3-dimensional structures. **Fig. 1-7a**

Proteins that catalyze (facilitate) chemical reactions are called **enzymes**. **Fig. 1-7b**

Other proteins also serve as signaling molecules and provide structure to the cell.

All Cells Translate RNA into Protein in the Same Way

Three nucleotides at a time are translated into one amino acid. **Fig. 1-9**

A group of three nucleotides that code for an amino acid is called a **codon**.

There are 64 possible codons (4x4x4), that code for 20 amino acids.

Codons are translated by **transfer RNAs (tRNA)**.

The anticodon part of the tRNA pairs with the codon.

So, there are many different kinds of tRNAs

Each kind of tRNA is attached to a particular amino acid.

The **ribosome** uses mRNA as a template to align the tRNAs, which then allows the amino acids to be stitched together. **Fig. 1-10**

The ribosome is composed of mostly **ribosomal RNA (rRNA)** and ribosomal protein.

The Fragment of Genetic Information (DNA) Corresponding to One Protein (or One Functional RNA) Is One **Gene**

Not all of the DNA has genes.

Certain stretches of DNA regulate the 'expression' of genes.

When the cell needs to make a particular protein, it 'read' or 'expresses' the corresponding gene.

The entire sequence of DNA of an organism is called a **genome**.

Life Requires Free Energy

Cells take energy (food) from its environment and use it to build more of itself.

All Cells Function as Biochemical Factories Dealing with the Same Basic Molecular Building Blocks

All Cells Are Enclosed in a **Plasma Membrane** Across Which Nutrients and Waste Materials Must Pass

The plasma membrane (or cell membrane) is a 'sack' that keeps all cellular components together.

It keeps unwanted things out.

It allows food and other important things to enter. **Fig. 1-13**

It is composed of lipids (fats).

A Living Cell Can Exist with Fewer Than 500 Genes **Fig. 1-14**

Humans have over 30,000 genes.

Summary

The cell is the minimal operational unit.

All info to make a cell is stored in DNA.

Central Dogma: DNA makes RNA makes protein.

Proteins do much of the work in the cell.

THE DIVERSITY OF GENOMES AND THE TREE OF LIFE

Most of life on earth is **microorganisms** (single-celled),

Cells Can Be Powered by a Variety of Free Energy Sources

Where do cells get their energy to make more of themselves?

Inorganic chemicals

The sun

Other organisms

Some Cells Fix Nitrogen and Carbon Dioxide for Others

The Greatest Biochemical Diversity Is Seen Among Prokaryotic Cells

Prokaryotes have no nucleus. **Fig. 1-18**

They live in a wide variety of habitats (hydrothermal vents, Arctic, bogs, sea, dirt, other organisms)

Are microorganisms.

Eukaryotes have a nucleus. **Fig. 1-43b**

A **nucleus** is an intracellular compartment that houses DNA.

Eukaryotes can be microorganisms or multicellular.

The Tree of Life Has Three Primary Branches (domains): **Bacteria, Archaea, and Eucaryotes** **Fig. 1-21**

Bacteria and Archaea are prokaryotes.

Note how diverse the prokaryotes are.

Note that plants, animals, and fungi are highly related!

Some Genes Evolve Rapidly; Others Are Highly Conserved

When DNA is replicated, mistakes are made (albeit very rarely).

Most mistakes (**mutations**) have little effect on the organism.

Some are detrimental to the organism. **Fig. 1-22**

Such **mutant** organisms are eliminated by **natural selection** (ability/inability to thrive or compete with other organisms for survival).

Therefore the mutation is also lost.

Such regions of DNA are therefore highly conserved and are indicative of important genetic information.

In rare cases, the mistake changes the genetic code for a protein in a beneficial way.

This is the core of **evolution**.

Most Bacteria and Archaea Have 1000–4000 Genes.

Natural selection favors those organisms that can reproduce the fastest.

Small size

Small genome

Specific environmental niche

New Genes Are Generated from Preexisting Genes **Fig. 1-23**

Mutation

Nucleotides within a gene can mutate.

Happens very frequently with HIV (as an example).

Duplication

Duplicating a gene allows one to mutate, while the other provides the essential function.

Segment shuffling

Two or more genes can be broken up and pieced back together differently.

Horizontal transfer

A gene from one organism can be transferred to a related or unrelated organism.

Humans have bacterial DNA in their genomes!!

Gene Duplications Give Rise to Families of Related Genes Within a Single Cell

Related organisms have related genes. The related genes are said to be **orthologs**.

When a gene is duplicated within the same cell, allowing them to evolve separately, then these genes are **paralogs**.

Homologs refer to both orthologs and paralogs.

All homologs form a **gene family**. **Fig. 1-26**

Genes Can Be Transferred Between Organisms, Both in the Laboratory and in Nature

Bacterial viruses (**bacteriophages**) are mobile genetic vehicles that allow genes to move horizontally.

Bacteria can also take up DNA from their environment.

This and the rapid rate of replication, allow bacteria to evolve rapidly. **Fig. 1-21**

Think antibiotic resistance.

Horizontal Exchanges of Genetic Information Within a Species Are Brought About by Sex

Primordial life may have extensively used horizontal transfer.

Groups of genes might have moved together.

Bacteria and Archaea but not Eukaryotes have similar metabolic genes

Metabolic genes are involved in getting food

Archaea and Eukaryotes but not Bacteria have similar genes that control information flow

DNA replication, transcription, translation

Horizontal gene transfer is essentially bacterial sex.

The Function of a Gene Can Often Be Deduced from Its Sequence

Genes with similar sequence have similar function.

If you know the function of one homolog, you then know the function of all homologs.

More Than 200 Gene Families Are Common to All Three Primary Branches of the Tree of Life **Table 1-2**

Mutations Reveal the Functions of Genes

How do we figure out what the function of any given gene is?

We know the function of only a small percentage of genes.

Determine function through biochemistry.

Isolate the protein coded for by a gene, and determine what chemical reaction it carries out.

Determine function through genetics.

Mutate the gene and see what effects (phenotype) it has on the organism (e.g. growth rate).

Molecular Biologists Have Focused a Spotlight on **Model Organisms**

Bacteria – *E. coli*

Eukaryote – Yeast, *Arabidopsis* (plant), fruit flies, mice, and more

Summary

All life requires energy, via inorganic chemicals, sunlight, or other organisms.

Prokaryotes represent the bulk of life's diversity and mass on earth.

There are three domains of life: bacteria, archaea, and eukaryotes.

All life evolved through mutation, duplication, shuffling, and horizontal transfer of genes.

GENETIC INFORMATION IN EUKARYOTES

Humans are eukaryotes, so we have an interest in how eukaryotes work.
Eukaryotes are much more complex than prokaryotes.

But not more evolved!!

Eukaryotes have more complex genomes, cell organization, and can be multi-cellular.

Eukaryotic Cells May Have Originated as Predators

A eukaryotic cell is ~1000x larger than a prokaryotic cell. **Fig. 1-31**

Eukaryotes have a nucleus (nuclear membrane or nuclear envelope) that compartmentalizes the DNA.

Eukaryotes also have other internal membrane compartments.

Eukaryotes have a protein cytoskeleton that gives shape to the cell. **Fig. 1-32**

Prokaryotes use a cell wall.

By rearranging the cytoskeleton, eukaryotic cells can rapidly change shape.

Eukaryotic cells can engulf bacteria (think immune system).

A primitive eukaryotic-like cell might have eaten other bacteria.

Eukaryotic Cells Evolved from a Symbiosis

Eukaryotic cells have **mitochondria**. **Fig. 1-34**

Mitochondria are membrane compartments (organelles) that convert food energy into usable chemical energy (**respiration**).

Mitochondria also have a small genome.

Some eukaryotic cells (e.g., plants) have **chloroplasts**.

Chloroplasts are organelles and have a genome.

Chloroplasts convert light energy into usable chemical energy (**photosynthesis**).

Mitochondria and chloroplast were once free-living bacteria that were engulfed by primitive eukaryotes and formed a **symbiotic** relationship. **Fig. 1-35**

Is mitochondrial and chloroplast DNA more like bacteria or eukaryotes?

Eukaryotes Have Hybrid Genomes

Eukaryotic Genomes Are Big **Fig. 1-38**

~1000x longer than bacteria, but only about 20x more genes.

99% of the eukaryotic genome does not code for genes!

Probably just 'junk'

When was the last time you removed unnecessary files from your 80 gig hard drive?

Eukaryotic Genomes Are Rich in Regulatory DNA

The Genome Defines the Program of Multicellular Development

Multi-cellular organisms have a diversity of cell types, all derived from a single fertilized egg, and all having the same genome.

Humans have skin cells, liver cells, and brain cells (sometimes).

Plants have leaf cells, flower cells, and root cells.

Different cell types are made when different subsets of genes are expressed.

When you are listening to music it's like having a different favorite play lists, depending on your mood or what your doing (e.g., breaking up w/ your boy/girlfriend vs. having a party – maybe...).

Cells are constantly sending signals to each other.

Cells at different locations get different signals.

Signals trigger the expression of particular sets of genes (favorite play list). **Fig. 1-40**

No signal – no expression. **Fig. 1-41**

Many Eucaryotes Live as Solitary Cells: the Protists

A Yeast Serves as a Minimal Model Eucaryote

Saccharomyces cerevisiae – bakers yeast and brewers yeast. **Fig. 1-43**

Fungi

Advantages:

Small genome

Easy genetics and biochemistry

Rapidly grows and divides

Inexpensive

Most cellular functions highly conserved with humans.

The Expression Levels of All The Genes of An Organism Can Be Monitored Simultaneously

As scientists, we used to study one gene at a time.

Now that the entire yeast genome has been sequenced we can study the expression of all

6300 yeast genes at a time, using DNA microarrays. **Fig. 1-45**

Each spot corresponds to the DNA of a particular gene. 6300 spots for yeast.

mRNA from a particular gene will bind (hybridize) to it's cognate spot.

If we color the mRNA first, the spot turns color.

To see how the yeast genetic program changes when they are hungry vs. when they are well-fed (glucose):

Isolate mRNA from hungry cells and color it red.

Isolate mRNA from fed cells and color it green.

Mix both together and hybridize to the DNA spots.

Red spots mean the gene was turned on in hungry cells and off in fed cells.

What do the green, yellow and black spots mean?

Arabidopsis Has Been Chosen Out of 300,000 Species As a Model Plant

The World of Animal Cells Is Represented By a Worm, a Fly, a Mouse, and a Human

worm: *Caenorhabditis elegans*

fly: *Drosophila melanogaster*

mouse: *Mus musculus*

human: *Homo sapien*

Studies in *Drosophila* Provide a Key to Vertebrate Development

The Vertebrate Genome Is a Product of Repeated Duplication

Genetic Redundancy Is a Problem for Geneticists, But It Creates Opportunities for Evolving Organisms

Mutating a gene might knock out its function, but if another gene serves the same function, there will be no phenotype.

Gene duplication provides great opportunities for the duplicated gene to evolve new functions. **Fig. 1-51**

The Mouse Serves as a Model for Mammals

Able to knock out specific genes.

Humans Report on Their Own Peculiarities

We Are All Different in Detail

Mouse – Human 90% identical

Chimp – Human 99%

Human – Human 99.9%

Summary

Eukaryotes are evolutionary less diverse than prokaryotes, but are way more complex.

Eukaryotic cells arose by symbiosis with prokaryotes.

“Playing” different parts of the genome gives rise to different cell types.

Model organisms are used as experimental surrogates to humans.

The more complex the organism the harder it is to work with (but may be a better proxy for humans).

CHAPTER 2 CELL CHEMISTRY AND BIOSYNTHESIS

THE CHEMICAL COMPONENTS OF A CELL

Cells Are Made of Relatively Few Types of Atoms

>99% of living matter is composed of six elements: P, S, C, O, H, N

(C,O,H,N) represent ~97%.

Other important elements: Cl, Ca, Mg, Si, Zn, Co, Mn, Fe, Se and others.

The Outermost Electrons Determine How Atoms Interact

Ionic Bonds Form by the Gain and Loss of Electrons

Know how a **covalent bond** differs from an **ionic bond**. Fig. 2-5

Covalent bonds are very strong and stable, essentially riveting atoms together.

Ionic bonds a positive charge interacting with a negative charge.

Ionic bonds are weak in water, because water interacts with the positive and negative charges making it hard for them to interact with each other. Fig. 2-14

Ions have gained or lost a charge (electron).

Cations are positively charged ions.

Anions are negatively charged ions.

Covalent Bonds Form by the Sharing of Electrons

Covalent bonds are strong.

Atoms form molecules through covalent bonds.

Covalent bonds also link repeating units of a polymer together

Defined as a stable chemical link between two atoms produced by sharing one or more pairs of electrons.

The amount of energy required to break a covalent bond varies depending on atoms and environment. (ave. ~90 kcal/mol.)

Enzymes are required to break covalent bonds (under normal physiological conditions).

There Are Different Types of Covalent Bonds

Polar vs. nonpolar

An Atom Often Behaves as if It Has a Fixed Radius

Different representation of molecules. Fig. 2-12

Water Is the Most Abundant Substance in Cells ~70% of a cell is water.

Some Polar Molecules Form Acids and Bases in Water

Four Types of Non-Covalent Interactions Help Bring Molecules Together in Cells **Table 2-2, and Panels 2-2 and 2-3 on pp. 112-115**

Ionic bonds

Weak; relative strength = 3

Cohesion between a positively charged atom and a negatively charged atom. **Fig. 2-14**

Water and salts are polar or charged and so can compete with these interactions, thereby weakening them.

By measuring the interactions between two molecules as a function of salt (NaCl, KCl, etc.) conc. one can get a quantitative handle on the extent of ionic interactions.

Examples:

DNA phosphates and protein lysine side chains make ionic bonds.

Protein side chains glutamate and arginine make ionic interactions.

Hydrogen bonds

Weak; relative strength = 1

Hydrogen atom with partial positive charge interacts with two electronegative atoms. **Fig. 2-15**

Two electronegative atoms (such as N and O) can share a hydrogen atom, even though a H atom can only form a single covalent bond.

Water can compete with H-bonds, thereby weakening them.

Example:

Protein secondary structure

DNA base-pairing

van der Waals interactions

Very weak; relative strength = 0.1

A large number of them can add up to generate strong interactions.

Due to asymmetric electrical charges, two atoms at very close distances will attract each other.

“Hand-in-a-glove” fit

Hydrophobic interactions

Water interacts with itself via hydrogen bonds (surface tension).

Nonpolar groups cannot interact with water and so are excluded (oil and water don't mix).

Water exclusion causes nonpolar groups to self associate.

Example:

Interior of proteins

Two surfaces of proteins

A Cell Is Formed from Carbon Compounds

Carbon represents the core constituent of all life because it can make a variety of strong covalent bonds with other elements.

Chemical groups (Functional groups) **Panel 2-1 on page 111**

- $-\text{CH}_3$ methyl
- $-\text{OH}$ hydroxyl
- $-\text{COO}^-$ carboxylate
- $-\text{CO}$ carbonyl
- $-\text{NH}_2$ amino amine
- $-\text{SH}_2$ sulfhydryl
- $-\text{PO}_3^-$ phosphate

Cells Contain Four Major Families of Small Organic Molecules **Fig. 2-17** **Sugars, Fatty acids, Amino acids, Nucleic acids**

Sugars Are Energy Sources for Cells and Subunits of Polysaccharides **Fig. 2-18,19,20**

Panel 2-4, pp. 116-117

- Monosaccharides -> disaccharides -> oligosaccharides -> polysaccharides
-
- Hexose, pentose, triose
- Be able to distinguish glucose from ribose from glycerol
- Structure: linear, branched, ring
- Functional groups: $-\text{OH}$, $\text{C}=\text{O}$
- Sugars have a variety of functions:
 - food - glucose, sucrose, glycogen, starch (6 carbon sugars or hexoses)
 - nucleic acid backbone - ribose (5 carbon)
 - cell adhesion
 - cell recognition
 - cell wall

Fatty Acids Are Components of Cell Membranes Fig. 2-21,22

Panel 2-5, pp. 118-119

- Amphiphilic: hydrophobic repeat of $-\text{CH}_2-$
- One end terminates with a carboxylate (e.g. palmitate)
- Phospholipids end with phosphate
- Be able to recognize fatty acids, phospholipids, triglycerides, glycolipids
- Fatty acids have a variety of functions:
 - food - stored as triglycerides (fat droplets in cytoplasm)
 - cell membrane and internal membranes
 - consists of phospholipids (2 FA + phosphate head group)
 - lipid bilayer
 - internal signaling

Amino Acids Are the Subunits of Proteins Fig. 2-23, 24

- Functional groups: carboxylate + central carbon (alpha) + amino
- Be able to distinguish amino acids -> peptides -> polypeptides (proteins).
- Know what a peptide bond (amide linkage) is.
- There are 20 natural amino acids.
- They differ by the functional groups (or side chain) arrayed off of the alpha carbon
- Know the properties of the side chains: nonpolar (hydrophobic), acidic, basic, polar (uncharged)
- Amino acids chart
- Amino acids serve a variety of functions:
 - enzymatic catalysis
 - cell structure
 - energy source
 - cell-cell signaling
 - toxins

Nucleotides Are the Subunits of DNA and RNA Fig. 2-26, 27, 28

Panel 2-6, pp.120-121

- Composition: phosphate + ribose sugar + pyrimidine/purine base
- DNA (2' H) vs. RNA (2' OH)
- Purine vs. pyrimidine
- Four bases: adenosine (A), guanosine (G), cytosine (C), thymine (T) (DNA) or uracil (U) (RNA)
- N-glycosidic bond
- Nucleoside (base + sugar) vs. nucleotide (base + sugar + phosphate)
- phosphodiester linkage
- nucleotide -> oligonucleotide -> polynucleotide
- base pairing: A w/ T(U); G w/ C
- Nucleotides serve a variety of functions:
 - genetic information storage (DNA, RNA)
 - carriers of chemical energy (high energy phosphoanhydride bonds): ATP
 - enzyme cofactors: CoA, NAD
 - signaling molecules: cyclic AMP

The Chemistry of Cells is Dominated by Macromolecules with Remarkable Properties

Noncovalent Bonds Specify Both the Precise Shape of a Macromolecule and its Binding to Other Molecules Fig. 2-32

Summary

Cells are composed of primarily six elements: PSCOHN

Covalent bonds stably connect atoms to form molecules to form biopolymers.

Ionic, hydrogen-bonds, van der Waals, and hydrophobic interactions drive dynamic interactions between and among biopolymers and small molecules.

Specific interactions are provided through precisely positioned functional groups on the 3-D structure of the bio-molecules.

Sugars are the building blocks of polysaccharides.

Fatty acids are the building blocks of membranes.

Amino acids are the building blocks of proteins

Nucleotides are the building blocks of DNA and RNA.

The rest of this chapter will be covered in more detail in the second half of BMB 251. For now, just read the text relevant to the figures.

CATALYSIS AND THE USE OF ENERGY BY CELLS

Cell Metabolism Is Organized by Enzymes Fig. 2-34

Every enzyme has one particular function (division of labor).
A cell can have many copies of the same enzyme.

Biological Order Is Made Possible by the Release of Heat Energy from Cells

Photosynthetic Organisms Use Sunlight to Synthesize Organic Molecules

Cells Obtain Energy by the Oxidation of Organic Molecules

Oxidation and Reduction Involve Electron Transfers

Enzymes Lower the Barriers That Block Chemical Reactions Fig. 2-44, 46 (Fig. 2-46c doesn't make any sense).

Breaking covalent bonds requires activation energy.
Enzymes lower the activation energy.
Enzymes bind to substrates and convert them to products.

How Enzymes Find Their Substrates: The Importance of Rapid Diffusion

Life processes require molecules to interact.

They do so via random diffusion or active transport.

Diffusion is temperature dependent Fig. 2-48

The distance coverage only goes up by the square root of the allowed time.

Large molecules diffuse slower than small molecules.

Long distances movements may involve active transport.

Active transport requires the input of energy.

Diffusion-limited reactions occur as fast as the molecules collide.

More concentrated, the faster the reaction goes.

Why are many reactions not diffusion limited?

Nonproductive collisions

Conformational changes in the protein

Catalytic steps.

Dissociation of a regulatory subunit or molecule

The cell is very crowded with biopolymers and other biomolecules. Fig. 2-49

The Free-Energy Change for a Reaction Determines Whether It Can Occur

The Concentration of Reactants Influences ΔG

For Sequential Reactions, ΔG° Values Are Additive

Activated Carrier Molecules are Essential for Biosynthesis

When cells take in food, they break down the bonds, which releases energy.

Some of the energy dissipates as heat.

Some of the energy is coupled to the production of activated carriers.

Activated carriers have high energy covalent bonds that can be used to make an unfavorable reaction more favorable.

Important energy carriers: ATP, NADH (and NADPH).

The Formation of an Activated Carrier Is Coupled to an Energetically Favorable Reaction Fig. 2-56

High energy covalent bonds are very unstable, and thus their easy breakage can be coupled to the breakage of more stable covalent bonds.

ATP Is the Most Widely Used Activated Carrier Molecule Fig. 2-57

Be able to recognize ATP, and distinguish it from ADP.

Energy Stored in ATP Is Often Harnessed to Join Two Molecules Together Fig. 2-59

ATP (**adenosine triphosphate**) is **hydrolyzed** to ADP (**adenosine diphosphate**).

Hydrolysis as its name implies means using water to break bonds.

Know the ATP hydrolysis reaction.

NADH and NADPH Are Important Electron Carriers

There Are Many Other Activated Carrier Molecules in Cells

The Synthesis of Biological Polymers Requires an Energy Input

Example of polynucleotide synthesis Fig. 2-67

Summary

Cellular reactions (e.g. synthesis of more cell components) are carried out by enzymes.

Energetically unfavorable reactions can be coupled to favorable ones.

ATP is the major energy carrier in the cell (equivalent to \$ in our economy).

HOW CELLS OBTAIN ENERGY FROM FOOD

Food Molecules Are Broken Down in Three Stages to Produce ATP **Fig. 2-70**

1. **Enzymatic break down of food**
Covalent bonds are broken such that biopolymers are broken down to monomer units (e.g. glucose). This happens in the intestines, and is relevant to multi-cellular organisms. Monomer units are absorbed into the cell.
Transport proteins help the molecules traverse the membrane.
The molecules end up in the cell's **cytoplasm**.
2. **Glucose is cleaved into two molecules called pyruvate.**
This occurs over many steps, with many chemical intermediates. A series of enzymes catalyze these coupled reactions. ATP input is used to help lower the activation energy. Ultimately, more ATP produced.
The whole cascade of events is called glycolysis.
Pyruvate diffuses into mitochondria, where it becomes a substrate for respiratory enzymes.
3. **Pyruvate is converted to an activated molecule called acetyl-coA**
A series of very complex enzymatic reactions couples the break down of acetyl-coA to the production of ATP.

Glycolysis Is a Central ATP-producing Pathway

Fermentations Allow ATP to Be Produced in the Absence of Oxygen

Glycolysis Illustrates How Enzymes Couple Oxidation to Energy Storage

Sugars and Fats Are Both Degraded to Acetyl CoA in Mitochondria

The Citric Acid Cycle Generates NADH by Oxidizing Acetyl Groups to CO₂

Electron Transport Drives the Synthesis of the Majority of the ATP in Most Cells

Organisms Store Food Molecules in Special Reservoirs

Amino Acids and Nucleotides Are Part of the Nitrogen Cycle

Many Biosynthetic Pathways Begin with Glycolysis or the Citric Acid Cycle

Metabolism Is Organized and Regulated

Summary

Food is broken down into monomer units.

Breakage of covalent bonds release energy, which is re-coupled in the form of ATP (and other energy carriers).

CHAPTER 3 PROTEINS

THE SHAPE AND STRUCTURE OF PROTEINS

The Shape of a Protein Is Specified by Its Amino Acid Sequence

A **protein** is made from a polymerized chain of amino acids.

The covalent bond that links amino acids is called a **peptide bond**. Fig. 3-1

Small chains of amino acids (<100) are called **peptides**.

Long chains are called **polypeptides**.

Polypeptides and proteins mean the same thing, but often polypeptide refers to the unfolded chain.

There are 20 different amino acids. Fig. 3-2

The part that is similar among all 20 is called the backbone.

The part that is unique is called the side chain.

Be able to recognize amino acid names.

Know that some have basic, acidic, or nonpolar side chains.

Each type of amino acid serves a purpose in the context of a protein.

Amino acids with similar side chain properties have similar function.

If you are a serious BMB major, then you should memorize all 20 amino acid names and properties, and be able to identify their side chains. (It'll be needed to really understand 251 and related courses).

Fig. 3-3 and Panel 3-1.

A polypeptide chain folds back on itself giving the protein a unique 3-D structure.

Four noncovalent forces direct the folding: Fig. 3-5, 3-6, 3-7

Ionic, hydrogen-bonding, van der Waals, and hydrophobic

Proteins Fold into a Conformation of Lowest Energy Fig. 3-8

Although the sequence of amino acids dictate the folding pathway and the final folded conformation, in the cell, other proteins called chaperones, assist the folding process.

Proteins range in size from 50-2000 amino acids.

Large proteins have multiple independent folding **domains**. Fig. 3-12

Domains can be thought of as different proteins strung together (more below).

The α Helix and the β Sheet Are Common Folding Patterns Fig. 3-9

The Protein Domain Is a Fundamental Unit of Organization

There are four levels of organization:

Primary structure is the linear arrangement of amino acids.

Secondary structure involve a helices and β sheets.

The protein can be partially unfolded and still have secondary structure.

Tertiary structure represents the full 3-D structure of a protein. Fig. 3-12, 3-13

A protein can have one or more independently folding domains.

Domains range from 5-350 amino acids.

Quaternary structure refers to complexes of multiple proteins.

So, quaternary interactions involve more than one protein.

Tertiary structure of a multi-domain protein is analogous to quaternary structure.

Few of the Many Possible Polypeptide Chains Will Be Useful

Certain amino acids along the polypeptide chain are more important than others in determining the structure of a protein.

Proteins Can Be Classified into Many Families

Proteins of similar biochemical function are likely to have similar structure. [Fig. 3-14](#)
However, the proteins may or may not have nearly identical primary sequence. [Fig. 3-15](#)

Proteins Can Adopt a Limited Number of Different Protein Folds

Sequence Homology Searches Can Identify Close Relatives [Fig. 3-17](#)

Computers can be used to align the primary sequence of proteins, to determine if they are related.

Computational Methods Allow Amino Acid Sequences to Be Threaded into Known Protein Folds

Some Protein Domains, Called Modules, Form Parts of Many Different Proteins

Some proteins may be unrelated except in one domain (or **module**). [Fig. 3-18, 3-19, 3-21](#)

Remember, genes can evolve by shuffling parts of the gene.

Since we might not know if a conserved region (as defined by comparing primary sequence) meets the definition of a domain (independently folding unit), we call them modules.

Some proteins may be related only by a **motif**.

A motif is a small sequence of amino acids found in many proteins.

A motif is small than a domain, and probably does not fold independently of the rest of the protein.

Motifs typically represent an interface – a section of the protein that binds something else (like another protein or a small molecule.)

The Human Genome Encodes a Complex Set of Proteins, Revealing Much That Remains Unknown

Larger Protein Molecules Often Contain More Than One Polypeptide Chain

Proteins are generally thought of as having one polypeptide.

However, many proteins have more than one polypeptide, and can be thought of as protein complexes. [Fig. 3-21, 3-22, 3-23, 3-24](#)

Each polypeptide of a protein complex is called a **subunit**.

Some Proteins Form Long Helical Filaments [Fig. 3-25, 3-26](#)

A Protein Molecule Can Have an Elongated, Fibrous Shape

Extracellular Proteins Are Often Stabilized by Covalent Cross-Linkages **Fig. 3-28, 3-29**

Disulfide bonds between cysteine side chains stabilize protein tertiary and quaternary structure.

Protein Molecules Often Serve as Subunits for the Assembly of Large Structures

Advantages of using repeating identical subunits to build very large protein structures:

Less genetic information required.

Assembly and Disassembly can be easily controlled since interactions are repeated.

Single-domain small subunits allow mis-folded subunits not to be incorporated.

Some very large protein structures with repeated subunits: **Fig. 3-30**

Actin filaments

Tubulin filaments

Bacterial flagella

Some super-sized structures: **Fig. 3-31, 3-32**

Viral coats (capsids)

Many Structures in Cells Are Capable of Self-Assembly

The Formation of Complex Biological Structures Is Often Aided by Assembly Factors

Summary

The sequence of amino acids dictates the structure of a protein.

Proteins have substructure including α helices and β sheets.

α helices and β sheets fold into domains

Domains fits together to form the protein.

Certain proteins can coalesce to form complexes or large structures.

Proteins with similar function will have similar structure and may (or may not) have similar amino acid sequence.

PROTEIN FUNCTION

All Proteins Bind to Other Molecules **Fig. 3-37**

Molecules that bind to proteins are called **ligands**.

Proteins are very selective toward the ligands they bind.

Precise docking of the ligand provides specificity.

Driving forces: ionic, hydrogen-bonding, van der Waals, hydrophobic. **Fig. 3-38, 3-43**

The Details of a Protein's Conformation Determine Its Chemistry

Amino acid side chains can be made to be very reactive.

Good for catalyzing biochemical reactions.

Sequence Comparisons Between Protein Family Members Highlight Crucial Ligand Binding Sites **Fig. 3-40**

Proteins Bind to Other Proteins Through Several Types of Interfaces

The Binding Sites of Antibodies Are Especially Versatile

Binding Strength Is Measured by the Equilibrium Constant **Fig. 3-44**

Enzymes Are Powerful and Highly Specific Catalysts

Enzymes do not get used up in the reaction.

Enzymes do not alter the equilibrium ratio of substrate and product.

Enzymes speed up reaction rates.

Enzymes can catalyze the reverse reaction as well.

Substrate Binding Is the First Step in Enzyme Catalysis

The reaction that an enzyme catalyzes occurs in the enzyme active site.

- The 3-dimensional arrangement of amino acids in the active site defines the active site.
- The substrate and/or ligand precisely dock at the active site via noncovalent and sometimes covalent interactions.
- Some ligands bind at other sites on the protein and change the proteins conformation (and activity).

▪ These other sites are called allosteric sites.

The interplay of amino acid side chains in the active site can cause a side chain to be hyper reactive.

The first step in an enzyme-catalyzed reaction is the binding of a substrate to the enzyme's active site.

- In this example, a bond is being broken (e.g. $\text{ATP} \rightarrow \text{ADP} + \text{phosphate}$)
- $\text{E} + \text{S} \rightleftharpoons \text{ES} \rightleftharpoons \text{EP} \rightleftharpoons \text{E} + \text{P}$
- E = enzyme; S = substrate; P = products
- The substrate encounters the active site via random diffusion.
- Most of the time the substrate passes right by the active site or hits the protein at the wrong site.
- Fortunately, diffusion is rapid and so a molecule might make over a billion collisions with the enzyme every second.
- Of course, in the cell, there are many different kinds of molecules which also collide with the enzyme.
- The wrong molecule might enter the active site.
- Precise docking of the substrate in the active site keeps it there.
- But sometimes the substrate dissociates before it can go onto the second step.
- How often a substrate dissociates versus going on to step 2, depends upon how strong the interaction is between the protein and the substrate.
- When a substrate binds an active site it often induces the protein into a subtle change in conformation.

The second step is catalysis:

- $E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P$
- The reason why the substrate does not spontaneously convert to product in the absence of the enzyme is there is a major energy barrier in breaking covalent bonds.
- This is called the activation energy.
- When the substrate is halfway to product, its bond is very strained.
- This can cause a slight change in the conformation of the substrate.
- This state is very unstable and it called the transition state.
- The enzyme active site is actually configured to bind the transition state much better than to the substrate.
- So the substrate is chemically stable but binds to the active site relatively weakly.
- The transition state, which is chemically very unstable, binds to the active site very strongly.
- This results in the enzyme lowering the activation energy for the reaction by stabilizing the transition state.
- As a dramatic proof of principle on this, antibodies raised against synthetic artificially stable transition state analogs can be used to catalyze the native reaction.

The third step in the reaction is the dissociation of the products:

- $E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P$
- In order for the enzyme to catalyze another reaction, the product must first dissociate from the active site.
- Since in many cases, the product is not that much different from the substrate , it is important that neither the substrate nor the product have too high of an affinity for the enzyme. Otherwise it would get stuck.
- A complete reaction cycle is called a turnover.
- The turnover number is the number of reactions an enzyme can catalyze per unit time (e.g. 300 per sec.)

How fast an enzyme works depends on the concentration of substrate.

- At low $[S]$ (but $[S] \gg [E]$), V (reaction velocity) is proportional to $[S]$.
 - V has units of molar per sec ($M s^{-1}$)
 - As $[S]$ increases, it more frequently encounters an enzyme that already has an S bound.
 - The enzyme becomes saturated with S .
 - As $E \rightarrow$ saturation, $V \rightarrow V_{max}$ (called 'vee max')
 - Turnover number = $V_{max} / [E]$
 - $K_M = [S]$ at which reaction proceeds to half V_{max} .
- K_M is a measure of substrate affinity (as well as its tendency to react).
- Substrates in the cell are often at concentrations in the K_M range.
 - Slight changes in the enzymes K_M , caused by allosteric ligands, can cause dramatic changes in the amount of product produced.

Enzymes Speed Reactions by Selectively Stabilizing Transition States

Enzymes Can Use Simultaneous Acid and Base Catalysis

Lysozyme Illustrates How an Enzyme Works

Tightly Bound Small Molecules Add Extra Functions to Proteins

Where amino acid side chains are insufficient, coenzymes are employed.

- Coenzymes are not proteins.
- Coenzymes are often, but not always, derivatives of nucleotides.
- Remnants of an ancient RNA world?
- Many vitamins are precursors to coenzymes: Biotin, thiamine
- Coenzymes are very high affinity ligands.

Other small molecules add functionality

- Metals : zinc, iron,
- Many more...

Multi-enzyme Complexes Help to Increase the Rate of Cell Metabolism **Fig. 3-54**

Product of one reaction becomes the substrate for the next.

Think of a hand-off in a relay race.

No need for diffusion!

The Catalytic Activities of Enzymes Are Regulated **Fig. 3-55, 3-56**

Feed back inhibition: Product of one reaction inhibits the production of the substrate for that reaction.

When you eat and get full, and don't feel like eating anymore, this is like feed back inhibition.

Allosteric Enzymes Have Two or More Binding Sites That Interact **Fig. 3-57, 3-58**

An allosteric site binds a ligand, changing the conformation of the active site (an protein).

Two Ligands Whose Binding Sites Are Coupled Must Reciprocally Affect Each Other's Binding

Symmetric Protein Assemblies Produce Cooperative Allosteric Transitions **Fig. 3-60**

The Allosteric Transition in Aspartate Transcarbamoylase Is Understood in Atomic Detail

See movie 3.9 on Cell Biology Interactive to get a better feel for the dynamics of allosteric regulation.

Many Changes in Proteins Are Driven by Phosphorylation

Protein kinases add phosphates to proteins (called **phosphorylation**). **Fig. 6-63**

Phosphate is derived from ATP

Amino acid side chains that can get phosphorylated have –OH groups:

Tyrosine

Serine

Threonine

Protein phosphatases remove phosphates from proteins.

Done via hydrolysis.

A Eucaryotic Cell Contains a Large Collection of Protein Kinases and Protein Phosphatases

Cells contains hundreds of different kinds of kinases and phosphatases.

Each is very specific for a set of proteins.

Phosphorylation/dephosphorylation serves as a molecular switch, turning on and off the activity of a protein.

Phosphorylation/dephosphorylation events are dynamic.

The Regulation of Cdk and Src Protein Kinases Shows How a Protein Can Function as a Microchip **Fig. 3-66**

See movie 15.8 on Cell Biology Interactive to get a better feel for this.

Get a sense of the major regulatory themes, and not specific names.

Proteins That Bind and Hydrolyze GTP Are Ubiquitous Cellular Regulators **Fig. 3-70**

Some proteins are activated when they bind GTP (e.g. Ras).
The GDP-bound protein is inactive.

Regulatory Proteins Control the Activity of GTP-binding Proteins by Determining Whether GTP or GDP Is Bound

Regulatory proteins (e.g. GAP) can induce GTP hydrolysis in the GDP-bound protein.
Fig. 3-71

Guanine nucleotide exchange proteins induce GDP to dissociate, thereby allowing the protein (e.g. Ras) to be reactivated.

See movie 15.9 on Cell Biology Interactive to get a better feel for this.

Get a sense of the major regulatory themes, and not specific names.

You must know the major intracellular signaling pathways. **Fig. 3-72**

Large Protein Movements Can Be Generated from Small Ones

Motor Proteins Produce Large Movements in Cells **Fig. 3-75, 3-76**

ATP hydrolysis converts a random walk into a directional walk.

ATP hydrolysis is coupled to conformational changes in the protein that allow forward but not backward movement.

(Any backward movement must be coupled to ATP synthesis, which is energetically unfavorable, but could be accomplished if the ADP and phosphate concentrations were high enough and the ATP concentration low enough.)

Examples of directional movement

- Muscle contraction
- Mitosis
- Cell migration
- Pumping ions (e.g. sodium) out of a cell.
- DNA polymerization
- Protein synthesis

Membrane-bound Transporters Harness Energy to Pump Molecules Through Membranes

Also the reverse is true: Movement of ions (specifically hydrogen ions) through a membrane is coupled to ATP synthesis.

See movie 14.1 on Cell Biology Interactive to see an awesome example.

(See also 14.2, if you want to know what grad students do with their spare time.)

Proteins Often Form Large Complexes That Function as Protein Machines

We be seeing lots of this over the next several weeks:

- DNA replication
- Genetic recombination
- Transcription
- Translation
- RNA splicing

A Complex Network of Protein Interactions Underlies Cell Function

Summary

Proteins are designed to bind other molecules to elicit some biological change.

Enzymes are a type of protein that catalyze biochemical reactions.

To catalyze biochemical reactions, enzymes bind substrates, make or break covalent bonds, and release products.

Proteins are regulated by the binding of small molecules, other proteins, or by phosphorylation. These molecules cause their protein targets to change shape, making them more or less active.

Enzymes can use ATP hydrolysis to get work done.

Part II Basic Genetic Mechanisms

CHAPTER 4 DNA AND CHROMOSOMES

THE STRUCTURE AND FUNCTION OF DNA

A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A DNA double helix consists of two antiparallel polynucleotide chains. Fig. 4-3

The two chains are held together by hydrogen bonding of bases

A base pairs with T

C base pairs with G

A sugar + phosphate + base is called a what?

Must know how to identify antiparallel strands.

Must know what 5'-to- 3' direction is.

- The sugar consist of a four-carbon (called 1', 2', 3' and 4') and one-oxygen pentameric ring, plus an extra carbon (5') and numerous -OH (hydroxyl) groups hanging off the ring.
- Each sugar is linked by a phosphodiester bond attached at the 3' and 5' position.
- The chain therefore has directionality:
- You are moving in the 5' to 3' direction when you go from the 5' carbon to the 4' carbon to the 3' carbon on the same sugar.
- What comes after the 3' carbon?

The Structure of DNA Provides a Mechanism for Heredity

In Eucaryotes, DNA is Enclosed in a Cell Nucleus

Summary

Genetic information that defines an organism is carried in a linear sequence of nucleotides.

DNA is double helix of two antiparallel strands

A pairs with T, and C with G.

CHROMOSOMAL DNA AND ITS PACKAGING IN THE CHROMATIN FIBER

Eucaryotic DNA Is Packaged into a Set of Chromosomes **Fig. 4-10**

One **chromosome** corresponds to one continuous DNA double helix
Humans have 24 different chromosomes.

Two sets of 22 different autosomal chromosomes.

Males have one each of sex chromosomes: X and Y

Females have two X chromosomes.

So humans have a total of 46 chromosomes in every cell.

A chromosome can be as long as a hundred million nucleotides.

Eukaryotic chromosomes are linear.

Bacterial chromosomes are circular (one end is connected to the other end).

Chromosomes Contain Long Strings of Genes

The Nucleotide Sequence of the Human Genome Shows How Genes Are Arranged in Humans **Fig. 4-15**

The human genome has been sequenced

A composite of seven ethnically diverse people.

It contains over 3 billion base pairs and over 30,000 genes.

There could be as many as 60,000 genes, but we don't yet know how to recognize them.

Genes are arranged linearly along chromosomes.

Genes are split into pieces consisting of exons.

The DNA between exons is called introns.

Exons are about 5% of the length of introns.

Much of human DNA contains repetitive elements **Fig. 4-17**

Less than 2% of the human genome codes for protein!

Comparisons Between the DNAs of Related Organisms Distinguish Conserved and Nonconserved Regions of DNA Sequence

Since exons comprise <2% of the genome, and since we do not know how the DNA sequence dictates where exons start and stop, we don't really know what is coding and what is noncoding.

Introns are often considered to be junk since they are of little importance and are generally not conserved.

By comparing (aligning) the human genome with the mouse genome we should be able to determine what regions are conserved.

Conserved regions are likely to be important.

Human/mouse/other comparisons reveal that the linear arrangement of genes along a chromosome is not static. **Fig. 4-19**

Large chunks move from chromosome to chromosome.

Chromosomes Exist in Different States Throughout the Life of a Cell

Cells grow and divide. This repeated process is called the **cell cycle** (more on this later).

Fig. 4-20

At one stage of the cell cycle (**S-phase**) the chromosomes are replicated.

At the next stage (**mitosis**), the chromosomes get compacted and packaged so that they can be delivered to both cells after cell division.

These condensed chromosomes are what is normally shown in textbooks.

Normally, chromosomes are decondensed (and therefore barely visible by even the most powerful microscope). **Fig. 4-21**

Genes are expressed when the chromosomes are decondensed.

Each DNA Molecule That Forms a Linear Chromosome Must Contain a Centromere, Two Telomeres, and Replication Origins

Replication origin – locations along the chromosome where the replication machinery initiates chromosome duplication.

Centromere – a region of DNA the mitotic spindle attaches so that it can drag the chromosome to the daughter cells (during cell division).

Telomere – DNA sequences that act as ‘caps’, protecting the ends of chromosomes. **Fig. 4-22**

Sometimes chromosomes can get damaged, like when they break.

The cell has machinery to repair broken chromosomes.

Remember that normal chromosomes are linear.

The ends of normal chromosomes would be recognized as broken, if not for telomeres.

DNA Molecules Are Highly Condensed in Chromosomes

Parts of chromosomes are packaged to varying degrees.

The level of packaging is quite dynamic, reflecting the need to access/sequester the genetic information.

Nucleosomes Are the Basic Unit of Eucaryotic Chromosome Structure

Chromosomes are normally covered with many different proteins, each having a different role in managing the genetic information.

The DNA plus these proteins is generally referred to as **chromatin**. **Fig. 4-23**

The major proteins that packages DNA are called **histones**.

Histone H2A, histone H2B, histone H3, and histone H4

These are collectively referred to as **core** histones

Several histones get together to form a protein complex, in which 150-200 base pairs of DNA wrap around. **Fig. 4-25**

The histone complex plus the DNA is called a **nucleosome**.

Actually 146 bp (base pairs) are in contact with the histones.

~50 bp form a linker between adjacent nucleosomes.

The beginning part of each histone polypeptide chain reach out like arms, and help regulate the accessibility of the DNA.

Called **amino-terminal tails**. **Fig. 4-32**

Nucleosomes make chromosomal DNA look like beads on a string.

The Structure of the Nucleosome Core Particle Reveals How DNA Is

Packaged Fig. 4-25

The Positioning of Nucleosomes on DNA Is Determined by Both DNA Flexibility and Other DNA-bound Proteins

Nucleosomes Are Usually Packed Together into a Compact Chromatin Fiber Fig. 4-30

The 'beads-on-a-string' **10 nm fiber** are usually compacted further by histone H1 to form a **30 nm fiber**.

The chromatin fiber toggles between the 10 and 30 nm fiber, when it is generally decondensed (not in mitosis) and the genetic info needs to be read or not read. Fig. 4-31

The amino terminal tails of the core histones may also contribute to formation of the 30 nm fiber. Fig. 4-32

ATP-driven Chromatin Remodeling Machines Change Nucleosome Structure

To access the genetic information the histones must be moved or removed.

Chromatin remodeling complexes move and/or remove histones. Fig. 4-33

Chromatin remodeling complexes control gene expression. Fig. 4-34

There are different kinds of chromatin remodeling complexes.

Some use the energy of ATP hydrolysis to move histones around.

Covalent Modification of the Histone Tails Can Profoundly Affect Chromatin Fig. 4-35

The core histone amino terminal tails are subjected to covalent modification.

Covalent modification alters what the 'tails' can do to the DNA, resulting in a different functional state of the chromatin (i.e., accessible vs. not accessible).

Note that DNA is negatively charged (phosphate backbone).

Histones are positively charged (lysine and arginine side chains).

Modified histone tails lose their positive charge.

What is the consequence of this with regard to histone – DNA interactions?

Lysine side chains are acetylated by histone acetyltransferases (called HATs).

HAT are enzymes that convert lysine to acetyl lysine.

Enzymes that remove the acetyl group are called histone deacetylases (HDACs).

Other enzymes can methylate the lysines

Others phosphorylate serines on the tail.

***Proteins bind to modified histone tails!**

Summary

Eukaryotic DNA resides in a group of polynucleotide chains called chromosomes.

Genes are scattered and fragmented through the chromosomes, covering <2%.

Chromosomal DNA is packaged into chromatin by histones, forming nucleosomes.

Chromosomes are continuously condensing (compacting) and decondensing for purpose of packaging and transport (mitosis), and for accessing/sequestering genetic information.

Accessibility is regulated by enzymes that modify chromatin by physically moving histones or

by covalently modifying histones.

THE GLOBAL STRUCTURE OF CHROMOSOMES

Lampbrush Chromosomes Contain Loops of Decondensed Chromatin

Drosophila Polytene Chromosomes Are Arranged in Alternating Bands and Interbands

Both Bands and Interbands in Polytene Chromosomes Contain Genes

Individual Polytene Chromosome Bands Can Unfold and Refold as a Unit

Heterochromatin Is Highly Organized and Usually Resistant to Gene Expression

Highly condensed chromatin that is generally devoid of genes.

However, heterochromatin does play important roles in chromosome maintenance.

Genes that are placed in heterochromatin (by scientist) are generally inactive.

The same gene placed in normally active euchromatin is active

The Ends of Chromosomes Have a Special Form of Heterochromatin

Centromeres Are Also Packaged into Heterochromatin

Heterochromatin May Provide a Defense Mechanism Against Mobile DNA Elements

Mitotic Chromosomes Are Formed from Chromatin in Its Most Condensed State **Fig. 4-55**

Each Mitotic Chromosome Contains a Characteristic Pattern of Very Large Domains

Individual Chromosomes Occupy Discrete Territories in an Interphase Nucleus **Fig. 4-60**

Summary

Chromatin structure is quite diverse and largely unknown

CHAPTER 5 DNA REPLICATION, REPAIR AND RECOMBINATION

THE MAINTENANCE OF DNA SEQUENCES

DNA encodes all the information necessary to make an organism.

In order for an organism or cell to reproduce, it must make a nearly exact copy of its DNA.

I said “nearly”, since mistakes during DNA replication are the essence of evolutionary change.

A permanent change in the DNA is called a **mutation**.

Mutation Rates Are Extremely Low

E. coli makes a permanent mistake (mutation) about once every billion nucleotides of replicated DNA.

Many Mutations in Proteins Are Deleterious and Are Eliminated by Natural Selection

Low Mutation Rates Are Necessary for Life as We Know It

Mutations at very low rates are essential for evolution.

However, high mutation rates in germ cells are detrimental to the species

Germ cells are sperm and egg, which go to make the next generation.

High mutation rates in somatic cells cause a variety of diseases, including cancer.

Somatic cells are all non-germ cells (like, skin cells, liver cells, brain cells, etc.)

Cancer is an uncontrolled proliferation of cells.

Your body has built-in mechanisms to stop cells from dividing.

If it didn't, since human cells take about a day to duplicate themselves, you would be as large as the entire class and weight about 20 tons!

Mutations can inactivate those growth control mechanisms, which would lead to cancer.

Summary

Mutations are rare and unhealthy for the organism, but are the driving force behind evolution (adaptation to a changing environment).

DNA REPLICATION MECHANISMS

DNA replication is fundamental to all organisms.
All organisms replicate their DNA the same way.
E. coli has been used as the model system.

Base-pairing Underlies DNA Replication and DNA Repair

One strand acts as a template for the other strand. **Fig. 5-2**

Remember: A pairs with T, and C with G. **Fig. 5-3**

DNA polymerase is the name of the enzyme that replicates DNA. **Fig. 5-4**

The substrates for DNA polymerase are:

dATP (deoxy ATP)

dTTP

dCTP

dGTP

Collectively, these four nucleotide substrates are called dNTPs.

What's the difference between dATP and ATP?

How many phosphates does dGTP have?

Also need are a template and a 3' -OH from the growing polynucleotide chain.

The DNA Replication Fork Is Asymmetrical

Replication forks are the point at which DNA replication is occurring. **Fig. 5-6**

Replication proceeds in the 5' (five prime) to 3' (three prime) direction, NOT in the 3' to 5' direction! **Fig. 5-7**

Therefore the replication fork must be asymmetric. **Fig. 5-8**

Remember that the each strand in the DNA double helix is antiparallel.

Continuous polymerization in the same direction as the fork is called leading strand synthesis.

Polymerization in the direction opposite to the fork has to be discontinuous, and is called lagging strand synthesis (I guess because it takes so long).

Why must lagging strand synthesis be discontinuous (i.e., synthesized in short stretches)?

What the heck are Okazaki fragments?

What was the name of the scientist who discovered these fragments?

You can't just leave them there as fragments.

How do Okazaki fragments get stitched together?

The High Fidelity of DNA Replication Requires Several Proofreading Mechanisms

Normally base-pairing interactions (What kind? Hydrogen bonding – good) between the incoming nucleotide base and the template dictate which of the four possible nucleotides is accepted.

Random nucleotides of all sorts are constantly diffusing in and out of the active site. **Fig. 5-9** Two mechanisms keep the correct nucleotide in the active site long enough to react with the 3' end of the growing polynucleotide chain;

1. A and G have big bases; C and T have small bases. Only a big and small one can occupy the tight quarters of the active site at the same time.
2. Proper base-pairing keeps the correct nucleotide there long enough to react.

Believe it or not, the nucleotide bases can morph into other chemical structures.

This happens very rarely.

Called tautomerization

The morphed structures can fool DNA polymerase.

For example, C can morph into something that pairs with A instead of G.

After it gets incorporated into the growing polynucleotide chain, it can morph back.

This happens about once for every 100,000 nucleotides incorporated.

If left uncorrected, this mutation frequency would be lethal to the cell.

Before moving on and incorporating the next nucleotide, DNA polymerase looks back and checks to see whether the newly incorporated nucleotide is correctly base-paired.

In reality, its hard for the growing 3' end to stay in the enzyme active site if it is not properly base-paired (and thus positioned) with the template strand.

An unpaired nucleotide at the 3' end is 'flapping around in the breeze'. **Fig. 5-10**

This nucleotide will 'flap' into a different active site located next door to the polymerization active site.

This adjacent active site cuts off the unpaired nucleotide.

This enzyme activity is called a **3'-to-5' proofreading exonuclease**.

About one out of every 100 misincorporations is missed by this proofreading activity.

The result is continued DNA polymerization, with an incorporated mutation.

Fortunately, there is one final check. Another enzyme scans the DNA looking for mismatches.

If it finds one, it chops out the mismatched nucleotide (more on this later).

This 'mismatch repair' is able to find 99 out of 100 mismatches.

If you put together all the proofreading mechanisms, nucleotide misincorporation occurs about once every billion incorporation events. **Table 5-1**

Only DNA Replication in the 5'-to-3' Direction Allows Efficient Error Correction

A Special Nucleotide-Polymerizing Enzyme Synthesizes Short RNA Primer Molecules on the Lagging Strand

DNA polymerase can only polymerize off of a 3' end.

Remember, it needs an -OH to attack the phosphate on the incoming nucleotide triphosphate. Why? I don't know. Perhaps DNA polymerase can't proofread as well on the first nucleotide??

Anyway, the cell initiates DNA replication by first polymerizing RNA on the template instead of DNA. **Fig. 5-12**

This RNA is called a **primer**, because it primes DNA synthesis (i.e., gets it going).

The enzyme is called **DNA primase**, and is a kind of RNA polymerase.

The RNA is about 10 nucleotides long.

The RNA/DNA double helix looks about the same as a DNA double helix except for those two chemical properties unique to RNA. Remember what those two properties are?

On the lagging strand, each Okazaki fragment begins with an RNA primer.

Synthesis of Okazaki fragments initiates at intervals of about every 200 nucleotides along the template.

How many RNA primers are required for the leading strand?

Once the RNA primer is made, DNA polymerase displaces primase and begins polymerizing DNA. **Fig. 5-13**

The RNA is then erased (removed) by an **RNase**

RNases hydrolyze an RNA polymer into individual nucleotide monophosphates.

On the lagging strand, DNA polymerase polymerizes until it runs into the prior Okazaki fragment.

The fragments are joined by a **DNA ligase**. **Fig. 5-14**

Special Proteins Help to Open Up the DNA Double Helix in Front of the Replication Fork

The DNA double helix is very stable, meaning that it is very hard to pull apart the two strands.

What kind of interactions prevent strand separation?

Separation of the two strands is called denaturation.

In order for DNA polymerase to insert new nucleotides, the strands have to be separated, so that the bases can pair with the correct nucleotide.

A **DNA helicase** separates the DNA strands. **Fig. 5-15, 5-16**

This requires a lot of energy.

Guess where this energy comes from?

But the two strands can re-anneal (or renature, same thing) to form the double helix again.

Another problem with single stranded DNA is that it tends to form base-pairs with itself, particularly if a complementary sequence is nearby. **Fig. 5-17**

These intramolecular interactions are called **hairpins**.

Want to guess why they are called hairpins?

So what else might be important to keep the strands separated?

A Moving DNA Polymerase Molecule Stays Connected to the DNA by a Sliding Ring

After DNA polymerase adds a nucleotide to the growing polynucleotide chain, one of two things can happen:

1. DNA polymerase can step forward and add another nucleotide.
2. Or it can dissociate from the template, and diffuse away.

When would it want to step forward?

When would it want to dissociate?

DNA polymerase has the natural tendency to dissociate, so a **sliding clamp** holds it on the DNA. **Fig. 5-19**

The sliding clamp forms a ring around the DNA.

You can see why, on a very long piece of DNA how the sliding clamp cannot fall off the DNA.

So, how does the clamp get on the DNA?

A clamp loader separates the ring, allowing it to encircle the DNA.

Do you think energy is required to do this?

How is that energy supplied?

When DNA polymerase runs into the next Okazaki fragment, it dissociates. **Fig. 5-20**

The Proteins at a Replication Fork Cooperate to Form a Replication Machine

So lets review what we have so far:

Protein	Function
DNA helicase	Separates DNA strands
SSB	Keeps single-stranded DNA from re-annealing
Primase	Lays down the RNA primer
Clamp loader	Loads the sliding clamp onto DNA
Sliding clamp	Holds the DNA polymerase on the template
DNA polymerase	Makes the DNA
RNase	Removes the RNA primer
DNA ligase	Joins together Okazaki fragments

Actually there are a lot more proteins involved but we won't get in to them. Many of these proteins work closely together and so they form complexes.

For example, the primase and helicase form a complex called the primosome.

Actually, the whole enchilada is called a replisome. **Fig. 5-22**

A Strand-directed Mismatch Repair System Removes Replication Errors That Escape from the Replication Machine

A protein complex scans newly replicated DNA for bulges in the DNA helix. **Fig. 5-23**

You can imagine how a protein can sense a bulging DNA.

These bulges correspond to mismatches, due to incorrect nucleotide incorporation (after fooling DNA polymerases proofreading activity).

The scanning must happen very soon after the new strand is made

The complex is called the **mismatch repair** complex.

The mismatch repair complex, cuts out the mismatched nucleotide (and a bunch of surrounding nucleotides as well).

DNA polymerase then comes along a fills in the gap.

Big, big problem: How does the mismatch repair complex know which of the two opposing mismatched nucleotides is the wrongly incorporated (the other being the parental template)?

Solution: Have the parental strand marked in some way so that the mismatch repair complex know who the parent was.

How to mark?

In bacteria, methylate the 'A' of a GATC nucleotide sequence of the parental strand.

In eukaryotes, make a break (called a nick) in the newly synthesized strand.

Long before replication is initiated a DNA methylase goes along marking the bacterial DNA.

The newly synthesized strand is unmethylated.

So the mismatch repair complex finds the unmethylated strand and cleaves it. No problem!

So there is a race between the mismatch repair complex and the DNA methylase.

Without the mismatch repair complex you will get more mutations in you DNA and be more susceptible to certain cancers.

DNA Topoisomerases Prevent DNA Tangling During Replication

Strand separation by the DNA helicases causes a topological problem. **Fig. 5-24**

The DNA helix ahead of the replication fork get twisted upon itself.

When a helix twists upon itself this is called a **supercoil** (or **superhelical** turns).

An enzyme that removes superhelical twists is called a **topoisomerase**.

Some topoisomerases break (nick) on of the strands of the double helix, allowing the other strand to swivel. **Fig. 5-25**

Is energy required for this?

DNA Replication Is Similar in Eucaryotes and Bacteria

Summary

DNA replication begins by separating the DNA strands.

Next, an RNA primer laid down.

DNA polymerase initiates off the primer.

A sliding clamp keeps the DNA polymerase on the template.

DNA polymerase proceeds continuously on one strand and discontinuously on the other.

Topoisomerases keep down the supercoiling.

DNA polymerase proofreads the DNA to make sure it put in the correct nucleotide.

A mismatch repair complex further proofreads the DNA immediately after DNA polymerase has done its thing.

THE INITIATION AND COMPLETION OF DNA REPLICATION IN CHROMOSOMES

DNA Synthesis Begins at **Replication Origins**

DNA replication begins at very precise locations. **Fig. 5-29**

The replication machinery does not begin at the end of the chromosome, or at random locations. Why?

In bacteria, there are no chromosomal ends.

Eukaryotic chromosomes are too long to do end-to-end replication.

The replication machinery may physically interfere with gene expression.

Head-on collision of DNA polymerase and RNA polymerase – ouch!

Co-ordination of the two polymerases could alleviate this problem.

Certain regions of chromosomes need to replicate before other regions.

A specific sequence of nucleotides, comprising the replication origin, binds to proteins that specialize in recruiting the replication machinery.

In bacteria, the origin recognition protein is called dnaA protein (generally called ‘initiator proteins’ in the text).

In eukaryotes, the origin is recognized by the Origin Recognition Complex (ORC).

Another feature of replication origins is that part of the sequence has a lot of ‘A’ and ‘T’ nucleotides. Why?

Hint 1: A-T base pairs are bonded by two hydrogen bonds (G-C, by three).

Hint 2: DNA replication requires strand separation.

Bacterial Chromosomes Have a Single Origin of DNA Replication **Fig. 5-30**

Bacterial DNA replication is regulated at the point of initiation. **Fig. 5-31**

Once initiation begins, it continues until the whole chromosome is duplicated.

How do you prevent reinitiation? **Fig. 5-32**

Mark the parental strands, and only assemble dnaA if both strands are parental.

DNA methylases chemically modify (‘mark’) the DNA.

Eucaryotic Chromosomes Contain Multiple Origins of Replication

Eukaryotic chromosomes are large, so there is a need for multiple origins of replication.

A bacterial chromosome has a few million base pairs.

A eukaryotic chromosome has hundreds of millions of base pairs.

In Eucaryotes DNA Replication Takes Place During Only One Part of the **Cell Cycle**

There are four phases to the eukaryotic **cell cycle** (growth and duplication of a cell):

Fig. 5-34

- **G1**
- **S**
- **G2**
- **M**

DNA synthesis occurs during **S phase**.

Different Regions on the Same Chromosome Replicate at Distinct Times in S Phase

Highly Condensed Chromatin Replicates Late, While Genes in Less Condensed Chromatin Tend to Replicate Early

Well-defined DNA Sequences Serve as Replication Origins in a Simple Eucaryote, the Budding Yeast

A replication origin in yeast is called an ARS (autonomously replicating sequence).

An ARS is a DNA sequence of about 150 base pairs in length, and binds multiple protein complexes.

How might you use the power of genetics to isolate an ARS? [Fig. 5-36](#)

Sequencing of the entire yeast genome has revealed the location of ARSs [Fig. 5-37](#)

An ARS has binding sites for multiple protein complexes. [Fig. 5-38](#)

A Large Multisubunit Complex Binds to Eucaryotic Origins of Replication

This protein complex is called an **ORC** (origin recognition complex).

An ORC binds to a portion of an ARS.

An ORC is bound to the chromosome throughout the cell cycle.

During S-phase the ORC is phosphorylated, allowing it to recruit other initiation factors (helicase, etc.)

The Mammalian DNA Sequences That Specify the Initiation of Replication Have Been Difficult to Identify

New Nucleosomes Are Assembled Behind the Replication Fork

How does the replication machinery move through nucleosomes? [Fig. 5-41](#)

Remember chromatin remodelling complexes?

Where do the nucleosomes go after the replication fork has passed?

How do additional nucleosomes get assembled?

Chromatin assembly factors (CAFs) help assemble histones and DNA into a nucleosome.

Telomerase Replicates the Ends of Chromosomes

How do eukaryotes replicate the ends of their linear chromosomes? Fig. 5-43

Remember, the RNA primer is laid down, then removed.

How do you replicate that part of the chromosome?

If you don't, repeated cell divisions will lead to progressively shortened telomeres.

Telomerase extends the chromosomal end using an RNA template.

The RNA template is a component of telomerase.

Since the template is RNA, telomerase is an RNA-dependent DNA polymerase (a.k.a. **Reverse Transcriptase**)

Since telomerase repeatedly extends the chromosomal ends, eukaryotic chromosomes have repeating sequences at their telomeres.

A DNA polymerase – primase protein complex then performs lagging strand synthesis.

*Note: The problem of replicating the actual chromosomal end never actually gets resolved.

Telomeres are packaged into different protein complexes and have different structures than the rest of the chromosome.

Telomere Length Is Regulated by Cells and Organisms

Somatic cells are initially formed with the full complement of telomeric repeats.

Telomerase is not made in somatic cells.

Somatic cells can undergo only a limited number of cell division due to progressive telomere shortening.

In every cell division the telomere gets a bit shorter.

After a number of cell division, the telomeric repeats are gone, and subsequent cell divisions lead to progressive loss of coding information (genes) located near the telomeres.

Cells without these genes die (called replicative cell senescence).

This might explain in part why we stop growing.

It also might provide a mechanism to prevent uncontrolled cell growth (cancer).

Summary

DNA replication begins at precise locations on the chromosome called origins.

Bacteria have one origin, eukaryotes have multiple origins.

An origin recognition complex binds to origins and recruits the replication machinery.

In eukaryotes, DNA synthesis occurs during S-phase of the cell cycle.

Telomerase is used to maintain the ends of eukaryotic chromosomes.

Telomerase is a reverse transcriptase that uses RNA to extend chromosomal ends.

Telomerase is turned off in somatic cells and turned on in cancer cells.

DNA REPAIR

Genetic variability allows a species to evolve and adapt to a changing environment. Genetic stability is important for the functioning of an organism.

Genetic instability leads to cancer, aging, death, and other not-so-fun things. [Table 5-2](#)

Your body is constantly bombarded by solar radiation (nice tan!), environmental toxins (smoke that cigarette), and metabolic by-products (have a big lunch).

This stuff damages (mutates) your DNA.

Mutate means a chemical change in a nucleotide, which might alter its coding information.

Also this stuff can actually break your chromosomes.

Each cell of your body acquires thousands of mutations a day.

Fortunately for you, you have several potent **DNA repair** machines.

<1 mutation in a thousand escapes these repair machines.

Without DNA Repair, Spontaneous DNA Damage Would Rapidly Change DNA Sequences

Mutations can occur at a variety of locations. [Fig. 5-46](#)

Thymidine dimers [Fig. 5-48](#)

Depurination [Fig. 5-47](#)

Deamination [Fig. 5-52](#)

DNA replication prior to DNA repair propagates the mutation. [Fig. 5-49](#)

The DNA Double Helix Is Readily Repaired

The beauty of the DNA double helix is that each strand provides an information ‘backup’ for the other strand.

Damage one strand and information on the complimentary strand can be used to repair the damage.

Each and every cell contains many copies of a variety of DNA repair machines.

DNA Damage Can Be Removed by More Than One Pathway

Common features:

Damage is cut out.

Nondamaged strand is used as a template to restore the correct nucleotide sequence.

What enzyme would do this?

Base excision repair [Fig. 5-50A](#)

Initially just the damage base is removed

The enzyme that does this is called a DNA glycosylase

Every kind of base mutation has a particular kind of DNA glycosylase designed for its removal.

The enzyme scans the DNA, ‘flipping out’ each base and checking it for damage. [Fig. 5-51](#)

Then the sugar phosphate is removed by a different enzyme

AP endonuclease

Nucleotide excision repair [Fig. 5-50B](#)

Targets large bulky mutations

Thymine dimers form UV light.

Carcinogens in tobacco smoke covalently attach to DNA.

A long patch of damage DNA is excised (12 nucleotides stretch).

The Chemistry of the DNA Bases Facilitates Damage Detection

The four bases (GATC) were selected during evolution in part because, deamination does not lead to interconversion to another base.

Instead they are recognized as non-natural and are removed.

Double-Strand Breaks are Efficiently Repaired

Some types of DNA damaging agents break both strand of DNA. **Fig. 5-53**

If left unrepaired the chromosomes would fragment.

A DNA ligase can rejoin the fragments. Called **nonhomologous end-joining**

Typically there is a loss of a base.

The vast majority of the mammalian genome is noncoding, so losing a nucleotide is no problem.

Homologous chromosomes can be used for repair. Called **homologous end-joining**.

Remember that we get one set of chromosomes from Mom and the other from Dad.

Cells use the information from one of the chromosomes to repair the other chromosome.

If the double strand break occurs during the G2 phase of the cell cycle, sister chromosomes can be used.

Cells Can Produce DNA Repair Enzymes in Response to DNA Damage

When cells are bombarded by DNA and/or protein damaging agents they undergo a stress response.

Called heat shock response in the textbook, since the response pathway was first characterized by a response to high temperatures.

Stress proteins (or heat shock proteins) are produced that help stabilize the cell against damage.

SOS response occurs in response to DNA damage.

Single-stranded DNA is one indicator of DNA damage.

Results from UV light.

SOS response is initiated when the **recA protein** binds to the single-stranded DNA.

This leads the expression of a number of DNA repair genes.

One of the SOS induced proteins is an error-prone DNA polymerase.

It is used when there is so much DNA damage that the template strand cannot be used to restore the genetic info.

Because it is damaged too.

Better to put in any nucleotide and take your chances, rather than leave a lethal gap in the DNA.

DNA Damage Delays Progression of the Cell Cycle

When you have damaged DNA, the last thing you want to do is duplicate your chromosomes.

Mutations would get replicated.

Breaks and single-stranded gaps would be lethal.

Part of the SOS response is the production of proteins that stop the cell cycle.

Summary

DNA is constantly being damaged by every day living.

Bases get damaged (chemically modified), DNA strands get broken.

Organisms have several DNA repair mechanisms to faithfully maintain their genetic information.

Ultimately, accumulation of unrepaired DNA damage leads to disease and death.

GENERAL RECOMBINATION

The rearrangement or movement of DNA sequences within the genome is called genetic recombination.

There are two different mechanisms that lead to genetic recombination:

General (or Homologous) Recombination

Site-specific Recombination

(Recombinant DNA technology is another means of moving DNA around, but this only occurs in test tubes, being a product human handiwork)

General Recombination Is Guided by Base-pairing Interactions Between Two Homologous DNA Molecules

Two chromosomes or any two stretches of DNA are **homologous** if they are identical or substantially similar in their nucleotide sequence.

Two homologous pieces of DNA can swap equivalent strands of their double helix.

The swapped DNA is called a **heteroduplex**.

The original parental DNA is called a homoduplex.

All organisms from bacteria to humans have machinery to conduct homologous recombination.

The first step in swapping strands is having one strand 'invade' the duplex. Called **synapsis**.

Of course, proper base-pairing drives the proper alignment of the strands.

Strand swapping then proceeds like a zipper.

Can occur over thousands of nucleotides.

Homologous recombination is used to:

Repair DNA

Align chromosomes during meiosis **Fig. 20-6, 20-11**

Evolve chromosomes

Meiotic Recombination Is Initiated by Double-strand DNA Breaks

This is homologous recombination that occurs during meiosis (described on pp. 1130-1139).

During meiosis your germ cells (egg or sperm) must go from a diploid state (two sets of chromosomes – from Mom and Dad) to a haploid state (one set – when you become a Mom or Dad).

Each set of chromosomes must align or pair up so that they can segregate to opposite cells.

The aligning occurs via homologous recombination. **Fig. 5-56**

DNA Hybridization Reactions Provide a Simple Model for the Basepairing Step in General Recombination

As a scientist you can artificially swap DNA strands in a test tube:

Start by having a solution of identical or nearly identical DNA fragments.

Then, separate the strands by heating them to 100°C – called **melting** or **denaturation**.

Then, cool the them down to let them come back together.

Called: **renaturation, reanneal, or hybridization**

The two originally paired strands are unlikely to find each other again, and are likely to pair with a similar complimentary strand. **Fig. 5-57**

Cells use enzymes to melt out and denature DNA.

Cells cannot tolerate the conditions used to denature DNA in a test tube.

DNA helicases separate the strands.

SSB (single-stranded DNA binding protein) binds to the separated strands .

SSB forms a stiff nucleo-protein structure that allows intermolecular reannealing of two strands, but not intramolecular reannealing of neighboring sequences (called hairpins).

Some of this should be familiar to you from the DNA replication section.

The RecA Protein and its Homologs Enable a DNA Single Strand to Pair with a Homologous Region of DNA Double Helix

DNA strand exchange is catalyzed by a nucleoprotein filament. **Fig. 5-58**

In bacteria, the protein component is called **RecA**; in eukaryotes its called **Rad51**.

RecA/Rad51 can simultaneously bind single stranded DNA and double stranded DNA to form a three-stranded structure. **Fig. 5-59**

If the DNAs within the complex are homologous the strands will synapse (pair).

After synapsis, the strands exchange via a zippering action – called branch migration.

RecA/Rad51 use the energy of ATP hydrolysis to promote uni-directional branch migration.

Fig. 5-60.

There Are Multiple Homologs of the RecA Protein in Eucaryotes, Each Specialized for a Specific Function

Also these recombination protein get help from accessory proteins.

Bra1 and Bra2 are one such accessory protein.

I mention them because defects in these proteins cause breast cancer.

General Recombination Often Involves a Holliday Junction

Once strand exchange has initiated via single-stranded DNA, a reciprocal exchange can occur. Fig. 5-67.

The point at which the strands cross over is called a Holliday junction. Fig. 5-61, Fig. 5-62

The name comes from the guy who discovered them.

Holliday junctions are recognized, branch migrated, and cut by a complex of Ruv proteins. Fig. 5-63

The 'cutting' process is also called resolution or resolved.

Only two of the strands are cut.

Two very different outcomes result, depending which two strands get cut.

One is called **crossing over**, where two chromosome arms are exchanged. Fig. 5-64

The other is called **gene conversion**, where only the strand that formed the heteroduplex is swapped (usually thousands of nucleotides). Fig. 5-66

Selective DNA repair can convert any mismatches to either of the two parental strands.

General Recombination Can Cause Gene Conversion

General Recombination Events Have Different Preferred Outcomes in Mitotic and Meiotic Cells

Mismatch Proofreading Prevents Promiscuous Recombination Between Two Poorly Matched DNA Sequences

Summary

Homologous recombination allows large stretches of DNA to move from one chromosome to another.

Homologous recombination involves the swapping of identical or nearly identical strands. RecA protein in bacteria or Rad51 in eukaryotes forms a nucleoprotein filament on DNA. RecA/Rad51 aligns DNA molecules, using the energy of ATP hydrolysis to 'compare' two DNAs.

If they base pair they will stay put and RecA/Rad51 continue to exchange strands in a zipper action.

Resolution of the crossing over results in chromosomes exchange arms, or could result in DNA repair.

SITE-SPECIFIC RECOMBINATION

In homologous recombination, DNA rearrangement required alignment of homologs.

Strand exchange took place over thousands of nucleotides

In site-specific recombination, strand exchange occurs at a very short specific DNA sequence, typically less than 12 base pairs. In some cases any sequence will do.

Long stretches of homology are not required.

Allows for pieces of DNA to jump around the genome.

Both sides of mobile genetic element need to be cut.

Mobile genetic elements can be hundreds to thousands of nucleotides long.

Examples of **mobile genetic elements**:

Transposons (aka **transposable elements**)

Move around within a cell.

Generally cannot move between cells unless free DNA is picked up by a cell.

Viruses (eukaryotes) and **bacteriophages** or **phages** (bacteria)

DNA gets packaged so that it can move between cells.

About half of the human genome is composed of mobile genetic elements or their relics.

Relics being elements that have acquired so many mutations that they are no longer capable of moving.

Sometimes when genetic elements move, they take adjacent DNA with them.

These are mutations that can be detrimental to the cell.

Mobile Genetic Elements Can Move by Either Transpositional or Conservative Mechanisms

Transpositional Site-specific Recombination Can Insert a DNA Element into Any DNA Sequence

Genes located on the mobile genetic elements often code for enzymes that aid in the transposition of the DNA. **Fig. 5-69**

Transposons encode **transposases**.

Transposons also have genes that code for antibiotic resistance.

There are three major classes of transposable elements. **Table 5-3**

DNA-only transposons **Fig. 5-70**

Some kinds get excised from the chromosome and moved as DNA

Others don't get excised but get replicated during the course of movement.

Retroviral-like transposons

These elements first get transcribed into RNA by an RNA polymerase.

Then reverse-transcribed into DNA!

The enzyme is called **reverse transcriptase**. **Fig. 5-74**

Retroviruses like HIV do this.

These elements encode their own promoters.

Unlike retroviruses, the elements lack genes for coat proteins that would allow them to move from cell to cell. **Fig. 5-73**

Non-retroviral retrotransposons.

The promoter lies outside of the element, and thus is not transposed.

Uses an RNA intermediate. **Fig. 5-76**

DNA-only Transposons Move By DNA Breakage and Joining Reactions

Some Viruses Use Transpositional Site-specific Recombination to Move Themselves into Host Cell Chromosomes

Retroviral-like Retrotransposons Resemble Retroviruses, but Lack a Protein Coat

A Large Fraction of the Human Genome Is Composed of Nonretroviral Retrotransposons

Different Transposable Elements Predominate in Different Organisms

Genome Sequences Reveal the Approximate Times when Transposable Elements Have Moved

Conservative Site-specific Recombination Can Reversibly Rearrange DNA

Fig. 5-79

Conservative Site-Specific Recombination Can be Used to Turn Genes On or Off

Fig. 5-82

Summary

Mobile genetic elements move around via site specific recombination.

Site specific recombination does not require the alignment of homologous chromosomes.

Recombination occurs at selected an/or random locations.

Some transposable elements move by a cut-and-paste mechanism.

Other transposable elements move via an RNA intermediate.

Transposable elements primarily move throughout the genome.

Other mobile genetic elements like viruses are designed to move from cell to cell.

Some viruses integrate into chromosomal DNA.

The human genome is mostly composed of repeated sequences that come from mobile genetic elements.

CHAPTER 6 HOW CELLS READ THE GENOME: FROM DNA TO PROTEIN

Central dogma: DNA makes RNA makes protein **Fig. 6-2**

Bacteria: Genes are easy to find

Look for a start codon, followed by a long (>100 codons) open reading frame (ORF), followed by a stop codon.

In multicellular eukaryotes the genes are split into small fragments (exons) separated by long stretches of noncoding DNA (introns).

<i>Process</i>	<i>Name</i>
DNA → pre-mRNA	Transcription
pre-mRNA → mRNA	RNA splicing and 3' cleavage/polyadenylation (eukaryotes only)
mRNA → protein	Translation

FROM DNA TO RNA

The more of a certain protein the cell needs, the more it transcribes and translates the corresponding gene. **Fig. 6-3**

Portions of DNA Sequence Are Transcribed into RNA

Transcription is the process by which DNA information is chemically rewritten in a slightly different chemical form.

One of the strands acts as a template (the other does not). **Fig. 6-7**

RNA differs chemically from DNA in that it: **Fig. 6-4**

uses U instead of T (difference of a methyl group)

has a 2' -OH instead of a 2' -H

RNA consists of 4 monomer units: **Fig. 6-5**

U pairs with A; C pairs with G

<i>name</i>	<i>Pyrimidine base</i>	<i>Purine base</i>	<i>name</i>
uracil	U	A	adenine
cytosine	C	G	guanine

DNA is almost always a double-strand helix

RNA is single-stranded but forms intramolecular helices that fold upon each other much like proteins do. **Fig. 6-6**

Transcription Produces RNA Complementary to One Strand of DNA

While DNA polymerase makes DNA, RNA polymerase makes RNA. **Fig. 6-8**
The chemistry of nucleotide polymerization is the same for DNA replication and transcription.

However, the proteins and regulation are very different.

Only a small portion of the entire genome is transcribed.

RNA polymerization proceeds in the 5'-to- 3' direction.

No primer needed.

Substrates: UTP, ATP, CTP, GTP

Transcription can be repeatedly initiated on the same gene. **Fig. 6-9**

RNA polymerase exhibits modest proofreading capability.

Cells Produce Several Types of RNA **Table 6-1**

mRNA: Genes that code for protein are transcribed into mRNA (m is for messenger).

Many different kinds

Each with a different range of abundance (depending upon cellular needs).

Eukaryotes: one mRNA typically makes one protein

Bacteria: one mRNA can make multiple proteins.

Polycistronic mRNA
One gene follows another.

rRNA: RNA component of ribosomes (ribosomal RNA).

90% of all cellular RNA!

Only a few different kinds.

tRNA: RNAs used to translate codons into amino acids (transfer RNA).

Small, typically ~75 nucleotides

About 30 different kinds, all very similar in sequence and structure.

snRNA: small nuclear RNAs that direct RNA splicing.

snoRNA: RNAs that direct chemical modification of rRNA

Others: telomerase RNA, RNAs that inhibit transcription, RNAs that inhibit translation

Signals Encoded in DNA Tell RNA Polymerase Where to Start and Stop

The DNA sequence that directs RNA polymerase where to start transcription is called a **promoter**.

Every gene has its own promoter.

RNA polymerase does not directly recognize promoters.

In bacteria, a protein called **sigma factor** (σ) recognizes the promoter and brings in RNA polymerase. **Fig. 6-10**

The RNA polymerase then undergoes a conformation change leading to the separations of the DNA stands near the promoter.

One strand acts a template allowing RNA polymerase to make RNA.

RNA polymerase moves down the template, releasing sigma.

Eventually RNA polymerase encounter a sequence of nucleotides that causes release of the transcript from RNA polymerase resulting in transcription **termination**.

The structure of the RNA polymerase transcribing DNA has recently been determined.

Fig. 6-11.

Transcription Start and Stop Signals Are Heterogeneous in Nucleotide Sequence

When writing the sequence of a gene:

The coding (or transcribed strand) is written in a 5'-3' (five prime to three prime) direction from left to right.

The noncoding (or template strand) is written in the 3'-5' direction below the coding strand.

However, in the genome, not all genes are aligned in the same relative direction.

The transcription start site is called +1.

Do not confuse this with the translation start site, which is located within the mRNA transcript.

'Downstream' is a term used to imply the direction that RNA polymerase travels.

RNA polymerase starts at +1 and travels downstream.

One nucleotide upstream of +1 is called -1.

There is no '0'

35 nucleotides upstream of the transcriptional start site is called -35.

Bacterial promoters are always upstream of the transcriptional start site.

In bacterial promoters, sigma factor binds to a sequence of nucleotides located at -10 and -35.

The -10 and -35 region comprise a portion of bacterial promoters.

The generic term for DNA sequences that bind proteins involved in transcription is **promoter element**.

The generic term for proteins involved in transcription is **transcription factor**.

The -10 and -35 region of each bacterial promoter are very similar, but not necessarily identical.

Actually, there are several different sigma factors.

Sigma-70 (σ^{70}) is the one used at most genes.

A small number of genes use other kinds of sigma factors, each recognizing a different sequence at the -10 and -35 region.

A **consensus sequence** is the sequence of nucleotides that are most commonly found at each position of the sequence.

Thus the σ^{70} consensus sequence for the -10 region is: TATAAT

The σ^{70} consensus sequence for the -35 region is: TTGACA

Other proteins that recognize specific sequences will have a different consensus sequence.

The point is, proteins can recognize a variety of related DNA sequences – not just one and only one sequence.

Some of sequences bind sigma more tightly, while others bind sigma weakly.

Tighter binding corresponds to more frequent recruitment of RNA polymerase (and thus more gene expression).

The concept of promoter elements is applicable to eukaryotes as well.

Binding sites for hundreds of different transcription factors have been identified.

Most sites are only 6-8 nucleotides in length, and can vary slightly in their actual sequence.

Simple math indicates that these sites are likely to occur quite often by chance, at random locations throughout the genome.

A six nucleotide sequence will occur by chance every 4^6 nucleotides (~4000 nucleotides), and thus have over a million copies per genome (using human as an example).

There are only 30,000-60,000 human genes.

So is very difficult to look at the sequence of the genome and predict where promoters will lie.

Transcription Initiation in Eucaryotes Requires Many Proteins

Bacteria have only one kind of RNA polymerase.

Eukaryotes have three kinds of RNA polymerase. **Table 6-2**

RNA polymerase I transcribes rRNA genes.

RNA polymerase II transcribes mRNA genes.

RNA polymerase III transcribes tRNA genes.

All three are structurally similar to each other and to bacterial RNA polymerases **Fig. 6-15**

RNA polymerases I, II, and III have many stably associated subunits (~12)

Bacterial RNA polymerases have 5 subunits, with sigma being one of them.

RNA Polymerase II Requires General Transcription Factors

RNA polymerase II requires a set of **general transcription factors** in order to bind to a promoter.

The general transcription factors for RNA polymerase **II** are required at all promoters.

These factors are listed in the table below and are pronounced 'tee-ef-two-A' for TFIIA, etc.

Name	DNA recognition site	# of subunits	Function
TFIIA	none	3	Help TFIID bind promoters
TFIIB	BRE	1	Links TFIID to TFIIF
TFIID	TATA box, initiator, DPE	15 (<i>TBP</i> + 14 <i>TAFs</i>)	Recognizes promoter and nucleates transcription complex assembly
TFIIE	none	4	Helps separate DNA strands
TFIIF	none	4	Links TFIIB to pol II
TFIIH	none	5	Separates DNA strands and phosphorylates pol II

How all these proteins arrive at a promoter inside of a cell is not known for certainty.

When these proteins are isolated from a cell, and placed in a test tube containing promoter DNA, they will assemble as described in **Fig. 6-16** to form a transcription initiation complex.

Other scientist have isolated nearly the entire complex pre-assembled, but not yet bound to the promoter.

The pre-assembled complex of RNA polymerase II + many of the general transcription factors is called RNA polymerase II **holoenzyme**.

One important promoter element present at many, but not all, eukaryotic promoters is the **TATA box**.

The TATA box is typically located at -30 (*not -25, as stated in the text*).

It has the following consensus sequence: TATA(A/T)A(A/T)(A/G) *note (A/T) means that it can be either A or T.*

The TATA box is recognized by the **TATA binding protein, TBP**. **Fig. 6-18**

TBP is a subunit of TFIID.

Distortion of the DNA by TBP may be important for transcription complex assembly.

The largest subunit of RNA polymerase two has a repeating amino acid sequence at its carboxy-terminal end.

For the human RNA polymerase II, a sequence of seven amino acids is repeated 52 times! The carboxy-terminal domain (or **CTD**) is thought to help tether RNA polymerase II to the general transcription factors and other regulatory factors (described below).

Once the transcription initiation complex has assembled it must separate the two strands at the transcriptional start site (just like in bacteria).

TFIIH has a **helicase** subunit for this purpose.

TFIIH kicks off the initiation of RNA synthesis by phosphorylating the CTD of RNA polymerase II.

An enzyme that phosphorylates another protein is called a **kinase**. So TFIIH is also a kinase.

RNA polymerase II is no longer tethered to the other factors at the promoter and is now free to move down the template, transcribing RNA.

Polymerase II Also Requires Activator, Mediator, and Chromatin-modifying Proteins

This a very very important point.

In a test tube (also called **in vitro**), RNA polymerase II and the general transcription factors are sufficient to initiate RNA synthesis at a promoter.

In a cell (**in vivo**), the sequences that are recognized by the general transcription factors are necessary but are not sufficient to assembly a transcription initiation complex.

The reason for this is that there are a lot of things that antagonize transcription complex assembly:

Huge amounts of non-promoter DNA compete for transcription factor binding.

Although binding is weak, there is a million times more non-promoter DNA than promoter DNA.

So basic laws of chemistry (mass action) say that these transcription factors will be tied up at non-promoter sites.

Histones bind DNA to form nucleosomes, which can cover up important promoter elements.

Inhibitor proteins bind to the general transcription factors and prevent them from assembling.

Transcriptional activators bind to specific DNA sequences and help assemble the transcription complex. **Fig. 6-19**

This is so important that the entirety of Chapter 7 is devoted to it.

Transcriptional activators control gene expression, in all forms of life.

Transcriptional activators help remove nucleosomal inhibitors of promoter elements by recruiting chromatin remodelling complexes.

Transcriptional activators help remove inhibitors of the general transcription factors.

Transcriptional activators recruit RNA polymerase II and the general transcription factors through interactions with the following:

TAF subunits of TFIID

TFIIB

Mediator

Mediator is a multi-subunit complex that helps bridge interactions between activators and RNA polymerase II.

Like the general transcription factors, mediator is required for the expression of nearly all genes.

(Because mediator is not strictly required to reconstitute promoter-specific transcription in vitro, it was originally missed as a general transcription factor.)

Transcription Elongation Produces Superhelical Tension in DNA

The phase of transcription by which RNA polymerase moves down the template transcribing RNA is called transcription elongation.

The problem is, is that there are a lot of obstacles for RNA polymerase.

One obstacle in the template sequence itself.

The RNA/DNA hybrid need to separate, as RNA polymerase moves past.

RNA/DNA strand separation is the job of RNA polymerase.

Very stable RNA/DNA hybrids slow the polymerase down.

Nucleotide mis-incorporation slows RNA polymerase down.

Its difficult to polymerize off of a mis-incorporated nucleotide.

Nucleosomes and other chromatin protein present barriers.

In all forms of life, both RNA and DNA polymerases create superhelical twists in the DNA.

Fig. 6-20

Superhelical twist were discussed earlier in Chapter 5.

Superhelical twists impede the movement of the polymerases, because the twisted DNA is harder to unwind.

Topoisomerases relieve this tension and thus enhance elongation.

Elongation factors help RNA polymerase move through these obstacles.

True in all forms of life.

Transcription Elongation in Eucaryotes Is Tightly Coupled To RNA Processing

The RNA produced by RNA polymerase II is modified in a number ways: Fig. 6-21, Fig. 6-22a

1. The 5' end of the RNA is 'capped'.
2. Introns are spliced out.

Eukaryotic genes are split into coding exons, and noncoding introns.

Introns are spliced out from the pre-mRNA (or primary mRNA) transcript to make the mature mRNA transcript.

3. The 3' end of the RNA transcript is cleaved and polyadenylated.

The enzymes that carry out these reactions might be recruited through interactions with the phosphorylated CTD of RNA polymerase II. Fig. 6-23

Remember , the CTD is phosphorylated upon transcription initiation, causing initiation factors to dissociate.

After that, the mRNA is exported from the nucleus to the cytoplasm, where it is translated.

RNA Capping Is the First Modification of Eucaryotic Pre-mRNAs

As soon as the pre-mRNA emerges from RNA polymerase II it is 'capped'. **Fig. 6-22b**
The cap is a modified guanine nucleotide that is added to the 5' end of the transcript.

Its actually a 5'-to- 5' linkage, as opposed to the normal 5' -to- 3' linkage in a standard polynucleotide (RNA or DNA) chain.

The cap protects the RNA from being degraded by RNases that recognize unprotected 5' ends.

5' caps do NOT exist in bacteria or on transcripts made by RNA polymerase I and III.

RNA Splicing Removes Intron Sequences from Newly Transcribed Pre-mRNAs

Eukaryotic genes are split into coding exons, and noncoding introns. **Fig. 6-25**

Introns are spliced out from the pre-mRNA (or primary mRNA) transcript to make the mature mRNA transcript.

Exons constitute a very small portion of the genomic sequence of a gene.

'Split' genes and RNA splicing is primarily associated with multi-cellular eukaryotes.

RNA splicing involves two sequential transesterification reactions.

There are different kinds of splicing mechanisms: Group II is shown in **Fig. 6-26**

One advantage of having 'split' genes is that different subsets of exons can be spliced together to make different proteins. **Fig. 6-27**

So the 30,000 different human 'genes' may actually code for >100,000 different proteins.

Nucleotide Sequences Signal Where Splicing Occurs

As with everything else the location of the splice sites is determined by the nucleotide sequence at the splice sites, as well as a sequence element located internal to the intron.

Fig. 6-28

Intron range in size from 10 to 10,000 nucleotides.

RNA Splicing Is Performed by the Spliceosome

snRNA are used to direct the splicing reaction.

The **spliceosome** is a large RNA/protein complex that conducts pre-mRNA splicing. **Fig. 6-29.**

The spliceosome is composed of several snRNPs ('snurps') and each snRNP is composed of an snRNA:

U1 snRNP = U1 snRNA + splicing proteins

U2 snRNP = U2 snRNA + splicing proteins

U4 snRNP = U4 snRNA + splicing proteins

U5 snRNP = U5 snRNA + splicing proteins

U6 snRNP = U6 snRNA + splicing proteins

snRNAs recognize the splice sites through base-pairing interactions. **Fig. 6-30**

The Spliceosome Uses ATP Hydrolysis to Produce a Complex Series of RNA–RNA Rearrangements

Some of the splicing proteins are RNA helicases, which use the energy of ATP hydrolysis to rearrange RNA-RNA interactions. **Fig. 6-30**

The active site for the transesterification reaction of splicing are made up of RNA!

Ordering Influences in the Pre-mRNA Help to Explain How the Proper Splice Sites Are Chosen

A Second Set of snRNPs Splice a Small Fraction of Intron Sequences

RNA Splicing Shows Remarkable Plasticity

Spliceosome-catalyzed RNA Splicing Probably Evolved from Self-splicing Mechanisms

In lower eukaryotes, certain splicing reactions are autocatalytic, requiring no protein.

There are two different kinds of self-splicing reactions: **Fig. 6-36**

Group I

The RNA intron forms a tertiary structure that binds to a free guanine (G) nucleotide cofactor in its active site.

The 3' –OH of the free G initiates the first transesterification reaction.

Group II

The RNA intron forms a tertiary structure that positions an adenosine nucleotide in the active site.

The adenosine (A) nucleotide is part of the intron sequence.

I call it the 'intronic A'.

The 2' –OH of the intronic A initiates the first transesterification reaction, forming a **lariat** structure.

A lariat has both 2'-5' and 3'-5' phosphodiester bonds.

Splicing of pre-mRNA (transcribed by RNA polymerase II) using the spliceosome, evolved from the group II mechanisms to include the use of protein and non-intronic RNA (i.e., snRNA), probably to assist in the regulation of splicing and alternative splice site selection.

Since the intron sequence was no longer required to provide the tertiary structure, intron sequence rapidly diverged.

Less constraints on intronic structure also allowed the number of introns and exons to proliferate.

It is interesting that the complex human beings don't have much more genes than worms, flies, or plants. But they do have the capacity to produce many more proteins through the use of additional exons.

RNA-Processing Enzymes Generate the 3' End of Eucaryotic mRNAs

mRNA that is not protected is rapidly degraded by RNases.

RNases can chew in from the 5' and 3' ends.

The guanosine cap protects the 5' end.

The 3' end is protected by **polyadenylation**. Fig. 6-37

Both reactions are specific to transcripts made by RNA polymerase II.

First the mRNA is **cleaved**. Fig. 6-38

As RNA polymerase II transcribes through the end of a gene, it encounters the sequence (AAUAAA).

I call this sequence 'A2UA3'.

A multi-subunit cleavage complex that loaded onto RNA polymerase II CTD via TFIID, scans the emerging RNA and binds to the A2UA3 sequence.

The nascent (newly made) RNA is then cleaved about 30 nucleotides downstream of the A2UA3 sequence.

Second a **polyA polymerase** associates with the 3' end of the mRNA transcript, and polyadenylates it.

The polyA polymerase is a template-independent RNA polymerase that uses only ATP as its substrate. Approximate 200 'A' nucleotides are added.

PolyA binding proteins bind to this polyA tail, forming an RNA/protein complex.

Binding of the polyA binding protein regulates polyA polymerase.

A protein/RNA structure that accommodates ~200 A's is probably the most stable, thus limiting the amount of polyadenylation.

Mature Eucaryotic mRNAs Are Selectively Exported from the Nucleus

RNA processing of pre-mRNA occurs in the nucleus.

Most of the RNA in the nucleus is the discarded introns, which are selectively degraded.

Probably because they lack a cap and are not polyadenylated, and thus are susceptible to RNase attack.

Also introns remain bound to a class of proteins, that form an hnRNP (heterogeneous nuclear ribonuclear protein) complex.

So named because they were varied in size as view under the electron microscope.

The proteins that are bound to the mRNA's cap, exons, and polyA tail are recognized by the cells export machinery.

The RNA is transported through the nuclear envelope into the cytoplasm via **nuclear pores**.

Fig. 6-39

Nuclear pores are aqueous channels, lined with protein, that traverse the nuclear envelope.

As the mRNA traverses the nuclear envelope, many of the associated proteins are replaced with cytosolic counterparts. Fig. 6-40

The cytosolic factors protect the mRNA from degradation and aid in the assembly of the ribosome.

The ribosome is where the mRNA is translated.

OK, lets take a break from the exciting travails of the mRNA journey so that we can find out more about what other non-mRNAs do....

Many Noncoding RNAs Are Also Synthesized and Processed in the Nucleus

Ribosomal RNA (rRNA) is made in the nucleolus by RNA polymerase I.

Most of the RNA in a cell is rRNA.

The nucleolus is described below.

RNA polymerase I lacks the CTD found on pol II, and so rRNA is not processed in the same way as pol II transcripts.

NO cap, NO splicing, NO polyadenylation. Nada

The ribosome contains one copy of 4 different kinds of rRNA. **Fig. 6-42**

Three of the four are made by cutting up a pre-rRNA transcript.

Remember, this is rRNA cutting and not mRNA splicing.

Their names are 18S, 5.8S, and 28S, which is a reflection of their size.

snoRNPs direct the cutting up of the pre-rRNA.

They are NOT the same kind of snRNPs involved in splicing.

The fourth (called 5S rRNA) is synthesized by RNA polymerase III, outside of the nucleolus.

The rRNAs are methylated at their 2'-OH. **Fig. 6-43**

Only certain rRNA nucleotides get methylated.

snoRNAs are small nucleolar RNAs that bind to specific rRNA sequences, providing a guide for a methylase.

Purpose: aid in folding, stability from degradation.

As an interesting twist, snoRNAs are often derived from introns of genes encoding ribosomal proteins.

The Nucleolus Is a Ribosome-Producing Factory

The **nucleolus** is a substructure within the nucleus. **Fig. 6-44**

It is where ribosomes are made.

It is formed from the coalescence of rRNA genes, RNA polymerase I, rRNA, tRNA, ribosomal proteins, snoRNPs, and partially assembled ribosomes.

The Nucleus Contains a Variety of Subnuclear Structures

Summary

Proteins and RNA do all the work in the cell, making us who we are. Have you thanked your proteins and RNA today?



The instructions for making protein and RNA is encoded in genes located on chromosomes. RNA polymerase and other help proteins recognize specific sequences of nucleotides that demarcate the beginning of the gene – called a promoter.

RNA polymerase transcribes the genes into RNA.

The genes of multicellular genes are composed for coding exons, and noncoding introns that are spliced out and generally discarded.

A spliceosome is an RNA/protein complex that splices out introns.

Eukaryotic mRNAs are modified,

having a guanosine cap  at its 5' end, and a poly A tail  at its 3' end.

Processed mRNA are exported to the cytoplasm where they are translated.

Non-mRNAs such as snRNA assist with a nuclear functions including RNA splicing.

The nucleolus is where rRNA is made, modified, and packaged into ribosomes.

FROM RNA TO PROTEIN

An mRNA Sequence Is Decoded in Sets of Three Nucleotides

The **genetic code** posits that the sequence of three nucleotides (codon) is the code ‘word’ for a particular amino acid. **Fig. 6-50**

The nucleotide sequence of mRNA is read as a series of codons.

There is a one-to-one correspondence between a codon and an amino acid. **Fig. 6-51.**

Actually there are three codons that signify translation to stop – called **stop codons**.

The codon for the amino acid methionine (AUG) is where translation starts – called the **start codon**.

An **open reading frame (ORF)** is a series of codons (typically >75) that are bounded by a start codon and stop codon (when read the mRNA is read in the 5’-3’ direction).

All forms of life use the same genetic code.

One notable exception: mitochondria have their own genome and translation machinery, so the genetic code differs slightly – but still 3 nucleotides to a codon.

Since a codon has three nucleotides, an RNA has three theoretical reading frames, of which only one is used.

There are $4 \times 4 \times 4 = 64$ possible codons, of which 61 code for 20 different amino acids.

Obviously, some amino acids are specified by more than one codon.

The other 3 codons are stop codons.

The mRNA is translated by a ribosome.

tRNA Molecules Match Amino Acids to Codons in mRNA

A codon does not directly match up to an amino acid.

A particular amino acid is covalently attached to a particular tRNA. **Fig. 6-52**

Since there are 61 different codons for 20 amino acids, are there 61 different tRNAs?

No! The number is somewhere between 20 and 61 depending upon the organism.

Reason: Some tRNAs can recognize more than one kind of codon. **Fig. 6-53**

These tRNAs require only proper base-pairing at the first two nucleotides of the codon.

Historically, called a ‘wobble’.

So, the code is often **degenerate** at the third position: For example, GGA, GGC, GGG, and GGU all code for glycine.

A 3-nucleotide sequence in the tRNA, called an **anticodon**, pairs with the appropriate codon.

What type of interactions are responsible for the pairing?

tRNAs Are Covalently Modified Before They Exit from the Nucleus

Eukaryotic tRNAs are made by RNA polymerase III.

Some contain introns, that are spliced out by mechanisms totally unrelated to mRNA splicing.

Specific Enzymes Couple Each Amino Acid to Its Appropriate tRNA Molecule

Aminoacyl-tRNA synthetases covalently attach the proper amino acid to the 3' end of the tRNA. **Fig. 6-56, 6-57**

Every type of amino acid has its own kind of aminoacyl-tRNA synthetase.

So, there are 20 different but structurally similar aminoacyl-tRNA synthetase.

Editing by RNA Synthetases Ensures Accuracy

It is critical that the correct amino acid be attached to the correct tRNA.

Aminoacyl-tRNA synthetases employ editing mechanisms analogous to that of DNA polymerase during DNA replication. **Fig. 6-59.**

If the correct amino acid is not covalently attached, it is removed.

Aminoacyl-tRNA synthetases recognize the correct tRNA through interactions with the anticodon. **Fig. 6-60**

Amino Acids Are Added to the C-terminal End of a Growing Polypeptide Chain

Like polynucleotide chains, amino acids and polypeptides have a directionality. **Fig. 6-61**

The polypeptide chain is synthesized in the N-terminal to C-terminal direction.

Every new amino acid is added at the C-terminal end.

The RNA Message Is Decoded on Ribosomes **Fig. 6-68** *Please marvel at its beauty!*

The ribosome is an RNA machine, that is decorated with a bit of protein.

It contains a large subunit and a small subunit, each of which is composed of many subunits.

Fig. 6-63

Bacterial and eukaryotic ribosomes are very similar.

A single mRNA can be translated simultaneously by many ribosomes.

The small subunit functions to pair the tRNA with the mRNA.

The large subunit catalyzes the **peptidyltransferase** reaction (protein synthesis).

The small and large subunit associate on the mRNA, near the 5' end, with the small subunit entering first.

The ribosome moves down the mRNA, recruiting the appropriate aminoacylated tRNA and synthesizing the polypeptide.

A ribosome can do 20 addition per second in bacteria!

There are three adjacent tRNA binding sites, each with a different function. **Fig. 6-64**

There are three major steps to each addition of an amino acid. **Fig. 6-65**

1. The middle 'P' site is bound by a tRNA having the growing polypeptide chain covalently attached to its 3' end. The adjacent 'A' site receives a new amino-acylated tRNA, as directed by the codon:anticodon interactions.
2. The amino group of the amino acid in the 'A' site, is adjacent to the aminoacyl bond of the C-terminus of the growing polypeptide. The amino group attacks the unstable protein-tRNA ester linkage, transferring the polypeptide from the 'P' site to the 'A' site. But the large subunit quickly moves forward, placing the two tRNAs into the 'E' and 'P' sites.

This central activity of the ribosome is called **peptidyltransferase**.

3. Using the energy of GTP hydrolysis, the small subunit translocates 1 codon downstream.

When a stop codon is encountered, a release factor (that looks like a tRNA but is really protein), enters and leads to peptide release.

Elongation Factors Drive Translation Forward **Fig. 6-66**

EF-Tu delivers the tRNAs to the ribosome, using the energy of GTP hydrolysis.

Accuracy is achieved by allowing the amino acylated tRNA to dwell in the A site before it is allowed to attack the polypeptide chain.

Improperly paired tRNAs dissociate more quickly than properly paired ones.

Because anticodon:codon base-pairing is what holds the tRNA in the 'A' site.

EF-G hydrolyze GTP to power the forward movement of the ribosome large subunit.

The Ribosome Is a Ribozyme

Fig. 6-68 illustrates that the ‘meat’ of the ribosome is RNA.

In fact, if all the proteins are removed (artificially in a test tube), the RNA still has peptidyltransferase activity.

The proteins seem to help hold the ribosome together, and assist in some of its conformational changes.

RNA molecules that catalyze chemical reactions are called **ribozymes**.

What other ribozymes have we heard of?

Fig. 6-70

Nucleotide Sequences in mRNA Signal Where to Start Protein Synthesis

Where on the mRNA does the ribosome know to begin translation?

The translation start site sets the reading frame.

Not to be confused with the transcription start site.

Translation always begins at an AUG.

For Eukaryotes: **Fig. 6-71**

A methionine-charged initiator tRNA is delivered to a free ribosomal small subunit through interactions with a protein complex called **eIF-2** (eukaryotic initiation factor 2)

This complex then recognizes the 5' end of the mRNA through interaction with factors (eIF-4E and eIF-4G) bound to the mRNA guanosine cap.

Using the energy of ATP hydrolysis the small subunit and its entourage move along the mRNA.

Once the first AUG is found (via anticodon:codon interactions with the met-tRNA), everything stop and waits until the ‘big guy’ (large subunit) comes along.

Now, translation proceeds as described above.

For Bacteria: **Fig. 6-72**

Same sort of thing as for eukaryotes, except there is no cap complex to recognize.

The rRNA in small ribosomal subunit base-pairs with an RNA sequence (called **Shine-Dalgarno** sequence) just upstream of the AUG start codon.

So the Shine-Dalgarno sequence directs where translation begins in bacterial mRNAs.

This allows for internal starts as would be necessary for polycistronic messages.

Stop Codons Mark the End of Translation

There is no tRNA that is designed to bind to one of the three stop codons.

Instead, a protein called **eRF** (eukaryotic release factor) binds. **Fig. 6-73**

eRF looks very much like a tRNA. **Fig. 6-74**

Called *molecular mimicry*.

Instead of adding an amino acid to the growing polypeptide chain, as tRNAs do, eRF adds H₂O.

The polypeptide is now no longer attached to anything, and so it diffuses away.

Proteins Are Made on Polyribosomes

Once a ribosome starts translating, another ribosome can start right behind it. **Fig. 6-75**
eIF-4G also binds to the polyA binding protein, making the 5' and 3' end of the mRNA right next to each other.

Quality-Control Mechanisms Operate at Many Stages of Translation

There Are Minor Variations in the Standard Genetic Code

Many Inhibitors of Prokaryotic Protein Synthesis Are Useful as Antibiotics

A Protein Begins to Fold While It Is Still Being Synthesized

As the N-terminal end emerges from the ribosome, it starts to adopt secondary structure.

alpha helices and beta sheets

Tertiary interaction involving amino acid side-chains start to happen, but these interactions constantly rearrange, settling on the most stable set of interactions, before the final structure is obtained. **Fig. 6-80**

Protein domains fold sequentially. **Fig. 6-81**

Molecular Chaperones Help Guide the Folding of Many Proteins

While proteins can fold into their final 3-D structure by themselves, other proteins called molecular chaperones, make the folding process more efficient. **Fig. 6-82.**

Molecular chaperones are also important for helping cells deal with stress such as high heat.

Heat can cause proteins to misfold or not fold.

Many molecular chaperones are also **heat shock proteins** (hsp).

Different molecular chaperones work in different ways. **Fig. 6-84**

Exposed Hydrophobic Regions Provide Critical Signals for Protein Quality Control

Hydrophobic regions of the protein do not want to be exposed to water.

So they normally bury themselves deep within the protein.

Misfolded proteins tend to have a hydrophobic patches on their surface.

These hydrophobic patches are recognized by chaperones.

If chaperones are unable to help refold the protein, the misfolded protein is ultimately degraded (recycled) by **proteases**.

Proteases are enzymes that cut proteins into small peptides.

The Proteasome Degrades a Substantial Fraction of the Newly Synthesized Proteins in Cells

A **proteasome** is a large multi-subunit complex that degrades proteins. **Fig. 6-86**

It includes enzymes with protease activity.

An Elaborate Ubiquitin-conjugating System Marks Proteins for Destruction

Ubiquitin is a short protein.

Surveillance proteins recognize damaged and misfolded proteins.

These protein covalently attach ubiquitin onto the targeted protein.

The proteasome recognizes the ubiquitin and degrades anything bound to it.

Many Proteins Are Controlled by Regulated Destruction

Many proteins need to be rapidly disposed of in response to a changing cellular environment. For example, as a cell goes through the cell cycle certain protein must be rapidly removed.

This is done by regulated degradation.

Proteins can have a degradation signal that is hidden. **Fig. 6-88b**

Protein phosphorylation, dissociation of a subunit, or cleavage of the protein can expose its degradation signal.

Abnormally Folded Proteins Can Aggregate to Cause Destructive Human Diseases

Excessive misfolding of proteins can lead to their aggregation.

Aggregated proteins can cause diseases such as Alzheimer's and Mad Cow disease.

There Are Many Steps From DNA to Protein

Each step can be regulated (turned 'on' or 'off') by cellular signals.

Summary

mRNAs contain a series of three-nucleotide codons, which get translated into protein.

Each codon corresponds to a particular amino acid.

mRNA is recognized by the ribosome, located in the cytoplasm.

Proteins are synthesized on ribosomes.

tRNAs, attached to specific amino acids, recognize each codon.

The codons align the tRNAs allowing the amino acids to polymerize.

As the polypeptide chain emerges from the ribosome, it starts to fold, forming its final 3-D structure.

Chaperones are proteins that help other proteins fold properly.

Misfolded proteins or proteins bearing degradation signals are rapidly degraded in the proteasome.

All these step from the gene to the life-and-times of the protein are regulated by cellular signals.

THE RNA WORLD AND THE ORIGINS OF LIFE

RNA probably existed long before DNA.

This is a very interesting section of the text, but unfortunately we do not have time to cover it in class.

Please read it, if you are interested in current theories about how life came into existence.

Life Requires Autocatalysis

Polynucleotides Can Both Store Information and Catalyze Chemical Reactions

A Pre-RNA World Probably Predates the RNA World

Single-stranded RNA Molecules Can Fold into Highly Elaborate Structures

Self-Replicating Molecules Undergo Natural Selection

How Did Protein Synthesis Evolve?

All Present-day Cells Use DNA as Their Hereditary Material

Summary

CHAPTER 7 CONTROL OF GENE EXPRESSION

AN OVERVIEW OF GENE CONTROL

The Different Cell Types of a Multicellular Organism Contain the Same DNA

Nearly all cells in an organism contain the same set of genes.
Animals have been cloned by placing the nucleus of a skin cell into an enucleated egg.

Different Cell Types Synthesize Different Sets of Proteins

Cells within the same organism become different because they express a different set of genes.

Fundamentals about cell differentiation:

1. All cells have a set of proteins that are the same (e.g. ribosomal proteins, cytoskeleton, RNA polymerase, etc.)
2. Specialized cells have certain proteins that are highly abundant (e.g., hemoglobin in red blood cells), and not found in other cells.
3. Every type of cell has a characteristic pattern and intensity of gene expression. **Fig. 7-3**
4. This results in a characteristic pattern of protein production.

A Cell Can Change the Expression of Its Genes in Response to External Signals

Cells sense chemical signals from other cells or from the environment.
These signals elicit specific changes in gene expression.

Different types of cells might respond differently to the same signal.

Signal	Responding cell type	Response
glucocorticoid hormone	liver	increase expression of the tyrosine aminotransferase gene (as well as change expression of certain other genes)
glucocorticoid hormone	fat cell	decrease expression of the tyrosine aminotransferase gene (as well as change expression of certain other genes)

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

Points at which the activity of a protein (or its gene) can be regulated:

- Gene expression
- RNA processing (capping, splicing, polyadenylation)
- Subcellular localization of mRNA
- Translation
- Degradation of mRNA
- Modification of the protein (covalent modification, protein/small molecule ligand binding)
- Subcellular localization of the protein
- Destruction of the protein.

Summary

The genome of a cell can make a variety of cell types.

A cell type is determined by its complement of RNAs and protein.

The production and activity of RNAs/proteins is highly regulated by cellular/environmental signals.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does the cell selectively turn on or off genes?

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics

The Outside of the DNA Helix Can Be Read by Proteins

The DNA double helix has a minor and major groove. Fig. 7-6

Each of the four nucleotide bases presents a unique arrangement of hydrogen bond donor and acceptors, and hydrophobic patches in the major groove. Fig. 7-7

A series of nucleotides can present a unique docking site for a protein with a complementary interacting surface. Fig. 7-8., Table 7-1,

Sequences are often 6-20 base pairs.

The Geometry of the DNA Double Helix Depends on the Nucleotide Sequence

Short DNA Sequences Are Fundamental Components of Genetic Switches

Gene Regulatory Proteins Contain Structural Motifs That Can Read DNA Sequences

Protein-DNA interactions, just like any other interaction, depends upon an exact fit. Fig. 7-12

The 3-D surface of the protein is shaped to accommodate the binding of a DNA double helix. Amino acid side-chains make precise hydrogen-bond interactions with the bases.

Hydrophobic and van der Waals interactions also take place.

Ionic interaction take place between amino acids and the DNA phosphates.

Interaction with the sugar or phosphates generally increase **affinity** and not **specificity**.

Interactions with the bases provide both affinity and specificity.

Specificity is the ability to discriminate one DNA sequence from another.

You must know the difference between affinity and specificity.

The Helix–Turn–Helix Motif Is One of the Simplest and Most Common DNA-binding Motifs

Helix-turn-helix is a protein motif designed to fit into the major groove. Fig. 7-13

Composed of two alpha helices.

Many different gene regulatory proteins have a helix-turn-helix motif. Fig. 7-14

By changing some of the amino acids on its DNA recognition helix, you can change the DNA binding specificity of the protein.

A feature common to many sequence-specific DNA binding proteins is that they bind DNA as dimers.

A **dimer** is a protein complex having two subunit, which many times are identical.

The helix-turn-helix motif on each monomer is spaced apart at the same distance as one turn of the DNA double helix.

Proteins that bind DNA as dimers must recognize a symmetrical DNA sequence. Fig. 7-16

Homeodomain Proteins Constitute a Special Class of Helix–Turn–Helix Proteins

Homeodomains are DNA binding motifs found in gene regulatory proteins that direct eukaryotic development.

That's what makes them so 'special'.

Mutation in the homeodomain can cause severe developmental problems.

For example, a fly antennae might turn into a leg.

There Are Several Types of DNA-binding Zinc Finger Motifs

Another general group of DNA binding motifs use zinc to hold their structure together.

One member of the group is called a zinc finger. Fig. 7-17

Several zinc fingers are strung together to form a DNA binding domain. Fig. 7-18

Each zinc finger sticks an alpha helix into the major groove to achieve binding specificity. Fig. 7-19

β sheets Can Also Recognize DNA

The Leucine Zipper Motif Mediates Both DNA Binding and Protein Dimerization

Some proteins dimerize via interaction of two alpha helices. Fig. 7-21, 3-11

Called a leucine zipper.

So named because hydrophobic leucine side-chains protrude from each alpha helix and interdigitate like a zipper.

Beyond the interacting helices, the helices extend and bifurcate outward allowing them to bind the DNA major groove.

Think of a scissor grip.

Heterodimerization Expands the Repertoire of DNA Sequences Recognized by Gene Regulatory Proteins

A homodimer is a dimer with identical subunits.

A heterodimer has two different subunits.

Two different leucine-zipper containing proteins can form a heterodimer. **Fig. 7-22**

Since they have different DNA binding specificities, the heterodimer can bind sequences that neither homodimer could (and visa versa).

Thus fewer proteins can be used to generate a larger DNA binding repertoire.

Amino acids around the leucines provide specificity and affinity, whereas the leucines provide affinity for dimerization.

The Helix–Loop–Helix Motif Also Mediates Dimerization and DNA Binding

Don't confuse these with helix-turn-helix proteins. Fig. 7-25

Helix-loop-helix proteins are more like leucine zipper proteins.

A helix-loop-helix protein could heterodimerize with a different one that lacks a DNA binding domain.

The complex won't be able to bind DNA.

Example of negative regulation of the gene-specific regulator.

Read the next five sections if you plan to work in a laboratory that studies gene regulation.

It Is Not Yet Possible to Accurately Predict the DNA Sequences Recognized by All Gene Regulatory Proteins

A Gel-Mobility Shift Assay Allows Sequence-specific DNA-binding Proteins to Be Detected Readily

DNA Affinity Chromatography Facilitates the Purification of Sequence-specific DNA-binding Proteins

The DNA Sequence Recognized by a Gene Regulatory Protein Can Be Determined

A Chromatin Immunoprecipitation Technique Identifies DNA Sites Occupied by Gene Regulatory Proteins in Living Cells

Summary

Proteins recognize specific DNA sequences through precise docking of amino acid side-chains on the protein on to the DNA major groove.

Specificity is provided through hydrogen bond interactions between the DNA bases and the protein amino acid side chains.

Many sequence-specific DNA binding proteins use similar structural motifs. Sequence-specific DNA binding proteins regulate the expression of a gene.

HOW GENETIC SWITCHES WORK

Above, we had a look at gene regulatory proteins.
Now, we look at how they regulate gene expression.

The Tryptophan Repressor Is a Simple Switch That Turns Genes On and Off in Bacteria

This is about as simple as it gets.

E. coli has about 4000 genes.

5 of them are dedicated to the synthesis of the amino acid tryptophan.

All 5 genes are located within the tryptophan **operon**. Fig. 7-33

All genes within an operon are transcribed onto a single mRNA.

The production of that mRNA is controlled by a single promoter.

E. coli lives in your intestines. When you eat a hamburger, the protein is digested to amino acids (including tryptophan). So, the tryptophan operon can be turned off. How?

Tryptophan binds to a site on the tryptophan ('trip' for short) repressor causing the repressor to rotate its DNA recognition helices such that it can now recognize a specific sequence on the tryptophan promoter. Fig. 7-35

The tryptophan repressor bound to the promoter prevents RNA polymerase from binding.

Result: operon is shut off. Fig. 7-34

This is an example of **negative control** or **negative regulation**.

In this case, a ligand (tryptophan) caused a repressor to bind DNA.

In other cases, a ligand could cause a repressor to dissociate from DNA. Fig. 7-36a

In this case, the ligand is an inducer gene expression.

Transcriptional Activators Turn Genes On

Bacterial RNA polymerases can recognize promoters through interactions of its sigma subunit with specific promoter sequences.

Many bacterial promoters have suboptimal binding sites for sigma, and thus have an intrinsically low capacity to recruit RNA polymerase and transcribe the gene.

Promoters are designed this way so that RNA polymerase does not continually transcribed the gene.

Continual transcription gene is wasteful, if not needed, and could be harmful.

Its better to have the gene off in the default state.

Bacteria use regulatable sequence-specific binding proteins to recruit and stabilize the binding of RNA polymerase at the promoter.

The generic term for this type of protein is **transcriptional activator**.

The type of regulation is called **positive regulation** or **positive control**.

As with negative control, a ligand can help or hinder the binding of transcriptional activator to a promoter. Fig. 7-36b

Some gene regulatory proteins can act as both an activator or repressor depending upon where its DNA binding site is located Fig. 7-38, 7-37

A Transcriptional Activator and a Transcriptional Repressor Control the *lac* Operon

Rationale for regulation:

E. coli normally eat glucose to provide carbon and energy.

When no glucose is around in the environment they look around for other kinds of sugars (e.g. lactose).

But, only make the lactose-metabolizing enzymes if lactose is present.

Algorithm...

Algorithm			Mechanism Fig. 7-38	
If glucose is ...	and lactose is...	then turn <u>on/off</u> the <i>lac</i> operon	CAP bound to promoter:	lac repressor bound to promoter:
present	present	off	No	No
present	absent	off	No	Yes
absent	absent	off	Yes	Yes
absent	present	on	Yes	No

Mechanism...

Lac repressor binds to promoter and blocks RNA polymerase binding.

Lac repressor + **allolactose** ligand cannot bind to promoter.

CAP + **cAMP** ligand bind to promoter and recruits RNA polymerase.

Absence of glucose causes cAMP levels to rise.

Gene regulation in bacteria can be thought of as a series of genetic on/off switches.

Regulation of Transcription in Eucaryotic Cells Is Complex

The complexity of eukaryotic cells requires something more than just simple on/off switches.

Features of eukaryotic regulation:

- Regulatory proteins can be targeted to sequences located thousands base pairs upstream or downstream of the transcription start site.
- Multiple components must come together at a promoter (RNA polymerase, general transcription factors, mediator, etc.)
- Transcriptional start sites can be covered by histones (chromatin).

Eucaryotic Gene Regulatory Proteins Control Gene Expression from a Distance

DNA in a cell is compacted

So a protein bound at a distant site might be physically close to the transcription start site.

A Eucaryotic Gene Control Region Consists of a Promoter Plus Regulatory DNA Sequences

A **eukaryotic promoter** is considered to be the DNA sequence near the transcriptional start site that is required for the binding of the general transcription machinery. However, keep in mind that in almost all cases, the promoter is insufficient to bind the transcription machinery inside of a cell. Transcription complex assembly requires gene regulatory proteins bound to DNA regulatory sites. **Fig. 7-41**
 Different names often refer to the same thing...

DNA*	Protein*
<ul style="list-style-type: none"> • regulatory site • regulatory sequence • enhancer • activator binding sites 	<ul style="list-style-type: none"> • activators • transcriptional activators • transcription factors • sequence-specific activators
<ul style="list-style-type: none"> • UAS in yeast (upstream activating sequence) • transcription factor binding sites • DNA binding elements 	<ul style="list-style-type: none"> • gene regulatory proteins • gene activator proteins • enhancer binding proteins • trans-activators

*Some of these terms refer only to positively acting factors; other refer to both positive and negative factors.

There are a few hundred of different kinds of gene regulatory proteins.

Each target an overlapping set of genes.

Any one gene might be regulated by only a few of them.

So any one of them might not be present in a particular cell type.

If they are present, they are not abundant.

Eucaryotic Gene Activator Proteins Promote the Assembly of RNA Polymerase and the General Transcription Factors at the Startpoint of Transcription

Transcriptional activators have a modular design. **Fig. 7-42**

Transcriptional activators have a DNA binding domain.

Transcriptional activators have an activation domain.

Activation domain stimulate transcription

These two different types of domains can be mixed and matched artificially.

Activation domains 'plug' into the general transcription machinery and/or mediator.

Fig. 7-43, 7-44

Activation domains help localize and stabilize the binding of the general transcription machinery to the promoter.

Eucaryotic Gene Activator Proteins Modify Local Chromatin Structure

Activators also recruit chromatin remodelling factors. Fig. 7-45

Some remodelling complexes chemically modify the histones.

Others re-arrange the histones on the DNA.

The remodeled chromatin may allow more ready access of the transcription machinery to the DNA. Fig. 7-45

Gene Activator Proteins Work Synergistically

Transcriptional synergism occurs when single activators don't stimulate transcription, but multiple activator do. Fig. 7-47

Synergism can arise when activators function to accelerate more than one step in transcription initiation. Fig. 7-48

Eucaryotic Gene Repressor Proteins Can Inhibit Transcription in Various Ways Fig. 7-49

Block activator binding to DNA

Mask activation domain

Mask activator target in the general transcription machinery

Reposition nucleosomes over promoter

Remove chemical modification of histone that allow factor access.

Eucaryotic Gene Regulatory Proteins Often Assemble into Complexes on DNA

In reality, eukaryotic gene regulation is a continuous and dynamic interaction of a variety of positively and negatively acting regulatory proteins.

Cellular and environmental signals tip the pos/neg balance in one direction or the other to turn up or turn down the expression of a gene.

There are many protein involved, some bind DNA, others do not. Fig. 7-50

Complex Genetic Switches That Regulate *Drosophila* Development Are Built Up from Smaller Modules

The *Drosophila eve* Gene Is Regulated by Combinatorial Controls

Complex Mammalian Gene Control Regions Are Also Constructed from Simple Regulatory Modules

Gene regulatory proteins are themselves regulated by environmental/cellular signals. **Fig. 7-**

58

This is an important figure.

Insulators Are DNA Sequences That Prevent Eucaryotic Gene Regulatory Proteins from Influencing Distant Genes

Bacteria Use Interchangeable RNA Polymerase Subunits to Help Regulate Gene Transcription

Gene Switches Have Gradually Evolved

Summary

The expression of genes is regulated by proteins that bind within a few thousand base pairs of the transcription start site.

Transcriptional activators act positively by recruiting the transcription machinery to the promoter or by making DNA more accessible to the transcription machinery .

Transcriptional repressor act negatively by inhibiting the binding of positive factors and/or blocking access of the transcription machinery.

Gene regulation is a dynamic balance of positive and negative acting factors, whose activity in turn is regulated by cellular and environmental signals.

THE MOLECULAR GENETIC MECHANISMS THAT CREATE SPECIALIZED CELL TYPES

Here we will go over a few of some of the best known mechanisms by which an organism changes its gene expression program.

DNA Rearrangements Mediate Phase Variation in Bacteria

This strategy is NOT a normal gene regulatory mechanism, but its kind of neat.

Think about it next time you get food poisoning.

Salmonella is a bacteria that causes severe illnesses in humans and is commonly associated with food poisoning.

Your immune system reacts by attacking the bacterial flagella.

Salmonella counters this attack by changing its flagella protein.

This is done by inverting a segment of DNA containing the flagellin promoter. Fig. 7-64

In one direction, the promoter allows transcription of the H2 flagellin gene, and a repressor of the H1 flagellin gene.

In the other direction, H2 (and the H1 repressor) have no promoter is thus are not expressed. H1 is no longer repressed, and thus is expressed.

A Set of Gene Regulatory Proteins Determines Cell Type in Yeasts

Two Proteins That Repress Each Other's Synthesis Determine the Heritable State of Bacteriophage Lambda

- Upon infecting *E. coli*, phage λ prepares to integrate itself into the *E. coli* chromosome.
 - This is called lysogeny.
 - Why do this?
 - λ would rather go along for a free ride, then have to make more of itself.
 - The host makes more λ when it replicates its own chromosome.
 - When should it do this?
 - When cell conditions are favorable (the "good life").
 - i.e., the cell is not starving or being bombarded by heat or UV light.
 - When conditions are unfavorable, λ chooses the lytic cycle (either upon infection or if it is already a lysogen).
 - In the lytic cycle, λ just makes more of itself and gets out (lysis the cell).

- **The lysis/lysogeny decision lies in a race between cII protein turning on the lysogeny program and the lytic genes being expressed. (Fig. 7-67)**
 - 99% of the time lysis wins out.
 - Upon infection, the λ cII gene is expressed and must function before the lysis genes are expressed.
 - cII is unstable and is readily degraded if the cellular conditions are unfavorable.
 - The cII protein activates expression of the cI repressor (lambda repressor) gene by binding to the cI promoter (pRE).
 - cII protein also activates the integrase gene for integration into the chromosome.
 - The cI repressor binds to promoters pR and pL to shut down the expression of the lysis genes. **Fig. 7-37**
 - It does this by physically blocking access of RNA polymerase to the promoter.
 - The cI repressor does bind RNA polymerase and repositions it to transcribe the cI promoter.
 - This is positive feedback
 - cI repressor is therefore responsible for maintaining lysogeny.
 - The cI protein is susceptible to self-cleavage/proteolysis.
 - This is brought on by the recA protein - a signal that the environment is less favorable.
 - The Cro protein is one of the first lytic genes to be transcribed once the level of cI drops off.
 - Cro represses cI expression and activates the remaining lysis genes.
 - Cro, like cI, is both a DNA-sequence specific activator and repressor.

Gene Regulatory Circuits Can Be Used to Make Memory Devices As Well As Oscillators

Circadian Clocks Are Based on Feedback Loops in Gene Regulation

The Expression of Different Genes Can Be Coordinated by a Single Protein

Many proteins need to come together to turn on a gene.

Some cells are poised to turn on/off a whole program of genes.

They do this by having all the necessary factors present except for one.

That one factor is then delivered when the correct cellular/environmental signal is present.

For example, the glucocorticoid receptor turns on/off many genes, but it requires a steroid hormone called glucocorticoid. **Fig. 7-71.**

Expression of a Critical Gene Regulatory Protein Can Trigger
Expression of a Whole Battery of Downstream Genes

**Combinatorial Gene Control Creates Many Different Cell Types in
Eucaryotes**

The Formation of an Entire Organ Can Be Triggered by a Single Gene
Regulatory Protein

Stable Patterns of Gene Expression Can Be Transmitted to Daughter
Cells

Chromosome Wide Alterations in Chromatin Structure Can
Be Inherited

The Pattern of DNA Methylation Can Be Inherited When Vertebrate
Cells Divide

Vertebrates Use DNA Methylation to Lock Genes in a Silent State
Genomic Imprinting Requires DNA Methylation

CG-rich Islands Are Associated with About 20,000 Genes in Mammals

Summary

Expression of specific sets of genes elicits specific cellular fates.

Cellular/environmental signals set in motion the particular gene expression programs.

POSTTRANSCRIPTIONAL CONTROLS

Transcription Attenuation Causes the Premature Termination of Some RNA Molecules

Transcription attenuation is when transcription by RNA polymerase stops prematurely, such that a functional mRNA is not produced.

Usually attenuation is caused by the nascent RNA forming a hairpin, which interferes with RNA polymerase function.

Regulatory proteins negate this interference allowing polymerase to continue transcribing. The HIV virus uses transcription attenuation to regulate the expression of its genome.

Alternative RNA Splicing Can Produce Different Forms of a Protein from the Same Gene

Alternative splicing is when the splicing machinery splices together different exons. [Fig. 7-88](#)

The result is a different coding sequence for part of the gene.

Two alternatively spliced mRNAs produce two proteins having identical domains and different domains.

An extreme example of alternative splicing is shown in [Fig. 7-89](#)

It is not known how many different spliced messages the human genome can be produced, but it is fair to say that alternative splicing can increase the repertoire of genes from about 30,000 to several hundred thousand.

Often alternative splicing occurs in different cell types to create proteins that have specialized function in that cell type. [Fig. 6-27](#)

Just like everything else, alternative splicing can be regulated both positively and negatively by factors that bind at or near splice sites. [Fig. 7-90](#)

The Definition of a Gene Has Had to Be Modified Since the Discovery of Alternative RNA Splicing

Sex Determination in *Drosophila* Depends on a Regulated Series of RNA Splicing Events

In female flies, splicing of certain transcripts produces splicing regulators that ultimately produce a repressor of male-specific genes via alternative splicing. [Fig. 7-92](#)

In males, an alternative splicing pattern generates nonfunctional splicing regulators that ultimately produce a repressor of female-specific genes.

A Change in the Site of RNA Transcript Cleavage and Poly-A Addition Can Change the C-terminus of a Protein

RNA Editing Can Change the Meaning of the RNA Message

mRNA can be chemically modified to change its coding information.

Sounds heretical!

An even more bizarre form of editing is the insertion of Us at various but precise location of the mRNA.

Not only can this change the codons, but it can change the reading frame.

This happens in bizarre organism like trypanosomes (and mostly in their mitochondria).

RNA Transport from the Nucleus Can Be Regulated

Some mRNAs Are Localized to Specific Regions of the Cytoplasm

Proteins That Bind to the 5' and 3' Untranslated Regions of mRNAs Mediate Negative Translational Control

In bacteria, the Shine-Dalgarno sequence can be blocked by translational repressors.

Remember, the Shine-Dalgarno sequence is how the ribosome recognizes the beginning of the gene.

In eukaryotes, the interaction between the 5' and 3' end of the mRNA can be inhibited.

Remember, proteins that bind the 5' cap and 3' polyA tail interact.

The part of the mRNA upstream of the translation start site is called the 5' UTR (untranslated region).

The part downstream of the stop codon is called the 3' UTR.

The Phosphorylation of an Initiation Factor Globally Regulates Protein Synthesis

The cells of most multicellular organisms stop proliferating, once adulthood is reached.

The go into a cell cycle phase called G_0 .

Also when cells are stressed or are starved, they stop growing.

One way to quickly stop growth is to inhibit translation initiation.

Kinase sensors phosphorylate eIF-2, causing it to bind tightly, and thus tying up limiting amounts of a nucleotide exchange factor, that eIF-2 needs in order to function.

Initiation at AUG Codons Upstream of the Translation Start Can Regulate Eucaryotic Translation Initiation

Internal Ribosome Entry Sites Provide Opportunities for Translation Control

Gene Expression Can Be Controlled By a Change In mRNA Stability

Those proteins that rapidly turnover (degraded) have mRNA that are rapidly degraded. Stable proteins often have stable mRNA.

Proteins involved in directing the cell cycle tend to have unstable mRNAs.

Shortening of the polyA tail precedes mRNA degradation.

This is a slow process, but once most of the polyA tail is chewed away, so goes the cap.

The mRNA is then rapidly degraded.

Translation competes with proteins that want to degrade the mRNA. Fig. 7-104

Sequences in the 3' UTR regulate mRNA stability, via interaction with stabilizing/destabilizing proteins.

Cytoplasmic Poly-A Addition Can Regulate Translation

Nonsense-mediated mRNA Decay Is Used as an mRNA Surveillance System in Eucaryotes

RNA Interference Is Used by Cells to Silence Gene Expression

Remember when we talked about how the human genome (as well as other multicellular organisms) contains a lot of repeated sequences?

Remember that many of the repeated sequences transposed around the genome via an RNA intermediate?

Too much transposition could trash the coding information of the genome.

These RNA intermediates form a lot of secondary structure (i.e., are double-stranded), which are recognized by a protein complex containing an RNA nuclease and helicase. Fig. 7-107

The complex cuts up the RNA destroying it.

Once again the day is saved!

RNA fragment remain bound to the complex, which can base-pair to other complementary RNAs causing their destruction.

Scientist can use this surveillance system to design RNAs that can bind specific mRNAs, leading to their destruction.

This is called **RNAi** (pronounced 'RNA eye').

This is very new, easy and powerful method of performing gene knock-out (and thus determining gene function).

Summary

There are many ways that the expression of a gene can be controlled after it is transcribed but before it is translated.

Alternative splicing hooks up different exons to make novel proteins.

Regulatory proteins can bind to 5' and 3' UTRs to regulate the efficiency of translation.

Regulatory proteins often bind to 3' UTRs to regulate mRNA stability.

HOW GENOMES EVOLVE

Interesting stuff, but we won't have time to cover it. ☹

Genome Alterations are Caused by Failures of the Normal Mechanisms for Copying and Maintaining DNA

The Genome Sequences of Two Species Differ in Proportion to the Length of Time That They Have Separately Evolved

The Chromosomes of Humans and Chimpanzees Are Very Similar

A Comparison of Human and Mouse Chromosomes Shows How the Large-Scale Structures of Genomes Diverge

It Is Difficult to Reconstruct the Structure of Ancient Genomes

Gene Duplication and Divergence Provide a Critical Source of Genetic Novelty During Evolution

Duplicated Genes Diverge

The Evolution of the Globin Gene Family Shows How DNA Duplications Contribute to the Evolution of Organisms

Genes Encoding New Proteins Can Be Created by the Recombination of Exons

Genome Sequences Have Left Scientists with Many Mysteries to Be Solved

Genetic Variation within a Species Provides a Fine-Scale View of Genome Evolution

Summary

Part III Methods

CHAPTER 8 MANIPULATING PROTEINS, DNA, AND RNA

Strategies for examining how biological molecules work:

General Strategy	Approach
Biochemistry and structural biology	in vitro (test tube)
Molecular biology	in vitro (test tube)
Genetics	in vivo (in living cells)
Cytology	in vivo (in living cells)
Genomics and bioinformatics	in silico (computers)

Fundamental premise:

- We cannot see individual molecules in living cells at high enough resolution to figure out what they are doing.
- What we can do is ‘perturb’ the cells or biomolecules, and observe what happens.
 - ‘Perturb’ means for example: add a substrate to an enzyme, mutate a gene, add a chemical to the growth media.
- Any observation is usually a summation of ‘signals’ given off by billions of individual molecules in the system under study.
 - ‘Signal’ could mean any measurable property, for example:
 - A product of an enzyme reaction might...
 - be a different color than the substrate,
 - have a different mass than the substrate
 - A mutated gene might cause the cell to die
 - Adding a chemical to cells might cause it to produce certain mRNAs
- With each observation, it is our job as scientists to interpret what that observation means in terms of the molecular picture.
 - We could be right, we could be wrong.
 - Science does not provide absolute truth, but provides a higher level of knowledge that we hope believe to be on the path toward ultimate truth.
 - OK, enough philosophy.

Limitations:

- The ‘signal’ that you could be observing in living cells could be the result of a perturbation of the one thing that you think you are studying, or it could be the result of a perturbation of many different things. Its hard to be certain.
- One way around this problem is to isolate the molecule that you are studying.
 - Actually, you isolate a large population of a particular molecule.
- Then, any observation caused by the perturbation must be due to that molecule.

ISOLATING CELLS AND GROWING THEM IN CULTURE

Cells Can Be Isolated from a Tissue Suspension and Separated into Different Types

Tissue can have many different cell types with different properties.
Want your system to be as homogenous as possible.

A chorus sounds better than a bunch of people saying different things at the same time.

Cells can be sorted by a FACS (fluorescence activated cell sorter). **Fig. 8-2**

Separation is based upon a certain measurable property of the cell.

Cells Can Be Grown in a Culture Dish

A better way to get a homogeneous population of cells is to start with one cell and grow a **clonal population** in a culture dish. **Fig. 8-4**

Term	what a cell biologist says...	what a biochemist says...
cells in an animal	in vivo	in vivo
cells in a culture dish	in vitro	in vivo
molecules in a test tube	in vitro	in vitro

Differentiated cells often maintain their differentiated state.

Example: Fibroblast still secrete collagen.

Serum-free, Chemically Defined Media Permit Identification of Specific Growth Factors

Eukaryotic cells require growth factors for growth.

Generally, only cancer cells grow without the need for growth factors.

Embryonic stem (ES) cells have the potential to differentiate into any cell type, given the proper growth factors.

Eucaryotic Cell Lines Are a Widely Used Source of Homogeneous Cells

Cells Can Be Fused Together to Form Hybrid Cells

Hybridoma Cell Lines Provide a Permanent Source of Monoclonal Antibodies

Antibodies are normally produced by the immune system of vertebrate animals.

Antibodies are proteins that can be tailored by the cell to bind to virtually any molecule.

Normally they bind to foreign invaders (e.g. virus coat protein), targeting them for destruction.

However, scientist can use antibodies raised against specific proteins as a very selective way to detect or isolate that protein from a complex mixture of other cellular proteins.

You make an antibody by injecting a purified target molecule (antigen), usually a protein, into a bunny rabbit.

The rabbit makes antibodies to the antigen.

Draw blood from the bunny, and you have a source of **polyclonal antibody**.

Polyclonal means that the rabbit immune system made a variety of different antibodies against the same antigen (or different parts of the antigen).

Any one antibody-producing cell in the immune system produces the same antibody.

If you were to isolate that cell and grow it in culture, it would produce only one type of antibody.

Called **monoclonal antibody**.

Usually monoclonal antibodies are made in mice, not rabbits.

Summary

Individual vertebrate cells can be propagated in culture.

Cell cultures provide a homogeneous source for defined experimentation.

Antibodies are potent means of detecting or isolating specific molecules in a complex mixture.

FRACTIONATION OF CELLS

Purpose: To identify and study individual biomolecules in a purified state.

Why? Properties associated with biomolecule preparation are more likely to be properties of that biomolecule (e.g. the enzyme alcohol dehydrogenase) rather than a contaminant (other biomolecules that you are not interested in).

Organelles and Macromolecules Can Be Separated by Ultracentrifugation

Cells can be harvest from their media by low speed centrifugation. Fig. 8-7

Cells can be broken apart to release their contents.

Organelles and cell membranes can be removed by centrifugation.

The supernate is called the **crude extract**.

The liquid that can be poured off after any centrifugation is generically called the **supernate**.

The Molecular Details of Complex Cellular Processes Can Be Deciphered in Cell-Free Systems

Proteins Can Be Separated by Chromatography

Proteins have several properties that can used to separate and purify them. Fig. 8-10, 8-11

1. Different proteins have different distribution of surface charges and hydrophobicity.

Due to the distribution of basic and acidic amino acids.

Use ion exchange chromatography and hydrophobic exchange chromatography.

2. Proteins come in different sizes

Use gel filtration chromatography.

3. Proteins bind to specific ligands.

Use affinity chromatography.

The ligand is immobilized on the chromatography resin.

1 and 3 work on similar principles, 2 does not.

Affinity Chromatography Exploits the Specific Binding Sites on Proteins

If the protein does not bind to a specific ligand, then you can attached a ligand binding domain to the protein.

The ligand binding domain is called a **tag**.

To do this you need to have the gene for the protein.

You place the DNA coding sequence for the tag in the same reading frame as the gene.

Tag	Ligand
GST (glutathione-S-transferase)	glutathione
polyhistidine	nickel
HA (11 amino acid sequence)	anti-HA antibody

If two proteins are known to interact, then one of the proteins could be used as a ligand to purify the other by affinity chromatography.

The Size and Subunit Composition of a Protein Can Be Determined by SDS Polyacrylamide-Gel Electrophoresis

The concept of gel electrophoresis: Fig. 8-14

Opposite charges attract

Any charged molecule placed in an electrical field will migrate to the pole having the opposite charge.

Electric field is generated by putting a positive electrode at one end of a gel and a negative electrode at the other, then running electricity through it.

The gel is made up of the same stuff as Jell-O.

The molecule is then placed in the gel.

A gel is used to minimize diffusion.

If two molecules having the same charge and the same charge density, the smaller one will migrate to the electrode faster than the larger one.

Due to more frictional drag on the bigger molecule.

Gel electrophoresis can be used to separate proteins or DNA (and other molecules) by size.

DNA has a uniform negative charge density.

Every nucleotide has a negatively charged phosphate.

Protein has a heterogeneous charge density.

There are positively and negatively charged amino acids, as well as neutral amino acids.

SDS is used to give protein a uniform charge density.

SDS is a negatively charged detergent. Fig. 8-13

SDS denatures a protein into a random coil.

SDS binds uniformly along the polypeptide backbone.

SDS-PAGE is used to separate proteins according to size. Fig. 8-15

More Than 1000 Proteins Can Be Resolved on a Single Gel by Two-dimensional Polyacrylamide-Gel Electrophoresis

Selective Cleavage of a Protein Generates a Distinctive Set of Peptide Fragments

Mass Spectrometry Can Be Used to Sequence Peptide Fragments and Identify Proteins

The main point about mass spec is that it can measure the mass of a molecule extremely accurately.

For example, it can tell you that a certain polypeptide has a mass of 5,237 daltons, exactly.

1 proton = 1 Dalton

Only a certain selection of amino acids can give rise to a mass of exactly 5,237.

Comparing that number with all the masses of proteins from the source organism allows you to identify the polypeptide.

The study of all proteins in an organism is called **proteomics**.

Summary

Proteins can be isolated from cells.

A variety of chromatography methods allows a specific protein to be separated from all others.

Gel electrophoresis provides a powerful means for separating protein or DNA according to size.

ISOLATING, CLONING, AND SEQUENCING DNA

Purpose: To better understand how a protein works, you often need its gene so that you can:

- Determine the amino acid sequence of the protein (arrangement of codons).
- Use the gene to measure the amount of mRNA being transcribed from the gene.
- Make large amounts of the protein (placing the gene under a very potent promoter).
- Mutate the protein (via the mutations in the gene).

We are now in the post-genomic era, which means that if we want to find the nucleotide sequence of a gene we look it up in an online database.

We still do a lot of DNA sequencing, but that is now an automated process.

As experimentalists, we want to manipulate the DNA.

The general process of manipulating DNA is called:

Recombinant DNA Technology
Genetic Engineering
DNA cloning (subcloning)

Large DNA Molecules Are Cut into Fragments by Restriction Nucleases

Bacteria produce restriction nucleases (restriction endonucleases, or restriction enzymes) to protect itself from invasion by foreign DNA.

A particular restriction enzyme will recognize a specific DNA sequence and cut both strands of that DNA.

Properties of restriction enzymes: **Fig. 8-21**

They are often dimers, recognizing palindromic sites (inverted repeats).

There are hundred of different restriction enzymes each with a specific recognition sequence.

The recognition sequences are usually 4, 6, or 8 base pairs.

All bacteria (but not eukaryotes) have restriction nucleases.

Restriction nucleases are named after the organism:

Example: EcoRI (pronounced 'Echo R one') (from *Escherichia coli*, strain R, 'I' means it was the first one isolated)

Here are some other common ones: BamHI, ClaI, EcoRV, HindIII, HpaI, PstI,

Some produce staggered cuts (aka 'sticky ends')

Restriction sites occur randomly in DNA.

Or you could engineer in a site.

A written nucleic acid sequence that identifies the location of restriction sites is called a restriction map.

Two DNA molecules that have complementary ends or blunt ends can be ligated together.

Fig. 8-22

Related terms: Gene fusion, gene splicing (not be confused with RNA splicing).

Gel Electrophoresis Separates DNA Molecules of Different Sizes

We already talked about this above. **Fig. 8-23**

The DNA in the gel is colorless.

To see it, a dye that binds DNA is added.

The dye is called ethidium bromide.

It fluoresces under UV light.

There other ways of detecting specific DNA molecules in a complex mixture (see below).

The 'bands' in the gel correspond to the shape of the well, where the DNA sample was loaded into the gel.

Purified DNA Molecules Can Be Specifically Labeled with Radioisotopes or Chemical Markers *in vitro*

Just like antibodies can be used to recognize specific proteins, DNA can be used to recognize (**probe**) a specific DNA or RNA sequence.

Through base-pairing interactions a DNA sequence will **hybridize** with a complementary DNA or RNA sequence.

To be able to detect the nucleic acid, it must be chemically modified so that it gives off a detectable signal.

The probe can be radioactively labeled.

Polynucleotide kinase + radioactive ATP will transfer the radiolabel to the DNA. **Fig. 8-24b**

Actually, the terminal phosphorous on the ATP is radioactive (^{32}P , called 'pee thirty two')

Nucleic Acid Hybridization Reactions Provide a Sensitive Way of Detecting Specific Nucleotide Sequences

Northern and Southern Blotting Facilitate Hybridization with Electrophoretically Separated Nucleic Acid Molecules

A complex mixture of RNA or DNA is separated by size using gel electrophoresis. **Fig. 8-27**

Typically, it total RNA or DNA isolated from cells.

The DNA would have to be first cut up with a restriction enzyme.

If you were to stain this gel with ethidium bromide you would see thousands of bands, which would like one big smear.

The separated RNA or DNA is transferred to a solid support.

Do this to get it out of the gel, so that the probe can readily access the nucleic acid.

This **blot** is incubated with a labeled DNA probe.

The probe diffuses around and sticks to any complementary sequence.

After a while, wash away any unbound probe.

If the probe has been radiolabeled, then you can expose the blot to photographic film.

The radiation emitted from the radioactive probe will expose the film.

This is similar to an X-ray that you get in a doctors office.

Black bands appear on the film, corresponding to the band of the specific nucleic acid being detected.

If the mixture of nucleic acid on the blot is DNA, then it is called a **Southern blot**.

Named after Ed Southern who invented the technique.

If the mixture of nucleic acid on the blot is RNA, then it is called a **Northern blot**.

An analogous procedure electrophoresing and blotting proteins, and probing with a labeled antibody is called a **Western blot**.

Hybridization Techniques Locate Specific Nucleic Acid Sequences in Cells or on Chromosomes

This whole probing thing can be done on intact cells and chromosomes. **Fig. 8-28**

Called **in situ hybridization**.

Gel electrophoresis is not involved.

Genes Can Be Cloned from a DNA Library

Plasmid are circular pieces of DNA normally found in bacteria. **Fig. 8-30, 8-31**

Properties of plasmids:

They replicate in bacteria

Thy range in size, but are usually a few kilobase pairs in length.

A DNA fragment can be cloned into a plasmid and propagated.

Use restriction enzymes.

Cut → ligate → transform E. coli → grow on media that allow only transformants to grow.

The plasmid usually has a gene for an the antibiotic ampicillin.

Add ampicillin to the media.

Two Types of DNA Libraries Serve Different Purposes

cDNA Clones Contain Uninterrupted Coding Sequences

Isolated DNA Fragments Can Be Rapidly Sequenced

Nucleotide Sequences Are Used to Predict the Amino Acid Sequences of Proteins

A nucleotide sequence has three possible reading frames in each direction. **Fig. 8-38**
Stop codon occur frequently in the wrong reading frame.

The correct reading frame is always “open”

An **ORF** is an **open reading frame**

The Genomes of Many Organisms Have Been Fully Sequenced

human	yeast
mouse	<i>E. coli</i> and hundreds of other bacteria
fruit fly	hundreds of viruses
<i>Arabidopsis</i> (plant)	<i>C. elegans</i> (worm)

The human genome has a lot of DNA of bacterial origin!

Selected DNA Segments Can Be Cloned in a Test Tube by a Polymerase Chain Reaction

Very important stuff here.

PCR is a quick way to isolate and/or amplify DNA. **Fig. 8-39**

PCR involves the following components:

thermostable DNA polymerase (e.g. Taq polymerase, which need to survive 95°C temperatures)

template DNA (often in trace amounts)

Two DNA primers (to prime DNA replication on each strand in a converging direction)

dNTPs (all four deoxyribonucleotide triphosphates as substrates).

The PCR cycle:

1. Melt template DNA at high temperature. (e.g. 95°C)
 2. Cool to allow primers to anneal. (e.g. 42°C)
 3. Raise temperature to optimal temperature for the thermostable polymerase to replicate DNA. (72°C)
- Repeat (e.g. 30 times)

Result: exponential amplification 1→2→4→8→16→32→64→...

One major use of PCR is in forensic science **Fig. 8-41**

Cellular Proteins Can Be Made in Large Amounts Through the Use of Expression Vectors

We touched on this area earlier. So be familiar with it.

Summary

Restriction enzymes are used to cut DNA at specific locations.

Proteins can be purified and studied in isolation.

Proteins can be engineered to have affinity tags that are useful for purification.

Proteins and DNA can be separated by size by gel electrophoresis.

Specific nucleic acids can be detected by hybridization with a labeled probe.

Proteins can be detected using labeled antibodies raised against the protein.

Many genomes including human have been fully sequenced.

PCR is a reiterative method for amplifying small amounts of DNA.

ANALYZING PROTEIN STRUCTURE AND FUNCTION

How do proteins work? enzymology
What do proteins look like? structural biology (crystallography and NMR)
What do proteins interact with? – other protein and DNA

The Diffraction of X-rays by Protein Crystals Can Reveal a Protein's Exact Structure

Fig. 8-45

Molecular Structure Can Also Be Determined Using Nuclear Magnetic Resonance (NMR) Spectroscopy

Sequence Similarity Can Provide Clues About Protein Function

Comparing the amino acid sequence of a protein in two distantly related organisms (e.g. yeast and humans) shows you conserved regions. Fig. 8-47

BLAST is an online alignment program.

Submit a DNA or protein sequence via the web.

Similar sequences in the database are returned, and aligned with your sequence.

Conserved regions are likely to have important function since mutations in conserved regions have been evolutionarily selected against.

Fusion Proteins Can Be Used to Analyze Protein Function and to Track Proteins in Living Cells

We talked about tagging proteins by fusing the coding sequence of an affinity tag (or any other useful protein) to the gene of interest.

If the tag is GFP (green fluorescent protein from jellyfish), then it will emit light in the cell.

Allow you to localize a protein in real time in live cells.

Immunoprecipitation and Affinity Chromatography Allow Identification of Associated Proteins

Proteins tagged with an antigen tag (e.g. HA) will bind antibodies.

If the antibodies are immobilized to a solid support (like chromatography resin) then you can **immunopurify** the protein.

If your tagged protein is in a complex with other proteins it will be **co-immunoprecipitated** along with your protein.

Protein–Protein Interactions Can Be Identified by Use of the Two-Hybrid System

The yeast two hybrid is based upon the concept that an activation domain will activate transcription if brought near a promoter. **Fig. 8-51**

An activation domain by itself cannot find its way to a promoter.

A DNA binding domain will not activate transcription, but can find its way to a promoter.

Assuming the DNA recognition sequence for the DNA binding domain is present.

Fuse both the DBD and AD to two proteins that interact, and wah-la! The activation domain is brought to the promoter and transcription is activated.

Transcription is typically from some reporter gene that is easy to detect.

Phage Display Methods Also Detect Protein Interactions

Protein Interactions Can Be Monitored in Real Time Using Surface Plasmon Resonance

DNA Footprinting Reveals the Sites Where Proteins Bind on a DNA Molecule

Summary

The function of proteins are dissected by a combination of a variety of techniques:

- *Determining their structure via X-ray crystallography.*
- *Identifying conserved domains.*
- *Mutating parts of the protein and determining how loss of structure affects function.*
- *Determining what other proteins it interacts with.*

STUDYING GENE EXPRESSION AND FUNCTION

The Classical Approach Begins with Random Mutagenesis

The classical approach is to start with a phenotype

The genome is randomly mutated.

Then you look through the pile of mutants and look for any organisms that possess the phenotype that interest you.

Using strategies to sort through millions of mutant for the interesting ones, is called a **genetic screen**.

Example of a screen for mutations that allow cells to grow at 23°C but not at 36°C. **Fig. 8-57**

Get **temperature-sensitive (TS) mutants**.

A **phenotype** is the outward manifestation of the genotype.

Example: yeast cells that lack the *LEU2* gene have a Leu- (pronounced 'lew minus') phenotype, meaning that they cannot synthesize leucine.

Genotype is a written description of any gene variants of an organism.

Example: *spt15-21* is a shorthand notation for a mutation in the yeast *SPT15* gene.

Genetic Screens Identify Mutants Deficient in Cellular Processes

A Complementation Test Reveals Whether Two Mutations Are in the Same or in Different Genes

Genes Can Be Located by Linkage Analysis

Searching for Homology Can Help Predict a Gene's Function

Reporter Genes Reveal When and Where A Gene Is Expressed

Microarrays Monitor the Expression of Thousands of Genes At Once

Traditionally, gene expression has been studied one gene at a time.

Now, the expression of all genes in an organism can be studied simultaneously using DNA microarrays. **Fig. 8-62**

How microarrays work:

PCR amplify each gene individually

Spot it on a glass slide.

Each spot corresponds to the DNA of a particular gene.

A yeast microarray has 6000 spots corresponding to 6000 genes.

Isolate mRNA from two differently treated cells that you wish to compare.

Typically, one set of cells are untreated and the other is subjected to some environmental change.

One mRNA set is fluorescently colored **Red**; the other **Green**.

They are then mixed and hybridized to the microarray.

Unbound mRNA is washed away.

The fluorescence of the microarray is examined.

A red spot means the gene is expressed in set colored red, but not the other.

You can guess what a green spot means.

A yellow spot means the gene is expressed in both sets.

A dark spot means that the gene is not expressed in either set.

If you do 12 experiments (like a time course response to a treatment) you have 12 x 6000 data points = 72,000 data points!

Bioinformatics:

Handling and analyzing large amounts of biological data is the field of **bioinformatics**.

It's hard to see relationships among genes by looking at a sea of numbers.

Color patterns are easy to see. **Fig. 8-63**

So the data is recoded such that genes that go up in one set are red, and up in the other are green.

No change (yellow and dark spots) are coded black.

So you have a spreadsheet in which numbers are shown to you as shades of color.

Statistical software can rearrange rows of gene data according to their similarity.

Called cluster analysis.

Targeted Mutations Can Reveal Gene Function

Cells and Animals Containing Mutated Genes Can Be Made to Order

The Normal Gene in a Cell Can Be Directly Replaced by An Engineered Mutant Gene in Bacteria and Some Lower Eucaryotes

Engineered Genes Can Be Used to Create Specific Dominant Negative Mutations in Diploid Organisms

Gain-of-Function Mutations Provide Clues to the Role Genes Play in a Cell or Organism

Genes Can Be Redesigned to Produce Proteins of Any Desired Sequence

Site-directed mutagenesis allows a single codon to be changed in a gene. **Fig. 8-69**
An oligonucleotide primer is synthesized such that the mutation is contained in the middle of the primer.

The primer is then anneal to the cloned gene that is to be mutated.

DNA polymerase is used to replicate the DNA using the primer.

The replicated DNA is then amplified in *E. coli*.

Half of the transformants will contain the mutation.

You can screen for the mutation by sequencing the mutated region.

Engineered Genes Can Be Easily Inserted into the Germ Line of Many Animals

Gene Targeting Makes It Possible to Produce Transgenic Mice That Are Missing Specific Genes

It is now routine to add, delete (or modify) a mouse gene.

The result is a **transgenic mouse**.

Transgenic mice provide useful proxies for human diseases.

Transgenic Plants Are Important for Both Cell Biology and Agriculture

Large Collections of Tagged Knockouts Provide a Tool for Examining the Function of Every Gene in an Organism

Summary

There are a number of ways at getting at gene function in vivo.

*Randomly mutagenize the genome and look for interesting phenotypes – then find the gene that restores the normal **wild type** phenotype.*

Use DNA microarray technology to simultaneously survey the expression of all genes in a genome.

Use site-directed mutagenesis and gene replacement technology to knock out part or all of a gene.