MOLECULAR GENETICS 4606



Helen Chamberlin
The Ohio State University

LibreTexts™

The Ohio State University

Molecular Genetics 4606

Helen Chamberlin

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 03/04/2026

TABLE OF CONTENTS

Licensing

1: The Cell Cycle

- 1.1: The Cell Cycle and Changes in DNA Content
- 1.2: Bacterial Cell Division and the Eukaryotic Cell Cycle
- 1.3: Regulation of the Cell Cycle

2: Mitosis and Meiosis

- 2.1: Mitosis
- 2.2: Meiosis
- 2.3: Mitosis and Meiosis

3: Alleles and Probabilities

- 3.1: Genes are the Basic Units of Inheritance
- 3.2: Model Organisms Facilitate Genetic Advances
- 3.3: Mendel's First Law
- 3.4: Biochemical Basis of Dominance
- 3.5: Crossing Techniques Used in Classical Genetics
- 3.6: Dihybrid Crosses
- 3.7: Why Use Nomenclature?
- 3.8: Basic Nomenclature
- 3.9: Linked Genes

4: Hypothesis Testing

- 4.1: Phenotypic Ratios May Not Be As Expected
- 4.2: Probability
- 4.3: Chi-Square Test of Goodness-of-Fit

5: Sex Determination, Chromosomes and Linkage

- 5.1: Karyotypes Describe Chromosome Number and Structure
- 5.2: Sex-Linkage- An Exception to Mendel's First Law

6: Pedigrees

- 6.1: Pedigree Analysis
- 6.2: Inferring the Mode of Inheritance
- 6.3: Sporadic and Non-Heritable Diseases
- 6.4: Calculating Probabilities

7: Multiple Gene Alleles

- 7.1: Relationships Between Genes, Genotypes and Phenotypes
- 7.2: Complementation tests and Allelism

8: Gene Interactions and Epistasis

- 8.1: Epistasis and Other Gene Interactions
- 8.2: Example of Multiple Genes Affecting One Character

9: Genetic Pathways

- 9.1: Genetic Pathways

10: Nucleic Acids

- 10.1: DNA is the Genetic Material
- 10.2: The Structure of DNA
- 10.3: DNA- The Genetic Material

11: Amplifying and Manipulating DNA Fragments

- 11.1: Prelude to Molecular Genetics
- 11.2: Isolating Genomic DNA
- 11.3: Isolating or Detecting a Specific Sequence by PCR
- 11.4: PCR and Gel Electrophoresis
- 11.5: Origins of Molecular Polymorphisms
- 11.6: Classification and Detection of Molecular Markers
- 11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
- 11.8: Make and Screen a cDNA Library
- 11.9: DNA Sequencing
- 11.10: Genomic Libraries
- 11.11: The Polymerase Chain Reaction (PCR)

12: Evaluating Recombinant DNA

- 12.1: Cloning DNA - Plasmid Vectors
- 12.2: DNA Analysis - Gel Electrophoresis
- 12.3: Recombinant DNA Technology

13: Detecting Genes and Gene Products

- 13.1: DNA Analysis- Blotting and Hybridization

14: DNA Replication

- 14.1: DNA Structure
- 14.2: Genes and Chromatin in Eukaryotes
- 14.3: Introduction
- 14.4: DNA Replication
- 14.5: DNA- The Genetic Material

15: Transcription

- 15.1: Introduction
- 15.2: Overview of Transcription
- 15.3: Details of Transcription

16: Transcriptional Regulation (prokaryotes)

- 16.1: The lac Operon
- 16.2: The Use of Mutants to Study the lac Operon
- 16.3: Gene Regulation in Prokaryotes

17: RNA processing

- 17.1: Details of Eukaryotic mRNA Processing
- 17.2: Gene Expression- Transcription

18: Transcriptional Regulation (eukaryotes)

- 18.1: Eukaryotic Gene Regulation
- 18.2: Regulatory Elements in Evolution
- 18.3: Additional Levels of Regulating Transcription
- 18.4: Gene Regulation in Eukaryotes
- 18.5: Transgenic organisms

19: Small RNAs

- 19.1: Post-transcriptional Control of Gene Expression

20: The Genetic Code

- 20.1: The Function of Genes
- 20.2: An Overview of the Genetic Code
- 20.3: Gene and Protein Colinearity and Triplet Codons
- 20.4: Translation
- 20.5: Gene Expression- Translation
- 20.6: Gene Expression- Applied Example (Part 1)

21: Alleles and Gene Function

- 21.1: Types of Mutations
- 21.2: DNA Mutations

22: Cancer Genetics

- 22.1: Classification of Cancers
- 22.2: Cancer Cell Biology
- 22.3: Hallmarks of Cancer
- 22.4: Mutagens and Carcinogens
- 22.5: Oncogenes
- 22.6: Tumor Suppressor Genes
- 22.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

23: Genomics

- 23.1: ‘Omics Technologies
- 23.2: DNA Sequencing
- 23.3: Whole Genome Sequencing

24: Functional Genomics

- 24.1: Functional Genomics – Determining Function(s)
- 24.2: Genomic Approaches- The DNA Microarray

25: Genetic Linkage

- 25.1: Linkage
- 25.2: Recombination
- 25.3: Linkage Reduces Recombination Frequency
- 25.4: Crossovers Allow Recombination of Linked Loci
- 25.5: Inferring Recombination From Genetic Data
- 25.6: Genetic Mapping

26: Genetic Maps

- 26.1: Mapping With Three-Point Crosses

Index

Glossary

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: The Cell Cycle

1.1: The Cell Cycle and Changes in DNA Content

1.2: Bacterial Cell Division and the Eukaryotic Cell Cycle

1.3: Regulation of the Cell Cycle

Thumbnail: Life cycle of the cell. (CC BY-SA 4.0; BruceBlaus).

1: The Cell Cycle is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

1.1: The Cell Cycle and Changes in DNA Content

Four stages of a typical cell cycle

The life cycle of eukaryotic cells can generally be divided into four stages and a typical cell cycle is shown in Figure **1.1.13**. When a cell is produced through fertilization or cell division, there is usually a lag before it undergoes DNA synthesis (replication). This lag period is called Gap 1 (**G_1**), and ends with the onset of the DNA synthesis (**S**) phase, during which each chromosome is replicated. Following replication, there may be another lag, called Gap 2 (**G_2**), before mitosis (**M**). Cells undergoing meiosis do not usually have a G_2 phase. **Interphase** is as term used to include those phases of the cell cycle excluding mitosis and meiosis. Many variants of this generalized cell cycle also exist. Some cells never leave G_1 phase, and are said to enter a permanent, non-dividing stage called **G_0**. On the other hand, some cells undergo many rounds of DNA synthesis (S) without any mitosis or cell division, leading to endoreduplication. Understanding the control of the cell cycle is an active area of research, particularly because of the relationship between cell division and cancer.

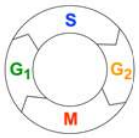


Figure **1.1.13**: A typical eukaryotic cell cycle.(Original-Deyholos-CC:AN)

Measures of DNA content and chromosome content

The amount of DNA within a cell changes following each of the following events: fertilization, DNA synthesis, mitosis, and meiosis (Fig 2.14). We use "**c**" to represent the DNA content in a cell, and "**n**" to represent the number of complete sets of chromosomes. In a gamete (i.e. sperm or egg), the amount of DNA is 1c, and the number of chromosomes is 1n. Upon fertilization, both the DNA content and the number of chromosomes doubles to 2c and 2n, respectively. Following DNA replication, the DNA content doubles again to 4c, but each pair of sister chromatids is still counted as a single chromosome (a **replicated chromosome**), so the number of chromosomes remains unchanged at 2n. If the cell undergoes mitosis, each daughter cell will return to 2c and 2n, because it will receive half of the DNA, and one of each pair of sister chromatids. In contrast, the 4 cells that come from meiosis of a 2n, 4c cell are each 1c and 1n, since each pair of sister chromatids, and each pair of homologous chromosomes, divides during meiosis.

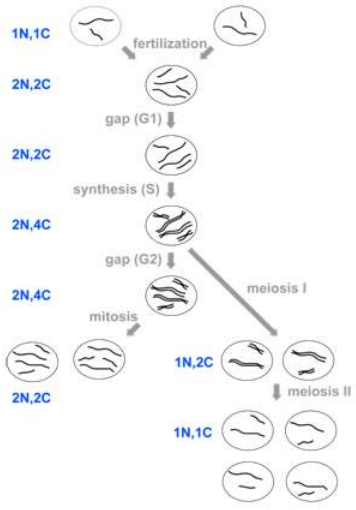


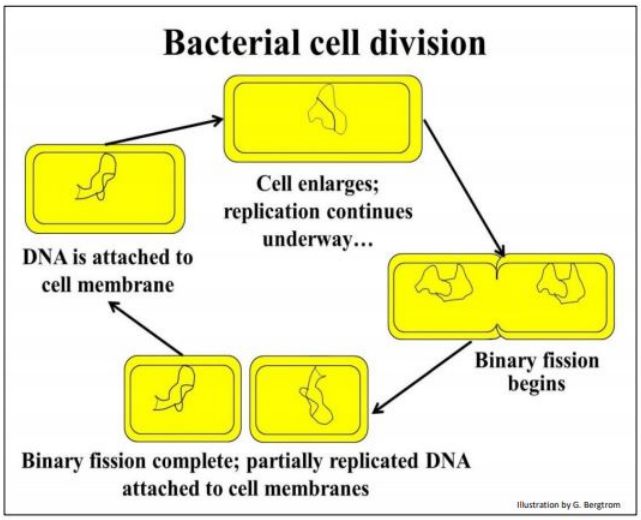
Figure **1.1.14**: Changes in DNA and chromosome content during the cell cycle. For simplicity, nuclear membranes are not shown, and all chromosomes are represented in a similar stage of condensation.*(Original-Deyholos-CC:AN)*

This page titled 1.1: The Cell Cycle and Changes in DNA Content is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: The Cell Cycle and Changes in DNA Content** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

1.2: Bacterial Cell Division and the Eukaryotic Cell Cycle

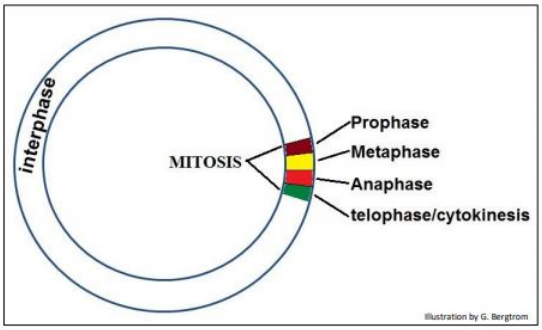
The life of actively growing bacteria is not separated into a time for duplicating genes (i.e., DNA synthesis) and one for **binary fission** (dividing and partitioning the duplicated DNA into new cells). Instead, the single circular chromosmome of a typical bacterium is replicating even before fission is complete, so that the new daughter cells already contained partially duplicated chromosomes. Cell growth, replication and fission are illustrated below



339 Binary Fission

The roughly 30-60 minute life cycle of an actively growing bacterium is not divided into discrete phases. On the other hand, typical eukaryotic cells have a roughly 16-24 hour cell cycle (depending on cell type) that is divided into four separate phases. In the late 1800s, light microscopy revealed that some cells lost their nuclei while forming **chromosomes** (from *chroma*, colored; *soma*, bodies). In **mitosis**, paired, attached chromosomes (**chromatids**) were seen to separate and to be drawn along **spindle fibers** to opposite poles of dividing cells. Thus homologous chromosomes were equally partioned to the daughter cells at the end of cell division. Because of the same chromosomal behavior was observed in **mitosis** in diverse organisms, chromosomes were soon recognized as the stuff of inheritance, the carrier of genes!

The short period of intense mitotic activity was in stark contrast to the much longer '*quiet*' time in the life of the cell, called **interphase**. The events of mitosis itself were described as occurring in 4 phases occupyiing a short time as shown below



Depending on whom you ask, **cytokinesis** (the cell movements of actually dividing a cell in two) is *not* part of mitosis. In that sense, we can think of three stages in the life of a cell: interphase, mitosis and cytokinesis. Of course, it turned out that interphase is not cellular 'quiet time' at all!

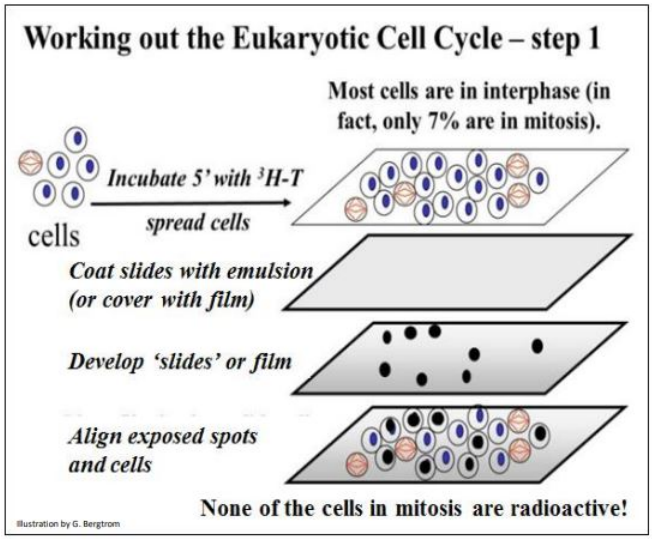
A. Defining the Phases of the Cell Cycle

Correlation of the inheritance of specific traits with that of chromosomes was demonstrated early in the early 20th century, most elegantly in genetic studies of the fruit fly, *Drosophila melanogaster*. At that time, chromosomes were assumed to contain the

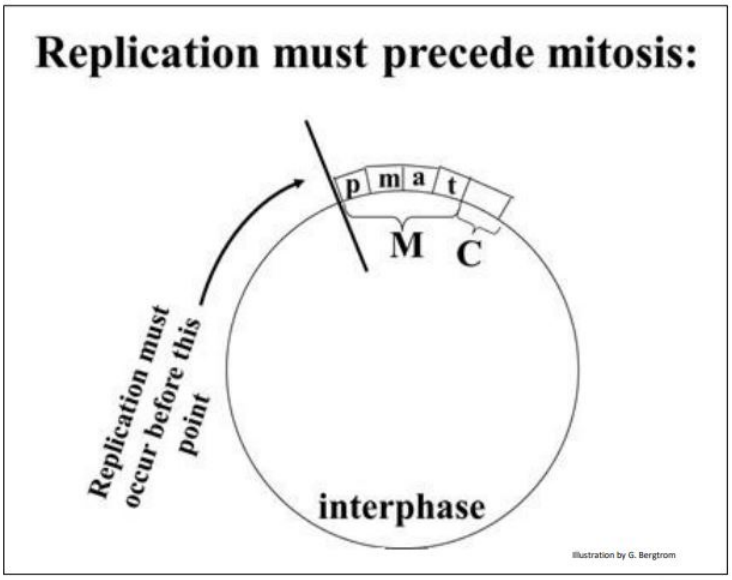
genetic material and that both were duplicated during mitosis. The first clue that this was not so came only after the discovery that DNA was in fact the chemical *stuff of genes*. The experiment distinguishing the time of chromosome formation from the time of DNA duplication is summarized below.

1. Cultured cells were incubated with **^{3}H-thymine**, the radioactive base that cells will incorporate into thymidine triphosphate (dTTP), and then into DNA.
2. Cultured cells were incubated with ^{3}H-thymine, the radioactive base that cells will incorporate into thymidine triphosphate (dTTP), and then into DNA.
3. Slides were dipped in a light-sensitive emulsion containing the same light sensitive chemicals as found in the emulsion-side of film.
4. After some time to allow the radioactivity on the slide to '*expose*' the emulsion, the slides were developed (in much the same way as developing film).
5. The resulting **autoradiographs** in the microscope revealed images in the form of dark spots created by exposure to **hot** (i.e., radioactive DNA.

If DNA replicates in chromosomes undergoing mitosis, then when the developed film is placed back over the slide, any dark spots should lie over the cells in mitosis, and not over cells that are not actively dividing. The experimental is illustrated below



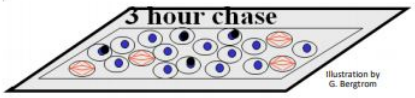
Observation of the autoradiographs show that none of the cells in mitosis is radioactively labeled. But some of the cells in interphase were! Therefore, DNA synthesis must take place sometime in interphase, before mitosis and cytokinesis (illustrated below).



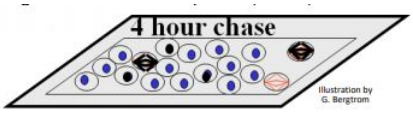
340 Experiments that Reveal Replication in Interphase of the Cell Cycle

Next a series of **pulse-chase** experiments were done to determine when in the cell cycle DNA synthesis actually takes place. Cultured cells given a short *pulse* (exposure) to ^{3}H-thymine and then allowed to grow in non-radioactive medium for different times (the *chase*). At the end of each chase time, cells were spread on a glass slide and again prepared for autoradiography. Analysis of the autoradiographs identified distinct periods of activity within interphase: Gap1 (**G_1**), a time of DNA *synthesis* (**S**) and Gap 2 (**G_2**). Here are the details of these very creative experiments, performed before it became possible to synchronize cells in culture so that they would all be growing and dividing at the same time.

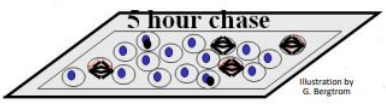
- Cells were exposed to 3H-thymine for just 5 minutes (the **pulse**) and then centrifuged. The radioactive supernatant was then discarded.
- The cells were rinsed and centrifuged again to remove as much labeled precursor as possible.
- The cells were re-suspended in fresh medium containing unlabeled (i.e., nonradioactive) thymine and further incubated for different times (the chase periods).After dipping the slides in light-sensitive emulsion, exposing and developing the film, the autoradiographs were examined, with the following results:
  - After a 3-hour (or less) chase period, the slides looked just like they would immediately after the pulse. That is, none of the 7% of the cells that were in mitosis is radioactively labeled, but many interphase cells showed labeled nuclei, as shown below.



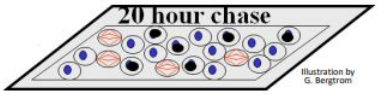
  - After 4 hours of chase, a few of the 7% of the cells that were in mitosis were labeled, along with others in interphase (below).



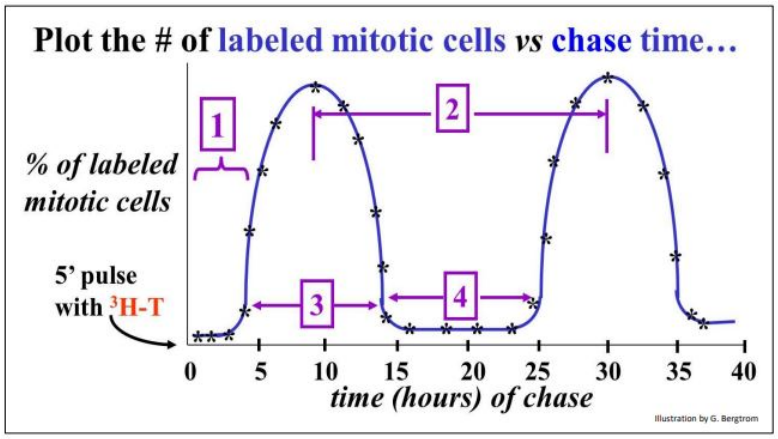
  - After a 5 hour chase, most cells in mitosis (still about 7% of cells on the slide) were labeled; many fewer cells in interphase were labeled (below).



  - After a 20 hour chase, none of the 7% of cells that were in mitosis is labeled. Instead, all of the labeled cells are in interphase (below).



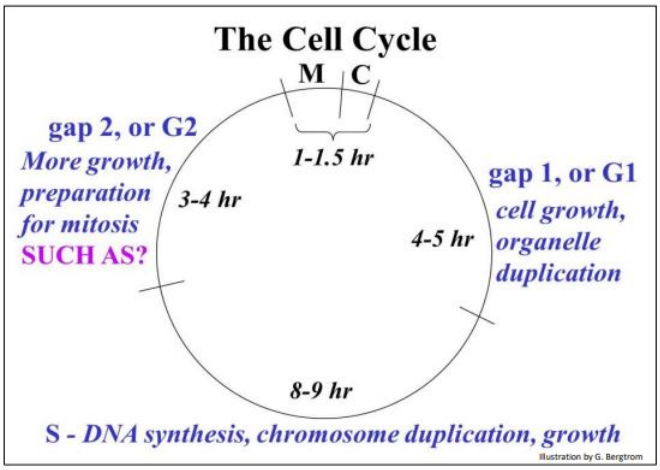
- The graph below plots a count of *radiolabeled mitotic cells* against *chase times*.



The plot defines the duration of events, or phases of the cell cycle as follows:

- The first phase (interval **#1** on the graph) must be the time between the end of DNA synthesis and the start of mitosis, defined as **Gap 2 (G_2)**.
- Cell doubling times are easily measured. Assume that the cells in this experiment doubled every 20 hours. This would be consistent with the time interval of 20 hours between successive peaks in the number of radiolabeled mitotic cells after the pulse (interval **#2**).
- Interval **#3** is easy enough to define. It is the time when DNA is synthesized, from start to finish; this is the **synthesis**, or **S** phase.
- One period of the cell cycle remains to be defined, but it is not on the graph! That would be the time between the end cell division (i.e., mitosis and cytokinesis) and the beginning of DNA synthesis (*replication*). That interval can be calculated from the graph as the time of the cell cycle (~20 hours) minus the sum of the other defined periods of the cycle. This phase is defined as the **Gap 1 (G_1)** phase of the cycle.

So at last, here is our cell cycle with a summary of events occurring in each phase.



During all of interphase (**G_1, S and G_2**) , the cell grows in size, preparing for the next cell division. Growth in **G_1** includes the synthesis of enzymes and other proteins that will be needed for replication.

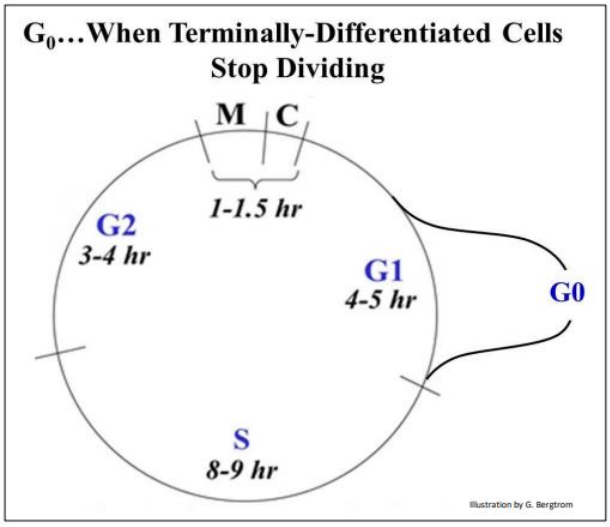
DNA is replicated during the S phase, along with the synthesis of new *histone* and other proteins that will be needed to assemble new *chromatin*. $\mathbf{G_2}$ is the shortest time of interphase and is largely devoted to preparing the cell for the next round of mitosis and cytokinesis. Among the proteins whose synthesis increases in this time are the *tubulins* and proteins responsible for condensing chromatin into the paired chromatids representing the duplicated chromosomes. **Cohesin** is a more recently discovered protein made in the run-up to mitosis. It holds centromeres of chromatids together until they are ready to separate.

341 Events in the Phases of the Cell Cycle

In a final note, typical dividing cells have generation times ranging from 16 to 24 hours. Atypical cells, like newly fertilized eggs, might divide every hour or so! In these cells, events that normally take many hours must be completed in just fractions of an hour.

B. When Cells Stop Dividing

Terminally differentiated cells are those that spend the rest of their lives performing a specific function. These cells no longer cycle. Instead, shortly after entering G_1 they are diverted into a phase called $\mathbf{G_0}$, as shown below.



Referred to as **terminally differentiated**, these cells normally never divide again. With a few exceptions (e.g., many neurons), most terminally differentiated cells have a finite lifespan, and must be replaced by stem cells. Examples include red blood cells. With a half-life of about 60 days, they are regularly replaced by reticulocytes produced in bone marrow

---

This page titled 1.2: Bacterial Cell Division and the Eukaryotic Cell Cycle is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

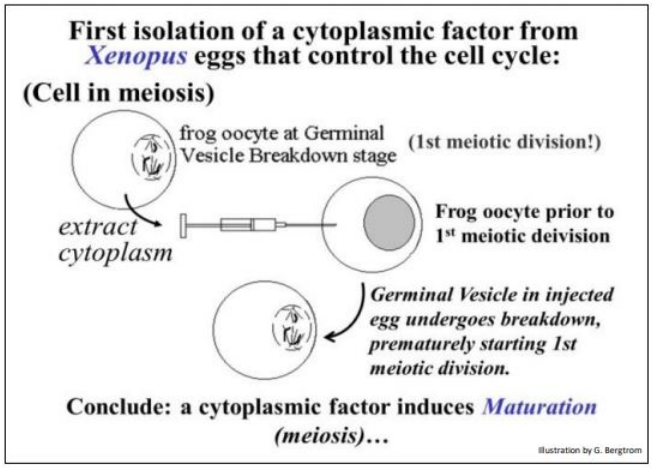
- **19.2: Bacterial Cell Division and the Eukaryotic Cell Cycle** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

1.3: Regulation of the Cell Cycle

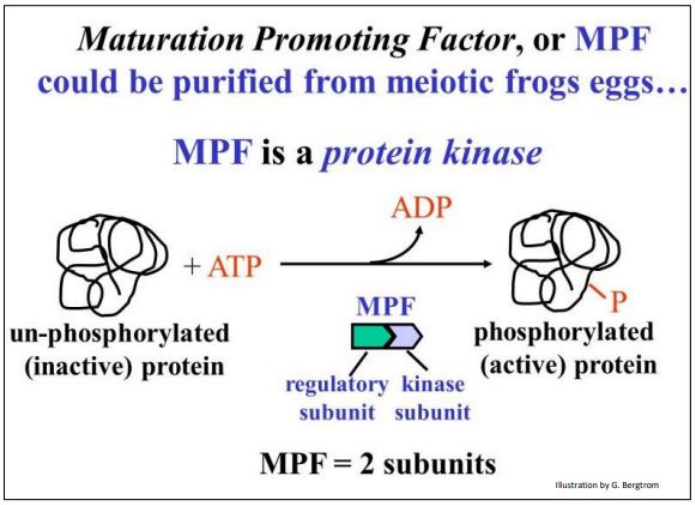
Progress through the cell cycle is regulated. The cycle can be controlled or put on 'pause' at any one of several phase transitions. Such **checkpoints** monitor whether the cell is on track to complete a successful cell division event. Superimposed on these controls are signals that promote cell differentiation. Embryonic cells *differentiate* as the embryo develops. Even after *terminal differentiation* of cells that form all adult tissues and organs, *adult stem cells* will divide and differentiate to replace worn out cells. Once differentiated, cells are typically signaled in $\mathbf{G_1}$ to enter $\mathbf{G_0}$ and stop cycling. In some circumstances cells in $\mathbf{G_0}$ are recruited to resume cycling. However, if this occurs to by mistake, the cells may be *transformed* to cancer cells. Here we consider how the normal transition between phases of the cell cycle is controlled.

A. Discovery and Characterization of *Maturation Promoting Factor (MPF)*

Growing, dividing cells monitor their progress through the phases. Cells produce internal chemical signals that tell them when it's time to begin replication or mitosis, or even when to enter into G_0 when they reach their terminally differentiated state. The experiment that first demonstrated a chemical regulator of the cell cycle involved fusing very large frog's eggs! The experiment is described below.



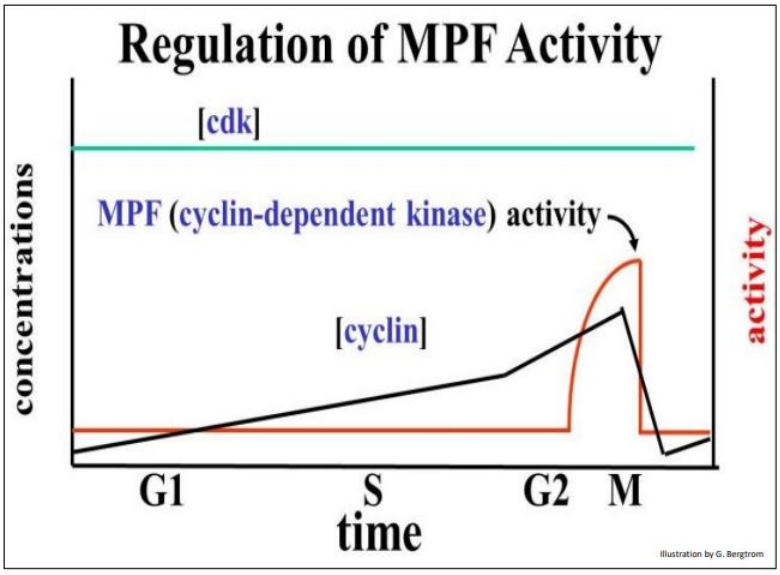
The hypothesis tested here was that frog oocyte cytoplasm from **germinal vesicle stage** oocytes (i.e., in mid-meiosis) contains a chemical that caused the cell to lose its nuclear membrane, condense its chromatin into chromosomes and enter meiosis. Cytoplasm was withdrawn from one of these mid-meiotic oocytes with a fine hypodermic needle, and then injected into a pre-meiotic oocyte. The mid-meiotic oocyte cytoplasm induced premature meiosis in the immature oocyte. A **maturation promoting factor** (**MPF**) could be isolated from the mid-meiotic cells and injected into pre-meiotic cells; it caused them to enter meiosis. **MPF** turns out to be a protein kinase made up of two polypeptide subunits as shown below.



MPF was then also shown to stimulate *somatic cells* in G_2 to enter premature mitosis. So conveniently, *MPF* can also be **Mitosis Promoting Factor**! Hereafter we will discuss the effects of *MPF* as being equivalent in mitosis and meiosis. When active, *MPF* targets many cellular proteins.

342 Discovery of MPF Kinase and Its Role in Meiosis and Mitosis

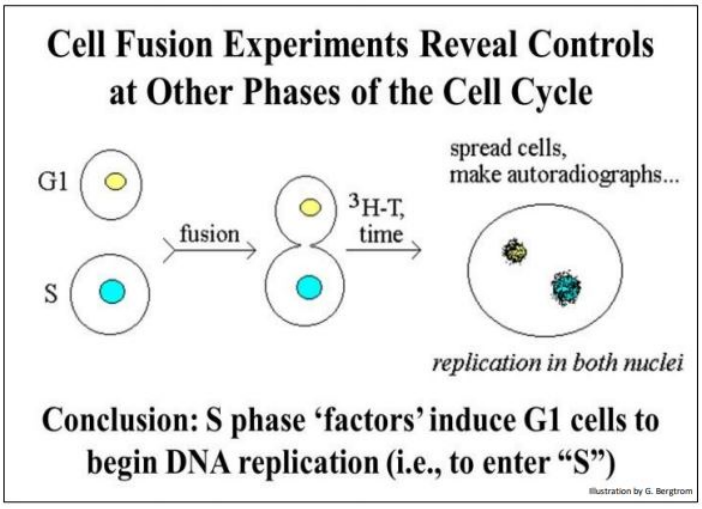
Assays of *MPF* activity as well as the actual levels of the two subunits over time during the cell cycle are graphed below.



One subunit of *MPF* is **cyclin**, a regulatory polypeptide. The other subunit, **cyclin-dependent kinase** (**cdk**), contains the kinase enzyme *active site*. Both subunits must be bound to make an active kinase. Cyclin was so-named because its levels rise gradually after cytokinesis, peak at the next mitosis, and then fall. Levels of the cdk subunit do not change significantly during the life of the cell. Because the kinase activity of MPF requires *cyclin*, it tracks the rise in cyclin near the end of the G_2, and its fall after mitosis. Cyclin begins to accumulate in G_1, rising gradually and binding to more and more *cdk* subunits. *MPF* reaches a threshold concentration in G_2 that triggers entry into mitosis. For their discovery of these central molecules Leland H. Hartwell, R. Timothy Hunt, and Paul M. Nurse won the 2001 Nobel Prize in Physiology or Medicine.

B. Other Cyclins, CDKs and Cell Cycle Checkpoints

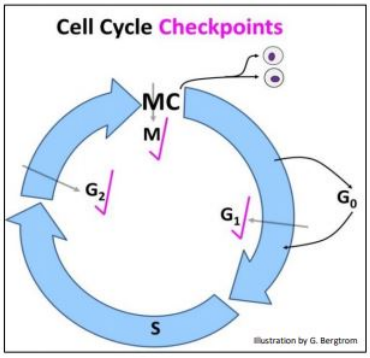
Other chemical signals accumulate at different points in the cell cycle. For example, when cells in S are fused with cells in G_1, the G_1 cells begin synthesizing DNA (visualized as ^{3}H-thymine incorporation). An experiment showing control of progress to different phases of the cell cycle is illustrated below.



An **S** phase factor could be isolated from the **S** phase cells. This factor also turns out to be a two-subunit protein kinase, albeit a *different* one from MPF. Just as MPF signals cells in $\mathbf{G_2}$ to begin mitosis, the **S** phase kinase signals cells in $\mathbf{G_1}$ to enter the **S** phase of the cell cycle. *MPF* and the **S** phase kinase govern activities at two of several cell cycle **checkpoints**. In each case, the activity of the kinases is governed by prior progress through the cell cycle. In other words, if the cell is not ready to begin mitosis, active *MPF* production is delayed until it is. Likewise, the **S** phase kinase will not be activated until the cell is ready to begin DNA synthesis.

343 Cell Cycle Control at Check Points and the Go "Phase"

The sequence of signals that control progress through the cell cycle is probably more intricate and extensive than we currently know, but the best-described checkpoints are in $\mathbf{G_1}$, $\mathbf{G_2}$ and **M** (below).



We generally envision checkpoints as monitoring and blocking progress until essential events of a current phase of the cell cycle phase are completed. These kinases are part of molecular sensing mechanisms that act by phosphorylating cytoplasmic and/or nuclear proteins required by upcoming phases of the cycle. Let's take a closer look at some events that are monitored at these *checkpoints* in more detail.

1. The G_1 Checkpoint

The G_1 checkpoint controls the transition from the G_1 to the S phase of the cell cycle. If actively dividing cells (e.g., stem cells) in G_1 fail to complete their preparation for replication, the S-phase kinase won't be produced and the cells won't proceed the S phase until the preparatory biochemistry catches up with the rest of the cycle. To enter S, a cell must be ready to make proteins of replication, like DNA polymerases, helicases, and primases among others. Only when these molecules have accumulated to (or become active at) appropriate levels, is it "safe" to enter S and begin replicating DNA. This description of $\mathbf{G_1}$ **checkpoint** activity is consistent with the idea that all checkpoints delay cycling until a prior phase is complete.

What about cells that are fully differentiated? Such *terminally differentiated* cells stop producing the active $\mathbf{G_1}$ *checkpoint* kinase and stop dividing. These cells are arrested in $\mathbf{G_0}$ (see below). As an interesting side-note, recall that somatic cells are **diploid** and germ cells (sperm, egg) are **haploid**. So, are cells in $\mathbf{G_2}$ that have already doubled their DNA 'tetraploid', however briefly? Whether or not we can call $\mathbf{G_2}$ cells tetraploid (officially, probably not), it is clear that $\mathbf{G_1}$ cells and $\mathbf{G_0}$ cells are diploid!

2. The G_2 Checkpoint

Passage through the $\mathbf{G_2}$ checkpoint is only possible if DNA made in the prior **S** phase is not damaged. Or if it was, that the damage has been (or can be) repaired (review the proofreading functions of DNA polymerase and the various DNA repair pathways). Cells that do successfully complete replication and pass the $\mathbf{G_2}$ checkpoint must prepare to make the proteins necessary for the upcoming mitotic phase. These include nuclear proteins necessary to *condense* chromatin into chromosomes, *tubulins* for making *microtubules*, etc. Only when levels of these and other required proteins reach a threshold can the cell begin mitosis. Consider the following two tasks required of the $\mathbf{G_2}$ checkpoint (in fact, any checkpoint):

- sensing whether prior phase activities have been successfully completed.
- delaying transition to the next phase if those activities are unfinished.

But what if sensing is imperfect and a checkpoint is leaky? A recent study suggests that either the $\mathbf{G_2}$ checkpoint **is** leaky, or at least, that incomplete activities in the **S** phase are tolerated, and that some DNA repair is not resolved until mitosis is underway in

M! Check it out at DNA repair and replication during mitosis.

3. M Checkpoint

The M checkpoint is monitored by the original *MPF* phosphorylation of proteins that: (a) bind to chromatin causing it to condense and form chromatids, (b) lead to the breakdown of the nuclear envelope, and (c) enable spindle fiber formation,. In addition, tension in the spindle apparatus at metaphase tugs at the kinetochores holding the duplicated chromatids together. When this tension reaches a threshold, MPF peaks and an activated **separase** enzyme causes the chromatids to separate at their centromeres. Beginning in *anaphase*, tension in the spindle apparatus draws the new chromosomes to opposite poles of the cell. Near the end of mitosis and cytokinesis, proteins phosphorylated by MPF initiate the breakdown of cyclin in the cell. Passing the **M** checkpoint means that the cell will complete mitosis and cytokinesis, and that each daughter cell will enter a new $\mathbf{G_1}$ phase.

Dividing yeast cells only seem to have the three checkpoints discussed here. More complex eukaryotes use more *cyclins* and *cdks* to control the cell cycle at additional checkpoints. Different *cyclins* show cyclic patterns of synthesis, while *cdks* remain at constant levels throughout the cell cycle (as in MPF). Different gene families encode evolutionarily conserved *cdks* or *cyclins*. But each *cyclin*/*cdk* pair has been coopted in evolution to monitor different cell cycle events and to catalyze phosphorylation of phase-specific proteins. To learn more, see Elledge SJ (1996) *Cell Cycle Checkpoints: Preventing an Identity Crisis*. Science 274:1664-1672.

344 Cyclin/cdk Checkpoint for Cell Cycle Phases

C. The G_0 State

This is not really a phase of the cell cycle, since cells in $\mathbf{G_0}$ have reached a terminally differentiated state and have stopped dividing. In development, terminally differentiated cells in tissues and organs no longer divide. Nevertheless, most cells have finite half-lives (recall our red blood cells that must be replaced every 60 days or so). Because cells in many tissues are in $\mathbf{G_0}$ and can't divide, they must be replaced by stem cells, which can divide and differentiate. Some cells live so long in $\mathbf{G_0}$ that they are nearly never replaced (muscle cells, neurons). Other cells live short lives in $\mathbf{G_0}$ (e.g., stem cells, some embryonic cells). For example, a **lymphocyte** is a differentiated *immune system* white blood cell type. However, exposure of lymphocytes to foreign chemicals or pathogens activates **mitogens** that cause them to re-enter the cell cycle from $\mathbf{G_0}$. The newly divided cells then make the antibodies that neutralize the chemicals and fight off the pathogens. The **retinoblastoma** (**Rb**) protein is an example of a mitogen. Like other mitogens, the Rb protein is a transcription factor that turns on genes that lead to cell proliferation. What if cells continue cycling when they aren't supposed to? Or, what if they are inappropriately signaled to exit G_o? Such cells are in trouble! Having escaped normal controls on cell division, they can become a focal point of cancer cell growth. You can guess from its name that the retinoblastoma gene was discovered as a mutation that causes retinal cancer. For more about the normal function of the Rb protein and its interaction with a G_1 cdk, check out the link below.

What if cells continue cycling when they aren't supposed to? Or, what if they are inappropriately signaled to exit $\mathbf{G_0}$? Such cells are in trouble! Having escaped normal controls on cell division, they can become a focal point of cancer cell growth. You can guess from its name that the retinoblastoma gene was discovered as a mutation that causes retinal cancer. For more about the normal function of the Rb protein and its interaction with a $\mathbf{G_1}$ *cdk*, check out the link below.

345 Rb Gene Encodes Transcription Factor Regulatory Subunit

This page titled 1.3: Regulation of the Cell Cycle is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **19.3: Regulation of the Cell Cycle** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

2: Mitosis and Meiosis

2.1: Mitosis

2.2: Meiosis

2.3: Mitosis and Meiosis

Thumbnail: Fluorescence image of mitosis in action. (CC BY 4.0; via PLOS ONE)

2: Mitosis and Meiosis is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

2.1: Mitosis

Cell growth and division is essential to asexual reproduction and the development of multicellular organisms. Accordingly, the primary function of mitosis is to ensure that at division each daughter cell inherits identical genetic material, i.e. exactly one copy of each chromosome. To make this happen, replicated chromosomes condense (**prophase**), and are positioned near the middle of the dividing cell (**metaphase**), and then each of the sister chromatids from each chromosome migrates towards opposite poles of the dividing cell (**anaphase**), until the identical sets of unreplicated chromosomes are completely separated from each other within the two newly formed daughter cells (**telophase**) (Figures 2.10 and 2.11). This is followed by the division of the cytoplasm (**cytokinesis**) to complete the process. The movement of chromosomes occurs through microtubules that attach to the chromosomes at the centromeres.

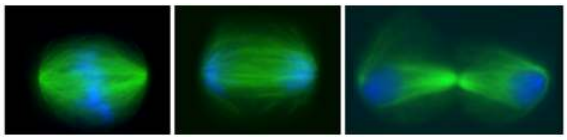


Figure **2.1.10**: Mitosis in arabidopsis showing fluorescently labeled chromosomes (blue) and microtubules (green) at metaphase, anaphase and telophase (from left to right). (PLoS Genetics-Somma MP et al. (2008) PLoS Genets 4(7): e1000126 – PD).

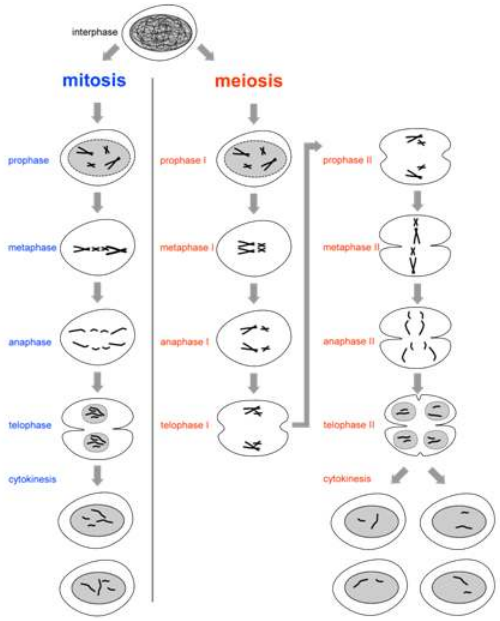


Figure **2.1.11**: Mitosis and meiosis. Note the similarities and differences between metaphase in mitosis and metaphase I and II of meiosis. (Original-Deyholos-CC:AN)

This page titled 2.1: Mitosis is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Mitosis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

2.2: Meiosis

Most eukaryotes replicate sexually - a cell from one individual joins with a cell from another to create the next generation. For this to be successful, the cells that fuse must contain half the number of chromosomes as in the adult organism. Otherwise, the number of chromosomes would double with each generation! The reduction in chromosome number is achieved by the process of **meiosis**. In meiosis, there are usually two steps, Meiosis I and II. In Meiosis I homologous chromosomes segregate, while in Meiosis II sister chromatids segregate. Most multicellular organisms use meiosis to produce **gametes**, the cells that fuse to make offspring. Some single celled eukaryotes such as yeast also use meiosis.

Meiosis I and II

Meiosis begins similarly to mitosis (a cell has replicated its chromosomes and grown large enough to divide), but requires two rounds of division (Figure **2.2.8**). In the first, known as meiosis I, the homologous chromosomes separate and segregate. During meiosis II the sister chromatids separate and segregate. Note how meosis I and II are both divided into prophase, metaphase, anaphase, and telophase. After two rounds of cytokinesis, four cells will be produced, each with a single copy of each chromosome.

Cells that will undergo the process of meiosis are called **meiocytes** and are diploid (2N). Meiosis is divided into two stages designated by the roman numerals I and II. Meiosis I is called a **reductional** division, because it reduces the number of chromosomes inherited by each of the daughter cells. Meiosis I is further divided into Prophase I, Metaphase I, Anaphase I, and Telophase I, which are roughly similar to the corresponding stages of mitosis, except that in Prophase I and Metaphase I, homologous chromosomes pair with each other, or **synapse**, and are called **bivalents** (Figs. 2.12). This is an important difference between mitosis and meiosis, because it affects the segregation of alleles, and also allows for recombination to occur through crossing-over, as described later. During Anaphase I, one member of each pair of homologous chromosomes migrates to each daughter cell (1N). Meiosis II resembles mitosis, with one sister chromatid from each chromosome separating to produce two daughter cells. Because Meiosis II, like mitosis, results in the segregation of sister chromatids, Meiosis II is called an **equational** division.

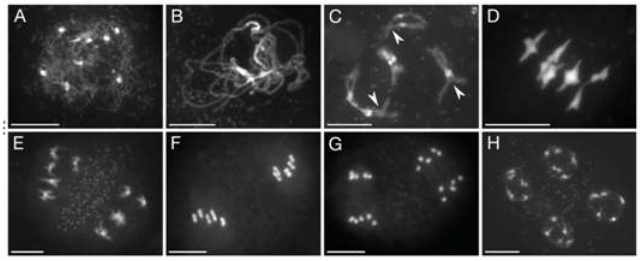


Figure **2.2.12**: Meiosis in Arabidopsis (n=5). Panels A-C show different stages of prophase I, each with an increasing degree of chromosome condensation. Subsequent phases are shown: metaphase I (D), telophase I (E), metaphase II (F), anaphase II (G), and telophase II (H). (PLoS Genetics-Chelysheva, L. et al (2008) PLoS Genetics-PD)

In meiosis I replicated, homologous chromosomes **pair up**, or **synapse**, during prophase I, lining up in the middle of the cell during metaphase I, and separating during anaphase I. For this to happen the homologous chromosomes need to be brought together while they condense during prophase I. These attachments are formed in two ways. Proteins bind to both homologous chromosomes along their entire length and form the **synaptonemal complex** (synapse means junction). These proteins hold the chromosomes in a transient structure called a **bivalent**. The proteins are released when the cell enters anaphase I.

Within the synaptonemal complex a second event, **crossingover**, occurs. These are places where DNA repair enzymes break the DNA two non-sister chromatids in similar locations and then covalently reattach non-sister chromatids together to create a crossover between non-sister chromatids. This reorganization of chromatids will persist for the remainder of meiosis and result in recombination of alleles in the gametes.

Crossovers function to hold homologous chromosomes together during meiosis I so they segregate successfully and they also cause the reshuffling of gene/allele combinations to create genetic diversity, which can have an important effect on evolution (see Chapter 7).

Stages of Prophase I

In meiosis, Prophase I is divided up into five visual stages, that are steps along a continuum of events. Leptotene, zygotene, pachytene, diplotene and diakinesis. From interphase, a cell enters **leptotene** as the nuclear material begins to condense into long visible threads (chromosomes). During **Zygotene** homologous chromosomes begin to pair up (synapse) and form an elaborate structure called the synaptonemal complex along their length. At **pachytene** homologous chromosomes are fully synapsed (two chromosomes and four chromatids) to form **bivalents**. **Crossing over** takes place in pachytene. After this, the pairing begins to loosen and individual chromatids become apparent in **diplotene**. This is when the consequences of each crossing over event can be seen as a **chiasma** (plural: **chiasmata**). **Diakinesis** follows as the chromosomes continue to condense and individualize. This is followed by metaphase I were the paired chromosomes orient on the metaphase plate in preparation for segregation (reductional).

Meiosis II and Gamete Maturation

At the completion of meiosis I there are two cells, each with one, replicated copy of each chromosome (1N). Because the number of chromosomes per cell has decreased (2->1), meiosis I is called a **reductional cell division**. In the second part of meiosis the chromosomes will once again be brought to the middle of the cell, but this time it is the sister chromatids that will segregate during anaphase.

After cytokinesis there will be four cells, each containing only one unreplicated chromosome of each type. Meiosis II resembles mitosis in that the number of chromosomes per cell is unchanged - both are **equational cell divisions –** but in meiosis II all cells won't have the same genetic composition. There will be allelic differences among the gametes.

In animals and plants the cells produced by meiosis need to mature before they become functional gametes. In male animals the four products of meiosis are called spermatids. They grow tails and become functional sperm cells. In female animals the gametes are eggs. In order that each egg contains the maximum amount of nutrients only one of the four products of meiosis becomes an egg. The other three cells end up as tiny disposable cells called **polar bodies**. In plants the products of meiosis reproduce a few times using mitosis as they develop into functional male or female gametes.

This page titled 2.2: Meiosis is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: Meiosis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

2.3: Mitosis and Meiosis

Learning Objectives

- Describe the cell cycle.
- Explain the events of all stages of mitosis.
- Track chromosome and chromatid number through all stages of mitosis.
- Explain the events of all stages of meiosis.
- Track chromosome and chromatid number through all stages of meiosis.
- Describe the role of meiosis in gamete formation and how it relates to inheritance.

Introduction

Multicellular organisms such as plants and animals are composed of millions to trillions (1,000,000,000) of cells that work together. The cells that make up different tissues have different shapes and perform different functions for the plant or animal. Even though they have diverse functions, each **somatic cell** in the organism normally has the same chromosomes and therefore the same genetic makeup. Furthermore, the millions of cells that makeup a mature organism originated from a single cell formed when the male and female gametes from the parents of the organism fused. This single cell established the life of the organism. Understanding multicellular organisms requires an understanding of the lifecycle of the cells that make up the organism.

The Cell Cycle

Let us think about the cell cycle from a personal point of view. Your age plus about nine months ago you were a **zygote**, a single cell formed when the sperm and egg from your biological parents fused in a fallopian tube (or possibly in a test tube if in vitro fertilization factored into your birth). You have come a long way since then, progressing one cell cycle at a time. The cell cycle is the life cycle of a single cell. We depict the cell cycle in a circular diagram although the cells in your body do not actually go around in circles. The main idea is that when new cells are made from existing cells, the new cells start their lifecycle, and the old cells end theirs.

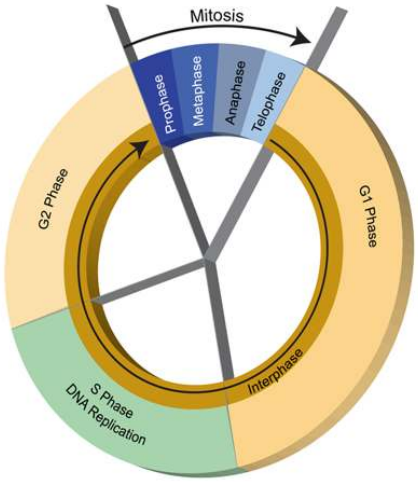


Figure 1. The cell cycle depicts the stages in the life of a cell. Image by NIH-NHGRI.

The first part of the cell cycle is the G1 phase. Cells can go through growth and development in this phase. Some cells differentiate into specialized cells and then never leave this phase. However, a **zygote** does not grow in size, but instead continues the cell cycle so it can quickly give rise to more cells.

The second part of the cell cycle is the S phase (synthesis phase). Here, the cell replicates its chromosomes so that it has a copy of each chromosome to pass on to daughter cells. The third part of the cell cycle is the G2 phase where the cell prepares for division. The G1, S and G2 phases together are called interphase. The M phase completes the cell cycle. 'M' could be mitosis or meiosis depending on the type of cell. For the zygote, the goal is to make more somatic cells. Therefore, it goes through mitosis and gives

rise to two daughter cells. This completes the life cycle of the zygote and starts the lifecycle of the new cells. Rounds of the cell cycle continued over and over to form the body you have today. As long as you live, some of your cells must be able to complete the cell cycle.

From a cytogenetics point of view, you have two types of cells in your body. You have cells with 46 **chromosomes** (**somatic** or body cells) and cells with 23 chromosomes (**gamete** or sex cells). Since you started as a single cell with 46 chromosomes, there must be two types of cell division taking place in your body to accommodate both somatic and gamete cells. Mitosis and meiosis are the two types of cell division.

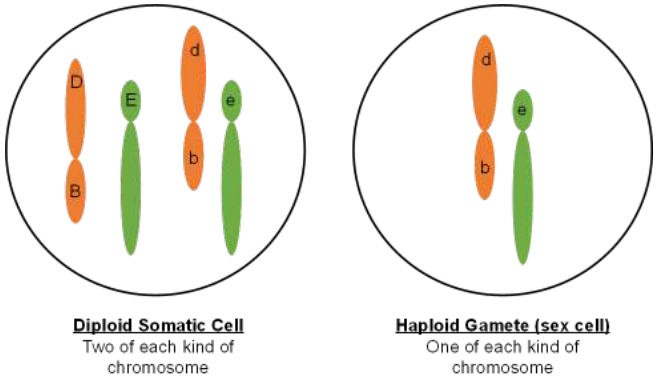


Figure 2. Multicellular organisms that sexually reproduce have diploid and haploid cells. Image by Marjorie Hanneman.

Mitosis: Somatic cell division

The objective of mitosis is to make two genetically identical cells from a single cell. In the cells of our body, we start with 46 chromosomes in a single cell and end up with 46 chromosomes in two cells. Obviously, replicating the chromosomes is a prerequisite to mitosis. Remember, replication takes place during interphase when the chromosomes are dispersed structures in the nucleus. Mitosis is an organized procession of activity in the cell that allows the replicated chromosomes to be properly divided into two identical cells. Chromosomes are important because they contain genes. Therefore, we will include **genes** on our chromosome diagrams and slide show. These pictures depict the four stages of mitosis. We will describe the events at each stage that are important in understanding the distribution of genes during cell division.

Mitosis: Prophase

Prophase is the beginning of mitosis (Figure 3).

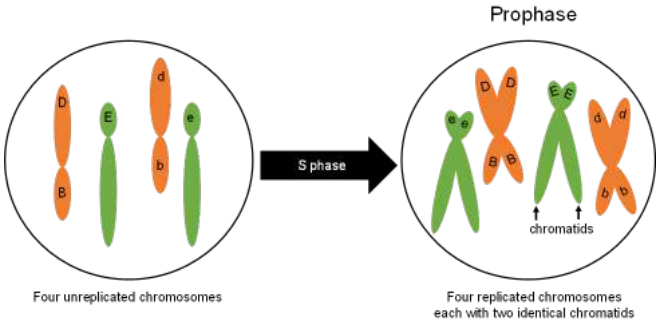


Figure 3. Prophase of mitosis. Image by Marjorie Hanneman.

During interphase, the chromosomes look like a plate of spaghetti in the nucleus. It is difficult to pick out an individual chromosome because they are each so spread out. The chromosomes in the nucleus change from being loosely dispersed to becoming more condensed. This change in chromosome structure makes them easier to move around the cell, an important issue for what is about to happen. As the chromosomes condense, they get shorter and thicker and can be seen through the microscope as individual structures (Figure 4). The chromosomes at prophase will consist of two identical parts called sister chromatids that stay connected at the centromere. It is now clear that the chromosomes have been replicated. Chromosome replication occurs during the

S-phase of interphase (Figure 1). Next, the nucleus is dissolved. At the end of prophase the **replicated** chromosomes are moved by the spindle apparatus to the center of the cell.

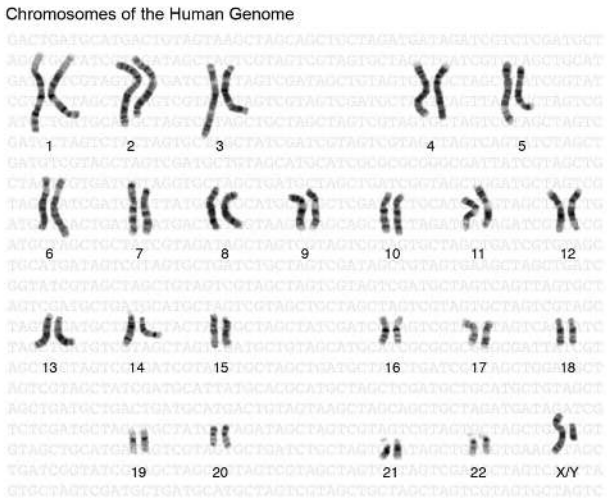


Figure 4. Chromosomes of the human genome. Image by NIH-NHGRI.

Mitosis: Metaphase

When the chromosomes are maneuvered to the center of the cell, metaphase begins (Figure 5). The spindle fiber network connects the centromere of the replicated chromosome to the outer part of each cell. The chromosomes now resemble the line of scrimmage of a football team, poised and waiting for some cellular signal to begin the next phase.

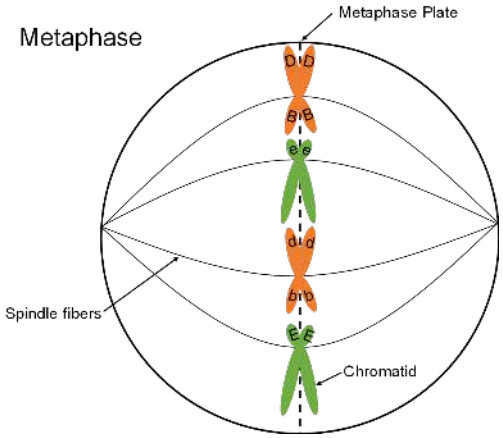


Figure 5. Metaphase of Mitosis. Image by Marjorie Hanneman.

Mitosis: Anaphase

At anaphase (Figure 6) the chromatids making up each chromosome are pulled apart and begin to move away from each other under the control of the spindle fibers. Once they are pulled apart, the chromatids are now considered individual chromosomes. As anaphase progresses, the chromosomes on each end of the cell are pulled into a bundle. When they stop moving, telophase begins.

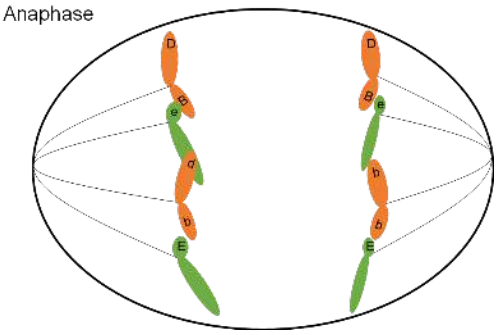


Figure 6. Anaphase of mitosis. Image by Marjorie Hanneman.

Mitosis: Telophase

The final stage of mitosis is telophase (Figure 7). A nuclear envelope will form around each bundle of chromosomes. The cell will undergo cytokinesis and the cytoplasm is split between the two identical daughter cells. Cell membranes (and cell walls in plants) form between the two cells. The chromosomes begin to decondense. They loosen up and spread around the nucleus because they no longer will need to move around. The new cells have now begun their lifecycle.

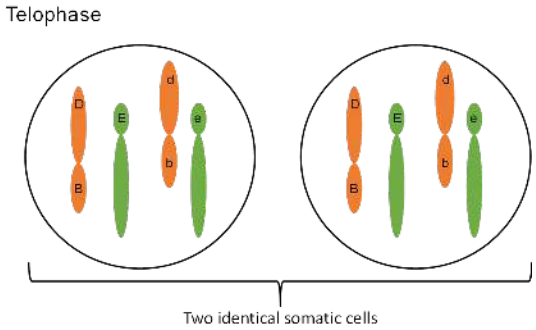


Figure 7: Telophase of Mitosis. Two cells produced from one cell. Image by Marjorie Hanneman.

Meiosis: Gamete formation

The objective of meiosis is to make four cells from a single somatic cell. The four cells each have half the chromosome number found in the somatic cell. In our human bodies, the four gametes will each have 23 chromosomes which means the 46 chromosomes in the somatic cell must replicate during interphase prior to meiosis just as they would before mitosis. Meiosis occurs in specialized cells of the body called germline cells.

To appreciate meiosis and gamete formation it is important to first understand two ideas, chromosome sets and homologous chromosomes.

Chromosome sets: The 46 chromosomes you have consist of two sets. You are a diploid organism ('di' means two and 'ploid' means sets). One set of chromosomes came from each parent when their gametes fused. Therefore, human gametes are haploid (one set).

Homologous chromosomes: The 46 chromosomes in a somatic cell can be arranged into 23 homologous or similar pairs. One chromosome from each pair came from the male parent, the other from the female. Homologous chromosomes have the same genes although in heterozygous people the genes would be different alleles (**A,a**). The exception to this would be the sex (X and Y) chromosomes. Passing on a complete set of human genes requires one chromosome from each pair to end up in each gamete.

There are several key differences between meiosis and mitosis that are summarized in the following table:

Table 1. The key events that happen in each of the stages of meiosis are summarized.

Mitosis

Meiosis

Chromosome number stays the same

Chromosome number is halved

One division occurs to make two cells. Four stages of this division.

Two divisions occur to make four cells. Eight stages in these divisions.

Similar or homologous chromosomes do not pair.

Homologous chromosomes pair during prophase l. Pairing is called synapsis.

Crossover exchanges between homologous chromosomes is rare.

Synapsis allows crossing over between homologous chromosomes.

Two cells made are genetically identical.

Four cells made are genetically different.

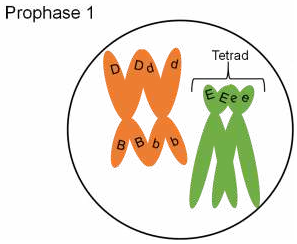


Figure 8. Prophase l of meiosis. Images by Marjorie Hanneman.

Meiosis: Prophase I

This stage starts meiosis and is the same as prophase of mitosis with one important change. As the chromosomes condense, they form groups of four chromatids called tetrads or bivalents. Close inspection reveals that each chromosome is replicated and consists of two sister chromatids. The two chromosomes in each cell that are homologous and have the same genes (but perhaps different alleles if the organism is heterozygous) will pair closely. This close association, or synapsis, allows the homologous chromosomes to crossover and exchange identical parts. The impact of crossing over is that genes that are on the same chromosome (Figure 8) can be recombined so that they are not always inherited together. The tetrad or bivalent formed during synapsis remains assembled as prophase progresses. The tetrad therefore moves as a unit to the center of the cell.

Meiosis: Metaphase I

Metaphase I starts when the tetrads are at the center of the cell (Figure 9). The tetrads have stayed together which insures that during the first division, each cell will get one chromosome from each homologous pair. The chromosomes remain at the center of the cell until the homologous pairs are ready to move away from each other.

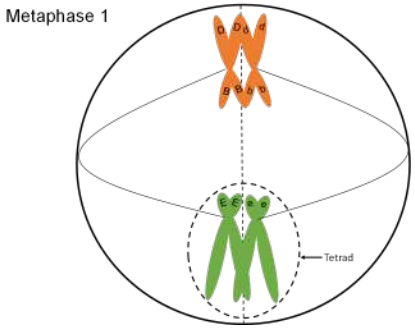


Figure 9. Metaphase I of meiosis. Image by Marjorie Hanneman.

Meiosis: Anaphase I

The chromosomes that make up each tetrad separate during anaphase I (Figure 10). However, the sister chromatids will stay connected at the centromere. Anaphase I proceeds until the chromosomes are pulled into a bundle at opposite ends of the cell.

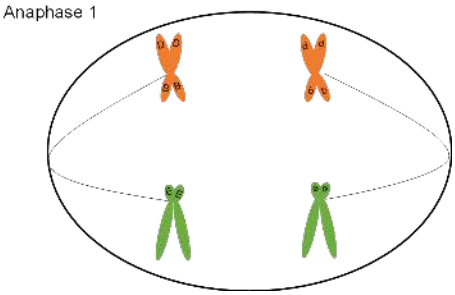


Figure 10. Anaphase l of meiosis. Image by Marjorie Hanneman.

Meiosis: Telophase l

The cell divides into two cells during telophase I (Figure 11). The bundle of chromosomes may have a nuclear envelope develop around them. The germline cells in some organisms such as human females, go through the first four stages of meiosis prior to birth. The **germline cells** remain at telophase I for some time. The second round of division occurs when the gamete is needed for reproduction. In other situations, telophase I is an abbreviated stage, and the second round of division proceeds without delay.

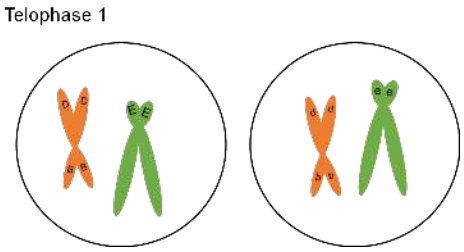


Figure 11: Telophase l of meiosis. Image by Marjorie Hanneman.

Meiosis: Prophase II

If the chromosomes became dispersed in telophase I, they will condense again at prophase II. The spindle apparatus moves the chromosomes to the middle of the cell. Look! The centromeres are still holding the sister chromatids together (Figure 12).

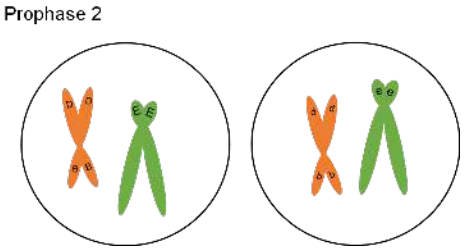


Figure 12. Prophase II of meiosis. Image by Marjorie Hanneman.

Meiosis: Metaphase II

In metaphase II the chromosomes are aligned at the center of the cell (Figure 13). This time there are not homologous chromosomes to be paired with. This metaphase looks similar to metaphase of mitosis but there is a key difference. What is the difference? (Compare Figures 5 and 13).

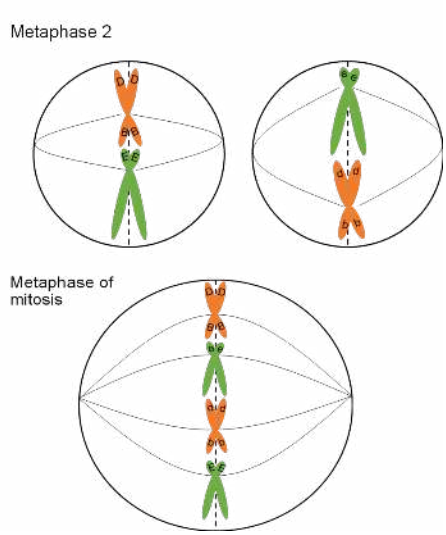


Figure 13. Metaphase II of meiosis and Metaphase of mitosis. Image by Marjorie Hanneman.

Meiosis: Anaphase II

During anaphase II, the chromatids are pulled apart by the spindle fibers. Now they are classified as chromosomes, not chromatids. The chromosomes move apart to opposite ends of the cell (Figure 14).

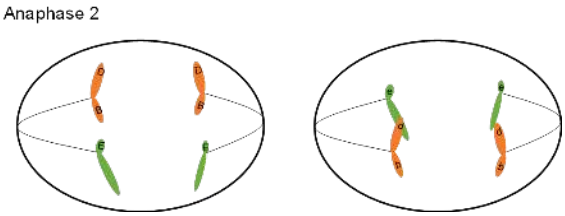


Figure 14. Anaphase II of meiosis. Image by Marjorie Hanneman.

Meiosis: Telophase II

In the final stage of meiosis, telophase II, the nucleus forms around the bundle of chromosomes (Figure 15). The cell divides. Now four cells exist that originated from one germline cell. Each cell is a gamete with half the number of chromosomes and genes as a somatic cell.

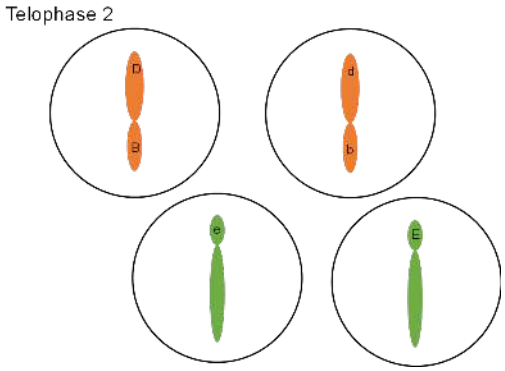


Figure 15. Telophase II of meiosis. Image by Marjorie Hanneman.

Parallel Behavior and the Chromosome Theory

The successful completion of mitosis or meiosis requires the cell to move large objects with precision and control many detailed events. The process surpasses any engineering accomplishments of NASA. The importance of mitosis and meiosis to an organism is obvious when we consider that genes are a part of the chromosome and the genes must be copied and distributed properly to produce viable daughter cells. The mechanisms of these events are far from being completely understood. From our current understanding, we can appreciate how the principles of segregation and independent assortment are controlled by the mechanics of meiosis.

When cell division was first observed and described by cytogeneticists, biologists were just beginning to accept the idea that genes were tiny objects that controlled traits and existed in the cells of living things. Two biologists, a German named Boveri and an American graduate student named Sutton, recognized that chromosome behavior during meiosis matched Mendel's principles of gene behavior. Both scientists proposed the idea that while genes had not yet been directly observed, they must be a part of the chromosome. Sutton and Boveri are both given credit for proposing this chromosome theory; genes are a part of chromosomes.

Segregation predicts gene behavior that matches the chromosome behavior observed by cytogeneticists. Genes are in pairs because chromosomes are in pairs. The gene pairs associate during gamete formation when the homologous chromosomes pair in prophase I. The gene pairs separate when the homologous pairs separate in anaphase I followed by chromatid separation in anaphase II. Thus, chromosome behavior dictates the gene's segregation behavior.

Independent assortment of gene pairs also correlates with chromosome behavior. Let us consider a dihybrid individual with the genotype **BbEe** (Figure 16). How many kinds of gametes can it make? The four possible combinations (**BE**, **bE**, **Be** and **be**) are made at equal frequencies. We can understand why this occurs if we think about what happens to chromosomes at metaphase I of meiosis. The chromosome pair that carries the '**E**' and '**e**' genes will be moved independently from the chromosome pair with the '**B**' and '**b**' chromosomes. When the chromosomes arrive at the center of the cell at metaphase I, the two tetrads may be aligned so that the '**E**' and '**B**' genes move to one cell and the '**e**' and '**b**' genes move to the other during the first division. In the other half of the cells that go through meiosis, the tetrads will line up so that the '**E**' genes will be passed on with the '**b**' genes and the '**e**' genes will go with the '**B**' genes. Keep in mind that organisms that make gametes make thousands of them. The chance alignment of the tetrads at metaphase I therefore dictates the overall frequency of gametes with different combinations of genes when the genes are on separate chromosomes.

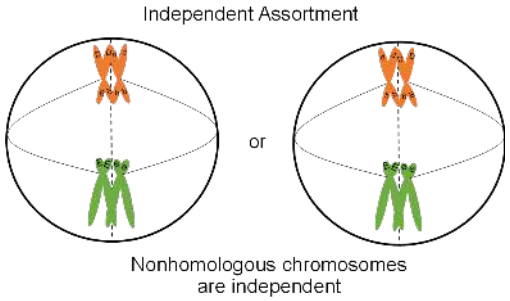


Figure 16. Alignment of chromosomes during metaphase I will determine the gene combinations that segregate into gametes. Image by Marjorie Hanneman.

When more gene pairs are considered, the same scenarios described above will be true as long as the genes are on separate chromosomes. When genes are on the same chromosome, the role of crossing over on gene inheritance needs to be considered in more detail. We will cover the inheritance of genes on the same chromosome in a later lesson.

Learning Activities

1. Watch this video about Mitosis from Daily Med Ed
2. Watch this video about Meiosis from Daily Med Ed

This page titled 2.3: Mitosis and Meiosis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, & Donald Lee (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.1: Mitosis and Meiosis** by Donald Lee, Marjorie Hanneman, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

3: Alleles and Probabilities

3.1: Genes are the Basic Units of Inheritance
3.2: Model Organisms Facilitate Genetic Advances
3.3: Mendel's First Law
3.4: Biochemical Basis of Dominance
3.5: Crossing Techniques Used in Classical Genetics
3.6: Dihybrid Crosses
3.7: Why Use Nomenclature?
3.8: Basic Nomenclature
3.9: Linked Genes

Thumbnail: Example of a Punnett square. In this example in peas, the color yellow is determined by the dominant allele Y and the color green is determined by a recessive allele y. (CC BY-SA 3.0 Unported; Pbroks13 via Wikipedia)

3: Alleles and Probabilities is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

3.1: Genes are the Basic Units of Inheritance

Blending vs Particulate inheritance

The once prevalent (but now discredited) concept of **blending inheritance** proposed that some undefined essence, in its entirety, contained all of the heritable information for an individual. It was thought that mating combined the essences from each parent, much like the mixing of two colors of paint. Once blended together, the individual characteristics of the parents could not be separated again. However, **Gregor Mendel** (Fig 1.10) was one of the first to take a quantitative, scientific approach to the study of heredity.



Figure **3.1.10**: Gregor Mendel. (Original-unknown-PD)

He started with well-characterized strains, repeated his experiments many times, and kept careful records of his observations. Working with peas, Mendel showed that white-flowered plants could be produced by crossing two purple-flowered plants, but only if the purple-flowered plants themselves had at least one white-flowered parent (Fig 1.11). This was evidence that the genetic factor that produced white-flowers had *not* blended irreversibly with the factor for purple-flowers. Mendel's observations disprove blending inheritance and favor an alternative concept, called **particulate inheritance**, in which heredity is the product of discrete factors that control independent traits.

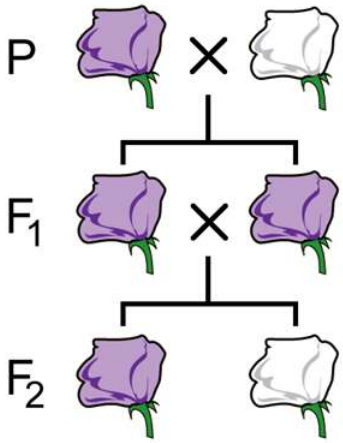


Figure **3.1.11**: Inheritance of flower color in peas. Mendel observed that a cross between pure breeding, white and purple peas (generation P) produced only progeny (generation F_1) with purple flowers. However, white flowered plant reappeared among the F_2 generation progeny of a mating between two F_1 plants. The symbols P, F_1 and F_2 are abbreviations for parental, first filial, and second filial generations, respectively. (Original-Deyholos-CC:AN)

This page titled 3.1: Genes are the Basic Units of Inheritance is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Genes are the Basic Units of Inheritance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.2: Model Organisms Facilitate Genetic Advances

Model organisms

Many of the great advances in genetics were made using species that are not especially important from a medical, economic, or even ecological perspective. Geneticists, from Mendel onwards, have sought the best organisms for their experiments. Today, a small number of species are widely used as **model organisms** in genetics (Fig 1.17). All of these species have specific characteristics that make large number of them easy to grow and analyze in laboratories: (1) they are small, (2) fast growing with a short generation time, (3) produce lots of progeny from matings that can be easily controlled, (4) have small genomes (small C-value), and (5) are diploid (i.e. chromosomes are present in pairs).

The most commonly used model organism are:

- The prokaryote bacterium, ***Escherichia coli,*** is the simplest genetic model organism and is often used to clone DNA sequences from other model species.
- Yeast (***Saccharomyces cerevisiae***) is a good general model for the basic functions of eukaryotic cells.
- The roundworm, ***Caenorhabditis elegans*** is a useful model for the development of multicellular organisms, in part because it is transparent throughout its life cycle, and its cells undergo a well-characterized series of divisions to produce the adult body.
- The fruit fly (***Drosophila melanogaster***) has been studied longer, and probably in more detail, than any of the other genetic model organisms still in use, and is a useful model for studying development as well as physiology and even behavior.
- The mouse (***Mus musculus***) is the model organism most closely related to humans, however there are some practical difficulties working with mice, such as cost, slow reproductive time, and ethical considerations.
- The zebrafish (***Danio rerio***) has more recently been developed by researchers as a genetic model for vertebrates.Unlike mice, zebrafish embryos develop quickly and externally to their mothers, and are transparent, making it easier to study the development of internal structures and organs.
- Finally, a small weed, ***Arabidopsis thaliana***, is the most widely studied plant genetic model organism. This provides knowledge that can be applied to other plant species, such as wheat, rice, and corn.



Figure **3.2.17**: Some of the most important genetic model organisms in use today. Clockwise from top left: yeast, fruit fly, arabidopsis, mouse, roundworm, zebrafish. (Original/Flickr/Wikipedia – Deyholos/M.Westby/D.Joly/Z.F.Altun/Masur/Azul – CC:AS/ANS/AN &GFDL)

This page titled 3.2: Model Organisms Facilitate Genetic Advances is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **1.7: Model Organisms Facilitate Genetic Advances** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.3: Mendel's First Law

Character Traits Exist in Pairs that Segregate at Meiosis

Through careful study of patterns of inheritance, Mendel recognized that a single trait could exist in different versions, or **alleles**, even within an individual plant or animal. For example, he found two allelic forms of a gene for seed color: one allele gave green seeds, and the other gave yellow seeds. Mendel also observed that although different alleles could influence a single trait, they remained indivisible and could be inherited separately. This is the basis of **Mendel's First Law**, also called **The Law of Equal Segregation**, which states: during gamete formation, the two alleles at a gene locus segregate from each other; each gamete has an equal probability of containing either allele.

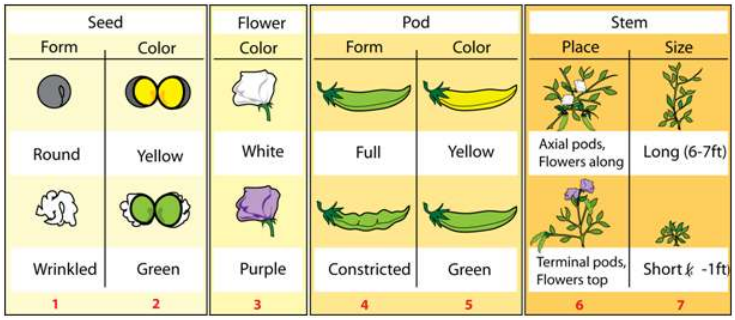


Figure **3.3.2**: Seven traits Mendel studied in peas. (Wikipedia-Mariana Ruiz-PD)

Hetero-, Homo-, Hemizygosity

Mendel's First Law is especially remarkable because he made his observations and conclusions (1865) without knowing about the relationships between genes, chromosomes, and DNA. We now know the reason why more than one allele of a gene can be present in an individual: most eukaryotic organisms have at least two sets of homologous chromosomes. For organisms that are predominantly diploid, such as humans or Mendel's peas, chromosomes exist as pairs, with one homolog inherited from each parent. Diploid cells therefore contain two different alleles of each gene, with one allele on each member of a pair of homologous chromosomes. If both alleles of a particular gene are identical, the individual is said to be **homozygous** for that gene. On the other hand, if the alleles are different from each other, the genotype is **heterozygous**. In cases where there is only one copy of a gene present, for example if there is a deletion on the homologous chromosome, we use the term **hemizygous**.

Although a typical diploid individual can have at most two different alleles of a particular gene, many more than two different alleles can exist in a population of individuals. In a natural population the most common allelic form is usually called the **wild-type** allele. However, in many populations there can be multiple **variants** at the DNA sequence level that are visibly indistinguishable as all exhibit a normal, wild type appearance. There can also be various **mutant** alleles (in wild populations and in lab strains) that vary from wild type in their appearance, each with a different change at the DNA sequence level. Such collections of mutations are known as an **allelic series**.

This page titled 3.3: Mendel's First Law is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.1: Mendel's First Law** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.4: Biochemical Basis of Dominance

Given that a heterozygote's phenotype cannot simply be predicted from the phenotype of homozygotes, what does the type of dominance tell us about the biochemical nature of the gene product? How does dominance work at the biochemical level? There are several different biochemical mechanisms that may make one allele dominant to another.

For the majority of genes studied, the normal (i.e. wild-type) alleles are **haplosufficient**. So in diploids, even with a mutation that causes a complete loss of function in one allele, the other allele, a wild-type allele, will provide sufficient normal biochemical activity to yield a wild type phenotype and thus be dominant and dictate the heterozygote phenotype.

On the other hand, in some biochemical pathways, a single wild-type allele is not enough protein and may be **haploinsufficient** to produce enough biochemical activity to result in a normal phenotype, when heterozygous with a non-functioning mutant allele. In this case, the non-functional mutant allele will be dominant (or semi-dominant) to a wild-type allele.

Mutant alleles may also encode products that have new and/or different biochemical activities instead of, or in addition to, the normal ones. These **novel activities** could cause a new phenotype that would be dominantly expressed.

This page titled 3.4: Biochemical Basis of Dominance is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.3: Biochemical Basis of Dominance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.5: Crossing Techniques Used in Classical Genetics

Classical Genetics

Not only did Mendel solve the mystery of inheritance as units (genes), he also invented several testing and analysis techniques still used today. **Classical genetics** is the science of solving biological questions using controlled matings of model organisms. It began with Mendel in 1865 but did not take off until Thomas Morgan began working with fruit flies in 1908. Later, starting with Watson and Crick's structure of DNA in 1953, classical genetics was joined by **molecular genetics**, the science of solving biological questions using DNA, RNA, and proteins isolated from organisms. The genetics of **DNA cloning** began in 1970 with the discovery of restriction enzymes.

True Breeding Lines

Geneticists make use of **true breeding lines** just as Mendel did (Figure **3.5.6**a). These are in-bred populations of plants or animals in which all parents and their offspring (over many generations) have the same phenotypes with respect to a particular trait. True breeding lines are useful, because they are typically assumed to be homozygous for the alleles that affect the trait of interest. When two individuals that are homozygous for the same alleles are crossed, all of their offspring will all also be homozygous. The continuation of such crosses constitutes a true breeding line or strain. A large variety of different strains, each with a different, true breeding character, can be collected and maintained for genetic research.

Monohybrid Crosses

A **monohybrid cross** is one in which both parents are heterozygous (or a hybrid) for a single (mono) trait. The trait might be petal colour in pea plants (Figure **3.5.6**b). Recall from chapter 1 that the generations in a cross are named P (parental), F_1 (first filial), F_2 (second filial), and so on.

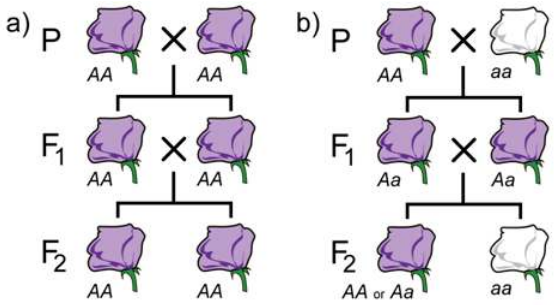


Figure **3.5.6**: (a) A true-breeding line (**b)** A monohybrid cross produced by mating two different pure-breeding lines. (Original-Deholos-CC:AN)

Punnett Squares

Given the genotypes of any two parents, we can predict all of the possible genotypes of the offspring. Furthermore, if we also know the dominance relationships for all of the alleles, we can predict the phenotypes of the offspring. A convenient method for calculating the expected genotypic and phenotypic ratios from a cross was invented by Reginald Punnett. A **Punnett square** is a matrix in which all of the possible gametes produced by one parent are listed along one axis, and the gametes from the other parent are listed along the other axis. Each possible combination of gametes is listed at the intersection of each row and column. The F_1 cross from Figure **3.5.6**b would be drawn as in Figure **3.5.7**. Punnett squares can also be used to calculate the frequency of offspring. The frequency of each offspring is the frequency of the male gametes multiplied by the frequency of the female gamete.

	A	***a***
A	*AA*	*Aa*
a	*Aa*	*aa*

***Figure* 3.5.7:** *A Punnett square showing a monohybrid cro*ss. (Original-Deholos (Fireworks)-CC:AN)

Test Crosses

Knowing the genotypes of an individual is usually an important part of a genetic experiment. However, genotypes cannot be observed directly; they must be inferred based on phenotypes. Because of dominance, it is often not possible to distinguish between a heterozygote and a homozgyote based on phenotype alone (e.g. see the purple-flowered F_2 plants in Figure **3.5.6**b). To determine the genotype of a specific individual, a **test cross** can be performed, in which the individual with an uncertain genotype is crossed with an individual that is homozygous recessive for all of the loci being tested.

For example, if you were given a pea plant with purple flowers it might be a homozygote (*AA*) or a heterozygote (*Aa*). You could cross this purple-flowered plant to a white-flowered plant as a **tester**, since you know the genotype of the tester is *aa*. Depending on the genotype of the purple-flowered parent (Figure **3.5.8**), you will observe different phenotypic ratios in the F_1 generation. If the purple-flowered parent was a homozgyote, all of the F_1 progeny will be purple. If the purple-flowered parent was a heterozygote, the F_1 progeny should segregate purple-flowered and white-flowered plants in a 1:1 ratio.

	A	***A***
a	*Aa*	*Aa*
a	*Aa*	*Aa*

	A	***a***
a	*Aa*	*aa*
a	*Aa*	*aa*

Figure* 3.5.8*: *Punnett Squares showing the two possible outcomes of a test cross.* (Original-Deholos (Fireworks)-CC:AN)

This page titled 3.5: Crossing Techniques Used in Classical Genetics is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.4: Crossing Techniques Used in Classical Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.6: Dihybrid Crosses

Mendel's Second Law

Before Mendel, it had not yet been established that heritable traits were controlled by discrete factors. Therefor an important question was therefore whether distinct traits were controlled by discrete factors that were inherited independently of each other? To answer this, Mendel took two apparently unrelated traits, such as seed shape and seed color, and studied their inheritance together in one individual. He studied two variants of each trait: seed color was either green or yellow, and seed shape was either round or wrinkled. (He studied seven traits in all.) When either of these traits was studied individually, the phenotypes segregated in the classical 3:1 ratio among the progeny of a monohybrid cross (Figure **3.6.2**), with ¾ of the seeds green and ¼ yellow in one cross, and ¾ round and ¼ wrinkled in the other cross. Would this be true when both were in the same individual?

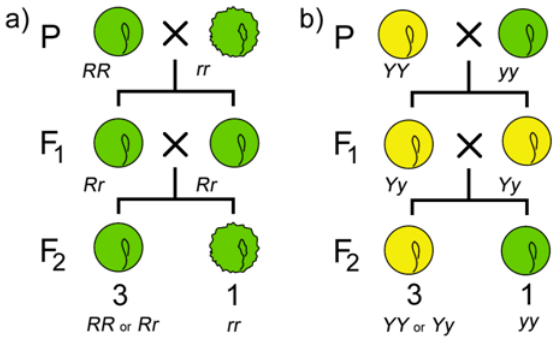


Figure **3.6.2**: Monohybrid crosses involving two distinct traits in peas. a) is R/r and b) is Y/y. *(Original-Deyholos-CC:AN)*

To analyze the segregation of both traits at the same time in the same individual, he crossed a pure breeding line of green, wrinkled peas with a pure breeding line of yellow, round peas to produce F_1 progeny that were all green and round, and which were also **dihybrids**; they carried two alleles at each of two loci (Figure **3.6.3**).

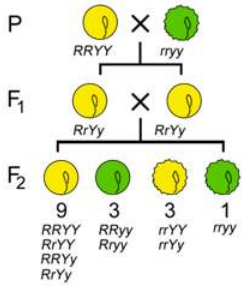


Figure **3.6.3**: Pure-breeding lines are crossed to produce dihybrids in the F_1 generation. The cross of these particular dihybrids produces four phenotypic classes. (Original-Deyholos-CC:AN)

If the inheritance of seed color was truly independent of seed shape, then when the F_1 dihybrids were crossed to each other, a 3:1 ratio of one trait should be observed within each phenotypic class of the other trait (Figure **3.6.3**). Using the product law, we would therefore predict that if ¾ of the progeny were green, and ¾ of the progeny were round, then ¾ × ¾ = 9/16 of the progeny would be both round and green (Table **3.6.1**). Likewise, ¾ × ¼ = 3/16 of the progeny would be both round and yellow, and so on. By applying the product rule to all of these combinations of phenotypes, we can predict a **9:3:3:1** phenotypic ratio among the progeny of a dihybrid cross, if certain conditions are met, including the independent segregation of the alleles at each locus. Indeed, 9:3:3:1 is very close to the ratio Mendel observed in his studies of dihybrid crosses, leading him to state his Second Law, the **Law of Independent Assortment**, which we now express as follows: two loci assort independently of each other during gamete formation.

Definition: Mendel's Second Law

Two loci assort *independently* of each other during gamete formation.

https://bio.libretexts.org/@go/page/42791

Table **3.6.1**: Phenotypic classes expected in monohybrid and dihybrid crosses for two seed traits in pea.

						Frequency of phenotypic crosses within separate monohybrid crosses: seed shape: ¾ round ¼ wrinkled seed color: ¾ yellow ¼ green **Frequency of phenotypic crosses within a dihybrid cross:** ¾ round × ¾ yellow = 9/16 round & yellow ¾ round × ¼ green = 3/16 round & green ¼ wrinkled × ¾ yellow = 3/16 wrinkled & yellow ¼ wrinkled × ¼ green = 1/16 wrinkled & green

The 9:3:3:1 phenotypic ratio that we calculated using the product rule can also be obtained using Punnett Square (Figure **3.6.4**). First, we list the genotypes of the possible gametes along each axis of the Punnett Square. In a diploid with two heterozygous genes of interest, there are up to four combinations of alleles in the gametes of each parent. The gametes from the respective rows and column are then combined in the each cell of the array. When working with two loci, genotypes are written with the symbols for both alleles of one locus, followed by both alleles of the next locus (e.g. *AaBb*, not *ABab*). Note that the order in which the loci are written does not imply anything about the actual position of the loci on the chromosomes.

	RY	***Ry***	***rY***	***ry***
RY	*RRYY*	*RRYy*	*RrYY*	*RrYy*
Ry	*RRYy*	*RRyy*	*RrYy*	*Rryy*
rY	*RrYY*	*RrYy*	*rrYY*	*rrYy*
ry	*RrYy*	*Rryy*	*rrYy*	*rryy*

Figure **3.6.3**. Each of the four phenotypic classes is represented by a different color of shading. (Original-Deyholos-CC:AN)

To calculate the expected phenotypic ratios, we assign a phenotype to each of the 16 genotypes in the Punnett Square, based on our knowledge of the alleles and their dominance relationships. In the case of Mendel's seeds, any genotype with at least one *R* allele and one *Y* allele will be round and yellow; these genotypes are shown in the nine, green-shaded cells in Figure **3.6.4**. We can represent all of four of the different genotypes shown in these cells with the notation (*R_Y_*), where the blank line (__), means "any

allele". The three offspring that have at least one R allele and are homozygous recessive for *y* (i.e. *R_yy*) will have a round, green phenotype. Conversely the three progeny that are homozygous recessive *r*, but have at least one *Y* allele (*rrY_*) will have wrinkled, yellow seeds. Finally, the rarest phenotypic class of wrinkled, yellow seeds is produced by the doubly homozygous recessive genotype, *rryy*, which is expected to occur in only one of the sixteen possible offspring represented in the square.

Assumptions of the 9:3:3:1 ratio

Both the product rule and the Punnett Square approaches showed that a 9:3:3:1 phenotypic ratio is expected among the progeny of a dihybrid cross such as Mendel's *RrYy* × *RrYy*. In making these calculations, we assumed that:

1. both loci assort independently;
2. one allele at each locus is completely dominant; and
3. each of four possible phenotypes can be distinguished unambiguously, with no interactions between the two genes that would alter the phenotypes.

Deviations from the 9:3:3:1 phenotypic ratio may indicate that one or more of the above conditions has not been met. Modified ratios in the progeny of a dihybrid cross can therefore reveal useful information about the genes involved.

Linkage is one of the most important reasons for distortion of the ratios expected from independent assortment. Linked genes are located close together on the same chromosome. This close proximity alters the frequency of allele combinations in the gametes. We will return to the concept of linkage in Chapter 7. Deviations from 9:3:3:1 ratios can also be due to interactions between genes. These interactions will be discussed in the remainder of this chapter. For simplicity, we will focus on examples that involve easily scored phenotypes, such as pigmentation. Nevertheless, keep in mind that the analysis of segregation ratios of any markers can provide insight into a wide range of biological processes they represent.

This page titled 3.6: Dihybrid Crosses is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **6.1: Dihybrid Crosses** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.7: Why Use Nomenclature?

Word and practical problems in biology can get confusing in a hurry, particularly if you're distracted by something like exam stress! When you are manipulating several ideas, it is good practice to be thoughtful and follow rules that keep you consistent in your interpretation. As an instructor, I have seen work in which the student clearly got flustered and forgot that the mutation he or she was working on was dominant. This often leads to an answer that is inconsistent with the data.

There are a few simple rules we can use for nomenclature. This first appendix uses a simplified system to communicate allele characteristics. The next appendix will show a more complicated system that carries even more information in the gene symbols.

This page titled 3.7: Why Use Nomenclature? is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **14.1: Why Use Nomenclature?** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.8: Basic Nomenclature

Single Letter System

Sometimes what you want to do is a little rough work for investigating your genetic model. A genetic model is a diagram of the logic that you propose for a particular type of inheritance. For instance, if you cross a true-breeding purple plant with a true-breeding white plant (Figure A1.2), you will get a heterozygote (the middle plant). If we name the gene after the mutation (*a* is the first letter in "albino"), we know that the heterozygote will have one capital letter "*A*" and a lower case "*a*". The heterozygote is the F_1 generation ("first filial", which means it's the first child from parents that are crossed). The F_1 is purple, which means the "*a*" allele is recessive; only one copy of the "*A*" allele is needed for enough purple pigment to make it identical to one true-breeding parent. This is complete dominance.

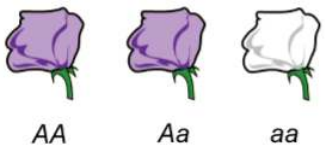


Figure A1.2

We cannot know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "*a*" allele is null and makes no pigment at all. One or two "*A*" alleles make enough protein to cause the plant to be purple.

We can't know from the information given which allele is wild type or mutant. One hypothesis, though, is that purple pigments are required to attract pollinators and therefore would help the plant in the wild. Albino plants could be a mutant and might not generate as many seeds for lack of pollinators. If the context of your genetics problem doesn't indicate which allele is wild-type, it's good practice to name your allele based on the recessive trait. Often the recessive allele is the mutation. We might consider that the "a" allele is null and makes no pigment at all. One or two "A" alleles make enough protein to cause the plant to be purple.

Figure A1.2 already assumes that the capital letter (*A*) stands for an allele that encodes a protein for purple pigment and the recessive allele (*a*) doesn't make pigment. Thus the *Aa* heterozygote is sufficient evidence to adopt upper- and lower-case letter "*A*"s to communicate the characteristics of purple and white alleles.

A note of caution. When you're writing down gene symbols for homework or on an exam, be sure to make the characters distinct. A typewritten "*y*" is easy to distinguish from the upper case "*Y*" but not as easy when writing it down. Instructors who ask you to show your work need to be able to follow your logic train. More important than that is YOU have to be able to follow your own reasoning. Students often switch symbols and come up with an answer that is inconsistent with the data given because of this. Consider underlining your capitals or putting a line through one of them to make it distinct (e.g. ~~Y~~ for the dominant allele; *y* for the recessive).

Name the gene after the mutant phenotype

Some instructors would accept "*P*" for "*purple*" for the previous cross. However, the better answer is to follow an established system. During "exam fog", it's easy to get lost if you are inconsistent with how you develop your symbols. During your study period and when you're practicing genetics problems, be thoughtful about the gene names you choose.

Let's always choose a letter based on the mutant phenotype for our gene symbol. If we are presented with a ladybug mutant that is small, we might choose "*d*" for "*dwarf*". Geneticists sometimes set up a research program based on unusual phenotypes of the organism they are studying. The fact that a mutant phenotype that is heritable exists tells us that there is a genetic control for the trait and that it might be isolated in the lab. When you look at your classmates, you don't necessarily note that none of them has an arm growing from the tops of their heads. If one student had this trait, however, you couldn't help but notice it. If you found out it showed up in that student's ancestry in a predictable fashion, you might reasonably suggest that there is a genetic basis for that. If it happened to be controlled by a single gene, you might call the gene "*extra arm*" or "*arm head*". If it happened to be a dominant trait, you might use the letter "~~*X*~~" or "*A*" for the mutant allele. The wild type allele would be "*x*" or "*a*", respectively.

you don't know which is mutant, use the recessive trait for the gene name

What if you don't know which allele is mutant? What if you're presented with two true-- breeding frogs: one that is gold and one that is yellow. If you don't know what the predominant colour in nature is you can't know which one is mutant. If you crossed them and all the progeny are gold, then you know the dominant allele encodes a protein to make it *gold*. The recessive, therefore, is "*yellow*" and you should name the gene "*y*" after the recessive phenotype. This means the dominant allele would be "*Y*". Your offspring would therefore be *Yy* and the gold parent would be *YY*. The yellow parent would be *yy*.

Apply these ideas at the Online Open Genetics exercises.

1.3 One Letter System

Superscripts

Sometimes a letter is used as the name of a gene, and superscripts can modify it to indicate the different alleles. One common single letter code for an allelic series is "*I*". Red blood cells can have their cell membranes modified by sugar tags that give rise to our blood type. One allele of *I* gives rise to blood type A and is therefore called I^A. An enzyme encoded by I^B modifies sugars to create blood type B. A heterozygote I^AI^B demonstrates *both* sugar tags because those alleles are expressed – they are codominant. People with blood type O only possess alleles for the *I* gene that don't work and are therefore recessive – they don't modify the extracellular sugar tags. Because it is recessive, individuals are homozygous for *i*: they are *ii*. Chapter 13, Section 7 has more detail on this allelic series.

Sometimes a superscript "plus sign" is used to denote the wild type allele. One might use the symbol W^+ to indicate a wild--type allele that promotes wing growth. Note that the generic "wing" gene name isn't a best practice – name the gene after the mutant phenotype! A wingless mutant would be W^-. You should never use a "+" and shift the case of the letter unless you are dealing with a special case such as the codominance in the blood type example above. The capital "*I*" letter indicates it is dominant to "*i*". The superscript A and B for the codominant alleles indicate the dominant alleles are different from each other.

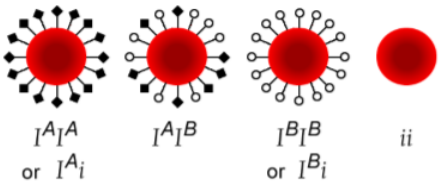


Figure A1.3: Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the i allele. The I^B allele is completely dominant to the i allele. (Original-Deholos -CC:AN – from Chapter 1)

Superscripts can be symbols, a single letter, or many letters. They modify the gene name only in the superscripted symbols: the regular--sized letters are identical between them (see Table 1). This means that Abc^+ and *abc* would be different genes (*i.e.* not allelic); Abc^+ and *Abc* are alleles, as are abc^+ and *abc*. Note that a superscript is not mandatory for all alleles of that gene, depending on the convention.

Alleles of bacterial genes are typically indicated with a superscript + or -. For example, a bacterial allele that creates an enzyme that makes methionine would be met^+, and a defective allele of that gene is met^-.

Table 1: Examples of genes using a superscript modifier.

white^{apr} *or white*^{apricot}	An allele of the *white* gene which has an “apricot” phenotype
Abc^{+} *Abc*	Two alleles for the *Abc* gene (wild-type and mutant, respectively). Note the mutant allele is dominant.
w^{+}; w^{a}; w	Three alleles in a series for the w gene. The first is wild type; the second two are different mutant alleles.
bio^{+}; bio^{-}	A wild-type allele of a biotin gene and its recessive counterpart. This is likely a bacterial gene due to the convention.

This page titled 3.8: Basic Nomenclature is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **14.3: Basic Nomenclature** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

3.9: Linked Genes

Mendel was lucky. He studied a variety of traits in pea plants and his data were consistent with his idea of traits being encoded by pairs of discrete heritable units. He did not call these “genes” and had no idea about their chromosomal origins or chemical makeup. It turns out that the genes he studied were either on different chromosomes and so assorted independently (Chapter 9), or so far apart on the same chromosome that linkage could not be detected (Chapter 10).

Symbols for a gene can be drawn on a page to communicate their position on a chromosome. To do this, we use a forward slash (/) to demonstrate what is on each chromosome. Figure A1.4 shows how we might conceptualize the position of a gene on two chromosomes by collapsing the chromosomes into a single line.

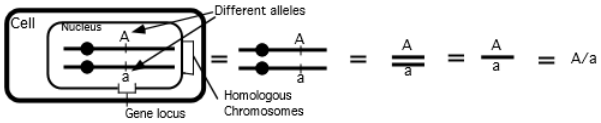


Figure A1.4: A diagram of how chromosomes, loci and alleles look in the cell, and how we depict them in written form. Note that only one gene is represented in this figure and chromosomes have not replicated. *(Original-J.Locke- CC BY-NC 3.0 from Chapter 10)*

There’s no question about where the gene is located when only one trait is under investigation: it will be at the same position on each homolog. Two genes, however, can be one of three possibilities. Each possibility has implications for gene mapping and predicting ratios from a dihybrid cross. Figure A1.5 shows the positions of genes for an unlinked situation as well as linked genes in coupling and repulsion configurations. If genes are unlinked, put the allele symbols for one gene on either side of one slash followed by a semicolon (indicating that it’s unlinked) and the other gene with the alleles separated by a second slash (*A/a*; *B/b*). When genes are linked, only one slash is used: remember, the slash stands for a pair of homologous chromosomes. Genes in coupling would have the dominant genes together on one side of the slash and recessives on the other side (*AB/ab*). Repulsion would represent the other arrangement (*Ab/aB*).

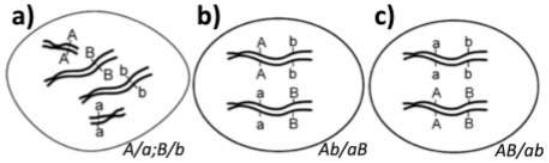


Figure A1.5: Three gene arrangements for cells of genotype AaBb. Chromosomes are replicated (shown with sister chromatids). a) demonstrates unlinked genes: A/a; B/b. Linked genes are shown in b) repulsion Ab/aB and c) coupling AB/ab. *(Modified by T. Nickle from an original by J.Locke- CC BY-NC 3.0 from Chapter 18)*

Practice your skills with identifying linked and unlinked genes online in module 1.5. Some examples of different forms of gene symbols are shown in Table 2. Keep in mind that sometimes you have flexibility in which system of nomenclature you use, but sometimes it is dictated to you, for example in publications or other formal submissions. You are discouraged from inventing your own system or mixing up different systems because it will confuse your readers (or graders!).

Table 2: Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used in publications because not all alleles show complete dominance and many genes have more than two alleles. It's quick and easy for you to use when working out genetics problems when you are sure each gene involves only two alleles.
a^+ and *a*	Superscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +. The mutant allele of gene a would be recessive.
met^+ and met^-	This is typical of a prokaryote gene symbol. It could be referring to wild-type (functional) and mutant (nonfunctional) alleles of a gene that makes a protein in the methionine synthesis pathway.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes. Both representations in this row are identical: it represents a homozygous dominant.
Aa/Aa or *Aa/aa*	Note that this example shows two alleles of the gene *Aa*. We know that the gene symbol is two letters because the slash separates the allele found on each of the homologous chromosomes. We cannot tell if the mutant phenotype is recessive because there's no indication which is wild type.
Grn^+ *shr/Grn shr*	The three-letter system is used here. "*Grn*" might mean that the phenotype is "green", but we can't be sure. What we do know is that the mutant allele codes for a protein leading to a dominant phenotype. The wild-type allele must be recessive to the mutant allele. Maybe "*shr*" means "shrunken" or "short", but we know that the mutant phenotype can only be seen in the homozygous recessive configuration. The phenotype for this organism is mutant for both *Grn* and *shr* traits. Final note: the genes are on the same chromosome based on the position of the slash.
bob^+*/bob*; *mia/mia*	This also uses the three-letter system. The organism is heterozygous for bob but shows the wild-type trait in its phenotype. It is homozygous recessive for *mia* and therefore shows that mutant phenotype. The genes are unlinked.

A more advanced system of nomenclature is outlined in Appendix 2. New rules are introduced to help you identify sex linked genes and predict phenotypes from gene symbols alone.

This page titled 3.9: Linked Genes is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **14.5: Linked Genes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

4: Hypothesis Testing

4.1: Phenotypic Ratios May Not Be As Expected

4.2: Probability

4.3: Chi-Square Test of Goodness-of-Fit

4: Hypothesis Testing is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

4.1: Phenotypic Ratios May Not Be As Expected

Explanation

For a variety of reasons, the phenotypic ratios observed from real crosses rarely match the exact ratios expected based on a Punnett Square or other prediction techniques. There are many possible explanations for deviations from expected ratios. Sometimes these deviations are due to **sampling effects**, in other words, the random selection of a non-representative subset of individuals for observation. On the other hand, it may be because certain genotypes have a less than 100% survival rate. For example, *Drosophila* crosses sometimes give unexpected results because the more mutant alleles a zygote has the less likely it is to survive to become an adult. Genotypes that cause death for embryos or larvae are underrepresented when adult flies are counted.

The χTest For Goodness-of-fit

A statistical procedure called the **chi-square** (χ^2) test can be used to help a geneticist decide whether the deviation between observed and expected ratios is due to sampling effects, or whether the difference is so large that some other explanation must be sought by re-examining the assumptions used to calculate the expected ratio. The procedure for performing a chi-square test is covered in the labs.

This page titled 4.1: Phenotypic Ratios May Not Be As Expected is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.7: Phenotypic Ratios May Not Be As Expected** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

4.2: Probability

Learning Objectives

- Simple rules about adding and multiplying probabilities

The basic idea of a statistical test is to identify a null hypothesis, collect some data, then estimate the probability of getting the observed data if the null hypothesis were true. If the probability of getting a result like the observed one is low under the null hypothesis, you conclude that the null hypothesis is probably not true. It is therefore useful to know a little about probability.

One way to think about probability is as the proportion of individuals in a population that have a particular characteristic. The probability of sampling a particular kind of individual is equal to the proportion of that kind of individual in the population. For example, in fall 2013 there were $22,166$ students at the University of Delaware, and $3,679$ of them were graduate students. If you sampled a single student at random, the probability that they would be a grad student would be $\frac{3,679}{22,166}$ or 0.166. In other words, 16.6% of students were grad students, so if you'd picked one student at random, the probability that they were a grad student would have been 16.6%.

When dealing with probabilities in biology, you are often working with theoretical expectations, not population samples. For example, in a genetic cross of two individual *Drosophila melanogaster* that are heterozygous at the *vestigial* locus, Mendel's theory predicts that the probability of an offspring individual being a recessive homozygote (having teeny-tiny wings) is one-fourth, or 0.25. This is equivalent to saying that one-fourth of a population of offspring will have tiny wings.

Multiplying probabilities

You could take a semester-long course on mathematical probability, but most biologists just need to know a few basic principles. You calculate the probability that an individual has one value of a nominal variable AND another value of a second nominal variable by multiplying the probabilities of each value together.



Fig. 1.3.1 Drosophila melanogaster with an allele at the Antennapedia locus that causes it to have legs where its antennae should be. Picture of Antennapedia from Mutations homéotiques de la drosophile/

For example, if the probability that a *Drosophila* in a cross has vestigial wings is one-fourth, AND the probability that it has legs where its antennae should be is three-fourths, the probability that it has vestigial wings AND leg-antennae is one-fourth times three-fourths, or 0.25×0.75, or 0.1875. This estimate assumes that the two values are independent, meaning that the probability of one value is not affected by the other value. In this case, independence would require that the two genetic loci were on different chromosomes, among other things.

Adding probabilities

The probability that an individual has one value OR another, MUTUALLY EXCLUSIVE, value is found by adding the probabilities of each value together. "Mutually exclusive" means that one individual could not have both values. For example, if the probability that a flower in a genetic cross is red is one-fourth, the probability that it is pink is one-half, and the probability that it is white is one-fourth, then the probability that it is red OR pink is one-fourth plus one-half, or three-fourths.

More complicated situations

When calculating the probability that an individual has one value OR another, and the two values are NOT MUTUALLY EXCLUSIVE, it is important to break things down into combinations that are mutually exclusive. For example, let's say you wanted to estimate the probability that a fly from the cross above had vestigial wings OR leg-antennae. You could calculate the

probability for each of the four kinds of flies: normal wings/normal antennae ($0.75 \times 0.25 = 0.1875$), normal wings/leg-antennae ($0.75 \times 0.75 = 0.5625$), vestigial wings/normal antennae ($0.25 \times 0.25 = 0.0625$), and vestigial wings/leg-antennae ($0.25 \times 0.75 = 0.1875$). Then, since the last three kinds of flies are the ones with vestigial wings or leg-antennae, you'd add those probabilities up ($0.5625 + 0.0625 + 0.1875 = 0.8125$).

When to calculate probabilities

While there are some kind of probability calculations underlying all statistical tests, it is rare that you'll have to use the rules listed above. About the only time you'll actually calculate probabilities by adding and multiplying is when figuring out the expected values for a goodness-of-fit test.

This page titled 4.2: Probability is shared under a not declared license and was authored, remixed, and/or curated by John H. McDonald.

- **1.3: Probability** by John H. McDonald has no license indicated. Original source: http://www.biostathandbook.com.

4.3: Chi-Square Test of Goodness-of-Fit

Learning Objectives

- Study the use of chi-square test of goodness-of-fit when you have one nominal variable
- To see if the number of observations in each category fits a theoretical expectation, and the sample size is large

When to use it

Use the chi-square test of goodness-of-fit when you have one nominal variable with two or more values (such as red, pink and white flowers). You compare the observed counts of observations in each category with the expected counts, which you calculate using some kind of theoretical expectation (such as a $1:1$ sex ratio or a $1:2:1$ ratio in a genetic cross).

If the expected number of observations in any category is too small, the chi-square test may give inaccurate results, and you should use an exact test instead. See the web page on small sample sizes for discussion of what "small" means.

The chi-square test of goodness-of-fit is an alternative to the *G*–test of goodness-of-fit; each of these tests has some advantages and some disadvantages, and the results of the two tests are usually very similar. You should read the section on "Chi-square vs. *G*–test" near the bottom of this page, pick either chi-square or *G*–test, then stick with that choice for the rest of your life. Much of the information and examples on this page are the same as on the *G*–test page, so once you've decided which test is better for you, you only need to read one.

Null hypothesis

The statistical null hypothesis is that the number of observations in each category is equal to that predicted by a biological theory, and the alternative hypothesis is that the observed numbers are different from the expected. The null hypothesis is usually an extrinsic hypothesis, where you knew the expected proportions before doing the experiment. Examples include a $1:1$ sex ratio or a $1:2:1$ ratio in a genetic cross. Another example would be looking at an area of shore that had 59% of the area covered in sand, 28% mud and 13% rocks; if you were investigating where seagulls like to stand, your null hypothesis would be that 59% of the seagulls were standing on sand, 28% on mud and 13% on rocks.

In some situations, you have an intrinsic hypothesis. This is a null hypothesis where you calculate the expected proportions after you do the experiment, using some of the information from the data. The best-known example of an intrinsic hypothesis is the Hardy-Weinberg proportions of population genetics: if the frequency of one allele in a population is p and the other allele is q, the null hypothesis is that expected frequencies of the three genotypes are p^2, $2pq$, and q^2. This is an intrinsic hypothesis, because you estimate p and q from the data after you collect the data, you can't predict p and q before the experiment.

How the test works

Unlike the exact test of goodness-of-fit, the chi-square test does not directly calculate the probability of obtaining the observed results or something more extreme. Instead, like almost all statistical tests, the chi-square test has an intermediate step; it uses the data to calculate a test statistic that measures how far the observed data are from the null expectation. You then use a mathematical relationship, in this case the chi-square distribution, to estimate the probability of obtaining that value of the test statistic.

You calculate the test statistic by taking an observed number (O), subtracting the expected number (E), then squaring this difference. The larger the deviation from the null hypothesis, the larger the difference is between observed and expected. Squaring the differences makes them all positive. You then divide each difference by the expected number, and you add up these standardized differences. The test statistic is approximately equal to the log-likelihood ratio used in the *G*–test. It is conventionally called a "chi-square" statistic, although this is somewhat confusing because it's just one of many test statistics that follows the theoretical chi-square distribution. The equation is:

$$\text{chi}^2 = \sum \frac{(O-E)^2}{E} \tag{4.3.1}$$

As with most test statistics, the larger the difference between observed and expected, the larger the test statistic becomes. To give an example, let's say your null hypothesis is a $3:1$ ratio of smooth wings to wrinkled wings in offspring from a bunch of

Drosophila crosses. You observe 770 flies with smooth wings and 230 flies with wrinkled wings; the expected values are 750 smooth-winged and 250 wrinkled-winged flies. Entering these numbers into the equation, the chi-square value is 2.13. If you had observed 760 smooth-winged flies and 240 wrinkled-wing flies, which is closer to the null hypothesis, your chi-square value would have been smaller, at 0.53; if you'd observed 800 smooth-winged and 200 wrinkled-wing flies, which is further from the null hypothesis, your chi-square value would have been 13.33.

The distribution of the test statistic under the null hypothesis is approximately the same as the theoretical chi-square distribution. This means that once you know the chi-square value and the number of degrees of freedom, you can calculate the probability of getting that value of chi-square using the chi-square distribution. The number of degrees of freedom is the number of categories minus one, so for our example there is one degree of freedom. Using the CHIDIST function in a spreadsheet, you enter **=CHIDIST(2.13, 1)** and calculate that the probability of getting a chi-square value of 2.13 with one degree of freedom is $P = 0.144$.

The shape of the chi-square distribution depends on the number of degrees of freedom. For an extrinsic null hypothesis (the much more common situation, where you know the proportions predicted by the null hypothesis before collecting the data), the number of degrees of freedom is simply the number of values of the variable, minus one. Thus if you are testing a null hypothesis of a $1:1$ sex ratio, there are two possible values (male and female), and therefore one degree of freedom. This is because once you know how many of the total are females (a number which is "free" to vary from 0 to the sample size), the number of males is determined. If there are three values of the variable (such as red, pink, and white), there are two degrees of freedom, and so on.

An intrinsic null hypothesis is one where you estimate one or more parameters from the data in order to get the numbers for your null hypothesis. As described above, one example is Hardy-Weinberg proportions. For an intrinsic null hypothesis, the number of degrees of freedom is calculated by taking the number of values of the variable, subtracting 1 for each parameter estimated from the data, then subtracting 1 more. Thus for Hardy-Weinberg proportions with two alleles and three genotypes, there are three values of the variable (the three genotypes); you subtract one for the parameter estimated from the data (the allele frequency, p); and then you subtract one more, yielding one degree of freedom. There are other statistical issues involved in testing fit to Hardy-Weinberg expectations, so if you need to do this, see Engels (2009) and the older references he cites.

Post-hoc test

If there are more than two categories and you want to find out which ones are significantly different from their null expectation, you can use the same method of testing each category vs. the sum of all other categories, with the Bonferroni correction, as I describe for the exact test. You use chi-square tests for each category, of course.

Assumptions

The chi-square of goodness-of-fit assumes independence, as described for the exact test

Examples

Extrinsic Hypothesis examples

Example

European crossbills (*Loxia curvirostra*) have the tip of the upper bill either right or left of the lower bill, which helps them extract seeds from pine cones. Some have hypothesized that frequency-dependent selection would keep the number of right and left-billed birds at a $1:1$ ratio. Groth (1992) observed 1752 right-billed and 1895 left-billed crossbills.



Fig. 2.3.1 Male red crossbills, Loxia curvirostra, showing the two bill types.

Calculate the expected frequency of right-billed birds by multiplying the total sample size (3647) by the expected proportion (0.5) to yield 1823.5. Do the same for left-billed birds. The number of degrees of freedom when an for an extrinsic hypothesis is

the number of classes minus one. In this case, there are two classes (right and left), so there is one degree of freedom.

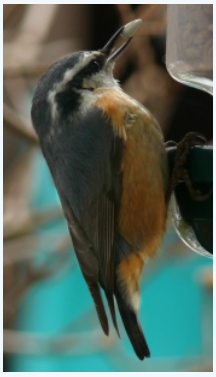
The result is chi-square=5.61, $1 d.f.$, $P = 0.018$, indicating that you can reject the null hypothesis; there are significantly more left-billed crossbills than right-billed.

Example

Shivrain et al. (2006) crossed clearfield rice, which are resistant to the herbicide imazethapyr, with red rice, which are susceptible to imazethapyr. They then crossed the hybrid offspring and examined the F_2 generation, where they found 772 resistant plants, 1611 moderately resistant plants, and 737 susceptible plants. If resistance is controlled by a single gene with two co-dominant alleles, you would expect a $1:2:1$ ratio. Comparing the observed numbers with the $1:2:1$ ratio, the chi-square value is 4.12. There are two degrees of freedom (the three categories, minus one), so the P value is 0.127; there is no significant difference from a $1:2:1$ ratio.

Example

Mannan and Meslow (1984) studied bird foraging behavior in a forest in Oregon. In a managed forest, 54% of the canopy volume was Douglas fir, 28% was ponderosa pine, 5% was grand fir, and 1% was western larch. They made 156 observations of foraging by red-breasted nuthatches; 70 observations (45% of the total) in Douglas fir, 79 (51%) in ponderosa pine, 3 (2%) in grand fir, and 4 (3%) in western larch. The biological null hypothesis is that the birds forage randomly, without regard to what species of tree they're in; the statistical null hypothesis is that the proportions of foraging events are equal to the proportions of canopy volume. The difference in proportions is significant (chi-square=13.59, $3 d.f.$, $P = 0.0035$).


Fig. 2.3.2 Female red-breasted nuthatch, Sitta canadensis.

The expected numbers in this example are pretty small, so it would be better to analyze it with an exact test. I'm leaving it here because it's a good example of an extrinsic hypothesis that comes from measuring something (canopy volume, in this case), not a mathematical theory; I've had a hard time finding good examples of this.

Intrinsic Hypothesis examples

Example

McDonald (1989) examined variation at the *Mpi* locus in the amphipod crustacean *Platorchestia platensis* collected from a single location on Long Island, New York. There were two alleles, Mpi^{90} and Mpi^{100} and the genotype frequencies in samples from multiple dates pooled together were 1203 $Mpi^{90/90}$, 2919 $Mpi^{90/100}$, and 1678 $Mpi^{100/100}$. The estimate of the Mpi^{90} allele proportion from the data is $5325/11600 = 0.459$. Using the Hardy-Weinberg formula and this estimated allele proportion, the expected genotype proportions are 0.211 $Mpi^{90/90}$, 0.497 $Mpi^{90/100}$, and 0.293 $Mpi^{100/100}$. There are three categories (the three genotypes) and one parameter estimated from the data (the Mpi^{90} allele proportion), so there is one degree of freedom. The result is chi-square=1.08, $1 d.f.$, $P = 0.299$, which is not significant. You cannot reject the null hypothesis that the data fit the expected Hardy-Weinberg proportions.

Graphing the results

If there are just two values of the nominal variable, you shouldn't display the result in a graph, as that would be a bar graph with just one bar. Instead, just report the proportion; for example, Groth (1992) found 52.0% left-billed crossbills.

With more than two values of the nominal variable, you should usually present the results of a goodness-of-fit test in a table of observed and expected proportions. If the expected values are obvious (such as 50%) or easily calculated from the data (such as Hardy–Weinberg proportions), you can omit the expected numbers from your table. For a presentation you'll probably want a graph showing both the observed and expected proportions, to give a visual impression of how far apart they are. You should use a bar graph for the observed proportions; the expected can be shown with a horizontal dashed line, or with bars of a different pattern.

If you want to add error bars to the graph, you should use confidence intervals for a proportion. Note that the confidence intervals will not be symmetrical, and this will be particularly obvious if the proportion is near 0 or 1.

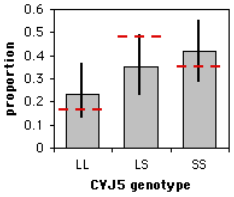


Fig. 2.3.3 Habitat use in the red-breasted nuthatch.. Gray bars are observed percentages of foraging events in each tree species, with 95% confidence intervals; black bars are the expected percentages.

Some people use a "stacked bar graph" to show proportions, especially if there are more than two categories. However, it can make it difficult to compare the sizes of the observed and expected values for the middle categories, since both their tops and bottoms are at different levels, so I don't recommend it.

Similar tests

You use the chi-square test of independence for two nominal variables, not one.

There are several tests that use chi-square statistics. The one described here is formally known as Pearson's chi-square. It is by far the most common chi-square test, so it is usually just called the chi-square test.

You have a choice of three goodness-of-fit tests: the exact test of goodness-of-fit, the *G*–test of goodness-of-fit,, or the chi-square test of goodness-of-fit. For small values of the expected numbers, the chi-square and *G*–tests are inaccurate, because the distributions of the test statistics do not fit the chi-square distribution very well.

The usual rule of thumb is that you should use the exact test when the smallest expected value is less than 5, and the chi-square and *G*–tests are accurate enough for larger expected values. This rule of thumb dates from the olden days when people had to do statistical calculations by hand, and the calculations for the exact test were very tedious and to be avoided if at all possible. Nowadays, computers make it just as easy to do the exact test as the computationally simpler chi-square or *G*–test, unless the sample size is so large that even computers can't handle it. I recommend that you use the exact test when the total sample size is less than 1000. With sample sizes between 50 and 1000 and expected values greater than 5, it generally doesn't make a big difference which test you use, so you shouldn't criticize someone for using the chi-square or *G*–test for experiments where I recommend the exact test. See the web page on small sample sizes for further discussion.

Chi-square vs. *G*–test

The chi-square test gives approximately the same results as the *G*–test. Unlike the chi-square test, the G-values are additive; you can conduct an elaborate experiment in which the G-values of different parts of the experiment add up to an overall G-value for the whole experiment. Chi-square values come close to this, but the chi-square values of subparts of an experiment don't add up exactly to the chi-square value for the whole experiment. *G*–tests are a subclass of likelihood ratio tests, a general category of tests that have many uses for testing the fit of data to mathematical models; the more elaborate versions of likelihood ratio tests don't have equivalent tests using the Pearson chi-square statistic. The ability to do more elaborate statistical analyses is one reason some people prefer the *G*–test, even for simpler designs. On the other hand, the chi-square test is more familiar to more people, and it's always a good idea to use statistics that your readers are familiar with when possible. You may want to look at the literature in your field and use whichever is more commonly used.

Of course, you should *not* analyze your data with both the *G*–test and the chi-square test, then pick whichever gives you the most interesting result; that would be cheating. Any time you try more than one statistical technique and just use the one that give the lowest *P* value, you're increasing your chance of a false positive.

How to do the test

Spreadsheet

I have set up a spreadsheet for the chi-square test of goodness-of-fit chigof.xls . It is largely self-explanatory. It will calculate the degrees of freedom for you if you're using an extrinsic null hypothesis; if you are using an intrinsic hypothesis, you must enter the degrees of freedom into the spreadsheet.

Web pages

There are web pages that will perform the chi-square test here and here. None of these web pages lets you set the degrees of freedom to the appropriate value for testing an intrinsic null hypothesis.

R

Salvatore Mangiafico's *R Companion* has a sample R program for the chi-square test of goodness-of-fit.

SAS

Here is a SAS program that uses PROC FREQ for a chi-square test. It uses the Mendel pea data from above. The "WEIGHT count" tells SAS that the "count" variable is the number of times each value of "texture" was observed. The ZEROS option tells it to include observations with counts of zero, for example if you had **20** smooth peas and **0** wrinkled peas; it doesn't hurt to always include the ZEROS option. CHISQ tells SAS to do a chi-square test, and TESTP=(75 25); tells it the expected percentages. The expected percentages must add up to **100**. You must give the expected percentages in alphabetical order: because "smooth" comes before "wrinkled," you give the expected frequencies for **75%** smooth, **25%** wrinkled.

```
DATA peas;
INPUT texture $ count;
DATALINES;
smooth 423
wrinkled 133
;
PROC FREQ DATA=peas;
WEIGHT count / ZEROS;
TABLES texture / CHISQ TESTP=(75 25);
RUN;
```

Here's a SAS program that uses PROC FREQ for a chi-square test on raw data, where you've listed each individual observation instead of counting them up yourself. I've used three dots to indicate that I haven't shown the complete data set.

```
DATA peas;
INPUT texture $;
DATALINES;
smooth
wrinkled
smooth
smooth
wrinkled
smooth
.
.
.
```

```
smooth
smooth
;
PROC FREQ DATA=peas;
TABLES texture / CHISQ TESTP=(75 25);
RUN;
```

The output includes the following:

```
Chi-Square Test
for Specified Proportions
-------------------------
Chi-Square 0.3453
DF 1
Pr > ChiSq 0.5568
```

You would report this as "chi-square=0.3453, 1 d.f., *P*=0.5568."

Power analysis

To do a power analysis using the G*Power program, choose "Goodness-of-fit tests: Contingency tables" from the Statistical Test menu, then choose "Chi-squared tests" from the Test Family menu. To calculate effect size, click on the Determine button and enter the null hypothesis proportions in the first column and the proportions you hope to see in the second column. Then click on the Calculate and Transfer to Main Window button. Set your alpha and power, and be sure to set the degrees of freedom (Df); for an extrinsic null hypothesis, that will be the number of rows minus one.

As an example, let's say you want to do a genetic cross of snapdragons with an expected **1 : 2 : 1** ratio, and you want to be able to detect a pattern with **5%** more heterozygotes that expected. Enter **0.25**, **0.50**, and **0.25** in the first column, enter **0.225**, **0.55**, and **0.225** in the second column, click on Calculate and Transfer to Main Window, enter **0.05** for alpha, **0.80** for power, and **2** for degrees of freedom. If you've done this correctly, your result should be a total sample size of **964**.

References

1. Picture of nuthatch from kendunn.smugmug.com.
2. Engels, W.R. 2009. Exact tests for Hardy-Weinberg proportions. Genetics 183: 1431-1441.
3. Groth, J.G. 1992. Further information on the genetics of bill crossing in crossbills. Auk 109:383–385.
4. Mannan, R.W., and E.C. Meslow. 1984. Bird populations and vegetation characteristics in managed and old-growth forests, northeastern Oregon. Journal of Wildlife Management 48: 1219-1238.
5. McDonald, J.H. 1989. Selection component analysis of the *Mpi* locus in the amphipod *Platorchestia platensis*. Heredity 62: 243-249.
6. Shivrain, V.K., N.R. Burgos, K.A.K. Moldenhauer, R.W. McNew, and T.L. Baldwin. 2006. Characterization of spontaneous crosses between Clearfield rice (*Oryza sativa*) and red rice (*Oryza sativa*). Weed Technology 20: 576-584.

This page titled 4.3: Chi-Square Test of Goodness-of-Fit is shared under a not declared license and was authored, remixed, and/or curated by John H. McDonald.

- **2.3: Chi-Square Test of Goodness-of-Fit** by John H. McDonald has no license indicated. Original source: http://www.biostathandbook.com.

CHAPTER OVERVIEW

5: Sex Determination, Chromosomes and Linkage

5.1: Karyotypes Describe Chromosome Number and Structure
5.2: Sex-Linkage- An Exception to Mendel’s First Law

5: Sex Determination, Chromosomes and Linkage is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

5.1: Karyotypes Describe Chromosome Number and Structure

Karyograms are images of real chromosomes

Each eukaryotic species has its nuclear genome divided among a number of chromosomes that is characteristic of that species. For example, a haploid human nucleus (i.e. sperm or egg) normally has 23 chromosomes (n=23), and a diploid human nucleus has 23 pairs of chromosomes (2n=46). A **karyotype** is the complete set of chromosomes of an individual. The cell was in metaphase so each of the 46 structures is a replicated chromosome even though it is hard to see the two sister chromatids for each chromosome at this resolution. As expected there are 46 chromosomes. Note that the chromosomes have different lengths. In fact, human chromosomes were named based upon this feature. Our largest chromosome is called 1, our next longest is 2, and so on. By convention the chromosomes are arranged into the pattern shown in Figure **5.1.15** and the resulting image is called a **karyogram**. A karyogram allows a geneticist to determine a person's karyotype - a written description of their chromosomes including anything out of the ordinary.

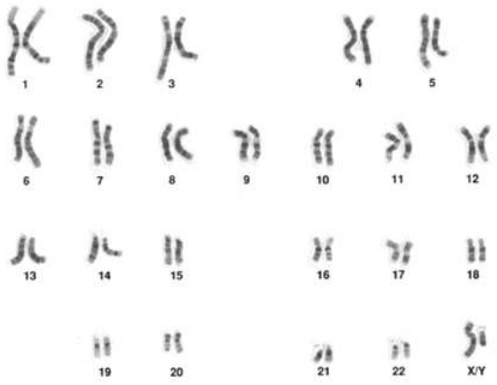


Figure* 5.1.15*: *Karyogram of a normal human male karytype.(Wikipedia-NHGRI-PD)*

Various stains and fluorescent dyes are used to produce characteristic banding patterns to distinguish all 23 chromosomes. The number of chromosomes varies between species, but there appears to be very little correlation between chromosome number and either the complexity of an organism or its total amount genomic DNA.

Autosomes and Sex Chromosomes

In the figure above note that most of the chromosomes are paired (same length, centromere location, and banding pattern). These chromosomes are called **autosomes**. However note that two of the chromosomes, the X and the Y do not look alike. These are **sex chromosomes**. In humans males have one of each while females have two X chromosomes. Autosomes are those chromosomes present in the same number in males and females while sex chromosomes are those that are not. When sex chromosomes were first discovered their function was unknown and the name X was used to indicate this mystery. The next ones were named Y, then Z, and then W.

The combination of sex chromosomes within a species is associated with either male or female individuals. In mammals, fruit flies, and some flowering plants embryos, those with two X chromosomes develop into females while those with an X and a Y become males. In birds, moths, and butterflies males are ZZ and females are ZW. Because sex chromosomes have arisen multiple times during evolution the molecular mechanism(s) through which they determine sex differs among those organisms. For example, although humans and Drosophila both have X and Y sex chromosomes, they have different mechanisms for determining sex .

In mammals, the sex chromosomes evolved just after the divergence of the monotreme lineage from the lineage that led to placental and marsupial mammals. Thus nearly every mammal species uses the same sex determination system. During embryogenesis the gonads will develop into either ovaries or testes. A gene present only on the Y chromosome called *TDF* encodes a protein that makes the gonads mature into testes. XX embryos do not have this gene and their gonads mature into ovaries instead (default). Once formed the testes produce sex hormones that direct the rest of the developing embryo to become male, while the

ovaries make different sex hormones that promote female development. The testes and ovaries are also the organs where gametes (sperm or eggs) are produced.

How do the sex chromosome behave during meiosis? Well, in those individuals with two of the same chromosome (i.e. **homogametic** sexes: XX females and ZZ males) the chromosomes pair and segregate during meiosis I the same as autosomes do. During meiosis in XY males or ZW females (**heterogametic** sexes) the sex chromosomes pair with each other (Figure 5.1.16). In mammals the consequence of this is that all egg cells will carry an X chromosome while the sperm cells will carry either an X or a Y chromosome. Half of the offspring will receive two X chromosomes and become female while half will receive an X and a Y and become male.

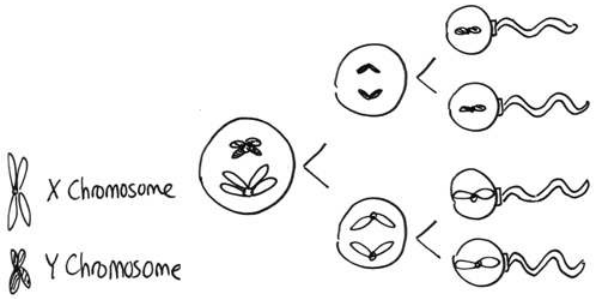


Figure* 5.1.16*: *Meiosis in an XY mammal. The stages shown are anaphase I, anaphase II, and mature sperm. Note how half of the sperm contain Y chromosomes and half contain X chromosomes. (Original-Harrington-CC:AN)*

Aneuploidy - Changes in Chromosome Number

Analysis of karyotypes can identify chromosomal abnormalities, including **aneuploidy**, which is the addition or subtraction of a chromosome from a pair of homologs. More specifically, the absence of one member of a pair of homologous chromosomes is called **monosomy** (only one remains). On the other hand, in a **trisomy**, there are three, rather than two (disomy), homologs of a particular chromosome. Different types of aneuploidy are sometimes represented symbolically; if **2n** symbolizes the normal number of chromosomes in a cell, then **2n-1** indicates monosomy and **2n+1** represents trisomy. The addition or loss of a whole chromosome is a mutation, a change in the genotype of a cell or organism.

The most familiar human aneuploidy is trisomy-21 (i.e. three copies of chromosome 21), which is one cause of **Down syndrome**. Most (but not all) other human aneuploidies are lethal at an early stage of embryonic development. Note that aneuploidy usually affects only one set of homologs within a karyotype, and is therefore distinct from **polyploidy**, in which the entire chromosome set is duplicated (see below). Aneuploidy is almost always deleterious, whereas polyploidy appears to be beneficial in some organisms, particularly many species of food plants.

Aneuploidy can arise due to a **non-disjunction** event, which is the failure of at least one pair of chromosomes or chromatids to segregate during mitosis or meiosis. Non-disjunction will generate gametes with extra and missing chromosomes.

Chromosomal abnormalities

Structural defects in chromosomes are another type of abnormality that can be detected in karyotypes (Fig 2.17). These defects include **deletions, duplications**, and **inversions**, which all involve changes in a segment of a single chromosome. **Insertions** and **translocations** involve two non-homologous chromosomes. In an insertion, DNA from one chromosome is moved to a non-homologous chromosome in a unidirectional manner. In a translocation, the transfer of chromosomal segments is bidirectional and reciprocal – a reciprocal translocation.

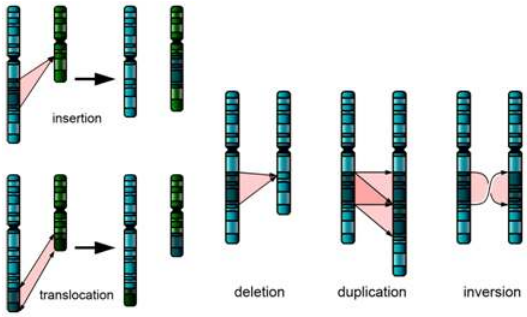


Figure* 5.1.17*: *Structural abberations in chromosomes.(Wikipedia-Zephyris-GFDL)*

Structural defects affect only part of a chromosome (a subset of genes), and so tend to be less harmful than aneuploidy. In fact, there are many examples of ancient chromosomal rearrangements in the genomes of species including our own. Duplications of some small chromosomal segments, in particular, may have some evolutionary advantage by providing extra copies of some genes, which can then evolve in new, potentially beneficial, ways.

Chromosomal abnormalities arise in many different ways, some of which can be traced to rare errors in natural cellular processes such as DNA replication. **Chromosome breakage** also occurs infrequently as the result of physical damage (such as ionizing radiation), movement of some types of transposons, and other factors. During the repair of a broken chromosome, deletions, insertions, translocations and even inversions can be introduced.

This page titled 5.1: Karyotypes Describe Chromosome Number and Structure is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: Karyotypes Describe Chromosome Number and Structure** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

5.2: Sex-Linkage- An Exception to Mendel's First Law

In the previous chapter we introduced sex chromosomes and autosomes. For loci on autosomes, the alleles follow the normal Mendelian pattern of inheritance. However, for loci on the sex chromosomes this is mostly not true, because most of the loci on the typical X-chromosome are absent from the Y-chromosome, even though they act as a homologous pair during meiosis. Instead, they will follow a **sex-linked** pattern of inheritance.

X-Linked Genes: the *white* gene in *Drosophila melanogaster*

A well-studied sex-linked gene is the *white* gene on the X chromosome of *Drosophila melanogaster*. Normally flies have red eyes but flies with a mutant allele of this gene called *white*$^-$ (w^-) have white eyes because the red pigments are absent. Because this mutation is recessive to the wild type w^+ allele females that are heterozygous have normal red eyes. Female flies that are homozygous for the mutant allele have white eyes. Because there is no *white* gene on the Y chromosome, male flies can only be hemizygous for the wild type allele or the mutant allele.

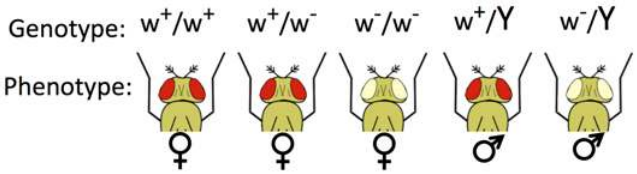


Figure **5.2.9**: Relationship between genotype and phenotype for a the white gene on the X-linked gene in Drosophila melanogaster. The Y chromosome is indicated with a capital Y because it does not have a copy of the white gene. (Original-Harrington/Locke-CC:AN)

A researcher may not know beforehand whether a novel mutation is sex-linked. The definitive method to test for sex-linkage is **reciprocal crosses** (Figure **5.2.10**). This means to cross a male and a female that have different phenotypes, and then conduct a second set of crosses, in which the phenotypes are reversed relative to the sex of the parents in the first cross. For example, if you were to set up reciprocal crosses with flies from pure-breeding w^+ and w^- strains the results would be as shown in Figure **5.2.10**. Whenever reciprocal crosses give different results in the F1 and F2 and whenever the male and female offspring have different phenotypes the usual explanation is sex-linkage. Remember, if the locus were autosomal the F1 and F2 progeny would be different from either of these crosses.

A similar pattern of sex-linked inheritance is seen for X-chromosome loci in other species with an XX-XY sex chromosome system, including mammals and humans. The ZZ-ZW system is similar, but reversed (see below).

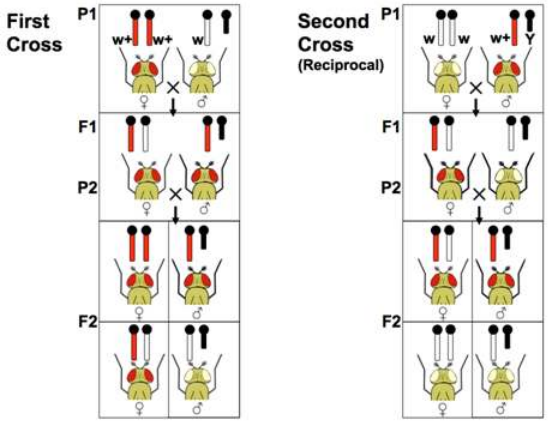


Figure **5.2.10:** Reciprocal crosses involving an X-linked gene in Drosophila melanogaster. In the first cross (left) all of the offspring have red eyes. In second (reciprocal) cross (right) all of the female offspring have red eyes and the male offspring all have white eyes. If the F1 progeny are crossed (to make the P2), the F2 progeny will be different in each cross. The first cross has all red-eyed females and half red-eyed males. The reciprocal cross has half red-eyed males and females. Thomas Morgan won the Nobel Prize for using these crosses to demonstrate that genes (such as white) were on chromosomes (in this case the X-chromosome). (Wikipedia-PAR-PD)

Sex Determination in animals.

There are various mechanisms for sex determination in animals. These include sex chromosomes, chromosome dosage, and environment.

For example in humans and other mammals XY embryos develop as males while XX embryos become females. This difference in development is due to the presence of only a **single gene**, the TDF-Y gene, on the Y-chromosome. Its presence and expression dictates that the sex of the individual will be male. Its absence results in a female phenotype.

Although *Drosophila melanogaster* also has an XX-XY sex chromosomes, its sex determination system uses a different method, that of **X:Autosome (X:A) ratio**. In this system it is the ratio of autosome chromosome sets (A) relative to the number of X-chromosomes (X) that determines the sex. Individuals with two autosome sets and two X-chromosomes (2A:2X) will develop as females, while those with only one X-chromosome (2A:1X) will develop as males. The presence/absence of the Y-chromosome and its genes are not significant.

In other species of animals the number of chromosome sets can determine sex. For example the **haploid-diploid system** is used in bees, ants, and wasps. Typically haploids are male and diploids are female.

In other species, the environment can determine an individuals sex. In alligators (and some other reptiles) the temperature of development dictates the sex, while in many reef fish, the population sex ratio can cause some individuals to change sex.

Dosage Compensation for Loci on Sex Chromosomes.

Mammals and *Drosophila* both have XX - XY sex determination systems. However, because these systems evolved independently they work differently with regard to compensating for the difference in gene dosage (and sex determination – see above). Remember, in most cases the sex chromosomes act as a homologous pair even though the Y-chromosome has lost most of the loci when compared to the X-chromosome. Typically, the X and the Y chromosomes were once similar but, for unclear reasons, the Y chromosomes have degenerated, slowly mutating and loosing its loci. In modern day mammals the Y chromosomes have very few genes left while the X chromosomes remain as they were. This is a general feature of all organisms that use chromosome based sex determination systems. Chromosomes found in both sexes (the X or the Z) have retained their genes while the chromosome found in only one sex (the Y or the W) have lost most of their genes. In either case there is a gene dosage difference between the sexes: e.g. XX females have two doses of X-chromosome genes while XY males only have one. This gene dosage needs to be compensated in a process called **dosage compensation**. There are two major mechanisms.

In Drosophila and many other insects, to make up for the males only having a single X chromosome the genes on it are expressed at twice the normal rate. This mechanism of dosage compensation restores a balance between proteins encoded by **X-linked genes** and those made by **autosomal genes**.

In mammals a different mechanism is used, called **X-chromosome inactivation**.

X-chromosome Inactivation in Mammals

In mammals the dosage compensation system operates in females, not males. In XX embryos one X in each cell is randomly chosen and marked for inactivation. From this point forward this chromosome will be inactive, hence its name $X_{inactive}$ (X_i). The other X chromosome, the X_{active} (X_a), is unaffected. The X_i is replicated during S phase and transmitted during mitosis the same as any other chromosome but most of its genes are never allowed to turn on. The chromosome appears as a condensed mass within interphase nuclei called the Barr body. With the inactivation of genes on one X-chromosome, females have the same number of functioning X-linked genes as males.

This random inactivation of one X-chromosome leads to a commonly observe phenomenon in cats. A familiar X-linked gene is the *Orange* gene (*O*) in cats. The O^O allele encodes an enzyme that results in orange pigment for the hair. The O^B allele causes the hairs to be black. The phenotypes of various genotypes of cats are shown in Figure **5.2.11**. Note that the heterozygous females have an orange and black mottled phenotype known as tortoiseshell. This is due to patches of skin cells having different X-chromosomes inactivated. In each orange hair the X_i chromosome carrying the O^B allele is inactivated. The O^O allele on the X_a is functional and orange pigments are made. In black hairs the reverse is true, the X_i chromosome with the O^O allele is inactive and the X_a chromosome with the O^B allele is active. Because the inactivation decision happens early during embryogenesis, the cells continue to divide to make large patches on the adult cat skin where one or the other X is inactivated.

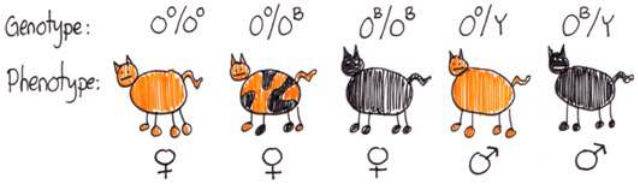


Figure **5.2.11**: Relationship between genotype and phenotype for an X-linked gene in cats. The O^O allele = orange while the O^B *allele = black. (Original-Harrington-CC:AN)*

The *Orange* gene in cats is a good demonstration of how the mammalian dosage compensation system affects gene expression. However, most X-linked genes do not produce such dramatic mosaic phenotypes in heterozygous females. A more typical example is the *F8* gene in humans. It makes Factor VIII blood clotting proteins in liver cells. If a male is hemizygous for a mutant allele the result is hemophilia type A. Females homozygous for mutant alleles will also have hemophilia. Heterozygous females, those people who are $F8^+/F8^-$, do not have hemophilia because even though half of their liver cells do not make Factor VIII (because the X with the $F8^+$ allele is inactive) the other 50% can (Figure **5.2.12**). Because some of their liver cells are exporting Factor VIII proteins into the blood stream they have the ability to form blood clots throughout their bodies. The genetic mosaicism in the cells of their bodies does not produce a visible mosaic phenotype.

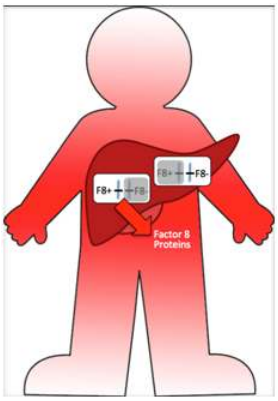


Figure **5.2.12**: This figure shows the two types of liver cells in females heterozygous for an F8 mutation. Because people with the $F8^+/F8^-$ genotype have the same phenotype, normal blood clotting, as $F8^+/F8^+$ people the $F8^-$ mutation is classified as recessive. (Original-Harrington/Locke-CC:AN)

Other Sex-Linked Genes – Z-linked genes

One last example is a **Z-linked gene** that influences feather colour in turkeys. Turkeys are birds, which use the ZZ-ZW sex chromosome system. The *E* allele makes the feathers bronze and the *e* allele makes the feathers brown (Figure **5.2.13**). Only male turkeys can be heterozygous for this locus, because they have two Z chromosomes. They are also uniformly bronze because the *E* allele is completely dominant to the *e* allele and birds use a dosage compensation system similar to *Drosophila* and not mammals. Reciprocal crosses between turkeys from pure-breeding bronze and brown breeds would reveal that this gene is in fact Z-linked.

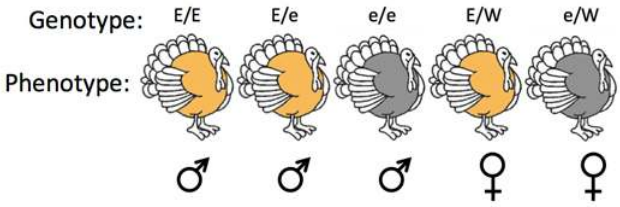


Figure **5.2.13**: Relationship between genotype and phenotype for a Z-linked gene in turkeys. The W chromosome does not have an E/e-gene so it is just indicated with a capital W. (Original-Harrington/Locke-CC:AN)

Mechanisms of Sex Determination Systems

Sex is a phenotype. Typically, in most species, there are multiple characteristics, in addition to sex organs, that distinguish male from female individuals (although some species are normally **hermaphrodites** where both sex organs are present in the same

individual). The morphology and physiology of male and females is a phenotype just like hair or eye colour or wing shape. The sex of an organism is part of its phenotype and can be genetically (or environmentally) determined.

For each species, the genetic determination relies on one of several gene or chromosome based mechanisms. See Figure **5.2.14** for a summary. There are, for other species, also a variety of environmental mechanisms, too (rearing temperature, social interactions, **parthenogenesis**). Whatever the sex choice mechanism, however, there are two different means by which the cells of an organism carry out this decision: **hormonal** or **cell-autonomous**

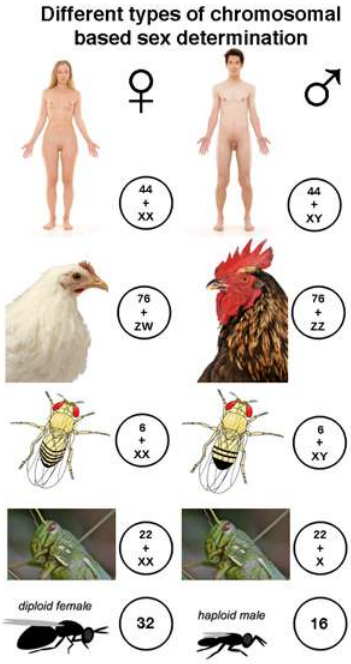


Figure **5.2.14**: – Different types of chromosomal (or gene) based sex determination. From top to bottom, there is the archetypal XX/XY system found in humans (and most mammals) with the TDF-Y gene leading to a male phenotype; the ZW/ZZ system found in chickens (birds, moths, and butterflies); the same XX/XY system in Drosophila (sex is determined by the X-chromosome:autosome ratio); the XX/XO system as found in grasshoppers; and the diploid/haploid system as found in bees (and ants, and wasps). Also, the hormonal mechanism is used in humans, while all the other examples use the cell-autonomous mechanism for development of the male or female sex phenotype. (Wikipedia+original - CFCF with additions and corrections by J. Locke- CC:AN)

Hormonal mechanism: With this system, used by mammals for example, including humans, the zygote initially develops into a sexually undifferentiated embryo that can become either sex. Then, depending on the sex choice of the genital ridge cells, they will grow and differentiate into male (testis) or female (ovary) gonads, which will then produce the appropriate hormones (e.g. testosterone or estrogen). This hormone will circulate throughout the body and cause all the other tissues to develop and differentiate accordingly, into a male or female phenotype for that individual. Thus, the circulating hormone "tells" all the cells and tissues what sex to be and which sexual phenotype to be.

A **freemartin** is a type of **chimera** found in cattle (and some other mammals). Externally it appears as a female but is infertile, and has masculinized behavior and non-functioning ovaries. The animal originates as a female (XX), but acquires male (XY) cells or tissues *in utero* by exchange of some cellular material from a male twin. The female reproductive development is altered by anti-Müllerian hormone from the male twin, acquired via vascular connections between placentas.

Cell-autonomous mechanism: With this system, used by many animals, including birds and insects, the zygote cell initially has a sex phenotype set at the cell level. All cells intrinsically know, individually, which sex they are and develop accordingly, giving the appropriate sexual characteristics and phenotype. Each cell is autonomous with respect to its sex; there are no sex hormone cues to determine the sex expressed. This autonomy can lead to **sexual gynandromorphs**, which are **mosaics** that display both male and female characteristics in a mosaic fasion, typically split down the midline of the organism. These rare individuals are thought to be the result of an improper sex chromosome segregation that occurs in a cell very early in development so that one half of the individual has cells with a male chromosome set while the other half has cells with a female set. If the species is **sexually**

dimorphic (external morphology easily distinguishs males from females) they are easily visible and are even sometimes seen in the wild. See Figure **5.2.15** for a local example. A search on the internet will bring up many more examples.



Figure **5.2.15**: – Drosophila sexual gynadromorph. The left side is female with the most distal abdominal segments being not heavily pigmented, while the right is male and the two most distal segments are heavily pigmeted (arrow). This example was found by a U. of Alberta student in the GENET 375 lab course, Introductions to Molecular Genetic Techniques. (Original – Locke – CC:AN)

While gynandromorphs are seen in cell-autonomous species, such as insects and birds, they are not seen in hormonally determined species, such as mammals, because all the cells display the same sex phenotype caused by the circulating sex hormones. Sexual gynandromorphs appear to be absent in reptiles, amphibians, and fish indicating that they don't use a cell-autonomous mechanism. Nevertheless, there are genetic mosaic individuals in these groups but they do not appear to involve sex determined traits, which is required for a true gynandromorph. They often involve mosaicism of alleles at a single gene locus that affect external morphology (e.g. color).

- A **gynandromorph** is an organism that made up of mosaic tissues of male and female genotypes and displays both male and female characteristics.
- A **mosaic** is an organism or a tissue that contains two or more types of genetically different cells derived from the same zygote.
- A **chimera** is a single organism composed of genetically distinct cells derived from different zygotes.

This page titled 5.2: Sex-Linkage- An Exception to Mendel's First Law is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.5: Sex-Linkage- An Exception to Mendel's First Law** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

6: Pedigrees

6.1: Pedigree Analysis

6.2: Inferring the Mode of Inheritance

6.3: Sporadic and Non-Heritable Diseases

6.4: Calculating Probabilities

6: Pedigrees is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

6.1: Pedigree Analysis

Pedigree charts are diagrams that show the phenotypes and/or genotypes for a particular organism and its ancestors. While commonly used in human families to track genetic diseases, they can be used for any species and any inherited trait. Geneticists use a standardized set of symbols to represent an individual's sex, family relationships and phenotype. These diagrams are used to determine the **mode of inheritance** of a particular disease or trait, and to predict the probability of its appearance among offspring. Pedigree analysis is therefore an important tool in both basic research and **genetic counseling**.

Each pedigree chart represents all of the available information about the inheritance of a single trait (most often a disease) within a family. The pedigree chart is therefore drawn using factual information, but there is always some possibility of errors in this information, especially when relying on family members' recollections or even clinical diagnoses. In real pedigrees, further complications can arise due to **incomplete penetrance** (including age of onset) and **variable expressivity** of disease alleles, but for the examples presented in this book, we will presume complete accuracy of the pedigrees. A pedigree may be drawn when trying to determine the nature of a newly discovered disease, or when an individual with a family history of a disease wants to know the probability of passing the disease on to their children. In either case, a tree is drawn, as shown in Figure **6.1.2**, with circles to represent females, and squares to represent males. Matings are drawn as a line joining a male and female, while a consanguineous mating (closely related is two lines.

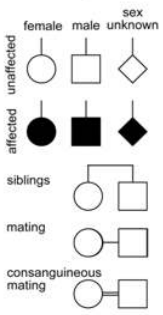


Figure **6.1.2**: Symbols used in drawing a pedigree. (Original-Deyholos-CC:AN)

The affected individual that brings the family to the attention of a geneticist is called the **proband** (or propositus). If an individual is known to have symptoms of the disease (**affected**), the symbol is filled in. Sometimes a half-filled in symbol is used to indicate a known **carrier** of a disease; this is someone who does not have any symptoms of the disease, but who passed the disease on to subsequent generations because they are a heterozygote. Note that when a pedigree is constructed, it is often unknown whether a particular individual is a carrier or not, so not all carriers are always explicitly indicated in a pedigree. For simplicity, in this chapter we will assume that the pedigrees presented are accurate, and represent fully penetrant traits.

This page titled 6.1: Pedigree Analysis is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **5.2: Pedigree Analysis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

6.2: Inferring the Mode of Inheritance

Given a pedigree of an uncharacterized disease or trait, one of the first tasks is to determine which modes of inheritance are possible and then which mode of inheritance is most likely. This information is essential in calculating the probability that the trait will be inherited in any future offspring. We will mostly consider five major types of inheritance: autosomal dominant (AD), autosomal recessive (AR), X-linked dominant (XD), X-linked recessive (XR), and Y-linked (Y).

Autosomal Dominant (AD)

When a disease is caused by a dominant allele of a gene, every person with that allele will show symptoms of the disease (assuming complete penetrance), and only one disease allele needs to be inherited for an individual to be affected. Thus, every affected individual must have an affected parent. A pedigree with affected individuals in every generation is typical of AD diseases. However, beware that other modes of inheritance can also show the disease in every generation, as described below. It is also possible for an affected individual with an AD disease to have a family without any affected children, if the affected parent is a heterozygote. This is particularly true in small families, where the probability of every child inheriting the normal, rather than disease allele is not extremely small. Note that AD diseases are usually rare in populations, therefore affected individuals with AD diseases tend to be heterozygotes (otherwise, both parents would have had to been affected with the same rare disease). Achondroplastic dwarfism, and polydactyly are both examples of human conditions that may follow an AD mode of inheritance.

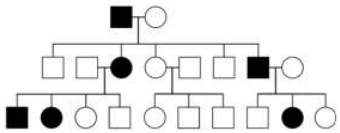


Figure 6.2.3: A pedigree consistent with AD inheritance. (Original-Deyholos-CC:AN)

X-linked dominant (XD)

In X-linked dominant inheritance, the gene responsible for the disease is located on the X-chromosome, and the allele that causes the disease is dominant to the normal allele in females. Because females have twice as many X-chromosomes as males, females tend to be more frequently affected than males in the population. However, not all pedigrees provide sufficient information to distinguish XD and AD. One definitive indication that a trait is inherited as AD, and not XD, is that an affected father passes the disease to a son; this type of transmission is not possible with XD, since males inherit their X chromosome from their mothers.

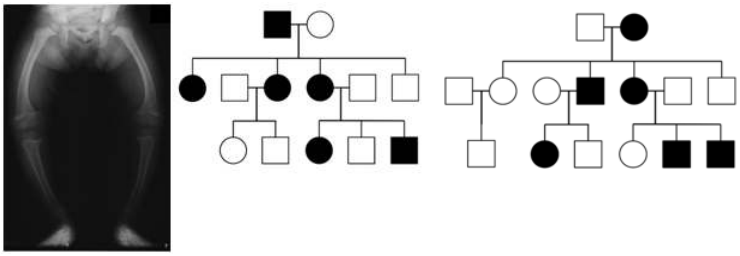


Figure 6.2.5: Some types of rickets may follow an XD mode of inheritance. (Wikipedia-Mrish-CC:AS) (middle and left) Two pedigrees consistent with XD inheritance. (Original-Deyholos_CC:AN)

Autosomal recessive (AR)

Diseases that are inherited in an autosomal recessive pattern require that both parents of an affected individual carry at least one copy of the disease allele. With AR traits, many individuals in a pedigree can be carriers, probably without knowing it. Compared to pedigrees of dominant traits, AR pedigrees tend to show fewer affected individuals and are more likely than AD or XD to “skip a generation”. Thus, the major feature that distinguishes AR from AD or XD is that unaffected individuals can have affected offspring.



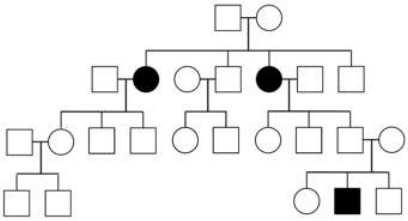
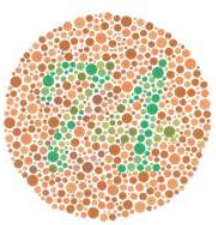


Figure **6.2.6**: (left) Many inborn errors of metabolism, such as phenylketonuria (PKU) are inherited as AR. Newborns are often tested for a few of the most common metabolic diseases. (Wikipedia-U.S. Air Force photo/Staff Sgt. Eric T. Sheler-PD) (right) A pedigree consistent with AR inheritance. (Original-Deyholos_CC:AN)

X-linked recessive (XR)

Because males have only one X-chromosome, any male that inherits an X-linked recessive disease allele will be affected by it (assuming complete penetrance). Therefore, in XR modes of inheritance, males tend to be affected more frequently than females in a population. This is in contrast to AR and AD, where both sexes tend to be affected equally, and XD, in which females are affected more frequently. Note, however, in the small sample sizes typical of human families, it is usually not possible to accurately determine whether one sex is affected more frequently than others. On the other hand, one feature of a pedigree that can be used to definitively establish that an inheritance pattern is not XR is the presence of an affected daughter from unaffected parents; because she would have had to inherit one X-chromosome from her father, he would also have been affected in XR.



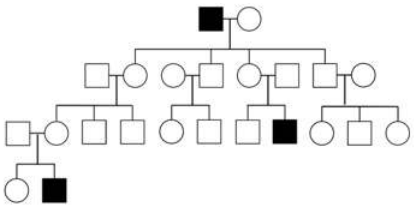


Figure **6.2.8**: Some forms of colour blindness are inherited as XR-traits. Colour blindness is diagnosed using tests such as this. Ishihara Test. (Wikipedia-unknown-PD) (right) A pedigree consistent with XR inheritance. (Original-Deyholos-CC:AN)

Y-linked and Mitochondrial Inheritance.

Two additional modes are **Y-linked** and **Mitochondrial inheritance**.

Only males are affected in human Y-linked inheritance (and other species with the X/Y sex determining system). There is only father to son transmission. This is the easiest mode of inheritance to identify, but it is one of the rarest because there are so few genes located on the Y-chromosome. An example of Y-linked inheritance is the **hairy-ear-rim** phenotype seen in some Indian families. As expected this trait is passed on from father to all sons and no daughters. Y-chromosome DNA polymorphisms can be used to follow the male lineage in large families or through ancient ancestral lineages. For example, the Y-chromosome of Mongolian ruler Genghis Khan (1162-1227 CE), and his male relatives, accounts for ~8% of the Y-chromosome lineage of men in Asia (0.5% world wide).

Mutations in Mitochondrial DNA are inherited through the maternal line (from your mother). There are some human diseases associated with mutations in mitochondria genes. These mutations can affect both males and females, but males cannot pass them on as the mitochondria are inherited via the egg, not the sperm. Mitochondrial DNA polymorphisms are also used to investigate evolutionary lineages, both ancient and recent. Because of the relative similarity of sequence mtDNA is also used in species identification in ecology studies.

This page titled 6.2: Inferring the Mode of Inheritance is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **5.3: Inferring the Mode of Inheritance** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

6.3: Sporadic and Non-Heritable Diseases

Not all of the characterized human traits and diseases are attributed to mutant alleles at a single gene locus. Many diseases that have a heritable component, have more complex inheritance patterns due to (1) the involvement of multiple genes, and/or (2) environmental factors. On the other hand, some non-genetic diseases may appear to be heritable because they affect multiple members of the same family, but this is due to the family members being exposed to the same toxins or other environmental factors (e.g. in their homes).

Finally, diseases with similar symptoms may have different causes, some of which may be genetic while others are not. One example of this is ALS (amyotrophic lateral sclerosis); approximately 5-10% of cases are inherited in an AD pattern, while the majority of the remaining cases appear to be **sporadic**, in other words, not caused by a mutation inherited from a parent. We now know that different genes or proteins are affected in the inherited and sporadic forms of ALS. The physicist Stephen Hawking (Figure **6.3.10**) and baseball player Lou Gehrig both suffered from sporadic ALS.



Figure **6.3.10**: Stephen Hawking(Wikipedia-NASA-PD)

This page titled 6.3: Sporadic and Non-Heritable Diseases is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **5.4: Sporadic and Non-Heritable Diseases** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

6.4: Calculating Probabilities

Once the mode of inheritance of a disease or trait is identified, some inferences about the genotype of individuals in a pedigree can be made, based on their phenotypes and where they appear in the family tree. Given these genotypes, it is possible to calculate the probability of a particular genotype being inherited in subsequent generations. This can be useful in genetic counseling, for example when prospective parents wish to know the likelihood of their offspring inheriting a disease for which they have a family history.

Probabilities in pedigrees are calculated using knowledge of Mendelian inheritance and the same basic methods as are used in other fields. The first formula is the **product rule**: the joint probability of two independent events is the product of their individual probabilities; this is the probability of one event AND another event occurring. For example, the probability of a rolling a "five" with a single throw of a single six-sided die is 1/6, and the probability of rolling "five" in each of three successive rolls is 1/6 x 1/6 x 1/6 = 1/216. The second useful formula is the **sum rule**, which states that the combined probability of two independent events is the sum of their individual probabilities. This is the probability of one event OR another event occurring. For example, the probability of rolling a five or six in a single throw of a dice is 1/6 + 1/6 = 1/3.

With these rules in mind, we can calculate the probability that two carriers (i.e. heterozygotes) of an AR disease will have a child affected with the disease as ½ x ½ = ¼, since for each parent, the probability of any gametes carrying the disease allele is ½. This is consistent with what we already know from calculating probabilities using a Punnett Square (e.g. in a monohybrid cross *Aa* x *Aa*, ¼ of the offspring are *aa*).

We can likewise calculate probabilities in the more complex pedigree shown in Figure **6.4.11**.

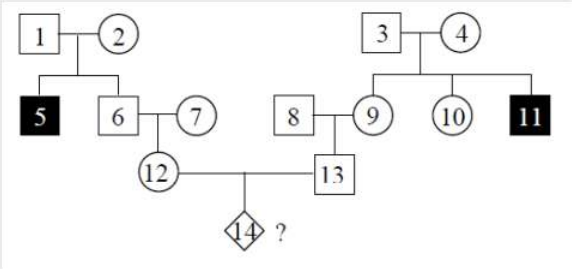


Figure **6.4.11**: Individuals in this pedigree are labeled with numbers to make discussion easier. (Original-Deyholos-CC:AN)

Assuming the disease has an AR pattern of inheritance, what is the probability that individual 14 will be affected? We can assume that individuals #1, #2, #3 and #4 are heterozygotes (*Aa*), because they each had at least one affected (*aa*) child, but they are not affected themselves. This means that there is a 2/3 chance that individual #6 is also *Aa*. This is because according to Mendelian inheritance, when two heterozygotes mate, there is a 1:2:1 distribution of genotypes *AA*:*Aa*:*aa*. However, because #6 is unaffected, he can't be *aa*, so he is either *Aa* or *AA*, but the probability of him being *Aa* is twice as likely as *AA*. By the same reasoning, there is likewise a 2/3 chance that #9 is a heterozygous carrier of the disease allele.

If individual 6 is a heterozygous for the disease allele, then there is a ½ chance that #12 will also be a heterozygote (i.e. if the mating of #6 and #7 is *Aa* × *AA*, half of the progeny will be *Aa*; we are also assuming that #7, who is unrelated, does not carry any disease alleles). Therefore, the combined probability that #12 is also a heterozygote is 2/3 x 1/2 = 1/3. This reasoning also applies to individual #13, i.e. there is a 1/3 probability that he is a heterozygote for the disease. Thus, the overall probability that both individual #12 and #13 are heterozygous, and that a particular offspring of theirs will be homozygous for the disease alleles is 1/3 x 1/3 x 1/4 = 1/36.

This page titled 6.4: Calculating Probabilities is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **5.5: Calculating Probabilities** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

7: Multiple Gene Alleles

7.1: Relationships Between Genes, Genotypes and Phenotypes
7.2: Complementation tests and Allelism

7: Multiple Gene Alleles is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

7.1: Relationships Between Genes, Genotypes and Phenotypes

Terminology

A specific position along a chromosome is called a **locus**. Each gene occupies a specific locus (so the terms locus and gene are often used interchangeably). Each locus will have an allelic form (allele). The complete set of alleles (at all loci of interest) in an individual is its **genotype**. Typically, when writing out a genotype, only the alleles at the locus (loci) of interest are considered – all the others are present and assumed to be wild type. The visible or detectable effect of these alleles on the structure or function of that individual is called its **phenotype** – what it looks like. The phenotype studied in any particular genetic experiment may range from simple, visible traits such as hair color, to more complex phenotypes including disease susceptibility or behavior. If two alleles are present in an individual, then various interactions between them may influence their expression in the phenotype.

Complete Dominance

Let us return to an example of a simple phenotype: flower color in Mendel's peas. We have already said that one allele as a homozygote produces purple flowers, while the other allele as a homozygote produces white flowers (see Figures 1.8 and 3.3). But what about an individual that has one purple allele and one white allele; what is the phenotype of an individual whose genotype is heterozygous? This can only be determined by experimental observation. We know from observation that individuals heterozygous for the purple and white alleles of the flower color gene have purple flowers. Thus, the allele associated with purple color is therefore said to be **dominant** to the allele that produces the white color. The white allele, whose phenotype is masked by the purple allele in a heterozygote, is **recessive** to the purple allele.

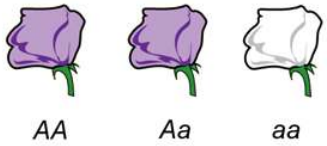


Figure **7.1.3**: Relationship between genotype and phenotype for an allele that is completely dominant to another allele. Original-Deholos (Fireworks)-CC:AN)

To represent this relationship, often, a dominant allele will be represented by a capital letter (e.g. *A*) while a recessive allele will be represented in lower case (e.g. *a*). However, many different systems of genetic symbols are in use. The most common are shown in Table **7.1.1**. Also note that genes and alleles are usually written in *italics* and chromosomes and proteins are not. For example, the *white* gene in *Drosophila melanogaster* on the X chromosome encodes a protein called WHITE.

Table **7.1.1**: Examples of symbols used to represent genes and alleles.

Examples	Interpretation
A and *a*	Uppercase letters represent dominant alleles and lowercase letters indicate recessive alleles. Mendel invented this system but it is not commonly used because not all alleles show complete dominance and many genes have more than two alleles.
a^+ and a^1	Superscripts or subscripts are used to indicate alleles. For wild type alleles the symbol is a superscript +.
AA or *A/A*	Sometimes a forward slash is used to indicate that the two symbols are alleles of the same gene, but on homologous chromosomes.

Incomplete Dominance

Besides dominance and recessivity, other relationships can exist between alleles. In **incomplete dominance** (also called **semi-dominance,** Figure **7.1.4**), both alleles affect the trait additively, and the phenotype of the heterozygote is intermediate between either of the homozygotes. For example, alleles for color in carnation flowers (and many other species) exhibit incomplete dominance. Plants with an allele for red petals (A_1) and an allele for white petals (A_2) have pink petals. We say that the A_1 and the A_2 alleles show incomplete dominance because neither allele is completely dominant over the other.

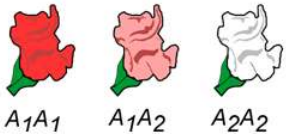


Figure **7.1.4**: Relationship between genotype and phenotype for incompletely dominant alleles affecting petal colour in carnations. (Original-Deholos-CC:AN)

Co-Dominance

Co-dominance is another type of allelic relationship, in which a heterozygous individual expresses the phenotype of both alleles simultaneously. An example of co-dominance is found within the **ABO blood group** of humans. The *ABO* gene has three common alleles which were named (for historical reasons) I^A, I^B, and *i*. People homozygous for I^A or I^B display only A or B type antigens, respectively, on the surface of their blood cells, and therefore have either type A or type B blood (Figure **7.1.5**). Heterozygous I^AI^B individuals have both A and B antigens on their cells, and so have type AB blood. Notice that the heterozygote expresses both alleles simultaneously, and is not some kind of novel intermediate between A and B. Co-dominance is therefore distinct from incomplete dominance, although they are sometimes confused.

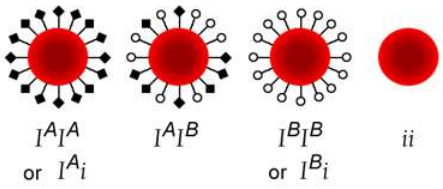


Figure* 7.1.5*: *Relationship between genotype and phenotype for three alleles of the human ABO gene. The I^A and I^B alleles show co-dominance. The I^A allele is completely dominant to the i allele. The I^B allele is completely dominant to the i allele. (Original-Deholos -CC:AN)*

It is also important to note that the third allele, *i*, does not make either antigen and is recessive to the other alleles. *I^A/i or I^B/i* individuals display A or B antigens respectively. People homozygous for the *i* allele have type O blood. This is a useful reminder that different types of dominance relationships can exist, even for alleles of the same gene. Many types of molecular markers, which we will discuss in a later chapter, display a co-dominant relationship among alleles.

This page titled 7.1: Relationships Between Genes, Genotypes and Phenotypes is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **3.2: Relationships Between Genes, Genotypes and Phenotypes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

7.2: Complementation tests and Allelism

– One, or more than one gene?

As explained earlier in this chapter, mutant screening is one of the beginning steps geneticists use to investigate biological processes. When geneticists obtain two independently derived mutants (either from natural populations or during a mutant screen) with similar phenotypes, an immediate question is whether or not the mutant phenotype is due to a loss of function in the same gene, or are they mutant in different genes that both affect the same phenotype (e.g., in the same pathway). That is, are they **allelic mutations**, or **non-allelic mutations**, respectively? This question can be resolved using **complementation tests**, which bring together, or combine, the two mutations under consideration into the same organism to assess the combined phenotype.

hypothetical example of purple flowers

The easiest way to understand a complementation test is by example (Fig.4.9). The pigment in a purple flower could depend on a biochemical pathway much like the **biochemical pathways** leading to the production of arginine in Neurospora (review in Chapter 1). A plant that lacks the function of gene A (genotype *aa*) would produce mutant, white flowers that looked just like the flowers of a plant that lacked the function of gene B (genotype *bb*). (The genetics of two loci are discussed more in the following chapters.) Both A and B are enzymes in the same pathway that leads from a colorless compound#1, thorough colorless compound#2, to the purple pigment. Blocks at either step will result in a mutant white, not wild type purple, flower.

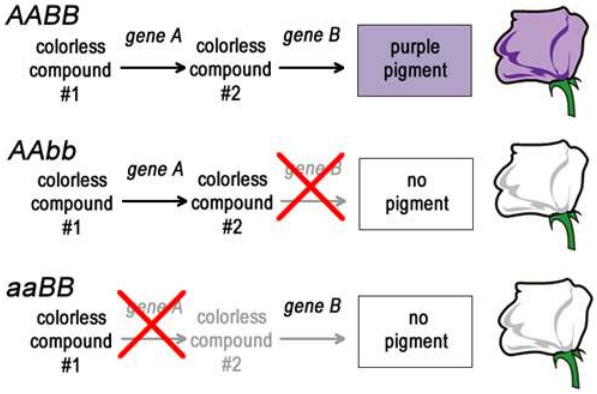


Figure 7.2.9: In this simplified biochemical pathway, two enzymes encoded by two different genes modify chemical compounds in two sequential reactions to produce a purple pigment. Loss of either of the enzymes disrupts the pathway and no pigment is produced. (Original-Deyholos-CC:AN)

Strains with mutations in gene A can be represented as the genotype *aa*, while strains with mutations in gene B can be represented as *bb*. Given that there are two genes here, A and B, then each of these mutant strains can be more completely represented as *aaBB* and *AAbb* . (LEARNING NOTE: Student often forget that genotypes usually only show mutant loci, however, one must remember all the other genes are assumed to be wild type.)

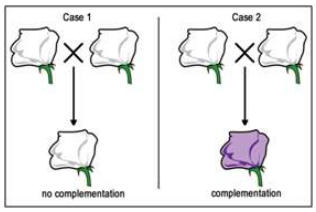
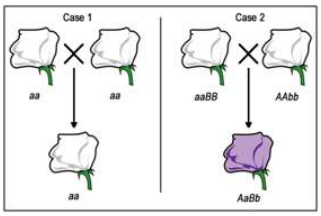
If these two strains are crossed together the resulting progeny will all be *AaBb*. They will have both a wild type, functional A gene and B gene and will thus have a pigmented, purple flower, a wild type phenotype. This is an example of **complementation**. Together, each strain provides what the other is lacking (*AaBb*). The mutations are in different genes and are thus called non-allelic mutations.

Now, if we are presented with a third pure-breeding, independently derived white-flower mutant strain, we won't initially know if it is mutant in gene A or gene B (or possibly some other gene altogether). We can use complementation testing to determine which gene is mutated. To perform a complementation test, two homozygous individuals with similar mutant phenotypes are crossed (Figure 7.2.10).

If the F1 progeny all have the same mutant phenotype (Case 1 - Figure 7.2.10A), then we infer that the same gene is mutated in each parent. These mutations would then be called allelic mutations - in the same gene locus. These mutations FAIL to COMPLEMENT one another (still mutant). These could be either the exact same mutant alleles, or different mutations in the same gene (allelic).

Conversely, if the F1 progeny all appear to be wild-type (Case 2 - Figure **7.2.10**B), then each of the parents most likely carries a mutation in a different gene. These mutations would then be called non-allelic mutations - in a different gene locus. These mutations do COMPLEMENT one another.

Note: For mutations to be used in complementation tests they are (1) usually true-breeding (homozygous at the mutant locus), and (2) must be recessive mutations. Dominant mutation CANNOT be used in complementation tests. Also, remember, some mutant strains may have more than one gene locus mutated and thus would fail to complement mutants from more than one other locus (or group).

A.		B.	
	Figure* 7.2.10*A* – *Observation: *In a typical complementation test, the genotypes of two parents are unknown (although they must be pure breeding, homozygous mutants). If the F1 progeny all have a mutant phenotype (Case 1), there is no complementation. If the F1 progeny are all wild-type, the mutations have successfully complemented each other.*		***Figure* 7.2.10*B* – *Interpretation:*** *The pure breeding, homozygous mutant parents had unknown genotypes before the complementation test, but it could be assumed that they had either mutations in the same genes (Case 1) or in different genes (Case 2). In Case 1, all of the progeny would have the mutant phenotype, because they would all have the same, homozygous genotype as the parents. In Case 2, each parent has a mutation in a different gene, therefore none of the F_1 progeny would be homozygous mutant at either locus. Note that the genotype in Case 1 could be written as either aa or aaBB. (Original-Deyholos-CC:AN)*

This page titled 7.2: Complementation tests and Allelism is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **4.6: Complementation tests and Allelism** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

8: Gene Interactions and Epistasis

8.1: Epistasis and Other Gene Interactions
8.2: Example of Multiple Genes Affecting One Character

8: Gene Interactions and Epistasis is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.1: Epistasis and Other Gene Interactions

Some dihybrid crosses produce a phenotypic ratio that differs from 9:3:3:1, such as 9:3:4, 12:3:1, 9:7, or 15:1. Note that each of these modified ratios can be obtained by summing one or more of the 9:3:3:1 classes expected from our original dihybrid cross. In the following sections, we will look at some modified phenotypic ratios obtained from dihybrid crosses and what they might tell us about interactions between genes.

Recessive epistasis

Epistasis (which means "standing upon") occurs when the phenotype of one locus masks, or prevents, the phenotype of another locus. Thus, following a dihybrid cross fewer than the typical four phenotypic classes will be observed with epistasis. As we have already discussed, in the absence of epistasis, there are four phenotypic classes among the progeny of a dihybrid cross. The four phenotypic classes correspond to the genotypes: *A_B_*, *A_bb*, *aaB_*, and *aabb*. If either of the singly homozygous recessive genotypes (i.e. *A_bb* or *aaB_*) has the same phenotype as the double homozygous recessive (*aabb*), then a **9:3:4** phenotypic ratio will be obtained. For example, in the Labrador Retriever breed of dogs (Figure 8.1.5), the B locus encodes a gene for an important step in the production of melanin. The dominant allele, *B* is more efficient at pigment production than the recessive *b* allele, thus *B_* hair appears black, and *bb* hair appears brown. A second locus, which we will call *E*, controls the deposition of melanin in the hairs. At least one functional *E* allele is required to deposit any pigment, whether it is black or brown. Thus, all retrievers that are *ee* fail to deposit any melanin (and so appear pale yellow), regardless of the genotype at the *B* locus (Figure 8.1.6).



Figure 8.1.5: Labrador Retrievers with different coat colors: (from left to right) black, chocolate, yellow: an example of recessive epistasis. (Flickr-John Curley/PhilRomans/Miss Chien-CC:AND)

The *ee* genotype is therefore said to be **epistatic** to both the B and b alleles, since the homozygous *ee* phenotype masks the phenotype of the B locus. The B/b locus is said to be **hypostatic** to the *ee* genotype. Because the masking allele is in this case is recessive, this is called **recessive epistasis**.

	EB	*Eb*	*eB*	*eb*
EB	*EEBB*	*EEBb*	*EeBB*	*EeBb*
Eb	*EEBb*	*EEbb*	*EeBb*	*Eebb*
eB	*EeBB*	*EeBb*	*eeBB*	*eeBb*
eb	*EeBb*	*Eebb*	*eeBb*	*eebb*

Figure 8.1.6: Genotypes and phenotypes among the progeny of a dihybrid cross of Labrador Retrievers heterozygous for two loci affecting coat color. The phenotypes of the progeny are indicated by the shading of the cells in the table: black coat (black, E_B_); chocolate coat (brown, E_bb); yellow coat (yellow, eeB_ or eebb). (Original-Locke-CC:AN)

Dominant epistasis

In some cases, a dominant allele at one locus may mask the phenotype of a second locus. This is called **dominant epistasis**, which produces a segregation ratio such as **12:3:1**, which can be viewed as a modification of the 9:3:3:1 ratio in which the *A_B_* class is combined with one of the other genotypic classes that contains a dominant allele. One of the best known examples of a 12:3:1 segregation ratio is fruit color in some types of squash (Figure 8.1.7). Alleles of a locus that we will call *B* produce either yellow (*B_*) or green (*bb*) fruit. However, in the presence of a dominant allele at a second locus that we call *A*, no pigment is produced at all, and fruit are white. The dominant *A* allele is therefore epistatic to both *B* and *bb* combinations (Figure 8.1.8). One possible biological interpretation of this segregation pattern is that the function of the *A* allele somehow blocks an early stage of pigment synthesis, before neither yellow or green pigments are produced.



Figure 8.1.7: Green, yellow, and white fruits of squash. (Flickr-Unknown-CC:AN)

	AB	***Ab***	***aB***	***ab***
AB	*AABB*	*AABb*	*AaBB*	*AaBb*
Ab	*AABb*	*AAbb*	*AaBb*	*Aabb*
aB	*AaBB*	*AaBb*	*aaBB*	*aaBb*
ab	*AaBb*	*Aabb*	*aaBb*	*aabb*

Figure 8.1.8***:*** *Genotypes and phenotypes among the progeny of a dihybrid cross of squash plants heterozygous for two loci affecting fruit color. (Original-Deyholos-CC:AN)*

This page titled 8.1: Epistasis and Other Gene Interactions is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **6.2: Epistasis and Other Gene Interactions** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

8.2: Example of Multiple Genes Affecting One Character

Cat Fur Genetics

Most aspects of the fur phenotypes of common cats can be explained by the action of just a few genes (Table 6-2). Other genes, not described here, may further modify these traits and account for the phenotypes seen in tabby cats and in more exotic breeds, such as Siamese.

For example, the X-linked ***Orange*** gene has two allelic forms. The O^O allele produces orange fur, while the O^B alleles produce non-orange (often black) fur. Note however, that because of X-chromosome inactivation the result is mosaicism in expression. In O^O / O^B female heterozygotes patches of black and orange are seen, which produces the tortoise shell pattern (Figure 6-13 A,B). This is a rare example of **co-dominance** since the phenotype of both alleles can be seen. Note that the cat in part A has short fur compared to the cat in part B; recessive alleles at an independent locus (L/l) produce long (*ll*) rather than short (*L_*) fur.

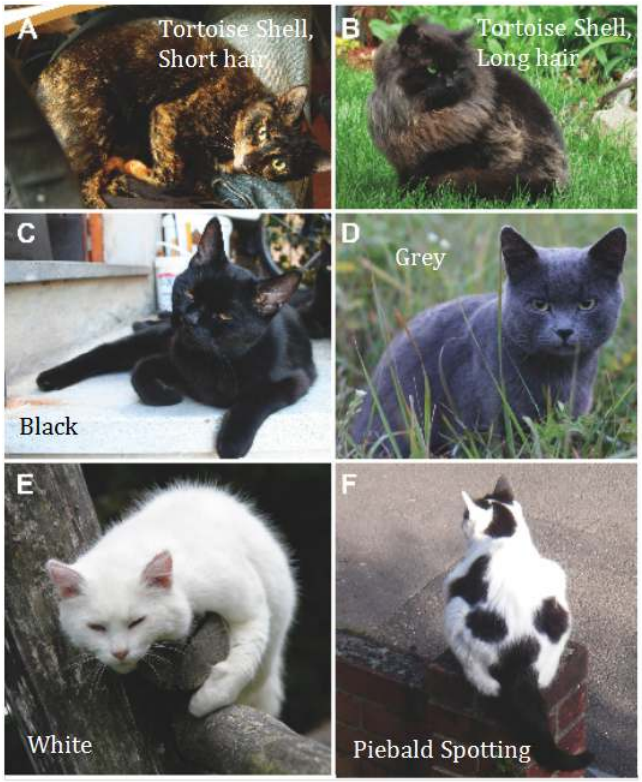


Figure **8.2.13**: **Representatives of various fur phenotypes in cats:** Tortoise shell (A,B) pigmentation in cats with short (A) and long (B) fur; black (C) and grey (D) cats that differ in genotype at the dilute locus. The pure white pattern (E) is distinct from piebald spotting (F). (flickr-cwbuecheler–CC:AN, flickr-=-.0=-CC:AND, flickr-Leonardo Zanchi CC:AS, flickr-malfet_ CC:AN, flickr- valentinastorti-CC:AN, flickr- Denni Schnapp-CC:AS)

Alleles of the ***dilute*** gene affect the intensity of pigmentation, regardless of whether that pigmentation is due to black or orange pigment. Part C shows a black cat with at least one dominant allele of *dilute* (*D_*), in contrast to the cat in D, which is grey rather than black, because it has the *dd* genotype.

Epistasis is demonstrated by an allele of only one of the genes in Table **8.2.2**. One dominant allele of ***white masking*** (*W*) prevents normal development of melanocytes (pigment producing cells). Therefore, cats with genotype (*W_*) will have entirely white fur regardless of the genotype at the *Orange* or *dilute* loci (part E). Although this locus produces a white colour, *W_* is not the same as albinism, which is a much rarer phenotype caused by mutations in other genes. Albino cats can be distinguished by having red eyes, while *W_* cats have eyes that are not red.

Piebald spotting is the occurrence of patches of white fur. These patches vary in size due to many reasons, including genotype. Homozygous cats with genotype *ss* do not have any patches of white, while cats of genotype *Ss* and *SS* do have patches of white,

and the homozygotes tend to have a larger proportion of white fur than heterozygotes (part F). The combination of piebald spotting and tortoise shell patterning produce a **calico cat**, which has separate patches of orange, black, and white fur.

Table **8.2.2**: Summary of simplified cat fur phenotypes and genotypes.

Trait	Phenotype	Genotype	Comments
fur length	short	*LL* or *Ll*	*L* is completely dominant
	long	*ll*	
all white fur (non-albino)	100% white fur	*WW* or *Ww*	If the cat has red eyes it is albino, not *W_*. *W* is epistatic to all other fur color genes; if cat is *W_*, can't infer genotypes for any other fur color genes.
		ww	
piebald spotting	> 50% white patches (but not 100%)	*SS*	*S* is incompletely dominant and shows variable expressivity
	< 50% white patches	*Ss*	
	no white patches	*ss*	
orange fur	all orange fur	X^OX^O *or* X^OY	*O* is X-linked
	tortoise shell variegation	X^OX^B	
	no orange fur (often black)	X^BX^B *or* X^BY	
dilute pigmentation	pigmentation is intense	*Dd* or *dd*	*D* is completely dominant
	pigmentation is dilute (e.g. gray rather than black; cream rather than orange; light brown rather than brown)	*dd*	
tabby	tabby pattern	*AA* or *Aa*	This is a simplification of the tabby phenotype, which involves multiple genes
	solid coloration	*aa*	
sex	female	XX	
	male	XY	

References

1. *Adapted from Christensen (2000) Genetics 155:999-1004*)

This page titled 8.2: Example of Multiple Genes Affecting One Character is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **6.3: Example of Multiple Genes Affecting One Character** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

9: Genetic Pathways

9.1: Genetic Pathways

9: Genetic Pathways is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.1: Genetic Pathways

Learning Objectives

- Understand how multiple genes often function together to control a trait
- Understand the concept of a biosynthetic pathway
- Understand the concept of a regulatory pathway
- Understand the concept of a network
- Understand how mutations in different genes of a pathway can cause the same phenotype
- Understand how mutations in different genes of a pathway might cause different phenotypes

Introduction

Genes do not function in isolation but rather suites of genes act in concert to perform biological functions. When different genes function in different sequential steps of a biological process, this is known as a **genetic pathway**. Perhaps the most conceptually intuitive type of pathway is a **biosynthetic pathway**, where a precursor molecule is chemically modified through a series of enzymatically catalyzed intermediate steps to produce a bioactive product. Since enzymes are (proteins), the genes that encode these enzymes are considered a genetic pathway. In a **regulatory pathway**, some type of stimulus leads to a change in the expression or activity of a particular gene product, which in turn acts to alter the expression or activity of another gene product or products, which in turn could regulate yet another level of activity.

Ultimately, these regulatory changes in expression or activity lead to a response to the stimulus, typically involving changes in gene expression. Again, the genes that encode the various types of regulatory molecules, mainly proteins, constitute a genetic pathway. Pathways are often targets in the biotechnological manipulation of traits, which requires an understanding of how pathways behave. Also note that many biosynthetic and regulatory systems are better described as networks because of often, pathways branch or converge and interact with other systems. However, for simplicity such complex networks will not be discussed here.

Biosynthetic pathways

Genes encode enzymes that catalyze steps in the synthesis of a compound. A **biosynthetic pathway** actually describes a process of converting a precursor substrate into a product. In addition to the precursor substrate and product, it includes the enzymes, the chemical reactions catalyzed by the enzymes and all the intermediate compounds. A pathway consists of a series of steps where a precursor molecule acts as a substrate for an enzyme, which catalyzes a chemical reaction to produce a product, which then serves as a substrate for a subsequent step.

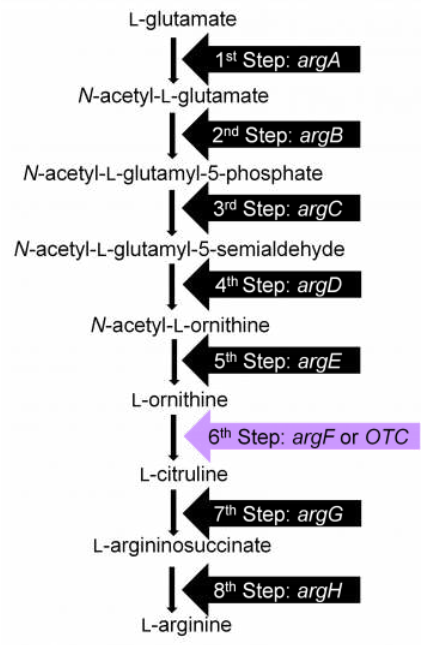


Figure 1. The arginine biosynthetic pathway. The labeled arrows represent chemical reaction steps. The italicized 4-letter symbol (e.g., argF) represents the gene that encodes the enzyme abbreviated to the right that catalyzes each step (e.g., OTC).

We have already discussed the one gene, one enzyme hypothesis which was an important advance in the discovery of how genes encode proteins. This analysis also provides a nice example of the concept of a biosynthetic pathway. This hypothesis was discovered by analyzing **auxotrophic** neurospora mutants that were defective in the production of amino acids. That is, the biosynthetic pathways producing particular amino acids were disrupted by mutations. The text discusses the analysis of the arginine pathway. Mutants that disrupt this pathway cannot grow on minimal medium lacking arginine but can grow if arginine is supplied in the medium. The same fundamental pathway also produces arginine in plants.

Mutations that disrupt the function of particular enzymes in the pathway block the progression of the pathway at that corresponding step. For example, a mutation in the argF gene would disrupt the enzyme ornithine carbamoyltransferase (OTC). This would block step 6 in the pathway, the conversion of L-ornithine to L-citruline. As a result, the intermediate compound L-Ornithine will likely accumulate, whereas all the compounds that occur later in the pathway (i.e., L-Citruline, L-argininosuccinate and L-arginine) will be deficient.

The anthocyanin pathway

The anthocyanin pathway is a slightly more complex pathway, although still simple as far as these things go. It is of significant interest to horticulturists because anthocyanin pigments are responsible for many plant colors, particularly in flowers and fruits (think red wine). Anthocyanin pigments range from orange to red, purple, and blue. In addition to the attractive colors, anthocyanins provide, there is also considerable recent interest in anthocyanins for their dietary value, acting as strong antioxidants and possibly anti-cancer agents.

Effects of mutations at different steps in anthocyanin biosynthesis.

Of course, each enzyme in the anthocyanin pathway is encoded by a gene. As such, mutations can generate variation or disrupt the pathway. The maize anthocyanin pathway and examples of mutants are shown in Figure 2. The ***a2*** mutant causes a deficiency in the anthocyanidin synthase enzyme, blocking the production of any pigmented compounds. The ***a1*** and ***c2*** mutants are similarly colorless. The ***bz2*** mutant is deficient in an enzyme related to glutathione-S-transferase, which is required for transport of anthocyanidins into the vacuole, resulting in the bronze-colored pigment due to the more alkaline environment of the cytosol. The ***bz1*** mutant (not shown) produces a similar phenotype. Mutation of the ***pr1*** gene eliminates just one class of anthocyanins. The ***pr1*** gene encodes flavanone 3′-hydroxylase, required for the production of cyanidin, which generates purple anthocyanin. In the mutant, only pelargonidin is synthesized, producing a red pigment.

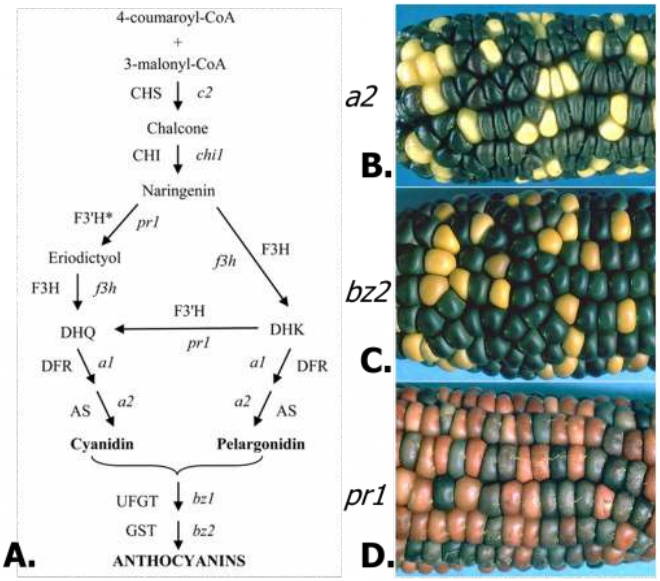


Figure 2. Maize anthocyanin mutants. A. The maize anthocyanin pathway. (Modified from Sharma et al., 2011). B-D. Ears segregating a2 (anthocyaninless2), bz2 (bronze2), and pr1 (red aleurone1) mutant kernels, respectively. Normal kernels are dark purple, nearly black, and the mutant kernels are the lighter colors. Images in B-D courtesy of M.G. Neuffer and MaizeGDB.

Regulatory pathways

A regulatory pathway refers to a situation where one gene or gene product controls the expression or activity of another gene or gene product, which may in turn also perform a regulatory function. As we discussed, gene function can be regulated at many points from transcriptional regulation to post-translational regulation. We will examine a couple of specific examples of regulatory pathways but first let's consider some of the general features of regulatory pathways. Understanding the logic of regulatory pathways will become important later when we seek to use biotechnology to manipulate traits by altering their regulation.

A simplistic but useful way to help understand the logic of regulatory pathways is to consider them as a series of switches with on and off states.

Positive vs. negative regulators

A key feature of regulatory pathways is that regulators might function either positively or negatively. A positive regulator is one that functions to activate the next component in the pathway, whereas a negative regulator would inhibit the next component. For example, a transcription factor that activated the transcription of a gene would be considered a positive regulator, while one that repressed the transcription of a gene would be a negative regulator. Another common type of regulator is a protein kinase. Protein kinases are proteins that add a phosphate group to another protein. Phosphorylation is a common type of post-translational modification that can regulate the activity of proteins; sometimes the phosphorylated protein becomes activated and sometimes it becomes repressed.

A positive regulatory step is represented with an arrow. For example, "protein A activates protein B" would be represented like:

$$\mathbf{A} \rightarrow \mathbf{B} \tag{9.1.1}$$

A negative regulatory step is represented with a bar. "Protein C represses protein D" would be represented like:

$$\mathbf{C} \dashv \mathbf{D} \tag{9.1.2}$$

Active vs. inactive states of regulators

A second point to consider in the logic of regulatory pathways is the activity state of each component. A useful way to think about this is that each component can have an "on" or an "off" state. Of course, in reality, many proteins can have intermediate levels of activity but for the sake of simplicity, we will just consider the on and off states.

The effect of a regulator being in the on or off state depends on whether it is a positive or negative regulator.

If a positive regulator is in the on state, it will function to activate the next factor. In the example of A → B, if **A** is in the on or active state, it will function to switch **B** to the on or active state.

If a negative regulator is in the on state, it will function to switch the next factor to the off or inactive state. For C ⊣ D, if **C** is in the on or active state, it will switch **D** to the off or inactive state.

Regulatory pathways sometimes contain many, sometimes few, steps. They can also contain a mixture of positive and negative regulators. Let us explore a few hypothetical examples to see how the pieces fit together to regulate plant responses to stimuli.

When a pathway is depicted, all the steps are shown. For the sake of simplicity, it is assumed that the default activity state of any given component is such that the preceding step functions to change it. We will begin with a simple example containing several positive regulatory steps that activate a cellular activity, **Activity 1**, in response to **Stimulus 1**. In the pathway shown below, the **Stimulus 1** activates **A**. Therefore, we assume that in the absence of this stimulus, **A** is in the inactive or off state. Since **A** is inactive, it is not functioning to activate **B**, which is therefore also in the inactive state, and likewise for **C**. Ultimately **Activity 1** does not occur in the absence of **Stimulus 1** but does occur in response to the stimulus.

Stimulus 1 → A → B → C → Activity 1

We can use color coding to help visualize this. Light grey represents the **OFF** state and blue represents the **ON** state.

No Stimulus 1 → A → B → C → Activity 1 (OFF)

Stimulus 1 → A → B → C → Activity 1 (ON)

Or we can make a table to help keep track of the states of each component in the presence or absence of stimulus.

Stimulus 1 →	A →	B →	C →	Activity 1
NO	OFF	OFF	OFF	NO
YES	ON	ON	ON	YES

Now let us look at another example containing both positive and negative regulators.

Stimulus 2 → D → E ⊣ F → Activity 2

In this case, components **D** and **E** would be in the inactive states in the absence of stimulus, because they would not have been activated. But **E** functions as a negative regulator of **F**. Since **E** is inactive, it is not functioning to inhibit **F**, which then remains in the active state promote cellular **Activity 2**. When **Stimulus 2** is present, **D** and **E** become activated, and **E** functions to inactivate **F**. Since **F** is off, **Activity 2** does not occur. Thus, the net response to **Stimulus 2** is the *repression* of **Activity 2**.

⊣ (9.1.3)

⊣ (9.1.4)

OR

Stimulus 2 →	D →	E ⊣	F →	Activity 2
NO	OFF	OFF	ON	YES
YES	ON	ON	OFF	NO

Now that we have considered regulatory pathways from a hypothetical perspective, let's consider a couple real ones. As mentioned, regulatory pathways can take on many forms. They can consist of a series of transcription factors that regulate the expression of one another's genes, ultimately resulting in the regulation of genes that effect some biological response. Here, we will look at a signal transduction pathway consisting of a series of post-translational modifications that occur response to a hormone stimulus and culminate in gene expression changes to effect a response.

Signal transduction pathway for hormone signaling

Brassinosteroids (BRs) are a class of plant steroid hormones that control many aspect of plant physiology. They are most noted for their role in promoting plant growth and mutants deficient in BR biosynthesis or signaling show dwarfism. BRs also promote a range of other responses, including yield and increased stress resistance.

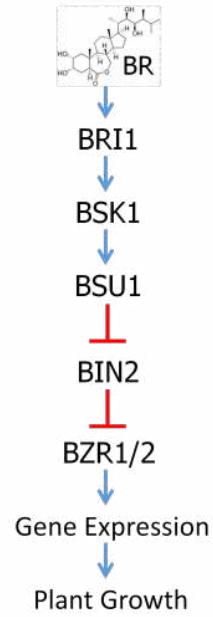


Figure 3. The brassinosteroid signaling pathway. Diagrammatic representation of the regulatory relationships among BR pathway components.

BR hormones, of which brassinolide (BL) is considered the most active, are recognized by a receptor called BRI1. BRI1 is a type of receptor known as a receptor kinase, which spans the cell membrane. On the outside of the cell is a receptor domain which specifically recognizes and binds to BR. There is also a membrane-spanning domain and then a protein kinase domain inside the cell. Binding to BR outside the cell then activates the protein kinase domain inside the cell. Protein kinases are enzymes that function to add phosphate groups to other proteins. As previously discussed, phosphorylation can alter the activity of a protein. Activation of the BRI1 receptor kinase then results in the activation of a signaling pathway that ultimately regulates the activity of transcription factors inside the nucleus, changing the expression of genes that regulate growth and stress resistance.

The BR signal transduction pathway is well studied and is summarized in Figure 3. Do not be concerned with memorizing the names of all these factors, but just focus on understanding the regulatory logic. The overall regulatory relationships of the pathway are depicted in Figure 3. The main target of regulation is a pair of closely related transcription factors called BZR1 and BZR2. BZR1/2 are negatively regulated by a protein called BIN2 in the absence of BR. BIN2 is a protein kinase that phosphorylates BZR1/2, which results in BZR1/2's exclusion from the nucleus and their proteolytic degradation. In the absence of BR, BIN2 contains a phosphate group that is required for its activity.

Binding of BR to the BRI1 receptor domain outside the cell results in the activation of the kinase domain inside the cell. Activated BRI1 then phosphorylates a protein called BSK. Phosphorylated BSK binds yet another protein called BSU1, which is a protein phosphatase. Protein phosphatases are enzymes that remove phosphate groups from other proteins. BSK binding activates BSU1 which then catalyzes the removal of the phosphate on BIN2, thereby inactivating BIN2. Thus, BSU1 is a negative regulator of BIN2. When BIN2 is inactivated, that allows BZR1/2 to accumulate, to enter the nucleus, and to regulate the expression of genes that effect the BR response, including promoting plant growth and yield.

Effects of mutation on different steps

Mutations can affect regulatory pathways in multiple ways. Most mutations are **loss-of-function** mutations where the function of the gene or gene product is partly or completely impaired. Considering the BR signaling pathway, the phenotypes of mutation will differ depending on whether the mutation affects a positive or negative regulator of the pathway. A loss-of-function mutation in the bri1 gene will block the perception of the hormone and therefore the pathway will not be activated; a dwarf phenotype will result.

On the other hand, a loss-of-function in the bin2 gene will release BZR1/2 from regulation. The result will be constant BZR1/2 activity regardless of whether BR hormone is present. This might be expected to generate giant plants but in fact the effects of deregulating hormone responses are more complex—let us just call it unregulated growth.

Mutations can also be **gain-of-function**, where the gene product assumes a form that is constantly in the active state. Such mutations are typically dominant and are much less common than loss-of-function. Again, the outcome of a gain-of-function mutation on the BR signaling pathway will vary depending on whether it affects a positive or negative regulator of the pathway. A gain-of-function mutation in a positive regulator would cause deregulated hormone responses, whereas such a mutation in a negative regulator would permanently shut down the pathway and cause dwarfism.

Complementary gene action and genetic epistasis in the context of pathways

Considering that biological processes are controlled by multiple genes acting together, several genetic principles make more sense. First, it becomes clear how mutations in different genes can produce the same phenotypic effects. Since multiple genes are often required for a single biological process, disruption of different genes in the process might have the same net effect. From this follows the concept of **complementary gene action**. Mutant plants with the same phenotype are crossed together and the offspring is normal looking. The mutations in the parent plants likely affected different genes in the same pathway. Finally, the concept of **epistasis** can be best understood in the context of pathways. Note that the term epistasis is used differently by different geneticists. In quantitative genetics, epistasis refers to any gene interaction. Here, the term epistasis refers to a specific type of gene interaction where the action of one gene is masked by the action of another gene. In other words, if two mutations are combined in an individual, only one of the phenotypes is apparent. The mutant whose phenotype is apparent is said to be epistatic to the one that is masked.

Complementary gene action in a regulatory pathway

The concept of complementary gene action is no different in a regulatory pathway than in a biosynthetic pathway. Mutations in different genes that have similar roles in the pathway (e.g., are both positive regulators of the pathway) would be expected to cause similar mutant phenotypes. If such mutants in different genes were crossed together, the F1 progeny would be normal because both genes would be heterozygous and therefore a functional copy of both factors would be present.

Epistasis in a regulatory pathway

Epistasis can be quite complex, especially in regulatory pathways. Unlike complementary gene action, epistasis is not restricted to recessive loss-of-function mutations. The situation is further complicated by the presence of positive and negative regulators in a pathway. In contrast to a biosynthetic pathway, the rule of thumb for a regulatory pathway is that the more “downstream” gene is epistatic to the more “upstream” one. Let’s turn again to the BR signaling pathway and look at a couple of examples to understand why.

First consider loss-of-function mutations in the *bri1* and *bin2* genes, which in single mutants cause a dwarf phenotype and an unregulated growth phenotype, respectively. If the two mutations are combined into a double mutant plant, what do we expect? The *bri1* mutations disrupt the receptor function required to initiate signaling through the pathway. However, BIN2 functions to inhibit the activity of pathway by repressing the function of BZR1/2. If BIN2 is rendered non-functional, then BZR1/2 will be active and promote the growth response. This will be true even if previous steps are blocked by the *bri1* mutation. So the double mutant would show the *bin2* unregulated growth phenotype making *bin2* epistatic to *bri1*.

What about a gain-of-function mutation in *BZR1* or *BZR2* combined with a loss of function mutation in *bri1*? A gain-of-function mutation would render *BZR1/2* constitutively active, regardless of the activity state of BIN2, resulting in the unregulated growth phenotype. Since BIN2 activity would be irrelevant, upstream activities would also be irrelevant. Thus, the double mutant would show unregulated growth and *BZR1/2* would be epistatic to *bri1*.

Lesson Summary

We saw that genes do not act alone but function in concert with other genes to perform various biological functions such as the biosynthesis of a compound or the regulation of a response to a stimulus. When genes control a sequential series of steps, that is called a pathway. But pathways typically branch and intersect with other pathways to form networks. Mutations affect the activity of pathways in various ways, depending on the nature of the mutation (loss- vs. gain-of-function) and the function of the gene product (positive vs. negative effector).

When mutations affect different steps in a pathway, they can often have a similar overall effect on the pathway and therefore lead to similar phenotypic consequences. Such mutations will complement one another even though their phenotypes are identical. It is also possible that mutations in different steps of a pathway can cause different phenotypes. When such mutations are combined into a double mutant individual, the phenotype of one is often masked by the phenotype of the other epistatic mutation. When pathways are linear, the “upstream” mutation is generally epistatic in biosynthetic pathways and in regulatory pathways it is generally the “downstream” mutation that is epistatic. In branched pathways or networks, it is more difficult to generalize.

Complementary gene action is important for geneticists trying to identify all the steps in a pathway. This is the basis of the classic complementation test used to determine if mutations are allelic or affect independent genes. Epistasis is also useful to geneticists who want to understand whether mutations with different phenotypes affect the same pathway, and if so, to determine the order of gene action.

When biotechnology is used to manipulate a trait, it is really the activity of a pathway or network that is being targeted. The same genetic principles outlined in this section are used to design biotechnological strategies to alter pathway activities so as to produce the desired outcome on the trait of interest.

Activity 1

To answer the following questions, refer to the pathways illustrated in Figure 3. They are labeled slightly differently but a1 encodes DFR, which catalyzes the production of leucoanthocyanidins and a2 encodes AS or LDOX/ANS, which subsequently catalyzes the production of anthocyanidins. Assume that normal, wild-type maize kernels are the deep purple/black color as in Figure 3.

1. What would you predict happens to the levels of leucoanthocyanidins in an *a2* mutant compared to normal?

a. no change
b. increase
c. decrease
d. can’t predict

Show Answer

Answer: b or d.

The expectation would be b, that leucoanthocyanidins would increase. Because the pathway is blocked at the subsequent step, leucoanthocyanidins cannot be converted to anthocyanidins. But since all the preceding enzymes remain functional, those reactions would be expected to continue resulting in the accumulation of leucoanthocyanidins.

d is also a correct response. Biosynthetic pathways do not always behave as expected. For example, there could be feedback inhibition of one or more enzymes by intermediate compounds. Bottom line, we won’t know for certain until the levels are measured experimentally.

2. What would you predict happens to the levels of pelargonidin in an *pr1* mutant compared to normal?

a. no change
b. increase
c. decrease
d. can’t predict

Show Answer

Answer: b or d.

Again, the expectation would be b, that pelargonidin would increase. Normally, some fraction of the intermediate compounds are channeled through the pathway branch leading to cyaniding. Because the pathway leading to cyanidin is blocked it might be expected that all the intermediates would be converted to pelargonidin, causing an increased accumulation.

d is also a correct response. As discussed in problem 1, biosynthetic pathways do not always behave as expected. Feedback inhibition could be a factor or it may be possible that the enzymatic capacity of the pelargonidin branch of the pathway is already saturated. To increase flux through the pathway, it could be required to increase the levels or activities of one or more rate-limiting enzymes. Again, we won’t know for certain until the levels are measured experimentally.

Activity 2

Predict the outcomes of the following mutations on the signaling output, or plant response for the BR signaling pathway.

1. A gain-of-function mutation in the BZR2 gene

a. increased output
b. no change
c. decreased output
d. can't tell

Show Answer

Answer: a. Since BZR2 is a positive regulator of BR signaling, increased output from the pathway would be expected from a gain-of-function mutation.

2. A loss-of-function mutation in the bsu1 gene. (Study the pathway and think about this one!)

a. increased output
b. no change
c. decreased output
d. can't tell

Show Answer

Answer: c. The output from the pathway would decrease causing decreased plant growth. Even though BSU1 is a negative regulator, its target is BIN2 which is another negative regulator. Without BSU1 function, BIN2 will always be in the ON state and inhibit the activities of BZR1/2, therefore decreasing the response to BRs.

Activity 3

1. Predict the phenotype of a BSK gain-of-function.

a. unregulated growth
b. no effect
c. dwarf
d. can't predict

Show Answer

Answer: a. BSK is a positive regulator of the pathway that functions to activate BSU1. Therefore a gain-of-function would cause it to positively activate the pathway, even in the absence of BR hormone, causing unregulated growth.

2. Now predict the phenotype of a double mutant between the BSK gain-of-function and a BIN2 gain-of-function.

a. unregulated growth
b. no effect
c. dwarf
d. can't predict

Show Answer

Answer: c. The gain-of-function BIN2 mutation locks BIN2 in the ON state where it would function to inhibit the activity of BZR1/2. It would be impervious to regulation by the upstream BSU1, so even if BSU1 were continually active due to the gain-of-function BSK, the BIN2 mutant would inhibit growth and therefore be epistatic.

Activity 4

Now you try a couple. Fill in each of the following tables to determine whether the hypothetical Activity will be activated or repressed in response to each Stimulus.

References

Sharma, M., Cortes-Cruz, M., Ahern, K.R., McMullen, M., Brutnell, T.P., and Chopra, S. (2011). Identification of the Pr1 Gene Product Completes the Anthocyanin Biosynthesis Pathway of Maize. Genetics 188, 69-79.

This page titled 9.1: Genetic Pathways is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Philip Becraft, & Philip Becraft (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.10: Genetic Pathways** by Donald Lee, Marjorie Hanneman, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

10: Nucleic Acids

10.1: DNA is the Genetic Material

10.2: The Structure of DNA

10.3: DNA- The Genetic Material

Thumbnail: DNA double helix. (public domain; NIH - Genome Research Institute).

10: Nucleic Acids is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

10.1: DNA is the Genetic Material

By the early 1900's, biochemists had isolated hundreds of different chemicals from living cells. Which of these was the genetic material? Proteins seemed like promising candidates, since they were abundant, diverse, and complex molecules. However, a few key experiments demonstrated that DNA, rather than protein, is the genetic material.

Griffith's Transformation Experiment (1928)

Microbiologists identified two strains of the bacterium *Streptococcus pneumoniae*. The R-strain produced rough colonies on a bacterial plate, while the other S-strain was smooth (Figure **10.1.2**). More importantly, the S-strain bacteria caused fatal infections when injected into mice, while the R-strain did not (top, Figure **10.1.3**). Neither did "heat-treated" S-strain cells. **Griffith** in 1929 noticed that upon mixing "heat-treated" S-strain cells together with some R-type bacteria (neither should kill the mice), the mice died and there were S-strain, pathogenic cells recoverable. Thus, some non-living component from the S-type strains contained genetic information that could be transferred to and **transform** the living R-type strain cells into S-type cells.

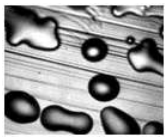


Figure **10.1.2**: Colonies of Rough (top) and Smooth (bottom) strains of S. pneumoniae. (J. Exp.Med.98:21, 1953-R. Austrian-Pending)

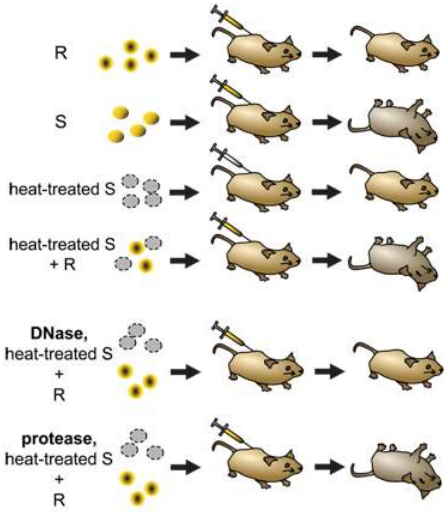


Figure **10.1.3**: Experiments of Griffith and of Avery , MacLeod and McCarty. R strains of S. pneumoniae do not cause lethality. However, DNA-containing extracts from pathogenic S strains are sufficient to make R strains pathogenic. (Original-Deyholos-CC:AN)

Avery, MacLeod and McCarty's Experiment (1944)

What kind of molecule from within the S-type cells was responsible for the transformation? To answer this, researchers named **Avery, MacLeod and McCarty** separated the S-type cells into various components, such as proteins, polysaccharides, lipids, and nucleic acids. Only the nucleic acids from S-type cells were able to make the R-strains smooth and fatal. Furthermore, when cellular extracts of S-type cells were treated with DNase (an enzyme that digests DNA), the transformation ability was lost. The researchers therefore concluded that DNA was the genetic material, which in this case controlled the appearance (smooth or rough) and pathogenicity of the bacteria.

Hershey and Chase's Experiment (1952)

Further evidence that DNA is the genetic material came from experiments conducted by **Hershey and Chase**. These researchers studied the transmission of genetic information in a virus called the T2 bacteriophage, which used *Escherichia coli* as its host

bacterium (Figure **10.1.4**). Like all viruses, T2 hijacks the cellular machinery of its host to manufacture more viruses. The T2 phage itself only contains both protein and DNA, but no other class of potential genetic material.



Figure **10.1.4**: Electronmicrograph of T2 bacteriophage on surface of E. coli. (Wikipedia-G. Colm-PD)

To determine which of these two types of molecules contained the genetic blueprint for the virus, Hershey and Chase grew viral cultures in the presence of radioactive isotopes of either phosphorus (^{32}P) or sulphur (^{35}S). The phage incorporated these isotopes into their DNA and proteins, respectively (Fig 1.5). The researchers then infected *E. coli* with the radiolabeled viruses, and looked to see whether ^{32}P or ^{35}S entered the bacteria. After ensuring that all viruses had been removed from the surface of the cells, the researchers observed that infection with ^{32}P labeled viruses (but not the ^{35}S labeled viruses) resulted in radioactive bacteria. This demonstrated that DNA was the material that contained genetic instructions.

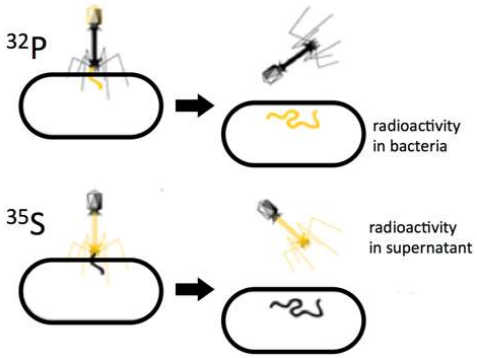


Figure **10.1.5**: When ^{32}P-labeled phage infects E. coli, radioactivity is found only in the bacteria, after the phage are removed by agitation and centrifugation. In contrast, after infection with ^{35}S-labeled phage, radioactivity is found only in the supernatant that remains after the bacteria are removed. (Original-Deyholos-CC:AN)

Meselson and Stahl experiment (1958)

From the complementary strands model of DNA, proposed by Watson and Crick in 1953, there were three straightforward possible mechanisms for DNA replication: (1) semi-conservative, (2) conservative, and (3) dispersive (Fig 1.6).

1. The **semi-conservative** model proposes the two strands of a DNA molecule separate during replication and then strand acts as a template for synthesis of a new, complementary strand.
2. The **conservative** model proposes that the entire DNA duplex acts as a single template for the synthesis of an entirely new duplex.
3. The **dispersive** model has the two strands of the double helix breaking into units that which are then replicated and reassembled, with the new duplexes containing alternating segments from one strand to the other.

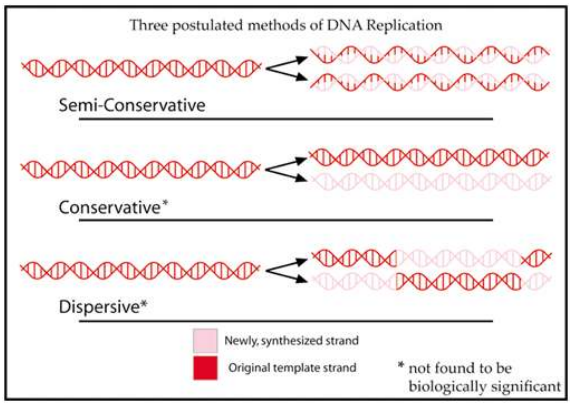


Figure **10.1.6**: The three models of DNA replication possible from the double helix model of DNA structure. (Wikipedia-Adenosine-CC:AN)

Each of these three models makes a different prediction about the how DNA strands should be distributed following two rounds of replication. These predictions can be tested in the following experiment by following the nitrogen component in DNA in ***E. coli*** as it goes through several rounds of replication. Meselson and Stahl used different isotopes of Nitrogen, which is a major component in DNA. **Nitrogen-14** (**^{14}N**) is the most abundant natural isotope, while **Nitrogen-15** (**^{15}N**) is rare, but also denser. Neither is radioactive; each can be followed by a difference in density – "**light**" 14 vs "**heavy**"15 atomic weight in a CsCl density gradient ultra-centrifugation of DNA.

The experiment starts with *E. coli* grown for several generations on medium containing only ^{15}N. It will have denser DNA. When extracted and separated in a CsCl density gradient tube, this "heavy" DNA will move to a position nearer the bottom of the tube in the more dense solution of CsCl (left side in Figure **10.1.7**). DNA extracted from *E. coli* grown on normal, ^{14} N containing medium will migrate more towards the less dense top of the tube.

If these *E. coli* cells are transferred to a medium containing only ^{14}N, the "light" isotope, and grown for one generation, then their DNA will be composed of one-half ^{15}N and one-half ^{14}N. If the this DNA is extracted and applied to a **CsCl gradient**, the observed result is that one band appears at the point midway between the locations predicted for wholly ^{15}N DNA and wholly ^{14}N DNA (Figure **10.1.7**). This "single-band" observation is inconsistent with the predicted outcome from the conservative model of DNA replication (disproves this model), but is consistent with both that expected for the semi-conservative and dispersive models.

If the *E. coli* is permitted to go through another round of replication in the ^{14}N medium, and the DNA extracted and separated on a CsCl gradient tube, then two bands were seen by Meselson and Shahl: one at the ^{14}N-^{15}N intermediate position and one at the wholly ^{14}N position (Figure **10.1.7**). This result is inconsistent with the dispersive model (a single band between the ^{14}N-^{15}N position and the wholly ^{14} N position) and thus disproves this model. The two band observation is consistent with the semi-conservative model which predicts one wholly ^{14} N duplex and one ^{14}N-^{15}N duplex. Additional rounds of replication also support the semi-conservative model/hypothesis of DNA replication. Thus, the semi-conservative model is the currently accepted mechanism for DNA replication. Note however, that we now also know from more recent experiments that whole chromosomes, which can be millions of bases in length, are also semi-conservatively replicated.

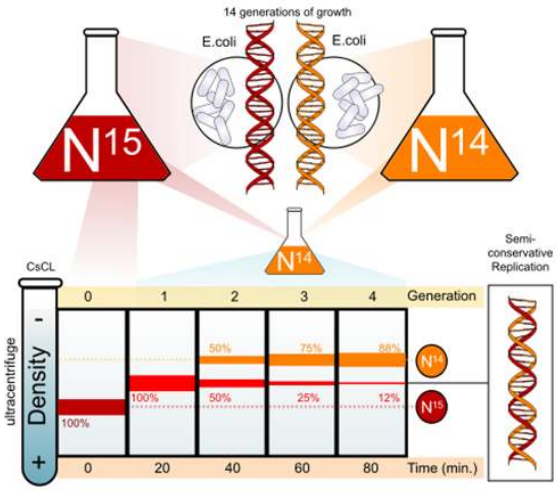


Figure **10.1.7**: The positions of the ^{14}N and ^{15}N containing DNA in the density gradient tube on the left. (Wikipedia-LadyofHats-CC:AN)

These experiments, published in 1958, are a wonderful example of how science works. Researchers start with three clearly defined models (hypotheses). These models were tested, and two (conservative and dispersive) were found to be inconsistent with the observations and thus disproven. The third hypothesis, semi-conservative, was consistent with the observations and thereby supported and accepted as mechanism of DNA replication. Note, however, this is not “proof” of the model, just strong evidence for it; hypotheses are not “proven”, only disproven or supported.

RNA and protein

While DNA is the genetic material for the vast majority of organisms, there are some viruses that use **RNA as their genetic material**. These viruses can be either single or double stranded and include SARS, influenza, hepatitis C and polio, as well as the retroviruses like HIV-AIDS. Typically there is DNA used at some stage in their life cycle to replicate their RNA genome.

Also, the case of **Prion** infections agents transmit characteristics via only a protein (no nucleic acid present). Prions infect by transmitting a misfolded protein state from one aberrant protein molecule to a normally folded molecule. These agents are responsible for bovine spongiform encephalopathy (BSE, also known as "mad cow disease") in cattle and deer and Creutzfeldt–Jakob disease (CJD) in humans. All known prion diseases act by altering the structure of the brain or other neural tissue and all are currently untreatable and ultimately fatal.

This page titled 10.1: DNA is the Genetic Material is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **1.2: DNA is the Genetic Material** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.2: The Structure of DNA

The experiments outlined in the previous sections proved that DNA was the genetic material, but very little was known about its structure at the time.

Chargaff's Rules

When **Watson and Crick** set out in the 1940's to determine the structure of DNA, it was already known that DNA is made up of a series four different types of molecules, called bases or nucleotides: adenine (A), cytosine (C), thymine (T), guanine (G). Watson and Crick also knew of **Chargaff's Rules**, which were a set of observations about the relative amount of each nucleotide that was present in almost any extract of DNA. Chargaff had observed that for any given species, the abundance of A was the same as T, and G was the same as C. This was essential to Watson & Crick's model.

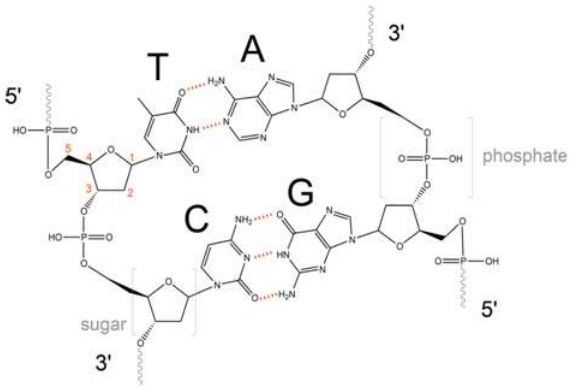


Figure **10.2.8**: Chemical structure of two pairs of nucleotides in a fragment of double-stranded DNA. Sugar, phosphate, and bases A,C,G,T are labeled. Hydrogen bonds between bases on opposite strands are shown by dashed lines. Note that the G-C pair has more hydrogen bonds than A-T. The numbering of carbons within sugars is indicated by red numbers. Based on this numbering the polarity of each strand is indicated by the labels 5' and 3'. (Wikipedia-M. Strock-GFDL)

The Double Helix

Using proportional metal models of the individual nucleotides, Watson and Crick deduced a structure for DNA that was consistent with Chargaff's Rules and with x-ray crystallography data that was obtained (with some controversy) from another researcher named Rosalind Franklin. In Watson and Crick's famous **double helix**, each of the two strands contains DNA bases connected through covalent bonds to a sugar-phosphate backbone (Fig 1.8, 1.9). Because one side of each sugar molecule is always connected to the opposite side of the next sugar molecule, each strand of DNA has polarity: these are called the 5' (5-prime) end and the 3' (3-prime) end, in accordance with the nomenclature of the carbons in the sugars. The two strands of the double helix run in **anti-parallel** (i.e. opposite) directions, with the 5' end of one strand adjacent to the 3' end of the other strand. The double helix has a **right-handed** twist, (rather than the left-handed twist that is often represented incorrectly in popular media). The DNA bases extend from the backbone towards the center of the helix, with a pair of bases from each strand forming hydrogen bonds that help to hold the two strands together. Under most conditions, the two strands are slightly offset, which creates a major groove on one face of the double helix, and a minor groove on the other. Because of the structure of the bases, A can only form hydrogen bonds with T, and G can only form hydrogen bonds with C (remember Chargaff's Rules). Each strand is therefore said to be complementary to the other, and so each strand also contains enough information to act as a template for the synthesis of the other. This complementary redundancy is important in DNA replication and repair.

How can this molecule, DNA, contain the genetic material?

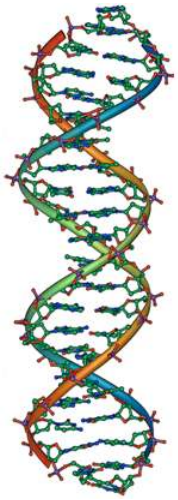


Figure **10.2.9**: DNA double helix structure. (Original-unknown-?)

This page titled 10.2: The Structure of DNA is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **1.3: The Structure of DNA** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

10.3: DNA- The Genetic Material

Learning Objectives

- Detail the molecular nature of a gene.
- Recognize when and where replication of the DNA that is packaged in an organism's chromosome is happening for growth, maintenance, and reproduction.
- Describe how chromosome replication speed can be increased with multiple origins of replication and bidirectional replication.
- Explain the role of each of the primary DNA replicating enzymes in conducting replication of a chromosome. Include the role of these enzyme in ensuring the accuracy of the replication process.
- Predict how the semiconservative replication of a double stranded DNA will happen given the structure of the molecule and the specific function of the DNA replicating enzymes.

Introduction

The genetic (hereditary) material for all living things is composed of DNA (deoxyribonucleic acid). The structure of DNA must enable this substance to store coded information that control the biological function of cells. The genetic material transmits this hereditary information in a stable form for the cell and organism through accurate **replication** of DNA. Although DNA is capable of change (we will discuss how this happens when we talk about DNA mutations), the replication process ensures high accuracy in copying the genetic information so that all progeny cells receive the same information. The DNA from all chromosomes in a human cell would be more than 6.5 feet long! Therefore, to fit inside the nucleus of a cell, DNA is packaged into chromosomes.

Read this article about Watson and Crick for more insight on the experimental facts discovered by chemists and biologist which contributed to the determination of the structure of DNA.

Chemical Structure of DNA Subunits

DNA is a polymer made of nucleotide subunits. A **nucleotide** consists of 3 chemical groups; a sugar, a phosphate and a nitrogenous base (Figure 1). In the case of DNA, the sugar is deoxyribose.

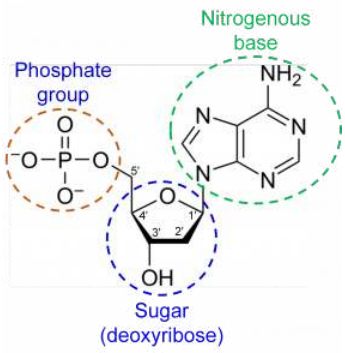


Figure 1. A nucleotide consists of three chemical groups. Image adapted from Wikipedia.

There are 4 different nucleotides in DNA that contain 4 different **bases** (Figure 2). These bases are named adenosine (A), cytosine (C), guanine (G) and thymine (T).

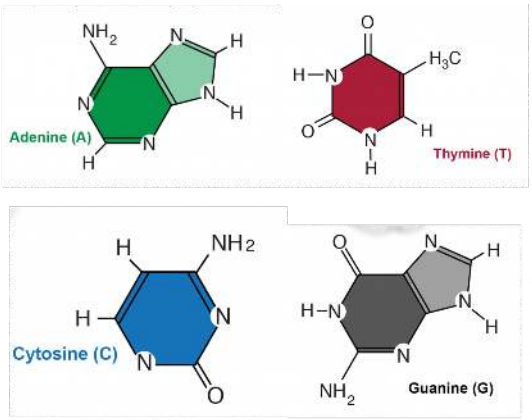


Figure 2. Adenine, Thymine, Cytosine, and Guanine are the nitrogenous bases of DNA. Adapted from NIH-NHGRI.

Complementary, antiparallel DNA strands

DNA codes information by ordering the sequence of the A,T,G and C nucleotides in a long polymer. The nucleotides are connected by phosphate bonds to form a strand. Two strands form a double stranded molecule (Figure 3). Note that in Figure 1, the carbon atoms on the deoxyribose sugar are numbered. The phosphate bond connects carbon #3 (the 3 prime or 3′ carbon) of one sugar to the 5′ carbon of the next. This gives a DNA strand directionality because the 5′ carbon faces one way and the 3′ carbon faces the other.

The bases on one strand form hydrogen bonds with the bases on the other. This is called "**base pairing**" and it is very specific, A only pairs with T and C only pairs with G. Thus, the sequences on the two strands of DNA are "**complementary**"; whenever there is an "A" on one strand, there is a "T" in the corresponding position of the complementary strand.

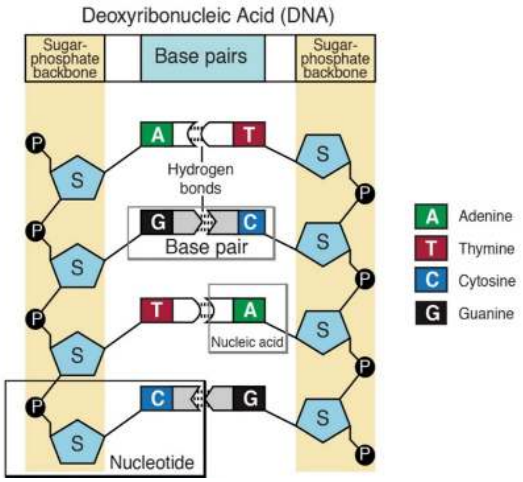


Figure 3. DNA is a double-stranded molecule. Illustration by NIH-NHGRI.

In addition to containing complementary nucleotide sequences, the strands are in opposite orientations. The 5′ end of one strand is oriented toward the 3′ end of the other. For this reason, the strands are referred to as "**antiparallel**".

Double helix

The final feature of the molecular structure is that DNA assumes a helical conformation (Figure 4). This is the most stable configuration that accommodates all the molecular structures and chemical bonds that make up DNA.

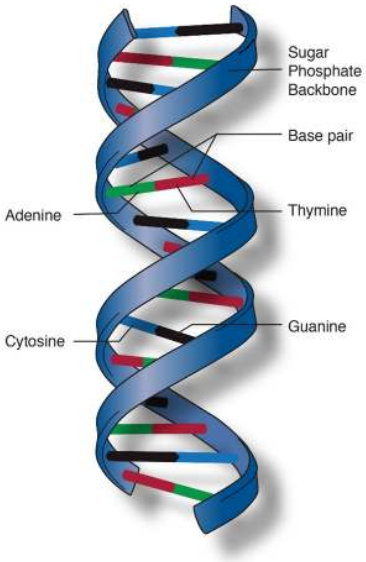


Figure 4. DNA is a double helix. Illustration by NIH-NHGRI.

Activity

Given the following sequence of a DNA strand, predict the sequence and orientation of the complementary strand.

5′-G-A-C-C-G-T-A-A-T-C-G-C-3′

Show Answer

Answer: 3′-C-T-G-G-C-A-T-T-A-G-C-G-5′

Packaging of DNA into chromatin and chromosomes

Chromosomes are long double stranded DNA molecules. Over 150,000,000 nucleotide pairs make up the human X chromosome. The complete replication and orderly transfer of something this large requires the chromosome to be packaged for stability and organization which creates genetic stability in the cells of a species. Every cell in a diploid organism contains two copies (2n) of every chromosome present in that organism (Figure 5). For example, humans have 46 chromosomes in their body, 23 were inherited from the father and 23 from the mother. **Gametes**, the reproductive cells of an organism, (egg or sperm), have only one set (1n) of chromosomes. When the two gametes unite, they form a living embryo with two sets of genetic information. Therefore, we actually have two copies of the genetic information for each trait. Sometimes one copy controls **trait expression**, and other times both copies influence a trait. As a result, the offspring will have characteristics of both the mother and the father.

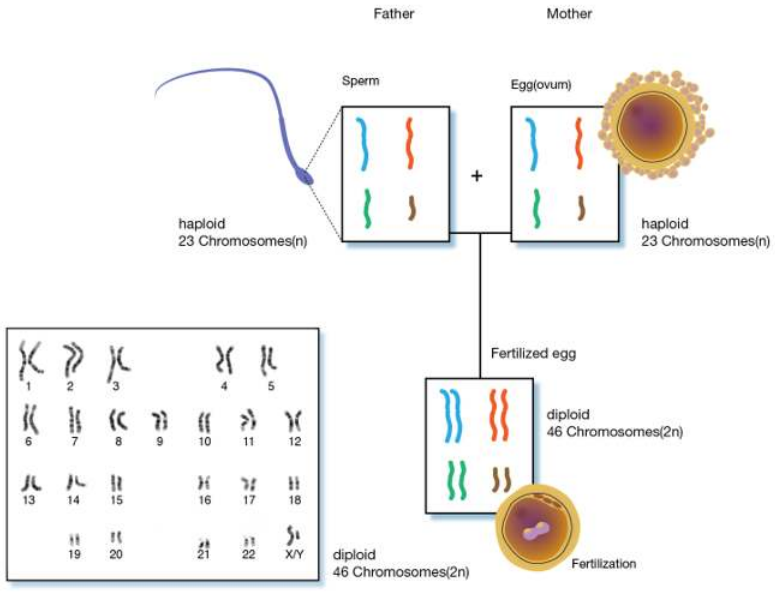


Figure 5. Human non-sex cells are diploid (2n) and contain 46 chromosomes. On the other hand, human sex cells or gametes (sperm and egg) are haploid (n) and contain single set of 23 chromosomes. Image from NIH-NHGRI.

Inside cell nuclei, DNA is assembled, or **packaged**, into material called chromatin, which is composed of both DNA and proteins (Figure 6). Chromatin functions to protect and regulate DNA, as well as to efficiently store the very long DNA molecules that make up chromosomes within the limited space of the nucleus.

DNA from higher eukaryotic organisms is about 109 to 1010 bp. Human and maize chromosomal DNA is about 109 bp, which is equivalent to 1.8 m if all chromosomal DNA were stretched end to end in a linear manner. The diameter of the nucleus is only about 4-6 μm making it a challenge to fit chromosomal DNA inside the cell nucleus. This is accomplished by the ability of DNA to assume a very condensed structure to fit inside the nucleus. Therefore, the organization of eukaryotic DNA into chromatin is an important aspect of DNA packaging.

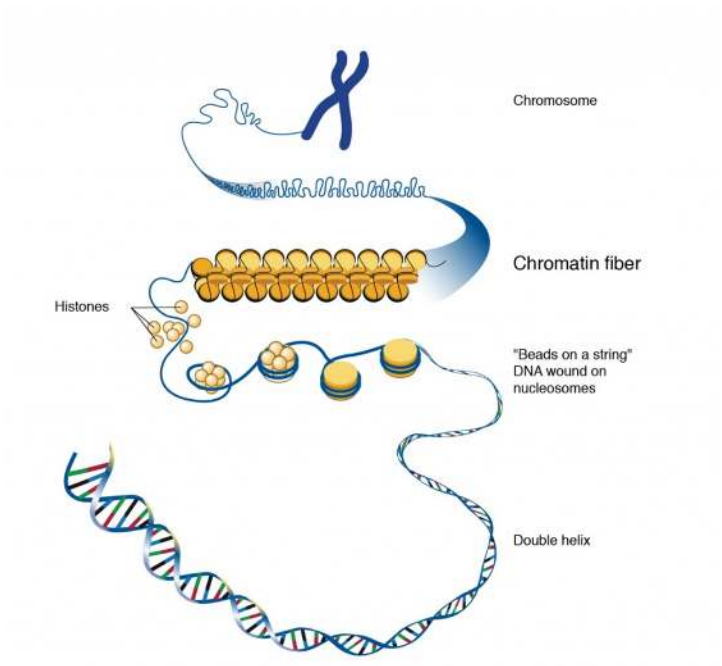


Figure 6. DNA is packed into chromatin and chromosomes. Illustration by NIH-NHGRI.

Nucleosomes and histones

The ordered coiling of chromosomal DNA around a histone protein core forms the **chromatin**. Chromatin is made up of **nucleosomes** (Figure 6) which represent the association of chromosomal DNA with histone proteins. A nucleosome is made up of about 145-147 base pairs of DNA coiled around each histone octamer, for about two complete turns. A histone octamer consists of two copies of each core histone. The total mass of the histones in the nucleus approaches that of the DNA – 2 molecules of each core histone to approximately 200 bp of DNA.

Higher order structures

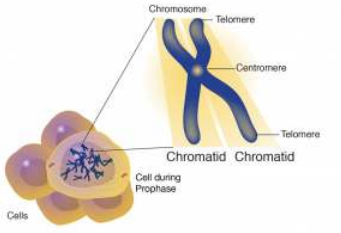


Figure 7. Chromosomes have centromeres and telomeres. Illustration by NIH-NHGRI.

If the higher order chromatin structure is disrupted, electron microscopy reveals the appearance of “beads on a string” with a diameter of about 10 nm. The “beads” represent the DNA wrapped around histones. Nucleosome formation results in a DNA fiber that is about 10 nm and a packaging ratio of about 7. The higher order chromatin structure results when the 10 nm fiber is coiled into a solenoid. The result of nucleosome coiling is a chromatin fiber of 30 nm that is observed by electron microscopy.

Centromeres

The centromere is a chromosomal region that controls chromosome segregation at mitosis and meiosis (Figure 7). Centromeres connect to microtubules of the spindle apparatus, which directs their movement to opposite poles (daughter cell nuclei) during cell division. Scientists were successful in isolating these centromeric sequences from yeast and engineered brewer’s yeast (*Saccharomyces cerevisiae*) plasmids that were able to replicate like the chromosomes. Through these efforts scientists were able to point centromeric function to a DNA stretch of about 120 bp that was resistant to DNAse and bound to a single microtubule.

Telomeres

The telomere lies at the end of the chromosome and confers stability by “sealing” the end of a chromosome (Figure 7). Telomeres consist of long series of short repeated DNA sequence that occurs in tandem arrays and are added to the end of the chromosome during DNA replication by an enzyme called telomerase. In most plant species the sequence **TTTAGGG** constitutes a conserved telomere motif.

The knowledge about the structure and function of centromeres and telomeres has had a profound impact in molecular genetics. For example, yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) have proven useful in the physical mapping of plant genomes requiring the cloning and multiplying large (>100 kb) DNA fragments in yeast or bacterial cells.

DNA Replication

In 1953, J.D. Watson and F.H.C. Crick published a note in NATURE; one of the world’s most widely read research journals. The note cited just six references.

Watson and Crick’s writing had limited chemistry details given the title of their note, “A Structure for Deoxyribonucleic Acid”. A later article would share more of the chemistry behind their proposed double helix structure for Deoxyribonucleic Acid (DNA). This note in Nature was written for a broader audience, including biologists who recognized that the structure of DNA was a prerequisite for understanding the function of the genetic material. They included a diagram to allow readers to visualize this unique double helix structure. To emphasize the structure determines function importance, Watson and Crick included this statement in their note…

> *“It has not escaped our notice that the specific pairing we have postulate immediately suggests a possible copying mechanism for the genetic material.”*

Accurate replication prior to cell division is one of the three key functions of the genetic material. While later experiments would reveal the details of how cells replicate DNA, Watson and Crick were obviously impressed that the proposed double helix structure would fit with the concept that living cells are needed to produce more living cells and the transmission of a complete set of biological instructions. The genetic material must have a structure that lends itself to being copied. Using old to help build new leads to a proposed semiconservative model for replicating the double stranded DNA molecule (Figure 8).

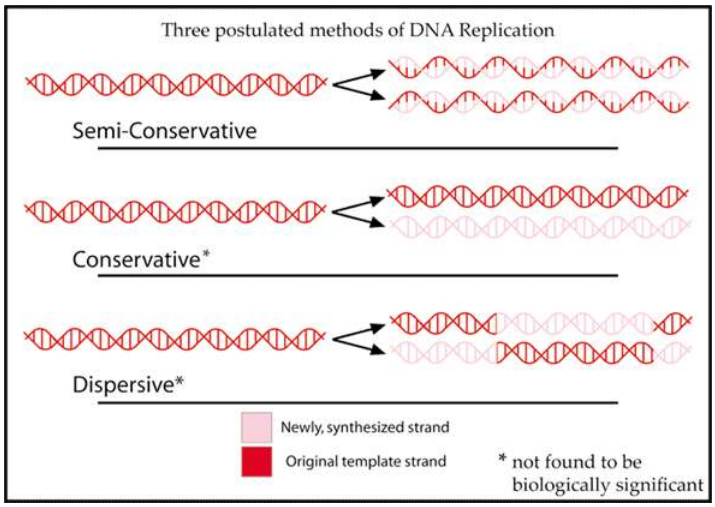


Figure 8. Watson and Crick predicted that the double helix molecule could be replicated in a living cell by the semi-conservative method. Each strand of the old double stranded molecule is read to build a new strand. Early experiments on replication confirmed this prediction. Image Source: Adenosine, CC BY-SA 2.5, via Wikimedia Commons.

Prior to both mitosis and meiosis cell division in multicellular organisms, and cell division in prokaryotes, the DNA inside the cell must be replicated. (see lesson Mitosis and Meiosis). This replication process generates the genetic information needed for either two cells, genetically identical with the original cell or in sexually reproducing organisms, four gamete cells with half of the original cell's genetic information.

When and Where:

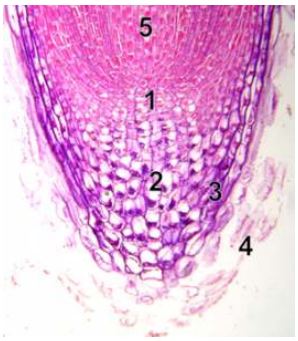


Figure 9. Plant root tip. Image Source: SuperManu, CC BY-SA 2.5, via Wikimedia Commons.

DNA replication will happen before cells divide. In hours old embryos that are developing into multicellular organisms, all the cells are replicating their DNA and dividing into new cells. Hours or days later, multicellular organisms have developed specific meristem regions where new cell production occurs. A visual example of meristem cells where DNA replication is needed is the tip of a plant root (Figure 9). New cells made at the root tip allow the root to grow. The root must also replace the cells that are damaged as the growing root moves through soil. In Figure 8, the meristem region (1) contains meristem cells that are replicating their chromosomes, then dividing into two identical cells by mitosis. Cells on the tip part of the meristem specialize in root cap cells (2,3) with the job of protecting the meristem. The fate of these cells is to die (4), and they need to be replaced by new cells made in the meristem. Cells made on the other half of the meristem (5) can elongate and eventually specialize. These root cells will then function for the plant, all season for annual plants, years for perennial plants. Replication of DNA will only happen in the meristem cells.

Speed of chromosome replication

Fifty-two years after Watson and Crick's Nature note, a group of about 50 molecular geneticists published the complete DNA sequence of the human X chromosome (Figure 5). This discovery revealed that X chromosome is one double-stranded DNA molecule that is about 155 million nucleotide pairs long. Every time a human cell divides, one or two of these 155,000,000 nucleotides must be replicated. This must be done with speed and accuracy.

The main DNA replication enzyme (DNA polII, see below) works fast. Estimates of its replicating speed are around 750 nucleotides per second.

Activity

Let us do the math.

155,000,000 nucleotides / chromosome X 1 second / 750 nucleotides

= 206,666 seconds per X chromosome

Let us convert these seconds to hours.

1 hour / 360 seconds X 206,666 seconds/chromosome = 547 hours per chromosome

That means one replication enzyme would take 547 hours or about 24 days to replicate an X chromosome. What are the biological ramifications of this 24 days? If you cut yourself, it would take 24 days to replicate the chromosomes before the skin cells could divide. The formation of new cells needed for healing cuts in our skin would keep us in a wounded state.

To speed up the chromosome replication process, two tactics are used by living cells. Both tactics are shown in Figure 10. First, replication will not start on just one end of the chromosome. Instead, there are hundreds of origins of replication (ori) along the chromosome. In addition, the replication process moves bidirectionally from the ori. This bidirectional replication creates forks of replication that move toward each other as the chromosome is replicated.

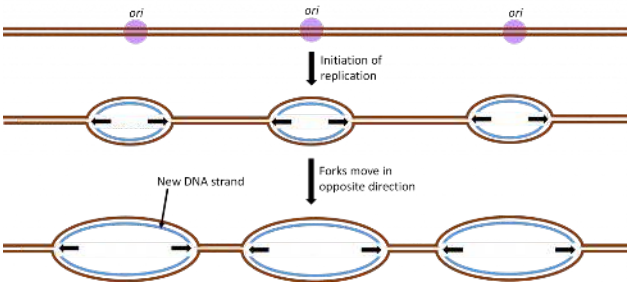


Figure 10. Origin of replication and replication forks. Image by Walter Suza.

As a result of these two tactics, a large chromosome can be replicated in hours rather than weeks and cell division can happen quickly when it is needed for growth or cell replacement.

The Chromosome Replicating Team

Like all processes in living cells, a collection of specific proteins works together to perform functions which control and complete chromosome replication. All the proteins involved in DNA replication make and break bonds, so they are enzymes. The function of the main DNA replicating enzymes is described below.

Five of the main DNA replicating enzymes

- **Helicase**: This is one of the enzymes that unwinds the double stranded DNA molecule. The unwound single strands no longer hydrogen bond to their complementary strand but the sugar-phosphate bonds remain intact.
- **Topoisomerase**: The unwinding work of Helicase creates tension in the double stranded DNA ahead of the replication fork. Topoisomerase relieves this tension by catalyzing a series of sugar phosphate bond breaking and making steps. Without this tension relief, the Double stranded DNA could break and the intact chromosome cannot finish replicating.
- **DNA Polymerase III (DNA pol III)**: This is the main DNA synthesizing enzyme. The enzyme reads the single strand as a template and places in the complementary deoxyribonucleotide. DNA pol III reads the template in the 3′ to 5′ direction and

builds the new strand in the 5′ to 3′ direction, adding the next nucleotide to the 3′ end by catalyzing sugar-phosphate bonding. DNA pol III illustrates the specificity of enzymes several ways; it reads and proofreads the placement of new nucleotides to insure accurate replication. It can only read and build in one direction. DNA pol III can only add nucleotides to a free 3′ end. This last specification means that DNA pol III cannot start the replication process on the single stranded template. Another enzyme needs to be part of the in vivo replication team to prime the work of DNA pol III.

- **Primase**: Biochemists named this enzyme to describe its role in starting or priming the replication process. DNA (or RNA) primase is a special RNA polymerase. The enzyme reads the single stranded DNA template 3′ to 5′ and adds ribonucleic acid (RNA) nucleotides in the 5′ to 3′ direction. Once a few hundred RNA nucleotides are added, primase falls off the template strand and leaves the 3′ end that DNA Pol III needs to continue the process.
- **DNA Polymerase I (DNA pol I)**: The DNA polymerase specializes in the removal of the RNA primers and replacing them with DNA nucleotides. The enzyme works with the same 3′ to 5′ reading and 5′ to 3′ building.
- **DNA Ligase**: After the five enzymes described above have completed their work, some sugar phosphate bonds need to be made to complete the double stranded molecule. DNA ligase has this backbone bond sealing assignment.

Now that we have introduced the replication enzymes, we can describe the step-by-step action of this team. We will focus on the action that happens at one origin of replication.

Step 1: Initiating replication at the Ori

Helicase enzymes bind to the ori sequences and start to unwind the double stranded DNA. This establishes two replication forks (one shown in Figure 10a) and a helicase enzyme will be working to unwind the DNA at each fork as the forks move away from each other. The unwinding exposes single strands of DNA that are the template for building a new strand. Topoisomerase can be seen working to relieve tension ahead of the replication fork in Figure 10a.

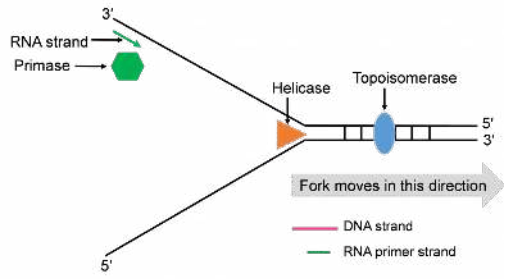


Figure 10a. Helicase and Topoisomerase work together to expose single strand DNA templates. Primase adds RNA nucleotides, reading the old DNA single strand template 3′ to 5′. Image by Walter Suza.

Step 2: Priming DNA with RNA

The process of priming with primase and then replicating with DNA pol III happens at both replication forks (Figure 10a and b). DNA pol III enzymes can replicate new strands as fast as the DNA is unwound at each replication fork. The unwinding creates two template strands at each fork so primase enzyme must work to prime both old strands.

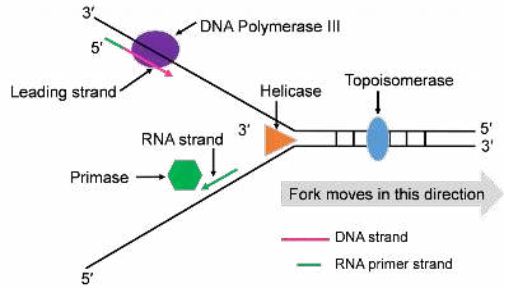


Figure 10b. DNA Polymerase III can add to the 3′ end of the RNA primer and start to read the template 3′ to 5′. Primase can prime the other template strand at this replication fork. Image by Walter Suza.

Step 3: Synthesis of leading and lagging strands

The double strands of DNA are anti-parallel and 'run' in opposite directions and DNA pol III can only work in one direction (Figure 10c). Thus, there is continuous and discontinuous replication happening at each replication fork. The strand that can be replicated as helicase unwinds is the leading strand. The DNA pol III must replicate the other strand in the opposite direction so this strand will lag behind. (Figure 10c).

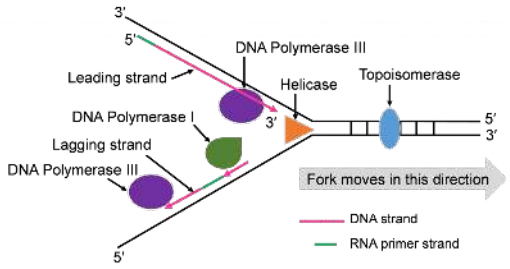


Figure 10c. Both leading and lagging strands are being made at this replication fork. DNA polymerase I removes RNA primer nucleotides and replaces them with DNA. Image by Walter Suza.

There is more priming needed to replicate the lagging strand and the priming leaves RNA-DNA hybrid stretches in the new double stranded molecule. This means there is work for the enzyme DNA pol I. This enzyme will be part of the removal of the RNA nucleotide primers, and replacement with DNA nucleotides (Figure 10c and d).

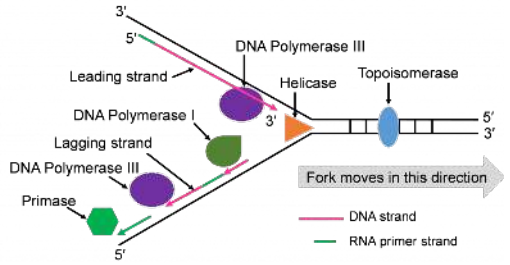


Figure 10d. Lagging strand synthesis must be primed multiple times with primase enzyme creating discontinuous replication. Leading strand just needs to be primed one time. Image by Walter Suza.

Both replication forks have a leading and a lagging strand of new DNA. The DNA nucleotide stretches between the red RNA primers would be the Okazaki fragments.

One of the first scientists to provide evidence to support this model of DNA replication was the team of Reiji and Tsuneko Okazaki in the 1960s. They used radioactive nucleotides to track the newly replicating fragments and found that many of the fragments were short. This fits with the discontinuous model.

As the four enzymes just described attend to their specific roles in replication, two double stranded DNA molecules with an old strand and a new strand are built. If the DNA pol III makes a mistake and adds the wrong, non-complementing nucleotide, the enzyme can proof read its work and replace these replication mistakes. When replication is perfect, the two double stranded DNA molecules will have identical sequences.

Step 4: Ligation of Okazaki fragments

The fifth enzyme that has a final role in completing replication is DNA ligase (not shown in Figure 10d and e). This enzyme will seal the sugar phosphate bond between the last replacement nucleotide added by DNA pol I and the first nucleotide that had been added after priming by DNA pol III. With this bond sealing work, the replication process is complete.

The two double stranded DNA molecules that get made will have identical sequences unless rare mistakes happen as the enzymes conduct their work.

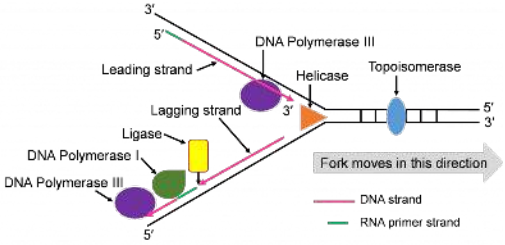


Figure 10e. Ligase must form a sugar-phosphate bond between the last nucleotide placed by DNA pol I and the first nucleotide placed by DNA pol III. Image by Walter Suza.

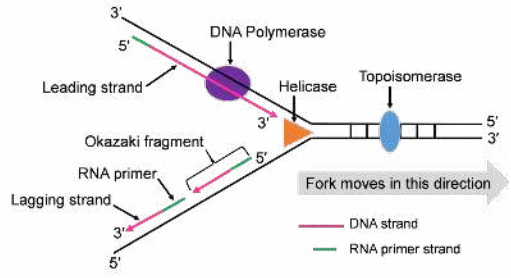


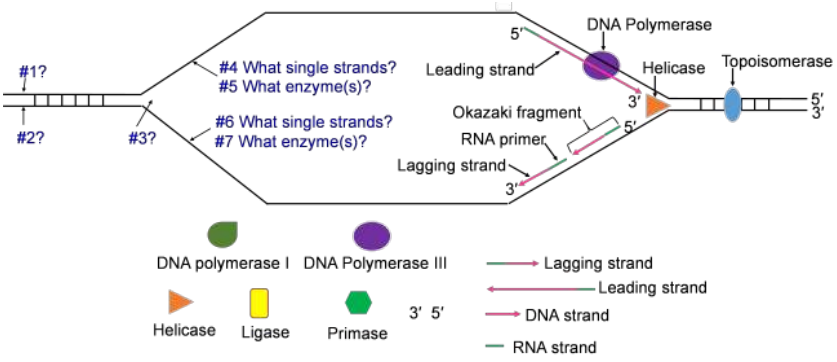
Figure 10f. Short fragments made by discontinuous replication were discovered by the Okazaki team and are called the Okazaki fragments. Image by Walter Suza.

Lesson Summary:

DNA is composed of nucleotides. Nucleotides are connected together by phosphate bonds to form a strand. The bases on one strand form hydrogen bonds with the bases on the other to form a double stranded molecule. The final feature of the molecular structure is that DNA assumes a helical conformation. To fit inside the nucleus, DNA assumes a very condensed structure. Therefore, chromosomal DNA is coiled around a histone protein core to form chromatin. The tight packaging of DNA in chromatin must be modified to allow DNA replication and transcription. In the process of replication, the two DNA strands separate and act as templates for the synthesis of complementary daughter strands.

Learning Activities

Place the single strands, replication enzymes, and 3′ / 5′ labels in the appropriate places (#1-#7)



This page titled 10.3: DNA- The Genetic Material is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Philip Becraft, Patricia Hain, & Patricia Hain (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.2: DNA- The Genetic Material** by Donald Lee, Marjorie Hanneman, Patricia Hain, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

11: Amplifying and Manipulating DNA Fragments

11.1: Prelude to Molecular Genetics
11.2: Isolating Genomic DNA
11.3: Isolating or Detecting a Specific Sequence by PCR
11.4: PCR and Gel Electrophoresis
11.5: Origins of Molecular Polymorphisms
11.6: Classification and Detection of Molecular Markers
11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
11.8: Make and Screen a cDNA Library
11.9: DNA Sequencing
11.10: Genomic Libraries
11.11: The Polymerase Chain Reaction (PCR)

11: Amplifying and Manipulating DNA Fragments is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

11.1: Prelude to Molecular Genetics

Genetics is the study of the inheritance and variation of biological traits. We have previously noted that it is possible to conduct genetic research without directly studying DNA. Indeed some of the greatest geneticists had no special knowledge of DNA at all, but relied instead on analysis of phenotypes, inheritance patterns, and their ratios in carefully designed crosses. Today, **classical genetics** is often complemented by **molecular biology**, to give **molecular genetics**, which involves the study of DNA and other **macromolecules** that have been isolated from an organism. Usually, molecular genetics experiments involve some combination of techniques to isolate and analyze the DNA or RNA transcribed from a particular gene. In some cases, the DNA may be subsequently manipulated by mutation or by recombination with other DNA fragments. Techniques of molecular genetics have wide application in many fields of biology, as well as forensics, biotechnology, and medicine.

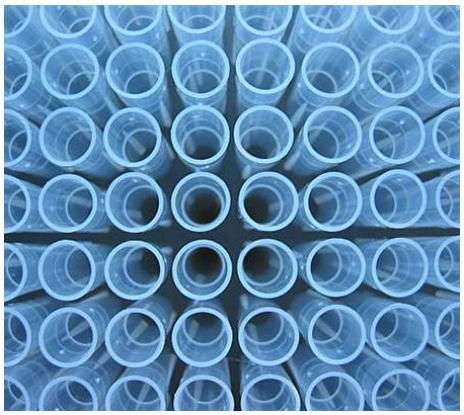


Figure* 11.1.1*: *Disposable tips for a pipette are used to distribute microliter volumes of liquid in molecular biology. (Flickr-esther ase-CC:ANS)*

This page titled 11.1: Prelude to Molecular Genetics is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.1: Prelude to Molecular Genetics** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.2: Isolating Genomic DNA

DNA purification strategies rely on the chemical properties of DNA that distinguish it from other molecules in the cell, namely that it is a very long, negatively charged molecule. To extract purified DNA from a tissue sample, cells are broken open by grinding or **lysing** in a solution that contains chemicals that protect the DNA while disrupting other components of the cell (Figure **11.2.2**). These chemicals may include **detergents**, which dissolve lipid membranes and denature proteins. A cation such as Na^+ helps to stabilize the negatively charged DNA and separate it from proteins such as histones. A chelating agent, such as EDTA, is added to protect DNA by sequestering Mg^{2+} ions, which can otherwise serve as a necessary co-factor for **nucleases** (enzymes that digest DNA). As a result, free, double-stranded DNA molecules are released from the chromatin into the extraction buffer, which also contains proteins and all other cellular components. (The basics of this procedure can be done with household chemicals and are presented on YouTube.)

The free DNA molecules are subsequently isolated by one of several methods. Commonly, proteins are removed by adjusting the salt concentration so they precipitate. The **supernatant**, which contains DNA and other, smaller metabolites, is then mixed with ethanol, which causes the DNA to precipitate. A small **pellet** of DNA can be collected by centrifugation, and after removal of the ethanol, the DNA pellet can be dissolved in water (usually with a small amount of EDTA and a pH buffer) for the use in other reactions. Note that this process has purified all of the DNA from a tissue sample; if we want to further isolate a specific gene or DNA fragment, we must use additional techniques, such as PCR.

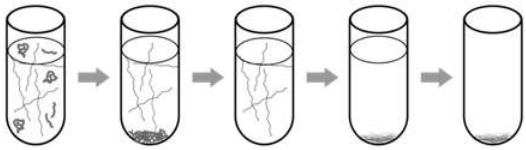


Figure **11.2.2**: Extraction of DNA from a mixture of solubilized cellular components by successive precipitations. Proteins are precipitated, then DNA (in the supernatant) is precipitated in ethanol, leaving a pellet of DNA. (Original-Deyholos-CC:AN)

This page titled 11.2: Isolating Genomic DNA is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.2: Isolating Genomic DNA** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.3: Isolating or Detecting a Specific Sequence by PCR

Components of the PCR Reaction

The **Polymerase Chain Reaction (PCR)** is a method of DNA replication that is performed in a test tube (i.e. *in vitro*). Here "polymerase" refers to a DNA polymerase enzyme extracted and purified from bacteria, and "chain reaction" refers to the ability of this technique produce millions of copies of a DNA molecule, by using each newly replicated double helix as a template to synthesize two new DNA double helices. PCR is therefore a very efficient method of amplifying DNA.

Besides its ability to make large amounts of DNA, there is a second characteristic of PCR that makes it extremely useful. Recall that most DNA polymerases can only add nucleotides to the end of an existing strand of DNA, and therefore require a **primer** to initiate the process of replication. For PCR, chemically synthesized primers of about 20 nucleotides are used. In an ideal PCR, primers only hybridize to their exact complementary sequence on the template strand (Figure **11.3.3**).



Figure **11.3.3**: The primer-template duplex at the top part of the figure is perfectly matched, and will be stable at a higher temperature than the duplex in the bottom part of the figure, which contains many mismatches and therefore fewer hydrogen bonds. If the annealing temperature is sufficiently high, only the perfectly matched primer will be able to initiate extension (grey arrow) from this site on the template. (Original-Deyholos-CC:AN)

The experimenter can therefore control exactly what region of a DNA template is amplified by controlling the sequence of the primers used in the reaction.

To conduct a PCR amplification, an experimenter combines in a small, thin-walled tube (Figure **11.3.4**), all of the necessary components for DNA replication, including DNA polymerase and solutions containing nucleotides (dATP, dCTP, dGTP, dTTP), a DNA template, DNA primers, a pH buffer, and ions (e.g. Mg^{2+}) required by the polymerase. Successful PCR reactions have been conducted using only a single DNA molecule as a template, but in practice, most PCR reactions contain many thousands of template molecules. The template DNA (e.g. total genomic DNA) has usually already been purified from cells or tissues using the techniques described above. However, in some situations it is possible to put whole cells directly in a PCR reaction for use as a template.



Figure **11.3.4**: A strip of *PCR tubes (Wikipedia-madprime-GFDL)*

An essential aspect of PCR is **thermal-cycling**, meaning the exposure of the reaction to a series of precisely defined temperatures (Figure **11.3.5**). The reaction mixture is first heated to 95°C. This causes the hydrogen bonds between the strands of the template DNA molecules to melt, or **denature**. This produces two single-stranded DNA molecules from each double helix (Figure **11.3.6**). In the next step (**annealing**), the mixture is cooled to 45-65°C. The exact temperature depends on the primer sequence used and the objectives of the experiment. This allows the formation of double stranded helices between complementary DNA molecules, including the annealing of primers to the template. In the final step (**extension**) the mixture is heated to 72°C. This is the temperature at which the particular DNA polymerase used in PCR is most active. During extension, the new DNA strand is synthesized, starting from the 3' end of the primer, along the length of the template strand. The entire PCR process is very quick, with each temperature phase usually lasting 30 seconds or less. Each cycle of three temperatures (denaturation, annealing, extension) is usually repeated about 30 times, amplifying the target region approximately 2^{30}-fold. Notice from the figure that most of the newly synthesized strands in PCR begin and end with sequences either identical to or complementary to the primer sequences; although a few strands are longer than this, they are in such a small minority that they can almost always be ignored.

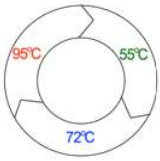


Figure **11.3.5**: Example of a thermalcycle, in which the annealing temperature is 55°C. (Original-Deyholos-CC:AN)

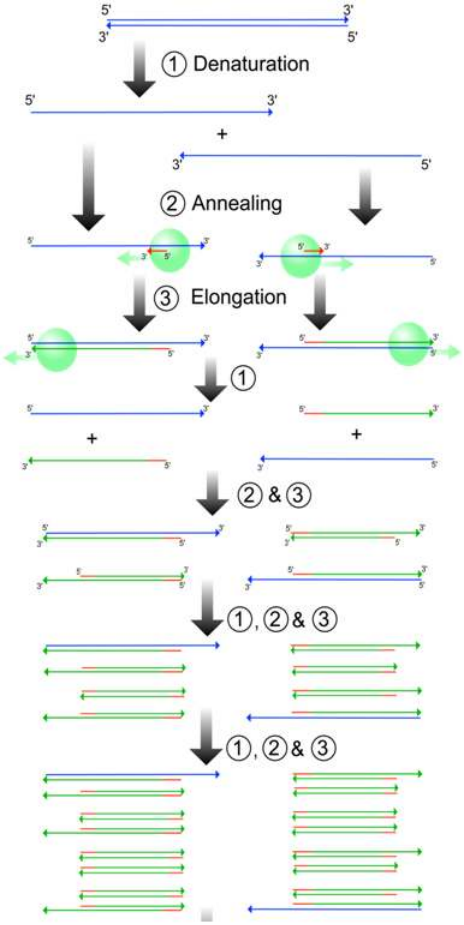


Figure* 11.3.6*: *PCR with the three phases of the thermalcycle numbered. The template strand (blue) is replicated from primers (red), with newly synthesized strands in green. The green strands flanked by two primer binding sites will increase in abundance exponentially through successive PCR cycles. (Wikipedia-madprime-GFDL)*

The earliest PCR reactions used a polymerase from *E. coli*. Because the high temperature of the denaturation step destroyed the enzyme, new polymerase had to be added after each cycle. To overcome this, researchers identified **thermostable** DNA polymerases such as **Taq DNA pol**, from *Thermus acquaticus*, a thermophilic bacterium that lives in hot springs. Taq, and similar thermostable polymerases from other hot environments, are able to remain functional in the repeated cycles of amplification. Taq polymerase cannot usually amplify fragments longer than about 3kbp, but under some specialized conditions, PCR can amplify fragments up to approximately 10kbp. Other polymerases, either by themselves or in combination with Taq, are used to increase the length of amplified fragments or to increase the fidelity of the replication.

After completion of the thermalcycling (amplification), an aliquot from the PCR reaction is usually loaded onto an **electrophoretic gel** (described below) to determine whether a DNA fragment of the expected length was successfully amplified or not. Usually, the original template DNA will be so dilute that it will not be visible on the gel, only the amplified PCR product. The presence of a sharp band of the expected length indicates that PCR was able to amplify its target. If the purpose of the PCR was to test for the presence of a particular template sequence, this is the end of the experiment. Otherwise, the remaining PCR product can be used as starting material for a variety of other techniques such as sequencing or cloning.

An Application of PCR: the StarLink Affair

PCR is very sensitive (meaning it can amplify very small starting amounts of DNA), and specific (meaning it can amplify only the target sequence from a mixture of many DNA sequences). This made PCR the perfect tool to test whether genetically modified corn was present in consumer products on supermarket shelves. Although currently (2013) 85% of corn in the United States is genetically modified, and contains genes that government regulators have approved for human consumption, back in 2000, environmental groups showed that a strain of genetically modified corn, which had only been approved for use as animal feed, had been mixed in with corn used in producing human food, like taco shells.

To do this, the groups purchased taco shells from stores in the Washington DC area, extracted DNA from the taco shells and used it as a template in a PCR reaction with primers specific to the unauthorized gene (*Cry9C*). Their suspicions were confirmed when they ran this PCR product on an agarose gel and saw a band of expected size. The PCR test was able to detect one transgenic kernel in a whole bushel of corn (1 per 100,000). The company (Aventis) that sold the transgenic seed to farmers had to pay for the destruction of large amounts of corn, and was the target of a class action law-suit by angry consumers who claimed they had been made sick by the taco shells. While no legitimate cases of harm were ever proven, and the plaintiffs were awarded $9 million, of which $3 million went to the legal fees, and the remainder of the judgment went to the consumers in the form of coupons for taco shells. The affair damaged the company, and exposed a weakness in the way the genetically modified crops were handled in the United States at the time.

This page titled 11.3: Isolating or Detecting a Specific Sequence by PCR is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.3: Isolating or Detecting a Specific Sequence by PCR** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.4: PCR and Gel Electrophoresis

Learning Objectives

At the completion of this PCR lesson, learners will be able to:

- List the 5 chemical components of a PCR reaction and describe their roles.
- List the functions of the 3 temperature cycles which are repeated during a PCR reaction.
- Describe the process of observing results and interpreting results of a PCR experiment.
- List possible uses of PCR in genetic testing and in research.

Overview

The polymerase chain reaction laboratory technique is used in a variety of applications to make copies of a specific DNA sequence. This lesson describes how a PCR reaction works, what it accomplishes, and its basic requirements for success. Examples of interpreting results are given. PCR's strengths, weaknesses, and applications to plant biotechnology are explained.

The Discovery of PCR

In 1983, Kary Mullis was driving along a Californian mountain road late one night. As a molecular biologist, Dr. Mullis was imagining a better way to study **DNA**. This late-night thinking led to a revolutionary way to make laboratory copies of DNA molecules (Saiki et al. 1985, Mullis 1990). In the decades since, the polymerase chain reaction or **PCR**, has become the standard method used for detecting specific DNA or RNA sequences. Selling the equipment and reagent kits for PCR is a multi-billion-dollar business because DNA and RNA detection is critical information in many applications.

In Vitro vs. *In Vivo* Replication

PCR is an *In Vitro* process; a series of chemical reactions that happen outside of a living cell. This laboratory technique is modeled after an In vivo process, the living cell's natural ability to replicate **DNA** during normal cell cycles (see Lesson on DNA: The Genetic Material). Every living cell makes a duplicate copy of each **chromosome** before the cell is ready to divide. Figure 1 below illustrates the key parts of *In vivo* DNA replication that are the basis for PCR success.

There are other enzymes that play an important role in in vivo replication. However, PCR works as an *in vitro* DNA replication process by using just one of these enzymes. Mullis imagined a chemical reagent and a temperature change step in the method that could perform the work of the other two enzymes. It should be noted that because Dr. Kerry Mullis had learned about the details of *in vivo* DNA replication, he could create this science changing *in vitro* method.

Before reading the description of PCR components and processes, watching this video can help you visualize the importance of each step,

Name and Chemical Components of PCR

The name 'Polymerase Chain Reaction' represents the nature of the process. 'Polymerase' because DNA Polymerase III is required for constructing new DNA strands, just like in a living cell. 'Chain Reaction' describes repeating cycles of replication which target a specific segment of a chromosome and use a "copy the copies" progression each cycle that doubles the amount of **DNA copies of a specific segment of DNA** present each cycle. In just 20 cycles of the chain reaction, over one million (2^{20}) copies of that specific segment of DNA can be produced. This is enough DNA to see with your naked eye. The goal of PCR is to make millions of copies of a specific segment of DNA that all originate from a single DNA sample.

The five chemical components that must be added to a test tube for the PCR reaction to work, include a **DNA template, DNA polymerase III enzyme**, single stranded DNA primers, nucleotides, and reaction buffer.

1. The DNA template is a sample of DNA that contains the target sequence of DNA for copying.
2. DNA pol III. There are two requirements for a suitable DNA polymerase enzyme for PCR. First, the enzyme must have a good activity rate around 75°C. Second, the enzyme should be able to withstand temperatures of 95-100°C without denaturing and losing activity.

3. Two primers. Primers are short oligonucleotides of DNA, usually around 20 base pairs in length. Because the purpose of PCR is to amplify a specific section of DNA in the genome, such as a known gene, then primers of specific sequences must be used. The geneticist planning the PCR reaction will design a forward primer to bind to one strand and a reverse primer that complements and binds to the other strand. The **primer design** process to select forward and reverse primers is requiring appropriate genetics thinking and is describe later in this reading.
4. The four different deoxyribonucleotide triphosphates (dNTPs). Adenine (A), guanine (G), cytosine (C), and thymine (T) are needed to provide the building blocks for DNA replication. DNA polymerase will add each **complementary** base to the new growing DNA strand according to the original strand's sequence following normal A-T and C-G pairings.
5. Finally, a reaction buffer. This creates a stable pH and provides the Mg^{2+} cofactor needed for DNA pol III activity.

Three Temperature Cycles

A key insight to the success of PCR as an in vitro DNA replication process which generates millions of specific sequence copies was a three-temperature cycle which accomplish three parts of DNA replication: **denaturation** of the double stranded template, **annealing** of the primers to the single strands and **extension** of new strand synthesis by DNA pol III.

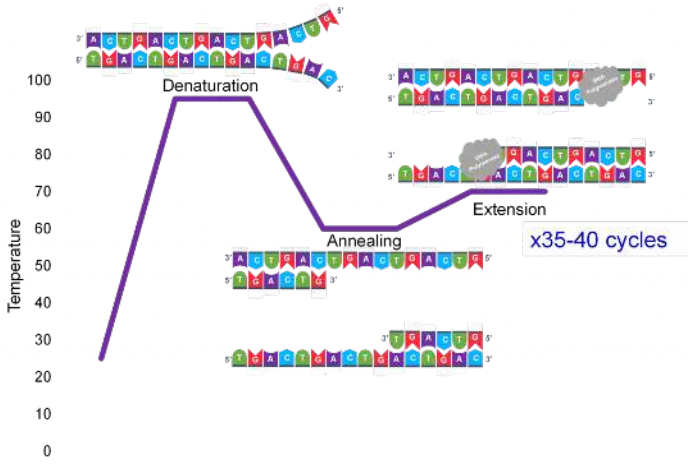


Figure 1. The three temperature steps in a PCR cycle. Image by Marjorie Hanneman.

In general, a single PCR run will undergo 25-35 cycles. The first step for a single cycle is the denaturation step, in which the double-stranded DNA template molecule (Figure 2) is made single-stranded(Figure 3). The temperature for this step is typically in the range of 95-100°C, near boiling. The high heat breaks the hydrogen bonds between the strands but does not break the sugar-phosphate bonds that hold the nucleotides of a single strand together (Figure 3).



Figure 2. DNA template in the PCR reaction test tube is a double stranded molecule. The targeted sequence of nucleotides for this illustration is shown. The template DNA could be fragments from the chromosome of various lengths. The 5′ and 3′ orientation of both strands is shown. Image by Marjorie Hanneman.

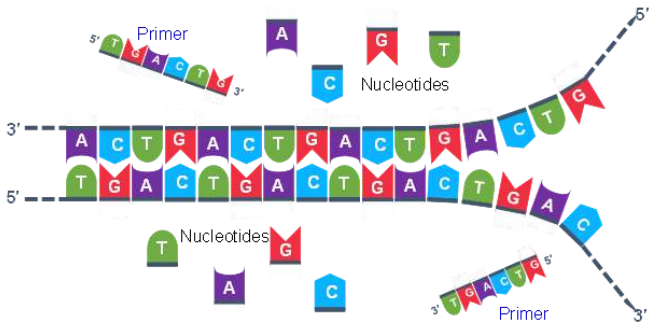


Figure 3. Denaturation: The hydrogen bonds holding the strands of the double stranded template together are broken by heating the test tube. Image by Marjorie Hanneman and Donald Lee.

Thousands of copies of the single stranded primers and the individual nucleotides were added to the test tube prior to beginning the cycles. Both the primers and nucleotides will become part of the new DNA strands. The second step in the PCR reaction is to cool the temperature in the test tube to 45-55°C. This is the primer annealing step in which the primers bind to **complementary** sequences in the single-stranded DNA template. The two primers are called the forward and the reverse primer and are designed because their sequences will target the desired segment of the DNA template for replication (Figure 4).

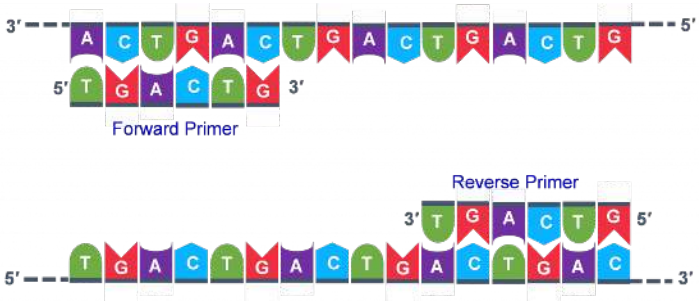


Figure 4. Annealing: The second step of PCR is to cool the test tube which allows the forward and reverse primer to bind to the template. Image by Marjorie Hanneman and Donald Lee.

The geneticist planning the PCR analysis must “design” the forward and reverse primers and then buy them from a vendor who can synthesize single stranded DNA that has a specific sequence and length. The two most important criteria for primer design are the following.

1. One primer must have a sequence that complements one of the template strands and the other primer must be complementary to the other strand. BOTH strands need to be primed for the replication process.
2. The primers must bind so that their 3′ ends are ‘pointing’ in the direction of the other primer. This ensures that the sequence between the primers is replicated in the PCR cycles.

Extension: The final PCR step is when the **DNA polymerase enzyme** reads the template and connects new nucleotidesto the primer’s 3’ end, extending a new complementary strand of DNA (Figure 5). Completion of the final step and the first cycle of PCR, will make two double stranded DNA copies from the original template DNA, doubling of the amount of DNA present. The test tube is heated to around 75°C, optimizing DNA pol. III activity and the newly synthesizing DNA strand is extended as the template strand is read by DNA pol. III. The Extension step will run for a few minutes and this step completes one PCR cycle.

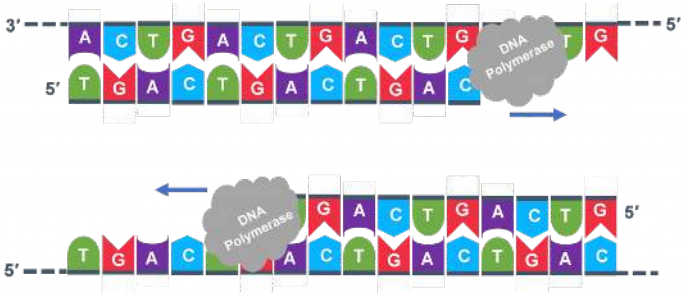


Figure 5. Extension: This step is catalyzed by DNA pol III. The time allowed for this step determines how much of the template DNA pol III can read and replicate. Image by Marjorie Hanneman.

For cycle 2, the denaturation, annealing, and extension steps are repeated (Figure 6 a, b, c). This time, though there will be twice as many DNA template molecules compared to what there was at the beginning of cycle 1. Copies are being made by reading the original template and copies are made by reading the copies made in the previous cycle. Therefore, if everything is working correctly, the DNA replication in the test tube is a chain reaction where at the end of a cycle, there is double the amount of that DNA sequence as what was found at the beginning of the cycle.



Figure 6a. Cycle 1 is complete; two double stranded DNA molecules are made from the original double stranded template. Image by Marjorie Hanneman and Donald Lee.



Figure 6b. Denaturation step generates four single strand templates. Image by Marjorie Hanneman

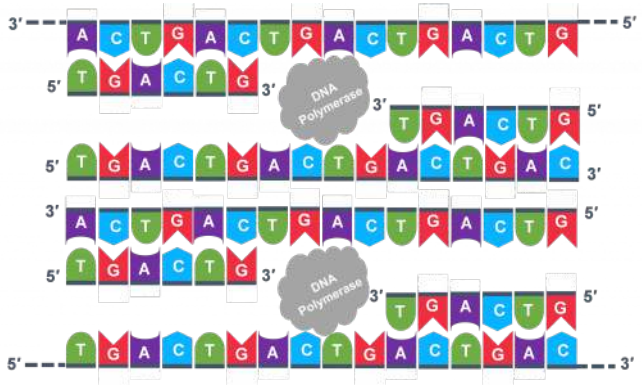


Figure 6c. Annealing and Extension. All templates are primed and four 3′ ends provide a place of DNA pol III extension. Image by Marjorie Hanneman

Because thousands of copies of the forward and reverse primer are added at the start of PCR, all the single strand templates, both the original, the copies in cycle 3 and beyond, and the copies of the copies made from previous cycles will be primed for the extension step of the cycles.

Thermal cycler



Figure 7. The thermal cycler is used to carry out PCR reaction. Image by Walter Suza.

When PCR was first invented, scientists used water baths set at different temperatures and the hand transfer of test tubes at timed intervals to run the PCR reaction. Once the technique became a proven technology for DNA analysis, engineers went to work to create PCR machines. The instrument which heats and cools the DNA samples is called a **thermal cycler** (Figure 7). Each small tube or sample well in a plate contains all the chemical components needed for a PCR reaction. Adding a specific sample to the reaction mix provides the template DNA. A thermal cycler can be programmed for specific temperatures and the amount of time spent at each temperature. The engineered design of thermal cyclers to maximize the accurate replication of the targeted DNA in a small sample volume with the minimum amount of time can be critical in many applications of PCR.

Taq DNA polymerase

When Dr. Kerry Mullis ran the first PCR experiments, he needed to add a new sample of DNA pol III after each denaturation step. This was because the high temperature needed to denature the double stranded DNA template also denatured the DNA pol III protein structure. The DNA pol III enzyme commonly available to molecular geneticists was from E. coli bacteria and this enzyme had no stability at near boiling temperatures. Fortunately, biologists had been investigating Thermus aquaticus, (Taq) a thermophilic eubacterium found in hot springs (Chien et al. 1976). The Taq version of DNA pol III does not easily denature in the hot temperatures required in PCR; plus, it has a good efficiency, able to add 60 base pairs/sec at 70°C. Like all other DNA polymerases, Taq DNA pol III cannot begin DNA replication without the addition of a starting primer. Thus, the discovery of Taq DNA pol III and the commercial availability of this enzyme made PCR a more reliable and doable technology which hastened its application to science investigation and diagnostic testing.

Visualizing the Results with Electrophoresis

Once a PCR reaction has been completed, we need to be able to see the results. To do this, a sample of the **PCR** mixture is loaded into an agarose gel for electrophoresis. The agarose gel contains a matrix of pores which enables it to separate DNA fragments based on their sizes. For details about setting up and running an electrophoresis gel, see Electrophoresis: How Scientist observe fragments of DNA

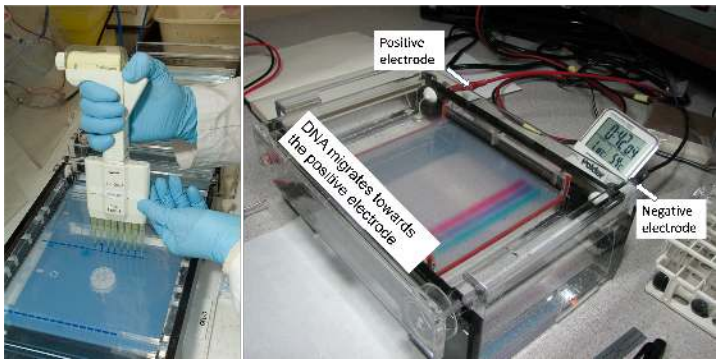


Figure 8. Electrophoresis: A gel electrophoresis set-up with agarose gel with DNA and loading dye on the left and the power supply on the right. Image Source: Michael, CC BY 2.0, via Wikimedia Commons and U. S. Department of Agriculture, CC BY 2.0, via Wikimedia Commons.

Figure 8 shows a picture of a gel electrophoresis gel that is running. The box on the right contains DNA loaded in the agarose gel. The gel placed in an aqueous solution of electrolytes. Depending on the type of dye used, color bands are a dye that was added to the PCR sample before it was loaded into the sample well. This allows for the tracking of the DNA's progression through the gel. Hooked up to this gel unit is an electrical power source which provides the force to move the DNA through the gel. Since DNA molecules are negatively charged, they will migrate towards the red, positive electrode. Shorter DNA fragments move faster than longer fragments through the pores in the gel.

After the gel is run, the DNA is stained with a chemical that binds specifically to DNA molecules and then will either reflect a specific color of visible light or fluoresce a specific color when viewed with ultraviolet light. A single 'band' contains 1000s of individual DNA fragments, all of the same length. Figure 10 illustrates the visual information that can be obtained from an electrophoresis gel after it has run. The electric current uniformly moves all the DNA fragments through a gel in the same direction. The sample wells at the top of the gel image thus establish lanes for the DNA samples to move.

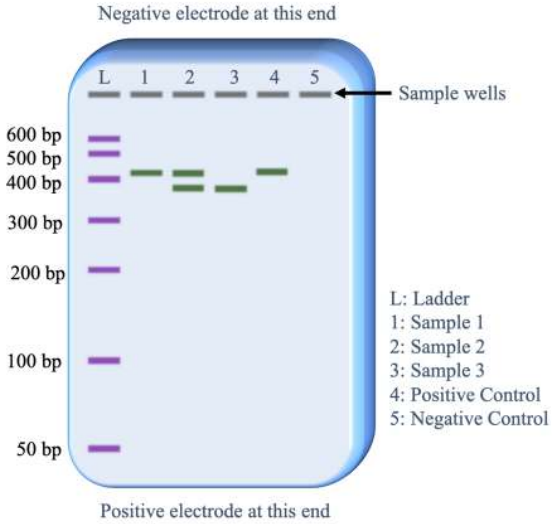


Figure 9. Depiction of an electrophoresis gel with six sample wells that were loaded with either a DNA size ladder (lane L) or a sample from a PCR run (1-5.) The gel was subjected to a DNA staining dye. Image by Marjorie Hanneman.

Below is a description of what information is revealed from each lane.

- Lane L: This was loaded with the DNA size ladder that contains copies of seven different lengths of DNA fragments. Commercial vendors of the DNA size ladder provide information on the lengths of each fragment in base pairs (bp). Running this lane provides an estimate of the DNA fragment lengths in the sample lanes (1-5).
- Lane 1: The PCR sample loaded in this lane has copies of a single length of DNA. The length is slightly more than 400 base pairs.
- Lane 2: The PCR sample loaded in this lane has copies of two lengths of DNA. One fragment is the same length as the fragments in lane 1 and the second fragment is slightly less than 400 bp.

- Lane 3: The PCR sample loaded in this lane has copies of one fragment that is the same length as the shorter fragments in lane 2.
- Lane 4: The positive control worked as predicted. The sample was set up with all the same reagents as the other PCR samples plus a DNA template that was known to contain the sequence targeted by the PCR primers that would generate a 410 base pair fragment.
- Lane 5: The negative control worked as predicted. The sample was set up with all the same reagents as the other PCR samples except no template DNA was added. Therefore, we would not expect the PCR reaction to work and the absence of a band of DNA fragments is as expected.

Because the positive and negative controls worked as expected, the biologist can be confident that the bands of DNA observed in lanes 1,2 and 3 reveal genetic information about the individual providing that DNA sample.

Advantages of PCR

PCR quickly became the method of choice for many types of DNA analysis because of several advantages over other DNA detection methods. The first; it is a simple procedure to set up and run. Fewer steps save time in getting a DNA analysis result. The second is the sensitivity. A very small amount of template DNA in the sample can be detected. Even just a few skin cells from one human hair contain enough DNA, making PCR useful in forensics. The third is that it can be designed to accurately differentiate genetic samples that are different by as little as single nucleotide in the targeted sequence. Finally, in application where large numbers of samples need to be subjected to the same PCR analysis, automation and robotic assistance allow the processing of many samples in a very short time. For example, an automated PCR would be critical for testing large numbers of people in hours during a pandemic.

Limitations of PCR

There are some drawbacks of using PCR that one should be aware of as well. First, the sequence of the gene or chromosome region being targeted is required. This limitation is rapidly diminishing as gene and genome sequencing technology and sharing of this sequence through Internet data bases has emerged as the norm in genetic analysis. Because of the size of the genomes of living things and a high conservation of gene sequences in many organisms, primers designed for a PCR test must be empirically tested with the proper controls. Biologists can easily generate false positive DNA from PCR that is caused by contamination or lack of specificity in primer design. There may also be a need to optimize concentrations of each chemical component. For example, changing the amount of DNA template, MgCl2 and Taq polymerase can affect both the quantity and quality of bands produced. Some studies have shown that even the brand of Taq polymerase can affect results (Holden et al. 2003). Likewise, the temperature cycles may need to be fine-tuned for a specific PCR test. Finally, as an in vitro DNA replication method, PCR cannot replicate entire chromosomes.

Summary

To briefly highlight the topics of this lesson, remember that **PCR** is a relatively easy lab technique which amplifies the amount of **DNA** present, much like living cells do in the beginning stages of a cell cycle. There are 5 chemical components of a PCR reaction: a **DNA template**, a **DNA polymerase**, primers, **nucleotides**, and a buffer. The 3 temperature steps for one cycle are the denaturation, primer annealing and extension steps. There are now many variations and uses of PCR ranging from forensics to genomic studies to identifying transgenic crops.

Lesson Activities

Watch these videos to learn more about PCR:

- Polymerase Chain Reaction (PCR) from DNA Learning Center
- Polymerase Chain Reaction (from UNL)

This page titled 11.4: PCR and Gel Electrophoresis is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, & Donald Lee (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: PCR and Gel Electrophoresis** by Donald Lee, Marjorie Hanneman, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

11.5: Origins of Molecular Polymorphisms

Mutations of DNA sequences can arise in many ways (Chapter 4). Some of these changes occur during DNA replication processes, resulting in an insertion, deletion, or substitution of one or a few nucleotides. Larger mutations can be caused by mobile genetic elements such as transposons, which are inserted more or less randomly into chromosomal DNA, sometimes occurring in clusters. In these and other types of **repetitive DNA** sequences, the number of repeated units is highly prone to change through unequal crossovers and other replication events.

This page titled 11.5: Origins of Molecular Polymorphisms is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **10.2: Origins of Molecular Polymorphisms** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.6: Classification and Detection of Molecular Markers

Regardless of their origins, molecular markers can be classified as polymorphisms that either vary in the length of a DNA sequence, or vary only in the identity of nucleotides at a particular position on a chromosome (Figure **11.6.1**). In both cases, because two or more alternative versions of the DNA sequence exist, we can treat each variant as a different allele of a single locus. Each allele gives a different **molecular phenotype**. For example, polymorphisms of **SSRs** (short sequence repeats) can be distinguished based on the length of PCR products: one allele of a particular SSR locus might produce a 100bp band, while the same primers used with a different allele as a template might produce a 120bp band (Figure **11.6.2**). A different type of marker, called a **SNP** (single nucleotide polymorphism), is an example of polymorphism that varies in nucleotide identity, but not length. SNPs are the most common of any molecular markers, and the genotypes of thousands of SNP loci can be determined in parallel, using new, hybridization based instruments. Note that the alleles of most molecular markers are co-dominant, since it is possible to distinguish the molecular phenotype of a heterozygote from either homozygote.

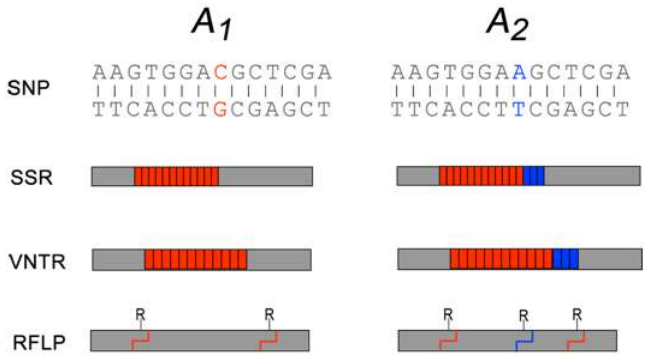


Figure **11.6.1**: Some examples of DNA polymorphisms. The variant region is marked in blue, and each variant sequence is arbitrarily assigned one of two allele labels. Abbreviations: SNP (Single Nucleotide Polymorphism); SSR (Simple Sequence Repeat) = SSLP (Simple Sequence Length Polymorphism); **VNTR** (Variable Number of Tandem Repeats); **RFLP** (Restriction Fragment Length Polymorphisms. VNTRs and SSRs differ in the size of the repeat unit; VNTRs are larger than SSRs. (Original-Deyholos-CC:AN)

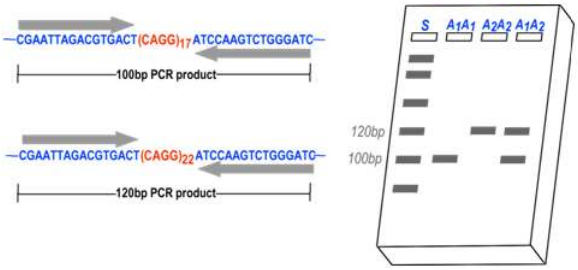


Figure **11.6.2**: Determining the genotype of an individual at a single SSR locus using a specific pair of PCR primers and agarose gel electrophoresis. S= size standard (Original-Deyholos-CC:AN)

Mutations that do not affect the function of protein sequences or gene expression are likely to persist in a population as polymorphisms, since there will be no selection either in favor or against them (i.e. they are **neutral**). Note that the although the rate of spontaneous mutation in natural populations is sufficiently high so as to generate millions of polymorphisms that accumulate over thousands of generations, the rate of mutation is on the other hand sufficiently low that existing polymorphisms are stable throughout the few generations we study in a typical genetic experiment.

This page titled 11.6: Classification and Detection of Molecular Markers is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **10.3: Classification and Detection of Molecular Markers** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation

Restriction Enzymes

Many bacteria have enzymes that recognize specific DNA sequences (usually 4 or 6 nucleotides) and then cut the double stranded DNA helix at this sequence (Figure **11.7.7**). These enzymes are called site-specific **restriction endonucleases**, or more simply "**restriction enzymes**", and they naturally function as part of bacterial defenses against viruses and other sources of foreign DNA. To cut DNA at known locations, researchers use restriction enzymes that have been purified from various bacterial species, and which can be purchased from various commercial sources. These enzymes are usually named after the bacterium from which they were first isolated. For example, ***EcoRI*** and *EcoRV* are both enzymes from *E. coli*. *EcoRI* cuts double stranded DNA at the sequence GAATTC, but note that this enzyme, like many others, does not cut in exactly the middle of the restriction sequence (Figure **11.7.8**). The ends of a molecule cut by *EcoRI* have an overhanging region of single stranded DNA, and so are sometimes called **sticky-ends**. On the other hand, *EcoRV* is an example of an enzyme that cuts both strands in exactly the middle of its recognition sequence, producing what are called **blunt-ends**, which lack overhangs.



Figure **11.7.7**: An EcoRI dimer (blue, purple) sits like a saddle on a double helix of DNA (one strand is green, one is brown). This image is looking down the center of the helix.(NCBI-?-PD)



Figure **11.7.8**: The recognition sequence for EcoRI (blue) is cleaved by the enzyme (grey). This particular enzyme cuts DNA at a position offset from the center of the restriction site. This creates an overhanging, sticky-end. (Original-Deyholos-CC:AN)

DNA Ligation

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end through the action of an enzyme called **DNA ligase**. Sticky-ended molecules with complementary overhanging sequences are said to have **compatible ends**, which facilitate their joining to form recombinant DNA. Likewise, two blunt-ended sequences are also considered compatible to join together, although they do not ligate together as efficiently as sticky-ends. Note: sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase. Ligation is therefore central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector.

This page titled 11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.4: Cutting and Pasting DNA- Restriction Digests and DNA Ligation** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

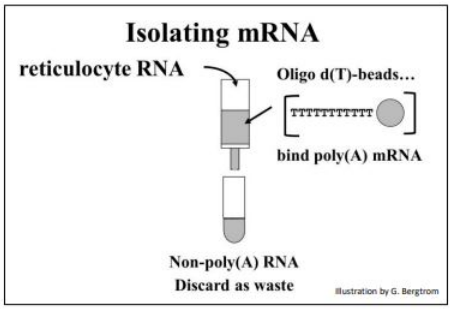
11.8: Make and Screen a cDNA Library

The first step in making a cDNA library is to isolate cellular mRNA. This mRNA extract should represent all of the transcripts in the cells at the time of isolation, or the cell's **transcriptome**. This term is used by analogy to genome. However, a genome is all of the genetic information of an organism. In contrast, a transcriptome (usually eukaryotic) reflects all of the genes expressed in a given cell type at a moment in time. Reversetranscribed cDNAs from an mRNA extract are also referred to as a transcriptome…, and likewise, a cDNA library. A cDNA library is a tube full of bacterial cells that have taken up (i.e., been **transformed** with) plasmids recombined with cDNAs. cDNA libraries made from mRNAs taken from different cell types or the same cells grown under different conditions are in effect, different transcriptomes. Each reflects mRNAs transcribed in cells at the moment of their extraction. When cells in a cDNA library are spread out on a nutrient agar petri dish, each cell grows into a colony of cells; each cell in the colony is a clone of a starting cell. cDNA libraries can be used isolate and sequence the DNA encoding a polypeptide that you are studying.

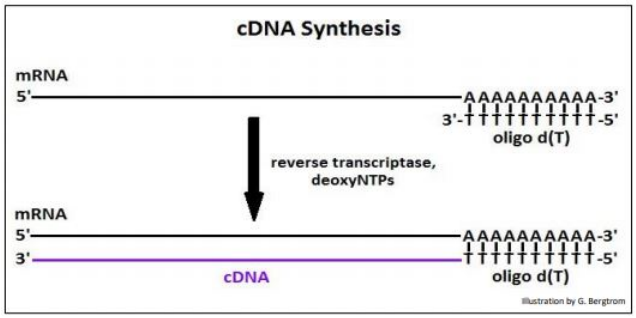
Recall that the mature mRNA in eukaryotic cells has been spliced. This means that cDNAs from eukaryotic cells do not include introns. Introns, as well as sequences of enhancers and other regulatory elements in and surrounding a gene must be studied in genomic libraries, to be discussed later. Here we look at how to make a cDNA library.

A. cDNA Construction

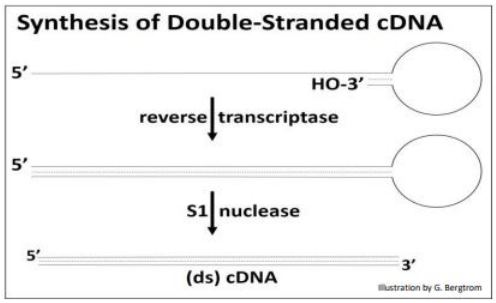
mRNA is only a few percent of a eukaryotic cell; most is rRNA. But that small amount of mRNA can be separated from other cellular RNAs by virtue of their 3' poly(A) tails. Simply pass a total RNA extract over an **oligo-d(T) column** (illustrated below).



The strings of thymidine (T) can H-bond with the poly(A) tails of mRNAs, tethering them to the column. All RNAs without a 3' poly(A) tail will flow through the column as waste. A second buffer is passed over the column to destabilize the A-T H-bonds to allow **elution** of an mRNA fraction. When free' oligo d(T) is added to the eluted mRNA, it forms H-bonds with the poly(A) tails of the mRNAs, serving as a primer for the synthesis of cDNA copies of the poly(A) mRNAs originally in the cells. Finally, four deoxynucleotide DNA precursors and **reverse transcriptase** (originally isolated from chicken retrovirus-infected cells) are added to start reverse transcription. The synthesis of a cDNA strand complementary to an mRNA is shown below.



After heating to separate the cDNAs from the mRNAs, the cDNA is replicated to produce double-stranded, or (ds)cDNA, as illustrated below.



Synthesis of the second cDNA strand is also catalyzed by reverse transcriptase! The enzyme recognizes DNA as well as RNA templates, and has the same 5'-to-3' DNA polymerizing activity as DNA polymerases. After 2nd cDNA strand synthesis, **S1 nuclease** (a *single-stranded endonuclease* originally isolated from an East Asian fungus!) is added to open the loop of the (ds) cDNA structure and trim the rest of the single-stranded DNA. What remains is the (ds) cDNA.

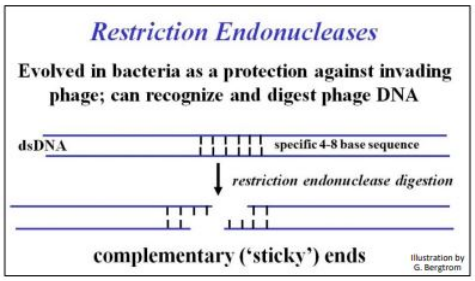
258 Isolate mRNA and Make cDNA

259 Reverse Transcriptase

B. Cloning cDNAs into Plasmid Vectors

To understand cDNA cloning and other aspects of making recombinant DNA, we need to talk a bit more about the recombinant DNA tool kit. In addition to *reverse transcriptase* and S1 nuclease, other necessary enzymes in the 'kit' include **restriction endonucleases** (*restriction enzymes*) and **DNA ligase**. The natural function of restriction enzymes in bacteria is to recognize specific restriction site sequences in phage DNA (most often **palindromic** DNA sequences), hydrolyze it and thus avoid infection.

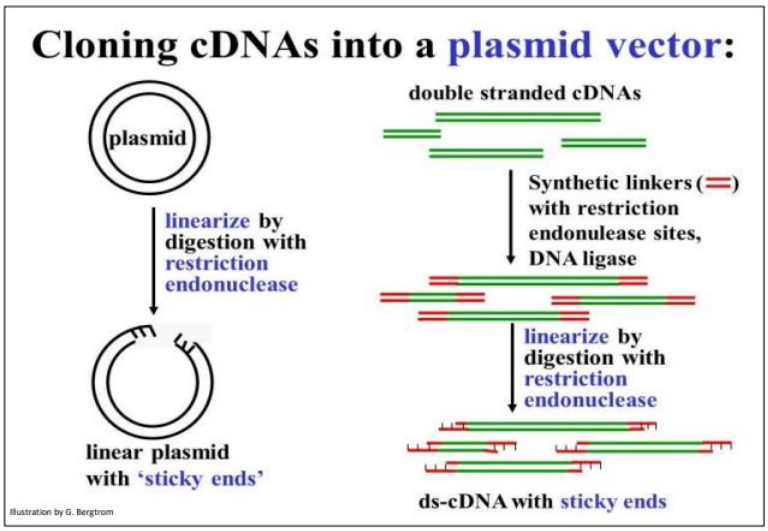
Restriction enzymes that make a scissors cut through the two strands of the double helix leaves *blunt ends*. Restriction enzymes that make a staggered cut on each strand at their restriction site leave behind *complementary* ('sticky') ends (below).



If you mix two of double-stranded DNA fragments with the same sticky ends from different sources (e.g., different species), they will form H-bonds at their complementary ends, making it easy to recombine plasmid DNA with (ds)cDNA, that have the same complementary 'sticky ends'. Using the language of recombinant DNA technologies, let's look at how plasmid vectors and cDNAs can be made to recombine.

1. Preparing Recombinant Plasmid Vectors Containing cDNA Inserts

Vectors are *carrier* DNAs engineered to recombine with foreign DNAs of interest. When a recombinant vector with its foreign **DNA insert** gets into a host cell, it can replicate many copies of itself, enough in fact for easy isolation and study. cDNAs are typically inserted into plasmid vectors that are usually purchased "off-the-shelf". They can be *cut* with a restriction enzyme at a suitable location, leaving those sticky ends. On the other hand, it would not do to digest (ds)cDNA with restriction endonucleases since the goal is not to clone cDNA fragments, but entire cDNA molecules. Therefore, it will be necessary to attach **linkers** to either end of the (ds)cDNAs. Plasmid DNAs and cDNA-linker constructs can then be digested with the same restriction enzyme to produce compatible 'sticky ends'. Steps in the preparation of vector and (ds)cDNA for recombination are shown below.

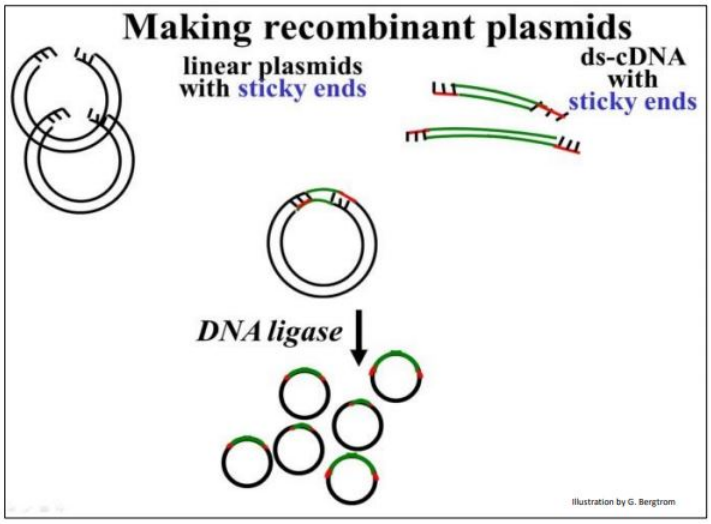


To prepare for recombination, a plasmid vector is digested with a **restriction enzyme** to open the DNA circle. To have compatible *sticky ends*, double-stranded cDNAs to be inserted are mixed with *linkers* and **DNA ligase** to put a linker DNA at both ends of the (ds) cDNA. DNA ligase is another tool in the recombinant DNA toolkit. Linkers are short, synthetic double-stranded DNA oligomers containing restriction sites recognized and cut by the same restriction enzyme as the plasmid. Once the linkers are attached to the ends of the plasmid DNAs, they are digested with the appropriate restriction enzyme. This leaves both the (ds)cDNAs and the plasmid vectors with complementary sticky ends.

260 Restriction Enzymes and Recombinant DNA

2. Recombining Plasmids and cDNA Inserts and Transforming Host Cells

The next step is to mix the cut plasmids with the digested linker-cDNAs in just the right proportions so that the most of the cDNA (linker) ends will anneal (form Hbonds) with the most of the sticky plasmid ends. Adding **DNA ligase** to the plasmid/linker-cDNA mixture forms phosphodiester bonds between plasmid and cDNA insert, completing the recombinant circle of DNA, as shown below.

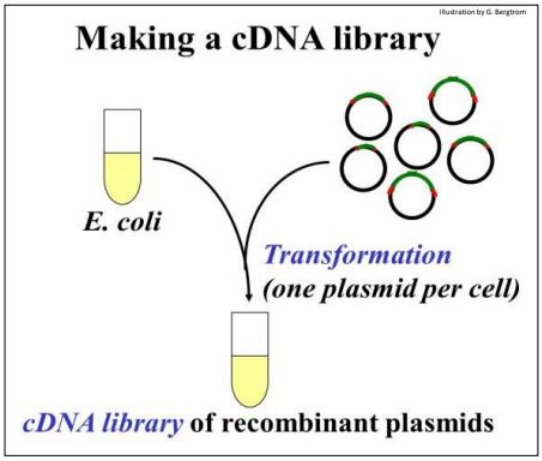


In early cloning experiments, an important consideration was to generate plasmids with only one copy of a given cDNA insert, rather than lots of re-ligated plasmids with no inserts or lots of plasmids with multiple inserts. Using betterengineered vector and linker combinations, this issue became less important.

261 Recombine a cDNA Insert with a Plasmid Vector

3. Transforming Host Cells with Recombinant Plasmids

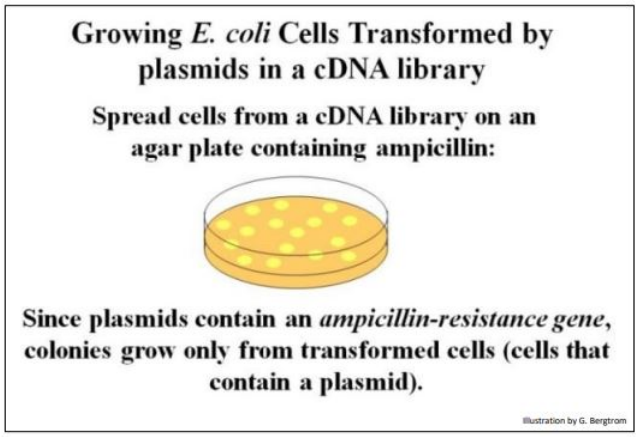
The recombinant DNA molecules are now ready for 'cloning'. They are added to E. coli (sometimes other host cells) *made permeable* so that they can be easily **transformed**. Recall that transformation as defined by Griffith is bacterial uptake of foreign DNA leading to a genetic change. The **transforming principle** in cloning is the recombinant plasmid! The transformation step is shown below.



The tube full of transformed cells is the **cDNA Library.**

262 Making the cDNA Library.

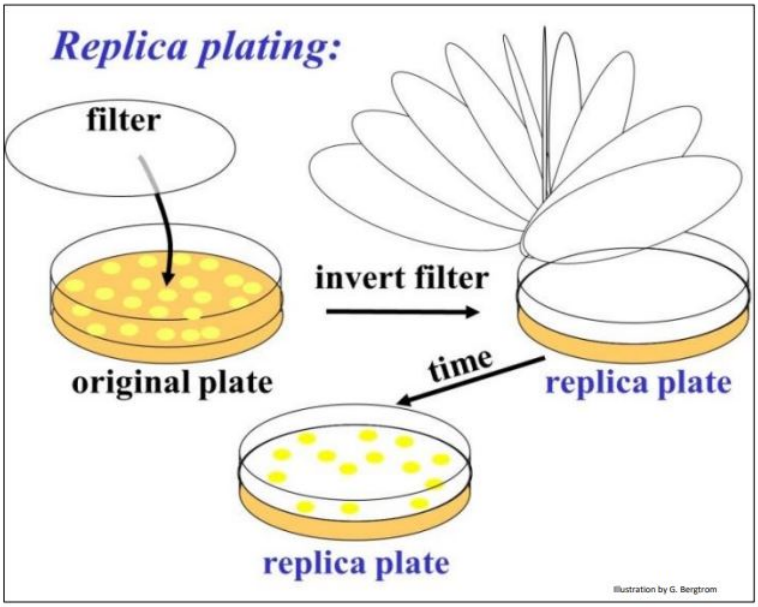
After all these treatments, not all plasmid molecules in the mix are recombinant; some cells in the mix haven't even taken up a plasmid. So when the recombinant cells are plated on agar, how do you tell which of the colonies that grow came from cells that took up a recombinant plasmid? Both the **host strain** of *E. coli* and plasmid vectors used these days were further engineered to solve this problem. One such plasmid vector carries an **antibiotic resistance gene**. In this case, ampicillin-sensitive cells would be transformed with recombinant plasmids containing the *resistance gene*. When these cells are plated on media containing **ampicillin** (a form of penicillin), they grow, as illustrated below.



Untransformed cells (cells that failed to take up a plasmid) lack the ampicillin resistance gene and thus, do not grow on ampicillin-medium. But, there is still a question. How can you tell whether the cells that grew were transformed by a recombinant plasmid containing a cDNA insert? It is possible that some of the transformants contain only non-recombinant plasmids that still have the ampicillin resistance gene!

To address this question, plasmids were further engineered with a **streptomycin resistance gene**. But this antibiotic resistance gene was also engineered to contain restriction enzyme sites in the middle of the gene. Thus, inserting a cDNA in this plasmid would disrupt and inactivate the gene. Here is how this second bit of genetic engineering enabled growth only of cells transformed with a

recombinant plasmid containing a cDNA insert. We can tell *transformants* containing recombinant plasmids apart from those containing non-recombinant plasmids by the technique of **replica plating** shown (illustrated below).

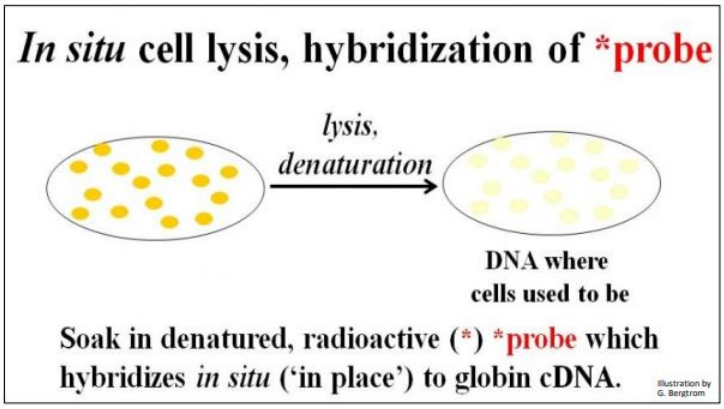


After colonies grow on the ampicillin agar plate, lay a filter over the plate. The filter will pick up a few cells from each colony, in effect becoming a replica (mirror image) of the colonies on the plate. Place the *replica filter* on a new agar plate containing streptomycin; the new colonies that grow on the filter must be streptomycin-resistant, containing only non-recombinant plasmids. Colonies containing recombinant plasmids, those that did not grow in streptomycin are easily identified on the original ampicillin agar plate. In practice, highly efficient recombination and transformation procedures typically reveal very streptomycinresistant cells (i.e., colonies) after replica plating. In this case, ampicillin-resistant cells constitute a good cDNA library, ready for screening.

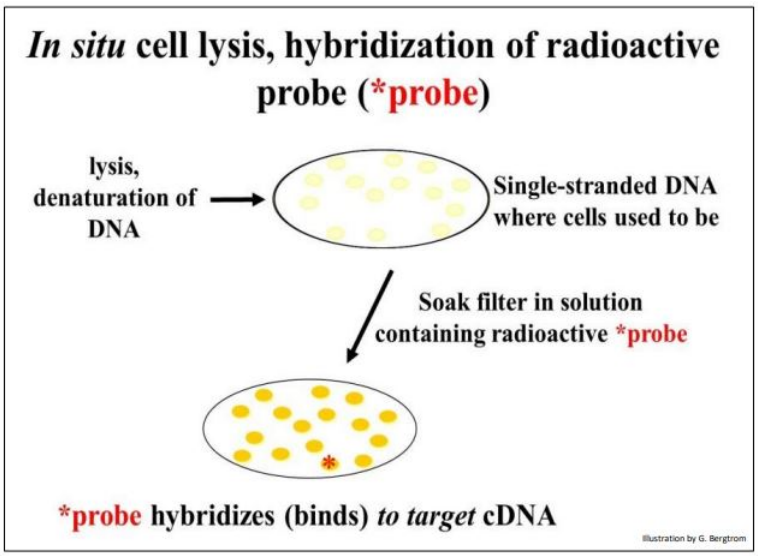
263 Making a Replica Plate Filter

4. Identifying Colonies Containing Plasmids with Inserts of Interest

The next step is to **screen** the colonies from the cDNA library for those containing the specific cDNA that you're after. This typically begins preparing multiple replica filters like the one above. Remember, these filters are replicas of bacterial cells containing recombinant plasmids that grow on ampicillin but not streptomycin. The number of replica filters that must be screened can be calculated from assumptions and formulas for estimating how many colonies must be screened to represent an entire transcriptome (i.e., the number of different mRNAs in the original cellular mRNA source). Once the requisite number of replica filters are made, they are subjected to **in situ lysis** to disrupt cell walls and membranes. The result is that the cell contents are released and the DNA is denatured (i.e., becomes single-stranded). The DNA then adheres to the filter *in place* (*in situ*, where the colonies were). The result of *in situ lysis* is a filter with faint traces of the original colony (below).

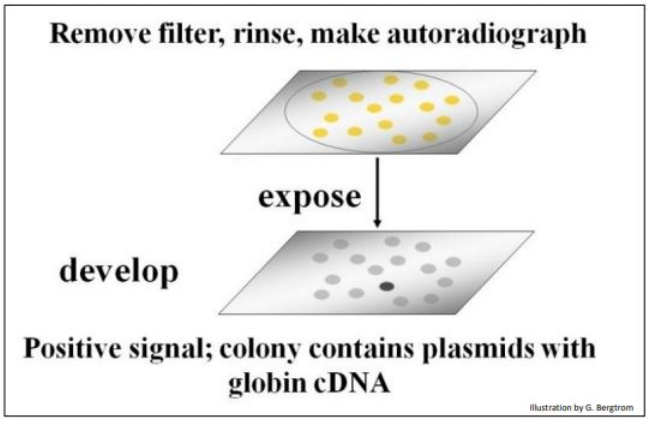


Next, a molecular **probe** is used to identify DNA containing the sequence of interest. The probe is often a **synthetic oligonucleotide** whose sequence was inferred from known amino acid sequences. These oligonucleotides are made radioactive and placed in a bag with the filter(s). DNA from cells that contained recombinant plasmids with a cDNA of interest will bind the complementary probe. The results of *in situ* lysis and **hybridization** of a radioactive probe to a replica filter are shown below.



264 Probing a Replica Plate Filter

The filters are rinsed to remove un-bound radioactive oligomer probe, and then placed on X-ray film. After a period of exposure, the film is developed. Black spots will form on the film from radioactive exposure, creating an autoradiograph of the filter. The black spots in the autoradiograph correspond to colonies on a filter that contain a recombinant plasmid with your target cDNA sequence (below).



Once a positive clone is identified on the film, the corresponding recombinant colony is located on the original plate. This colony is grown up in a liquid culture and the plasmid DNA is isolated. At that point, the cloned plasmid DNA can be sequenced and the amino acid sequence encoded by its cDNA can be inferred from the genetic code dictionary to verify that the cDNA *insert* in fact encodes the protein of interest. Once verified as the sequence of interest, a cloned plasmid cDNA can be made radioactive or fluorescent, and used to

- probe for the genes from which they originated.
- identify and quantitate the mRNA even locate the transcripts in the cells.
- quantitatively measure amounts of specific mRNAs.

Isolated plasmid cDNAs can even be expressed in suitable cells to make the encoded protein. These days, diabetics no longer receive pig insulin, but get synthetic human insulin human made from expressed human cDNAs. Moreover, while the introduction

of the *polymerase chain reaction* (**PCR**, see below) has superseded some uses of cDNAs, they still play a role in genome-level and transcriptome-level studies.

265 Pick a Clone From a Replica Filter and Play With It!

This page titled 11.8: Make and Screen a cDNA Library is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **15.2: Make and Screen a cDNA Library** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

11.9: DNA Sequencing

A. A Brief History of DNA Sequencing

RNA sequencing came first, when Robert Holley sequenced a tRNA in 1965. The direct sequencing of tRNAs was possible because tRNAs are small, short nucleic acids, and because many of the bases in tRNAs are chemically modified after transcription. An early method for DNA sequencing developed by Walter Gilbert and colleagues involved DNA fragmentation, sequencing of the small fragments of DNA, and then aligning the overlapping sequences of the short fragments to assemble longer sequences. Another method, the DNA synthesis-based 'dideoxy' DNA sequencing technique, was developed by Frederick Sanger in England. Sanger and Gilbert both won a Nobel Prize in Chemistry in 1983 for their DNA sequencing efforts. However, because of its simplicity, Sanger's method quickly became the standard for sequencing all manner of cloned DNAs.

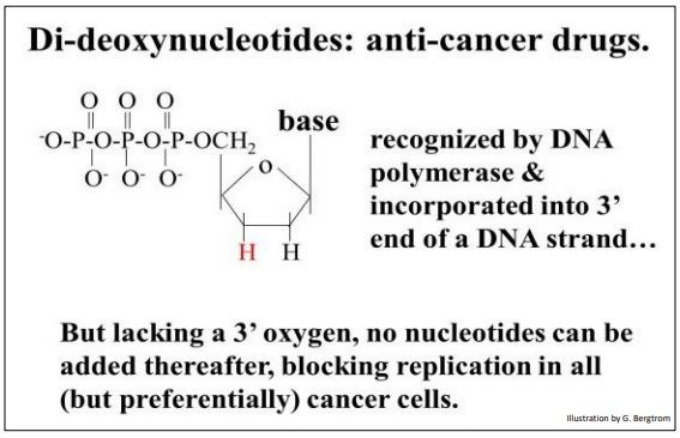
The first complete genome to be sequenced was that of a bacteriophage (bacterial virus) called φX174. At the same time as the advances in sequencing technology were occurring, so were some of the early developments in recombinant DNA technology. Together these led to more efficient and rapid cloning and sequencing of DNA from increasingly diverse sources. The first focus was of course on genes and genomes of important model organisms, such as *E. coli*, *C. elegans*, yeast (*S. cerevisiae*)…, and of course us! By 1995, Craig Venter and colleagues at the *Institute for Genomic Research* had completed the sequence of an entire bacterial genome (*Haemophilus influenzae*) and by 2001, Venter's private group along with Frances Collins and colleagues at the NIH had published a first draft of the sequence of the human genome. Venter had proven the efficacy of a whole-genome sequencing approach called *shotgun sequencing*, which was much faster than the gene-by-gene, fragment-by-fragment 'linear' sequencing strategy being used by other investigators (more later!). Since Sanger's dideoxynucleotide DNA sequencing method remains a common and economical methodology, let's consider the basics of the protocol.

B. Details of DiDeoxy Sequencing

Given a template DNA (e.g., a plasmid cDNA), Sanger used in vitro replication protocols to demonstrate that he could:

1. Replicate DNA under conditions that randomly stopped nucleotide addition at every possible position in growing strands.
2. *Separate and then detect* these DNA fragments of replicated DNA.

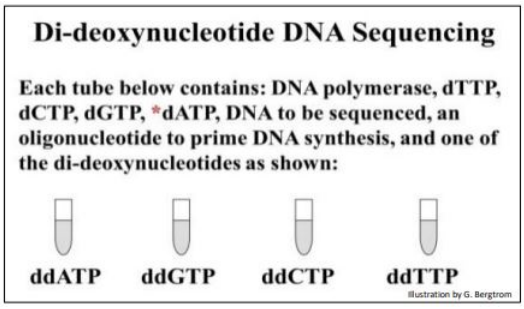
Recall that DNA polymerases catalyze the formation of *phosphodiester bonds* by linking the ***α*** *phosphate* of a **nucleotide triphosphate** to the free 3' OH of a deoxynucleotide at the end of a growing DNA strand. Recall also that the ribose sugar in the deoxynucleotide precursors of replication lack a 2' OH (hydroxyl) group. Sanger's trick was to add **dideoxynucleotide triphosphates** to his in vitro replication mix. The ribose on a **dideoxynucleotide triphosphate** (ddNTP) lacks a 3' OH, in addition to the 2' OH group (as shown below).



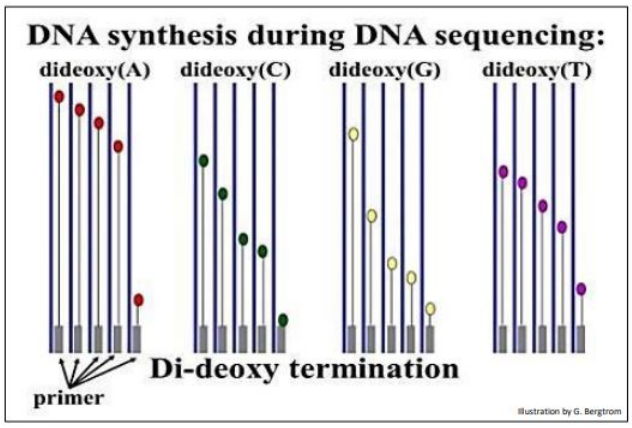
Adding a dideoxynucleotide to a growing DNA strand stops replication. No further nucleotides can add to the 3'-end of the replicating DNA strand because the 3'–OH necessary for the dehydration synthesis of the next phosphodiester bond is absent! Because they can stop replication in actively growing cells, ddNTPs such as dideoxyadenosine (tradename, **cordycepin**) are anti-cancer chemotherapeutic drugs.

A look at a manual DNA sequencing protocol reveals what is going on in the sequencing reactions. Four reaction tubes are set up, each containing the template DNA to be sequenced, a *primer* of known sequence and the four required deoxynucleotide precursors necessary for replication.

The set-up for manual DNA sequencing is shown below.

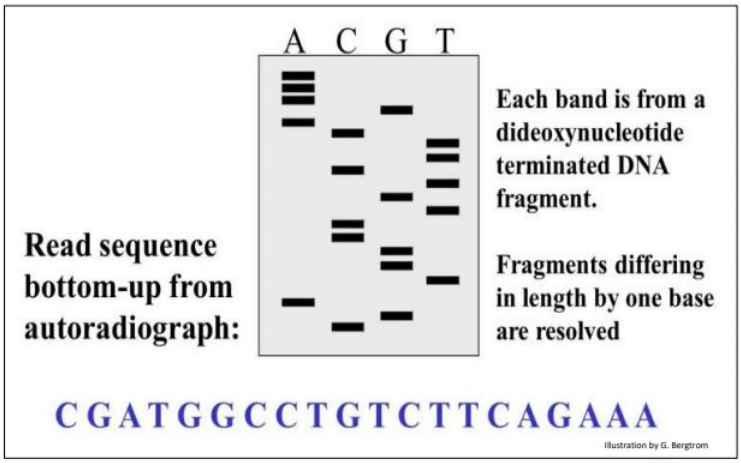


A different ddNTP, (ddATP, ddCTP, ddGTP or ddTTP) is added to each of the four tubes. Finally, DNA polymerase is added to each tube to start the DNA synthesis reaction. During DNA synthesis, different length fragments of new DNA accumulate as the ddNTPs incorporate randomly, opposite complementary bases in the template DNA being sequenced. The expectations of the didieoxy sequencing reactions in the four tubes are illustrated below.



A short time after adding the DNA polymerase to begin the reactions, the mixture is heated to separate the DNA strands and fresh DNA polymerase is added to repeat the synthesis reactions. These sequencing reactions are repeated as many as 30 times in order to produce enough radioactive DNA fragments to be detected. When the heat-stable **Taq DNA polymerase** from the thermophilic bacterium *Thermus aquaticus* became available (more later!), it was no longer necessary to add fresh DNA polymerase after each replication cycle. The many heating and cooling cycles required for what became known as *chain-termination DNA sequencing* were soon automated using inexpensive *programmable thermocyclers*.

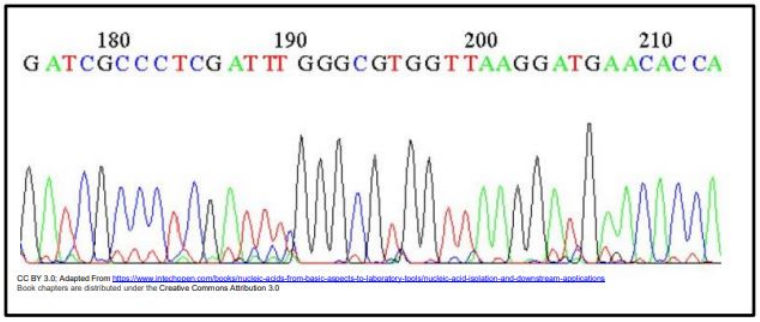
Since a small amount of a radioactive deoxynucleotide (usually 32P-labeled ATP) was present in each reaction tube, the newly made DNA fragments are radioactive. After electrophoresis to separate the new DNA fragments in each tube, autoradiography of the electrophoretic gel reveals the position of each terminated fragment. The DNA sequence can then be read from the gel as illustrated in the simulated autoradiograph below.



As shown in the cartoon, the DNA sequence can be read by reading the bases from the bottom of the gel, starting with the C at the bottom of the C lane. Try reading the sequence yourself!

267 Manual Dideoxy Sequencing

The first semi-automated DNA sequencing method was invented in Leroy Hood's California lab in 1986. Though still Sanger sequencing, the four dideoxynucleotides in the sequencing reaction were tagged for detection with a fluorescent dyes instead radioactive phosphate-tagged nucleotides. After the sequencing reactions, the reaction products are electrophoresed on an 'automated DNA sequencer'. UV light excites the migrating dye-terminated DNA fragments as they pass through a detector. The color of their fluorescence is detected, processed and sent to a computer, generating color-coded graph like the one below, showing the order (and therefore length) of fragments passing the detector and thus, the sequence of the strand.



A most useful feature of this sequencing method is that a template DNA could be sequenced in a single tube, containing all the required components, including *all* four dideoxynucleotides! That's because the fluorescence detector in the sequencing machine separately sees all the short ddNTP-terminated fragments as they move through the electrophoretic gel.

Hood's innovations were quickly commercialized making major sequencing projects possible, including whole genome sequencing. The rapidity of automated DNA sequencing led to the creation of large sequence databases in the U.S. and Europe.

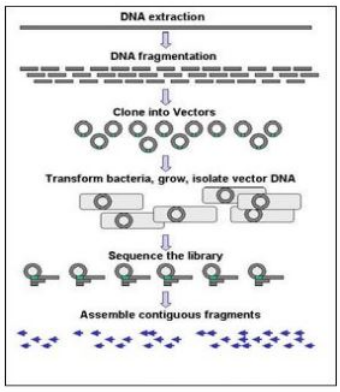
The NCBI (National Center for Biological Information) maintains the U.S. database. Despite its location, the NCBI archives virtually all DNA sequences determined worldwide. New 'tiny' DNA sequencers have made sequencing DNA so portable that in 2016, one was even used in the *International Space Station*. Expanding databases and new tools and protocols (some are described below) to find, compare and analyze DNA sequences have also grown rapidly.

268 Automated Sequencing Leads to Large Genome Projects

C. Large Scale Sequencing

Large-scale sequencing targets entire prokaryotic, and typically much larger eukaryotic genomes. The latter require strategies that either sequence long DNA fragments and/or sequencing DNA fragments more quickly. We already noted the shotgun sequencing used by Venter to sequence smaller and larger genomes (including our own… or more accurately, his own!). In shotgun

sequencing, cloned DNA fragments 1000 base pairs or longer are broken down at random into smaller, more easily sequenced fragments. The fragments are themselves cloned and sequenced and non-redundant sequences are assembled by aligning overlapping regions of sequence. Today's computer software is quite adept at rapid overlapping sequence alignment as well as connecting and displaying long contiguous DNA sequences. Shotgun sequencing is summarized below.



Sequence gaps that remain after shotgun sequencing can be filled in by primer walking, in which a known sequence near the gap is the basis of creating a sequencing primer to "walk" into the gap region on an intact DNA that has not been fragmented. Another 'gap-filling' technique involves **PCR** (the Polymerase Chain Reaction, to be described shortly). Briefly, two oligonucleotides are synthesized based on sequence information on either side of a gap. Then PCR is used to synthesize the missing fragment, and the fragment is sequenced to fill in the gap.

This page titled 11.9: DNA Sequencing is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **15.3: DNA Sequencing** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

11.10: Genomic Libraries

A **genomic library** might be a tube full of **recombinant bacteriophage**. Each phage DNA molecule contains a fragmentary insert of cellular DNA from a foreign organism. The library is made to contain a representation of all of possible fragments of that genome. Bacteriophage are often used to clone genomic DNA fragments because:

- phage genomes are bigger than plasmids and can be engineered to remove a large amount of DNA that is not needed for infection and replication in bacterial host cells.
- the missing DNA can thus be replaced by foreign insert DNA fragments as long as 18- 20kbp (kilobase pairs), nearly 20X as long as typical cDNA inserts in plasmids.
- purified phage coat proteins can be mixed with the recombined phage DNA to make infectious phage particles that would infect host bacteria, replicate lots of new recombinant phage, and then lyse the cells to release the phage.

The need for vectors like bacteriophage that can accommodate long inserts becomes obvious from the following bit of math. A typical mammalian genome consists of more than *2 billion base pairs*. Inserts in plasmids are very short, rarely exceeding 1000 base pairs. Dividing *2,000,000,000* by *1000, you get 2 million*, a minimum number of phage clones that must be screened to find a sequence of interest. In fact, you would need many more than this number of clones to find a gene (or parts of one!). Of course, part of the solution to this "needle in a haystack" dilemma is to clone larger DNA inserts in more accommodating vectors.

From this brief description, you may recognize the common strategy for genetically engineering a cloning vector: determine the minimum properties that your vector must have and remove non-essential DNA sequences. Consider the **Yeast Artificial Chromosome** (**YAC**), hosted by (replicated in) yeast cells. YACs can accept humongous foreign DNA inserts! This is because to be a chromosome that will replicate in a yeast cell requires **one centromere and two telomeres**… and little else!

Recall that telomeres are needed in replication to keep the chromosome from shortening during replication of the DNA. The centromere is needed to *attach chromatids to spindle fibers* so that they can separate during *anaphase* in *mitosis* (and *meiosis*). So along with a centromere and two telomeres, just include restriction sites to enable recombination with inserts as long as 2000 Kbp. That's a YAC! The tough part of course is keeping a 2000Kbp long DNA fragment intact long enough to get it into the YAC.

However a vector is engineered and chosen, sequencing its insert can tell us many things. They can show us how a gene is regulated by revealing known and uncovering new regulatory DNA sequences. They can tell us what other genes are nearby, and where genes are on chromosomes. Genomic DNA sequences from one species can probe for similar sequences in other species and comparative sequence analysis can then tell us a great deal about gene evolution and the evolution of species.

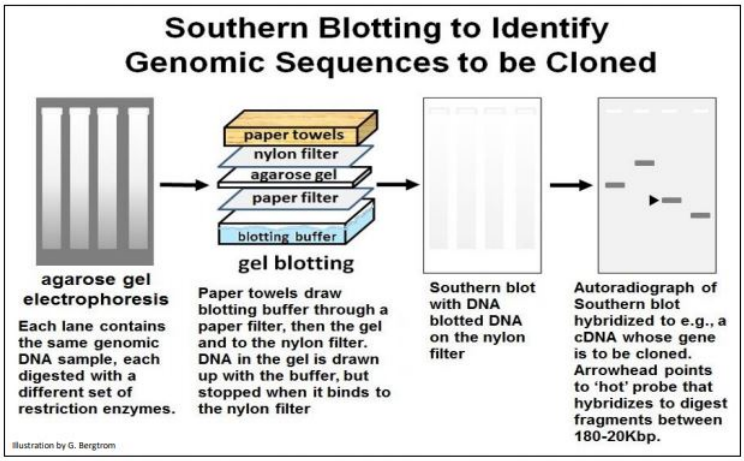
One early surprise from gene sequencing studies was that we share many common genes and DNA sequences with other species, from yeast to worms to flies… and of course vertebrates and our more closely related mammal friends. You may already know that the chimpanzee's and our genomes are 99% similar. Moreover, we have already seen comparative sequence analysis showing how proteins with different functions nevertheless share structural domains.

Let's look at cloning a genomic library in phage. As you will see, the principles are similar to cloning a foreign DNA into a plasmid, or in fact any other vector, but the numbers and details used here exemplify cloning in phage.

A. Preparing Genomic DNA of a Specific Length for Cloning

To begin with, high molecular weight (i.e., long molecules of) the desired genomic DNA are isolated, purified and then digested with a restriction enzyme. Usually, the digest is partial, aiming to generate overlapping DNA fragments of random length. When the digest is electrophoresed on agarose gels, the DNA (stained with ethidium bromide, a fluorescent dye that binds to DNA) looks like a bright smear on the gel. All of the DNA could be recombined with suitably digested vector DNA. But, to further reduce the number of clones to be screened for a sequence of interest, early cloners would generate a **Southern blot** (named after Edward Southern, the inventor of the technique) to determine the size of genomic DNA fragments most likely to contain a desired gene.

Beginning with a gel of genomic DNA restriction digests, the Southern blot protocol is illustrated below



To summarize the steps:

a) Digest genomic DNA with one or more restriction endonucleases.

b) Run the digest products on an agarose gel to separate fragments by size (length). The DNA appears as a smear when stained with a fluorescent dye.

c) Place a filter on the gel. The DNA transfers (blots) to the filter for e.g., 24 hours

d) Remove the blotted filter and place it in a bag containing a solution that can denature the DNA.

e) Add radioactive probe (e.g., cDNA) containing the gene or sequence of interest. The probe hybridizes (bind) to complementary genomic sequences on the filter

f) Prepare an autoradiograph of the filter and see a 'band' representing the size of genomic fragments of DNA that include the sequence of interest.

Once you know the size (or size range) of restriction digest fragments that contain the DNA you want to study, you are ready to:

a) run another gel of digested genomic DNA.

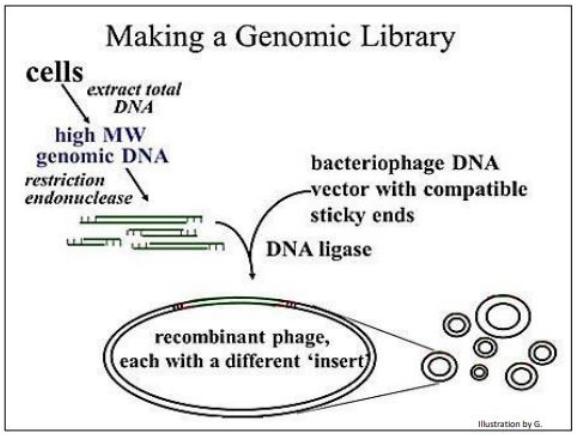
b) cut out the piece of gel containing the fragments that 'lit up' with your probe in the autoradiograph.

c) remove (elute) the DNA from the gel piece into a suitable buffer

d) prepare the DNA for insertion into (recombination with) a genomic cloning vector

B. Recombining Size-Restricted Genomic DNA with Phage DNA

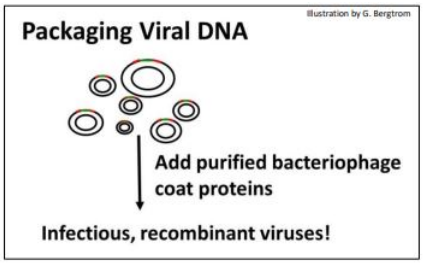
After elution of restriction digested DNA fragments of the right size range from the gels, the DNA is mixed with compatibly digested phage DNA at concentrations that favor the formation of H-bonds between the ends of the phage DNA and the genomic fragments. Addition of DNA ligase covalently links the recombined DNA molecules. These steps are abbreviated in the illustration below.



The recombinant phage that are made next will contain sequences that become the genomic library.

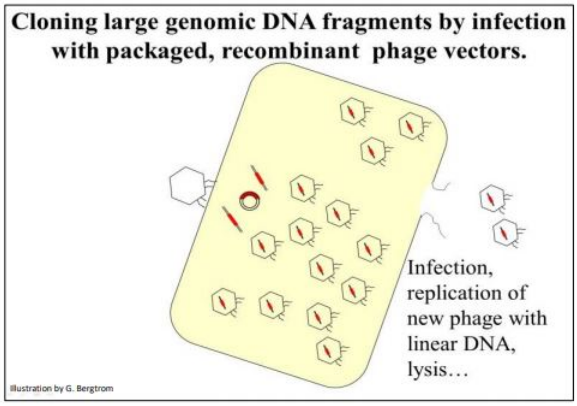
C. Creating Infectious Viral Particles with Recombinant Phage DNA

The next step is to **package** the recombined phage DNA with added purified viral coat proteins to make infectious phage *particles* (below)



269 Genomic Libraries: Make and Package Recombinant Phage DNA

Packaged phage are added to a culture tube full of host bacteria (typically *E. coli*). After infection, the recombinant DNA enters the cells where it replicates and directs the production of new phage that eventually lyse the host cell (illustrated below).



The recombined vector can also be introduced directly into the host cells by **transduction** (which is to phage DNA what transformation is to plasmid DNA). Whether by infection or *transduction*, the recombinant phage DNA ends up in host cells which produce new phage that eventually lyse the host cell. The released phages go on to infect more host cells until all cells have lysed. What remains is a tube full of **lysate** containing cell debris and lots of recombinant phage particles.

270 Infect Host with Recombinant Phage to Make a Genomic Library

D. A Note About Some Other Vectors

We've seen that phage vectors accommodate larger foreign DNA inserts than plasmid vectors, and YACs even more…, and that for larger genomes, the goal is to choose a vector able to house larger fragments of 'foreign' DNA so that you end up screening fewer clones. Given a large enough eukaryotic genome, it may be necessary to screen more than a hundred thousand clones in a phage-based genomic library. Apart from size-selection of genomic fragments before inserting them into a vector, selecting the appropriate vector is just as important. The table below lists commonly used vectors and the sizes of inserts they will accept.

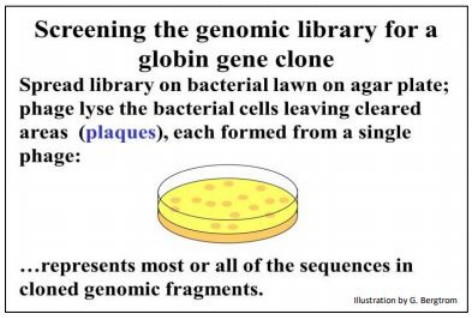
Vector Type	Insert Size (thousands of bases)
Plasmids	up to 15
Phage Lambda (λ)	up to 25
Cosmids	up to 45
Bacteriophage P1	70 to 100
P1 artificial chromosomes (PACs)	130 to 150
Bacterial artificial chromosomes (BACs)	120 to 300
Yeast artificial chromosomes (YACs)	250 to 2000

Click on the links to these vectors to learn more about them. We will continue this example by screening a phage lysate genomic library for a recombinant phage with a genomic sequence of interest.

E. Screening a Genomic Library; Titering Recombinant Phage Clones

A phage lysate is **titered** on a **bacterial lawn** to determine how many virus particles are present. A bacterial lawn is made by plating so many bacteria on the agar plate that they simply grow together rather than as separate colonies. In a typical **titration**, a lysate might be diluted 10-fold with a suitable medium and this dilution is further diluted 10-fold… and so on. Such **serial 10X dilutions** are then spread over bacterial (e.g., E. coli) lawns. What happens on such a culture plate?

Let's say that when 10 µl of one of the dilutions are spread on the bacterial lawn, they infect 500 E. coli cells on the bacterial lawn. After a day or so, there will be small clearings in the lawn called **plaques**…, 500 of them in this example. These are 500 tiny clear spaces on the bacterial lawn created by the lysis of first one infected cell, and then progressively more and more cells neighboring the original infected cell. Each plaque is thus a clone of a single virus, and each virus particle in a plaque contains a copy of the same recombinant phage DNA molecule (below).

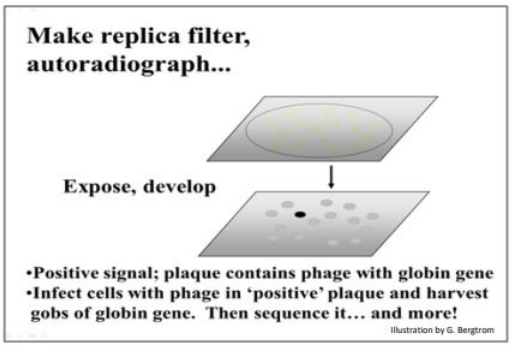


If you actually counted 500 plaques on the agar plate, then there must have been 500 virus particles in the 10 µl seeded onto the lawn. And, if this plate was the fourth dilution in a 10-fold serial dilution protocol, there must have been 2000 (4 X 500) phage particles in 10 µl of the original undiluted lysate.

F. Screening a Genomic Library; Probing the Genomic Library

In order to represent a *complete genomic library*, it is likely that many plates of such a dilution (~500 plaques per plate) will have to be created and then screened for a plaque containing a gene of interest. But, if only size-selected fragments were inserted into the

phage vectors in the first place, the plaques represent only a partial genomic library, requiring screening fewer clones to find the sequence of interest. For either kind of library, the next step is to make replica filters of the plaques. Replica plating of plaques is similar to making a replica filter bacterial colonies. While much of the phage DNA in a plaque is encased in viral proteins, there will also be DNA on the plaque replicas that were never packaged into viral particles. The filters can be treated to denature the latter DNA and then directly hybridized to a probe with a known sequence. In the early days of cloning, probes for screening a genomic library were usually an already isolated and sequenced cDNA clone, either from the same species as the genomic library, or from a cDNA library of a related species. After soaking the filters in a radioactively labeled probe, X-Ray film is placed over the filter, exposed and developed. Black spots will form where the film lay over a plaque containing genomic DNA complementary to the radioactive probe. In the example illustrated below, a globin cDNA might have been used to probe the genomic library (globin genes were among the first to be cloned!).



G. Isolating a Gene for Further Study

Cloned genomic DNA fragments are much longer than any gene of interest, and always longer than any cDNA from a cDNA library. They are also embedded in a genome that is thousands of times as long as the gene itself, making the selection of an appropriate vector necessary. If the genome can be screened among a reasonable number of cloned phage (~100,000 plaques for instance), the one plaque producing a positive signal on the autoradiograph would be further studied.

This plaque should contain the gene of interest. The next step is to find the gene within a genomic clone that can be as much a 20kbp long. The traditional strategy is to purify the cloned DNA, subject it to restriction endonuclease digestion, and separate of the digest particles by *agarose gel electrophoresis*. Using *Southern Blotting*, the separated DNA fragments are denatured and blotted to a nylon filter. The filter is then probed with the same tagged probe used to find the positive clone (plaque). The smallest DNA fragment containing the gene of interest can itself be *subcloned* in a suitable vector, and grown to provide enough DNA for further study of the gene.

271 Screen a Genomic Library, Pick and Grow a Phage Clone

This page titled 11.10: Genomic Libraries is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **15.4: Genomic Libraries** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

11.11: The Polymerase Chain Reaction (PCR)

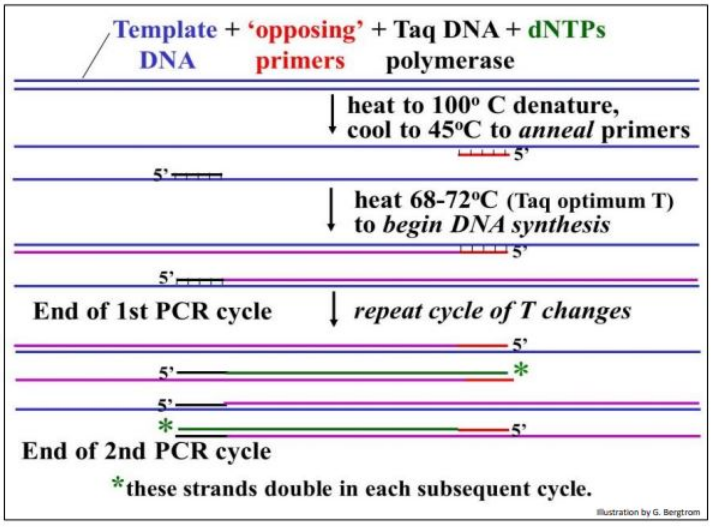
The polymerase chain reaction (PCR) can *amplify* a region of DNA from any source, even from a single cell's worth of DNA or from fragments of DNA obtained from a fossil. This amplification usually takes just a few hours, generating millions of copies of the desired *target* DNA sequence. The effect is to purify the DNA from surrounding sequences in a single reaction! Kary B. Mullis was awarded a Nobel Prize in 1993 for his development of PCR, which is now the basis of innumerable research studies of gene structure, function and evolution as well as applications in criminal forensics, medical diagnostics and other commercial uses. PCR is described in detail below.

A. PCR - the Basic Process

Typical PCR relies on knowing two bits of DNA sequence that will be used to design and synthesize short oligonucleotide sequences (*oligomers*) in the laboratory. The two oligomers are chosen to be complementary to sequences opposite strands of double-stranded DNA containing the gene to be studied. We say that the two oligomers face, or **oppose** each other. That just means that the 3' end of one oligomer faces the 3' end of the opposing oligomer. This way the two oligomers can serve as *primers* for the elongation replication of both strands of a double stranded target DNA sequence. Check out the link below for further explanation.

272 PCR: Design and Synthesize Opposing Oligonucleotide Primers

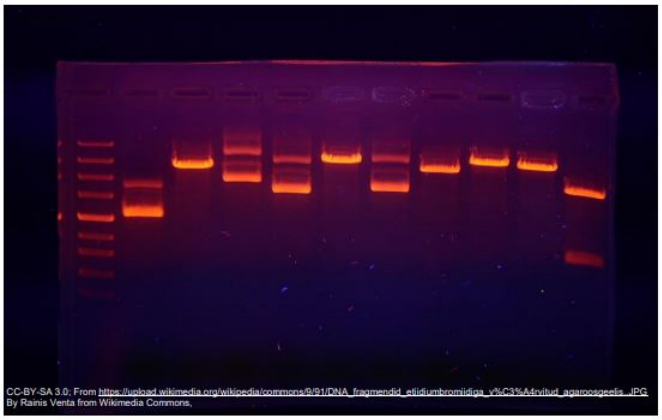
The first step in PCR is to add oligomer primers to the target DNA from which a gene (or other genomic sequence) is to be amplified. The mixture is then heated to denature the target DNA. The mixture is cooled to allow the primers to H-bond to complementary target DNA strands. Next, the four deoxynucleotide precursors to DNA (dATP, dCTP, dTTP and dGTP) are added along with a small amount of a DNA polymerase. New DNA strands will now lengthen from the oligonucleotide primers on the template DNAs. To make lots of the PCR product, this reaction cycle must be repeated many times. Therefore, after allowing elongation, the mixture is heated to denature (separate) all the DNA strands. When the mixture is again cooled, the oligomers again find complementary sequences with which to H-bond. Early versions of PCR originally relied on an *E. coli* DNA polymerase, which is inactivated by heating, and so had to be re-added to the PCR mixture for each elongation cycle. Just as with DNA sequencing, researchers very quickly switched to the heat-stable **Taq polymerase**, of *Thermus aquaticus*. The enzymes of *T. aquaticus* remain active at the very high temperatures at which these organisms live. Since heating does not destroy the Taq polymerase *in vitro*, PCR, like DNA sequencing reactions, could be automated with programmable *thermocylers* that raised and lowered temperature required by the PCR reactions. There was no longer a need to replenish a DNA polymerase once the reaction cycles were begun. Thermocyling in a typical PCR amplification is illustrated below for the first two PCR cycles, the second of which, produces the first strands of DNA that will actually be amplified exponentially.



You can see from the illustration that the second cycle of PCR has generated the two DNA strands that will be templates for doubling and re-doubling the desired product after each subsequent cycle. A typical PCR reaction might involve 30 PCR cycles, resulting in a nearly exponential amplification of the desired sequence.

> **Challenge:**
>
> Starting with a pair of complementary target DNA molecules (after the 3rd PCR cycle), how many double stranded PCR products should you theoretically have at the end of all 30 PCR cycles?

The amplified products of PCR amplification products are in such abundance that they can easily be seen under fluorescent illumination on an ethidium bromide-stained agarose gel (below).



In this gel, the first lane (on the left) contains a *DNA ladder*, a mixture of DNAs of known lengths that can be used to estimate the size of the PCR fragments in the 3 rd and 4th lanes (the gel lane next to the ladder is empty). The two bright bands in lanes 3 and 4 are PCR products generated with two different oligomer primer pairs. PCRamplified DNAs can be sequenced and used in many subsequent studies.

B. The Many Uses of PCR

PCR-amplified products can be labeled with radioactive or fluorescent tags to serve as hybridization probes for

- screening cDNA or genomic libraries and isolation of clones.
- determining migration position on a Southern blot.
- determining migration position on a northern blot (a fanciful name for RNAs that are separated by size on gels and blotted to filter).
- and more!

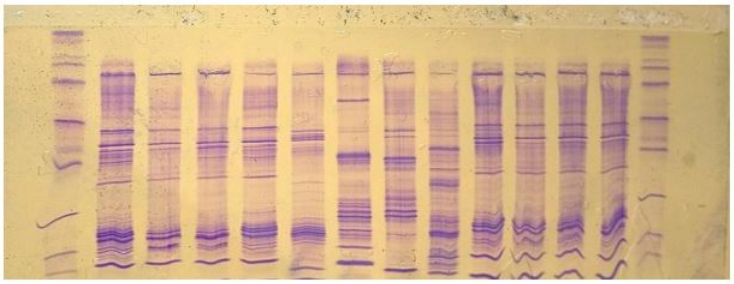
1. Quantitative PCR

We noted above that PCR has wide applications to research, medicine and other practical applications. A major advance was *Quantitative PCR*, applied to studies of differential gene expression and gene regulation. In Quantitative PCR, initial cDNAs are amplified to detect not only the presence, but also the relative amounts of specific transcripts being made in cells.

2. Forensics

Another application of PCR is in forensic science, to identify a person or organism by comparing its DNA to some standard, or control DNA. An example of one of these acrylamide gel DNA *fingerprints* is shown below.

DNA Fingerprint



Using this technology, it is now possible to detect genetic relationships between near and distant relatives (as well as to exclude such relationships), determine paternity, demonstrate evolutionary relationships between organisms, and on many occasions, solve recent and even 'cold-case' crimes. Click Sir Alec Jeffries to learn about the origins of DNA fingerprinting in real life …and on all those TV CSI programs! Check out here for a brief history of the birth of DNA fingerprinting, and to see how analysis of changes in gene activity that occur after death may even help ID criminals. For a video on DNA fingerprinting, click Alu and DNA fingerprinting. *Alu* is a highly repeated ~300bp DNA sequence found throughout the human genome. *Alu* sequences are *short interspersed elements*, or **SINES**, a retrotransposon we saw earlier. DNA fingerprinting is possible in part because each of us has a unique number and distribution of *Alu* SINEs in our genome. To read more about Alu sequences and human diversity, click Alu Sequences and Human Diversity.

Intriguing examples of the use of PCR for identification include establishing the identities of Egyptian mummies, the Russian Tsar deposed and killed during the Russian revolution (along with his family members), and the recently unearthed body of King Richard the 3rd of England. Variant PCR protocols and applications are manifold and often quite inventive! For a list, click Variations on Basic PCR.

274 The Power of PCR: Some Examples

3. Who are your Ancestors?

Tracing your ethnic, racial and regional ancestry is related to DNA fingerprinting, in that it relies on PCR amplification of genes and other DNA regions and comparison of these your sequences to distinguishing DNA markers in large sequence databases. The price of these services have come down, and as a result, their popularity has gone up in recent years. Typically, you provide spit or a salivary (buccal) swab to the service and they amplify and sequence the DNA in your samples. The analysis compares your DNA sequences to database sequences looking for patterns of ethnic and regional markers that you might share with the database(s). Based on these comparisons, you are provided with a (more…, or less) accurate map of your DNA-based ancestry. Folks who are spending around $100.00 (less when on sale!) often ask just how accurate are these analyses, and what do they actually mean. For example, what does it mean if your DNA says you are 5% native American? In fact, different services can sometimes give you different results! You can get some answers and explanations DNA Ancestry Testing.

This page titled 11.11: The Polymerase Chain Reaction (PCR) is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **15.5: The Polymerase Chain Reaction (PCR)** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

12: Evaluating Recombinant DNA

12.1: Cloning DNA - Plasmid Vectors

12.2: DNA Analysis - Gel Electrophoresis

12.3: Recombinant DNA Technology

12: Evaluating Recombinant DNA is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

12.1: Cloning DNA - Plasmid Vectors

Plasmids are Naturally Present in Some Bacteria

Many bacteria contain extra-chromosomal DNA elements called **plasmids**. These are usually small (a few 1000 bp), circular, double stranded molecules that replicate independently of the chromosome and can be present in high copy numbers within a cell. In the wild, plasmids can be transferred between individuals during bacterial mating and are sometimes even transferred between different species. Plasmids are particularly important in medicine because they often carry genes for pathogenicity and drug-resistance. In the lab, plasmids can be inserted into bacteria in a process called transformation.

Using Plasmids as Cloning Vectors

To insert a DNA fragment into a plasmid, both the fragment and the circular plasmid are cut using a restriction enzyme that produces compatible ends (Figure **12.1.1**). Given the large number of restriction enzymes that are currently available, it is usually not too difficult to find an enzyme for which corresponding recognition sequences are present in both the plasmid and the DNA fragment, particularly because most plasmid vectors used in molecular biology have been engineered to contain recognition sites for a large number of restriction endonucleases.

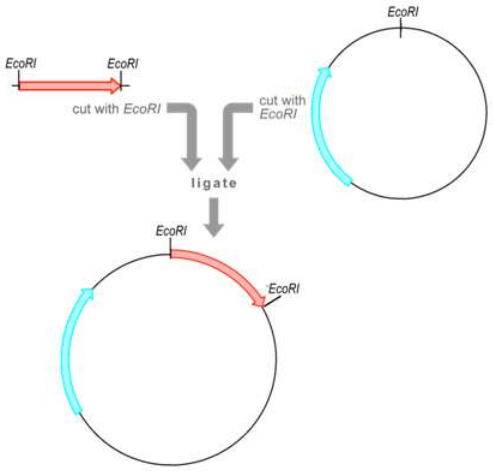


Figure **12.1.1**: Cloning of a DNA fragment (red) into a plasmid vector. The vector already contains a selectable marker gene (blue) such as an antibiotic resistance gene. (Original-Deyholos-CC:AN)

After restriction digestion, the desired fragments may be further purified or selected before they are mixed together with ligase to join them together. Following a short incubation, the newly ligated plasmids, containing the gene of interest are **transformed** into *E. coli*. Transformation is accomplished by mixing the ligated DNA with *E. coli* cells that have been specially prepared (i.e. made **competent**) to uptake DNA. Competent cells can be made by exposure to compounds such as $CaCl_2$ or to electrical fields (**electroporation**). Because only a small fraction of cells that are mixed with DNA will actually be transformed, a **selectable marker**, such as a gene for antibiotic resistance, is usually also present on the plasmid. After transformation (combining DNA with competent cells), bacteria are spread on a bacterial agar plate containing an appropriate antibiotic so that only those cells that have actually incorporated the plasmid will be able to grow and form colonies. This can then be picked and used for further study.

Molecular biologists use plasmids as **vectors** to contain, amplify, transfer, and sometimes express genes of interest that are present in the cloned DNA. Often, the first step in a molecular biology experiment is to **clone** (i.e. copy) a gene into a plasmid, then transform this recombinant plasmid back into bacteria so that essentially unlimited copies of the gene (and the plasmid that carries it) can be made as the bacteria reproduce. This is a practical necessity for further manipulations of the DNA, since most techniques of molecular biology are not sensitive enough to work with just a single molecule at a time. Many molecular cloning and recombination experiments are therefore iterative processes in which:

1. a DNA fragment (usually isolated by PCR and/or restriction digestion) is cloned into a plasmid cut with a compatible restriction enzyme
2. the recombinant plasmid is transformed into bacteria

3. the bacteria are allowed to multiply, usually in liquid culture
4. a large quantity of the recombinant plasmid DNA is isolated from the bacterial culture
5. further manipulations (such as site directed mutagenesis or the introduction of another piece of DNA) are conducted on the recombinant plasmid
6. the modified plasmid is again transformed into bacteria, prior to further manipulations, or for expression

An Application of Molecular Cloning: Production of Recombinant Insulin

Purified insulin protein is critical to the treatment of diabetes. Prior to ~1980, insulin for clinical use was isolated from human cadavers or from slaughtered animals such as pigs. Human-derived insulin generally had better pharmacological properties, but was in limited supply and carried risks of disease transmission. By cloning the human insulin gene and expressing it in *E. coli*, large quantities of insulin identical to the human hormone could be produced in fermenters, safely and efficiently. Production of recombinant insulin also allows specialized variants of the protein to be produced: for example, by changing a few amino acids, longer-acting forms of the hormone can be made. The active insulin hormone contains two peptide fragments of 21 and 30 amino acids, respectively. Today, essentially all insulin is produced from recombinant sources (Figure **12.1.2**), i.e. human genes and their derivatives expressed in bacteria or yeast.



Figure **12.1.2**: A vial of insulin. Note that the label lists the origin as "rDNA", which stands for recombinant DNA.(Flickr-DeathByBokeh-CC:AN)

This page titled 12.1: Cloning DNA - Plasmid Vectors is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.5: Cloning DNA - Plasmid Vectors** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.2: DNA Analysis - Gel Electrophoresis

A solution of DNA is colorless, and except for being viscous at high concentrations, is visually indistinguishable from water. Therefore, techniques such as **gel electrophoresis** have been developed to detect and analyze DNA (Figure 12.2.11).



Figure 12.2.11: Apparatus for agarose gel electrophoresis. A waterproof tank is used to pass current through a slab gel, which is submerged in a buffer in the tank. The current is supplied by an adjustable power supply. A gel (stained blue by a dye sometimes used when loading DNA on the gel) sits in a tray, awaiting further analysis, such as photography under a UV light source. (Flickr-457088634_585df11af5_o-pending)

This analysis starts when a solution of DNA is deposited at one end of a gel slab. This gel is made from polymers such as **agarose**, which is a polysaccharide isolated from seaweed. The DNA is then forced through the gel by an electrical current, with DNA molecules moving toward the positive electrode (Figure 12.2.12).

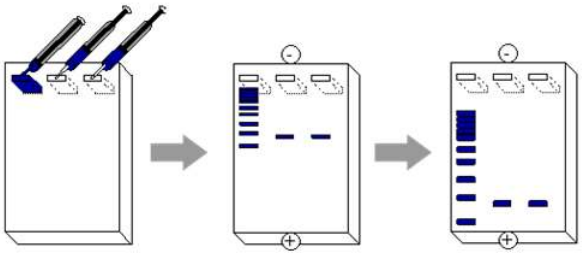


Figure 12.2.12: Agarose gel electrophoresis. DNA is loaded into wells at the top of a gel. A current is passed through the gel, pulling DNA towards the positively charged electrode. The DNA fragments are separated by size, with smaller fragments moving fastest towards the electrode. (Wikipedia-Magnus Manske_PD)

As it migrates, each piece of DNA threads its way through the pores, which form between the polymers in the gel. Because shorter pieces can move through these pores faster than longer pieces, gel electrophoresis separates molecules based on their size (length), with smaller DNA pieces moving faster than long ones. DNA molecules of a similar size migrate to a similar location in each gel, called a **band**. This feature makes it easy to see DNA after staining the DNA with a fluorescent dye such as **ethidium bromide** (Figure 12.2.13). By separating a mixture of DNA molecules of known size (**size markers**) in adjacent lanes on the same gel, the length of an uncharacterized DNA fragment can be estimated. Gel segments containing the DNA bands can also be cut out of the gel, and the size-selected DNA extracted and used in other types of reactions, such as sequencing and cloning.

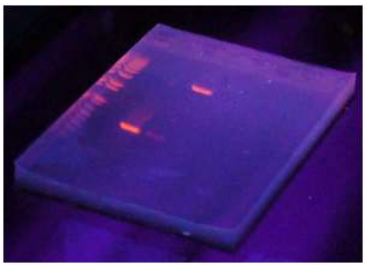


Figure 12.2.13: An agarose gel stained with ethidium bromide and illuminated by UV light. The stain associated with DNA is fluorescent. (Wikipedia-Transcontrol-GFDL)

This page titled 12.2: DNA Analysis - Gel Electrophoresis is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.6: DNA Analysis - Gel Electrophoresis** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

12.3: Recombinant DNA Technology

Learning Objectives

- Understand the importance of recombinant DNA technology.
- Learn isolation of DNA and its separation on an agarose gel.
- Understand restriction and ligase enzymes and their application in gene cloning.
- Understand vectors and their application in gene cloning and expression.
- Understand polymerase chain termination reaction (PCR).

Introduction

Recombinant DNA (rDNA) technology has resulted in breakthroughs in crop and animal biotechnology. The power of rDNA technology comes from our ability to study and modify gene function by manipulating genes and transform them into cells of plant and animals. To arrive at this several tools of molecular biology are used including, DNA isolation and analysis, molecular cloning, quantification of gene expression, determination of gene copy number, transformation of the appropriate host for replication or transfer into crop plants and analyses of transgenic plants.

Definition and background

Recombinant rDNA technology involves procedures for analyzing or combining DNA fragments from one or several organisms (Figure 1) including the introduction of the rDNA molecule into a cell for its replication, or integration into the genome of the target cell.

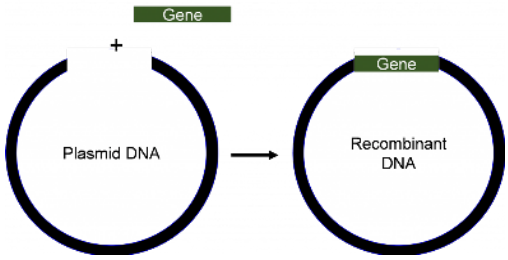


Figure 1. Recombinant DNA is made from combining DNA from different sources. Image by Walter Suza.

Advances in molecular biology in the early 1970s, including the success in creating, and transferring DNA molecules into cells, revolutionized both science and industry. The first genetically modified organisms were bacteria that made simple proteins of pharmaceutical interest, for example, insulin. As the technologies improved, other organisms including plants became amenable for improvement by rDNA technology. Table 1 provides important milestones in the development and application of rDNA technology.

Table 1.Events related to the development and application of recombinant DNA technology.

Event	Year
Mendel’s experiments published	1866
DNA discovered in cell	1869
Mutation of genes by x-rays	1927
One gene-one enzyme hypothesis	1941

DNA is identified as the genetic material

1944

Structure of DNA determined

1953

Ribosomes synthesize protein

1954

Function of mRNA proposed

1961

Genetic code determined

1961-64

Isolation of a restriction enzyme

1970

Recombinant DNA techniques developed

Early 1970s

Isolation of a single copy gene from higher eukaryote

1977

Rapid method of DNA sequencing developed

1977

Plant transformation

1983

Field testing of transformed plants

ca. 1986

Release of engineered plants to general public in the US

1995-96

Transformation of cells with rDNA produces organisms called bioengineered or genetically modified organisms (GMOs). The GMOs contain new traits from another organism. The first GMOs were *Escherichia coli* cells that were transformed with genes from human to produce various proteins for pharmaceutical purposes.

Isolation of DNA, restriction digestions and its separation on an agarose gel:

Preparation of DNA

For recombinant DNA procedures to work, a pure DNA sample must be obtained. The challenge is that plant and animal cells produce numerous other compounds that often act as contaminants and may inhibit cloning or sequencing of the DNA. Also, tissues and organs from the same plant, or different plants often contain different composition of metabolites, for example, proteins, lipids, and carbohydrates. These compounds must be separated from the DNA during isolation. To achieve this, scientists take advantage of the chemical and physical properties of different molecules inside the cell. For example, DNA is negatively charged, making it soluble in aqueous solution. However, the polar sugar phosphate groups of the DNA are repelled by non-polar solutions. Therefore, the final step in many DNA purifications protocols involves precipitation using alcohol. Other compounds, for example, proteins can also be easily separated from DNA by altering the concentration of salt in the extraction buffer.

Digestion of DNA with restriction endonucleases

Restriction endonucleases are a group of enzymes derived (primarily) from bacteria. Although there are several different types of **restriction enzymes**, those most useful for rDNA technology recognize specific short sequences in DNA and cleave the DNA at that site to produce cohesive (sticky) or blunt-ended fragments (Figure 2).

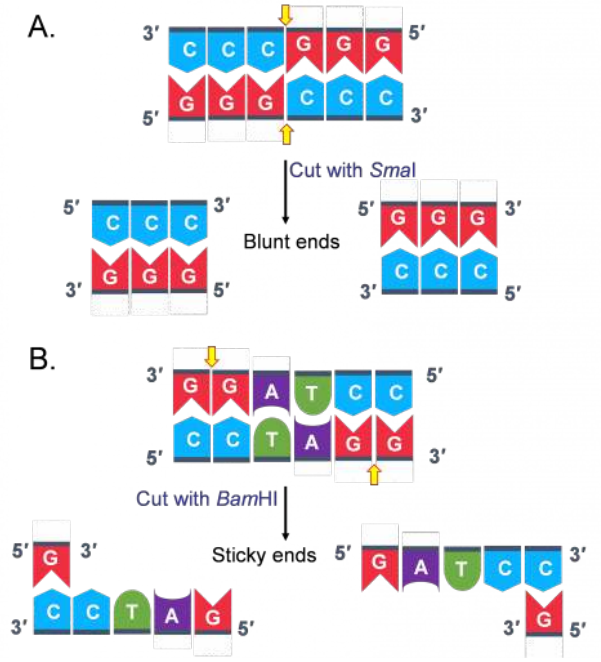


Figure 2. Example of how restriction enzymes cut DNA. (A) Treating the DNA with SmaI results in fragments with blunt ends. (B) Whereas treatment with BamHI produces fragments with “cohesive” or “sticky” ends. Image by Walter Suza.

More than 500 different restriction enzymes have been identified and can be purchased commercially. Thus, one may ask the question, how often does a restriction enzyme cut within a genomic sequence? It is not possible to give an exact answer for this question. However, let us assume that there are 4 bases on any strand of DNA. This means the probability of detecting an A (adenine) at a particular location is 1/4. Now, since most restriction enzymes recognize specific sequences of 6 bases long, the probability of finding such a site is (1/4)6 = 1 site in every 4,096 base pairs (bp). Assuming you have isolated genomic DNA from maize, and you want to digest it with a restriction enzyme that cuts every 4,096 (4,100) bp, how many fragments will you obtain? To answer this question, you need to have an idea of the size of the genome of the plant you are working with. The approximate size of the maize genome is 2,500,000,000 bp. Thus, using an enzyme that cuts every 4,100 nucleotides one would expect to obtain 2,500,000,000 bp/4,100 bp, approximately 610, 000 fragments. Note that nucleotide distribution is not always random and thus frequency may be different for given DNA. Also, methylation of specific bases in genomic DNA can prevent cleavage at some site.

Separation of digested genomic DNA on agarose gel

The only physical features of nucleic acid fragments that are routinely used for their characterization are their size and nucleotide sequence. Molecular weight of DNA is most conveniently evaluated by electrophoresis in agarose gels. Agarose forms a gel by hydrogen bonding when cooled from the melted state. This gel, interwoven network of agarose chains, interferes with the movement of DNA through the gel (see **Lesson on PCR and Gel Electrophoresis**). Pore size, which affects rate of movement of DNA fragments of a given size, depends on the concentration of agarose. The gel is submerged in electrolyte solution; sample is loaded into wells on one end and current is applied to facilitate movement of DNA fragments. Since DNA is negatively charged it will migrate in the electrical field. Fragments separate according to size. The distance of the migration in each time is proportional to 1/log MW. Following gel electrophoresis DNA can be visualized by staining with ethidium bromide (see **Lesson on PCR and Gel Electrophoresis**) or other DNA stains.

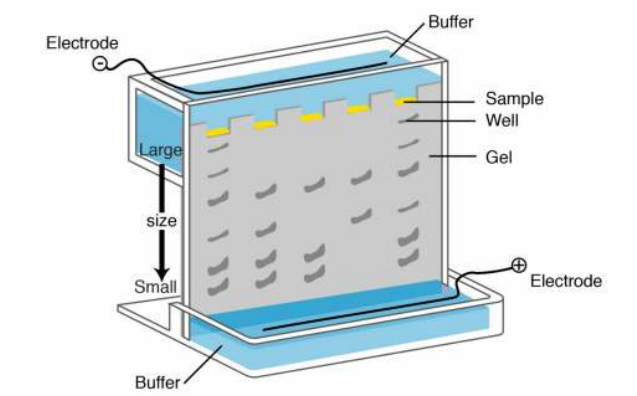


Figure 3. Gel electrophoresis is used to separate nucleic acid (DNA and RNA) or protein molecules by their size. For DNA, the sample is placed in an electric field and the negatively charged DNA will migrate to the positive electrode. Migration speed depends on the size of the DNA fragment. Use of semisolid matrix such as agarose or polyacrylamide gel facilitates separation of the DNA molecules. Image from NIH-NHGRI.

Tools for gene (DNA) cloning

Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) is a method by which millions of copies of a DNA fragment are produced in a test tube in a matter of minutes or a few hours. The basic steps in PCR reaction were discussed in the Lesson on PCR and Gel Electrophoresis.

Once the sequence of a particular gene is known, it becomes possible to use PCR to isolate that gene from any DNA sample. Recall in **Lesson on Gene Transcription** you learned that at the genomic level a gene is made up of regulatory sequences, coding, and non-coding sequences. Thus, if the goal is to use PCR to isolate both regulatory and non-regulatory sequences of a gene, the approach would be to use DNA as the starting material.

In cloning genes by PCR, restriction enzyme sites are added at the 5′-end of the primers to facilitate cloning of PCR fragments. A few additional nucleotides (~6 nucleotides) added at the 5′-end of the restriction sites to facilitate restriction digestion of PCR products prior to cloning in a plasmid vector. Alternatively, PCR products may be cloned directly into a T-vector without restriction digestions in *E. coli* (read more about Promega cloning vector systems). After cloning into *E. coli*, the fragment is analyzed by sequencing and then sub-cloned into suitable vectors for expression studies.

Like all other biochemical processes, DNA synthesis by PCR is not a perfect process, and occasionally the polymerase enzyme will add an incorrect base to the growing DNA strand. In the context of DNA replication in a cell, the errors are corrected by the DNA polymerase, this is called “proofreading”. Commercially available polymerases may or may not have proofreading capability.

Another important consideration in PCR analysis is contamination.

Minor contamination of the starting material can have serious consequences. Recall that minute amounts of starting DNA can be amplified to millions of copies through PCR. If one inadvertently (or carelessly) mixed DNA from two different sources, the results will be confounding making it impossible to distinguish lines and may cost a laboratory time and money. Ensure that proper procedures are followed in preparing PCR assays. One common source of contamination in plant biology is the products from previous amplification processes. A completed PCR reaction will contain millions of copies of amplified fragments so that even a minute droplet or aerosol from a pipette tip will contain an enormous number of amplifiable molecules. It is always essential to run negative controls which will reveal the presence of contaminating DNA in your PCR assays.

Cloning vector definition and requirements

A cloning vector is a specialized DNA sequence that can enter a living cell and provide means for detection of its presence to a researcher by conferring a selectable property on the host cell (e.g., resistance to antibiotics), and possess means for self-replication. A vector must also possess easily distinguishable physical traits, such as size, or shape, to allow purification away from the host cell’s genome.

Ligase enzyme and gene cloning

Cutting and joining together of vector and DNA fragments from different origins results in rDNA. Recall that restriction endonucleases are used to cut DNA. To join DNA molecules together, an enzyme called DNA ligase is used. The enzyme DNA ligase is used to seal together restriction fragments by forming new phosphodiester bonds. The ligated vector and DNA fragment can now be transformed into a host cell for replication and expression.

The transformation of *E. coli* takes several steps. First, a gene of interest is inserted into a plasmid that contains a selectable marker usually encoding for resistance to an antibiotic. Second, the plasmid construct containing the gene of interest is transformed into bacterial cells by briefly exposing the mixture of ligated plasmid-DNA fragment (rDNA molecule) and bacterial cells to cold (0^oC) and heat ($37\text{-}42^oC$). The next step is to grow the transformed cells on selection media containing an antibiotic. Only the cells that have been transformed with the plasmid containing the gene of interest and the marker for resistance to the antibiotic will survive. In addition to using an antibiotic, plasmid vector systems that contain the *lacZ* gene encoding β-galactosidase allow for easier selection of positive colonies that may harbor the rDNA molecule of interest.

Types of cloning vectors

An example of a cloning vector is a **plasmid** (Figure 4), defined as an autonomously replicating extra chromosomal circular DNA which is faithfully passed on to progeny. Plasmids are double stranded circular DNA and range in size from about 1 kb – 200 kb. The most useful for cloning are 2 – 10 kb because smaller plasmids are easier to manipulate and usually produce higher copy numbers when grown in bacterial host cells.

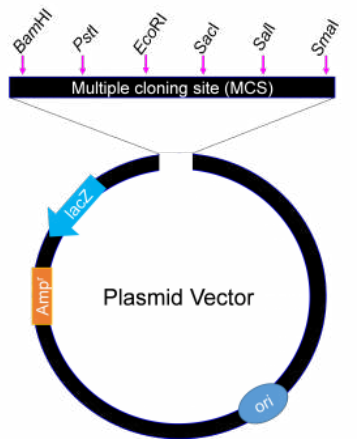


Figure 4. Basic map of a plasmid vector. The MCS contains several restriction enzymes sites and few are shown here as example. Image by Walter Suza.

Generally, no more than 10 kb is cloned into plasmids. For cloning large DNA fragments with high efficiency, a vector called **bacteriophage lambda** is used. Large chromosomal DNA fragments close to 23 kb are stable when introduced into a lambda phage vector.

mRNA as starting material for gene cloning

Recall in the **Lesson on Transcription** you learned about synthesis of RNA from DNA through the process of **transcription**. Transcription is an important step in gene expression. The mRNA produced can be isolated and "copied" back to DNA by a process called **reverse-transcription** (Figure 5). The first step of reverse transcription mimics a strategy used by retroviruses (e.g., HIV) that have RNA genomes. As part of their gene-transmission package, retroviruses also contain an enzyme called RNA-dependent DNA Polymerases, commonly referred to as **reverse transcriptase**. After infecting a host cell the retrovirus uses its reverse transcriptase to copy its single stranded RNA genome into a strand of complementary DNA (cDNA). The reverse transcriptase then synthesizes the second DNA strand from the first strand to make a double stranded-DNA copy which integrates into the host genome.

If only a gene's coding sequence is required, isolating the gene from cDNAs would be the strategy. It is important that cDNAs are synthesized from tissues expressing the gene of interest. Thus, prior knowledge of where the gene is functional is important in constructing the cDNA molecules for cloning the gene of interest.

The cDNAs produced *in vitro* (Figure 4) can be used for PCR analysis, similar to chromosomal DNA. The combination of reverse-transcription and PCR (RT-PCR) is a valuable tool in gene cloning and quantification of mRNA.

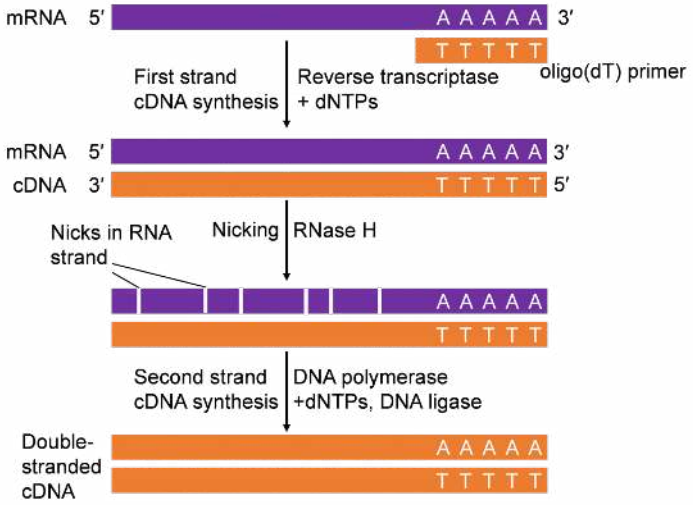


Figure 5. A population of mRNA isolated from plant tissues is combined with oligo(dT) primers that anneal to the poly(A) tail of the mRNAs to initiate the reverse transcription of cDNA from mRNA template with dNTPs. The result is hybrid molecules (mRNA-DNA). Treatment with RNAse H causes degradation of the mRNA leaving an intact single stranded DNA (first strand). DNA polymerase synthesizes the second strand by adding complementary dNTPs to the growing chain. The reagents for RT usually come in pre-mixed kits making such an assay easy to carry out routinely in most laboratories. Most mRNAs from cells in tissues analyzed are converted to cDNAs. Thus, using primers specific to a sequence, that double stranded cDNA can be amplified by PCR for various purposes for example, gene cloning. Image by Walter Suza.

Lesson Summary

Recombinant DNA technology has contributed significantly to development of agricultural biotechnology. Transformation of cells with rDNA produces organisms called bioengineered or genetically modified organisms. The tools for plant rDNA technology include, vectors, restriction enzyme, ligation enzymes, bacterial hosts, methods to isolate and multiply nucleic acids, methods to quantify nucleic acids, Agrobacterium as a vector to insert foreign DNA into plants.

Lesson Activities

Activity 1

Below is a hypothetical DNA fragment containing restriction sites for EcoRI and BamHI.

1. How many fragments will be produced after cutting the DNA with BamHI? What size will these fragments be in bp?
2. How many fragments will be produced after cutting the DNA with both BamHI and EcoRI at the same time? What size will these fragments be in bp?
3. How many fragments will be produced after cutting the DNA with EcoRI? What size will these fragments be in bp?

Activity 2

1. You need to clone an EcoRI fragment in the EcoRI site of an expression plasmid vector in E coli. The fragment is 1,809 bp long. There is a BamHI site in the fragment at 1,204 bp. In the plasmid vector, there is a single BamHI site at the promoter region for expression of the gene in E. coli. The BamHI site is 100 bp upstream of EcoRI cloning site.

Following cloning the EcoRI fragment you will get two classes of clones, only one of which will produce the protein. Describe the approach you will take to identify the correct class of clones for expression of the gene in E. coli.

This page titled 12.3: Recombinant DNA Technology is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Patricia Hain, & Patricia Hain (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.11: Recombinant DNA Technology** by Donald Lee, Marjorie Hanneman, Patricia Hain, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

13: Detecting Genes and Gene Products

13.1: DNA Analysis- Blotting and Hybridization

13: Detecting Genes and Gene Products is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

13.1: DNA Analysis- Blotting and Hybridization

Bands of DNA in an electrophoretic gel form only if most of the DNA molecules are of the same size, such as following a PCR reaction, or restriction digestion of a plasmid. In other situations, such as after restriction digestion of chromosomal (genomic) DNA, there will be a large number of variable size fragments in the digest and it will appear as a continuous smear of DNA, rather than distinct bands. In this case, it is necessary to use additional techniques to detect the presence of a specific DNA sequence within the smear of DNA separated on an electrophoretic gel. This can be done using a “Southern Blot”.

Southern Blots

A **Southern blot** (also called a **Southern Transfer**) is named after Ed Southern, its inventor. In the first step, DNA is digested with restriction enzymes and separated by gel electrophoresis (as discussed above). Then a sheet or **membrane** of nylon or similar material is laid under the gel and the DNA, in its separated position (bands or smear), is transferred to the membrane by drawing the liquid out of the gel, in a process called **blotting** (Figure **13.1.1**). The blotted DNA is usually covalently attached to the nylon membrane by briefly exposing the blot to UV light. Transferring the DNA to the sturdy membrane is necessary because the fragile gel would fall apart during the next two steps in the process. Next, the membrane is bathed in a solution to **denature** (double stranded made single stranded) the attached DNA. Then a **hybridization** solution containing a small amount of single-stranded **probe** DNA that is complementary in sequence to a target molecule on the membrane. This probe DNA is labeled using fluorescent or radioactive molecules, and if the hybridization is performed properly, the probe DNA will form a stable duplex only with those DNA molecules on the membrane that are exactly complementary to it. Then, the unhybridized probe is washed off and remaining radioactive or fluorescent signal will appear in a distinct band when appropriately detected. The band represents the presence of a particular DNA sequence within the mixture of DNA fragments.

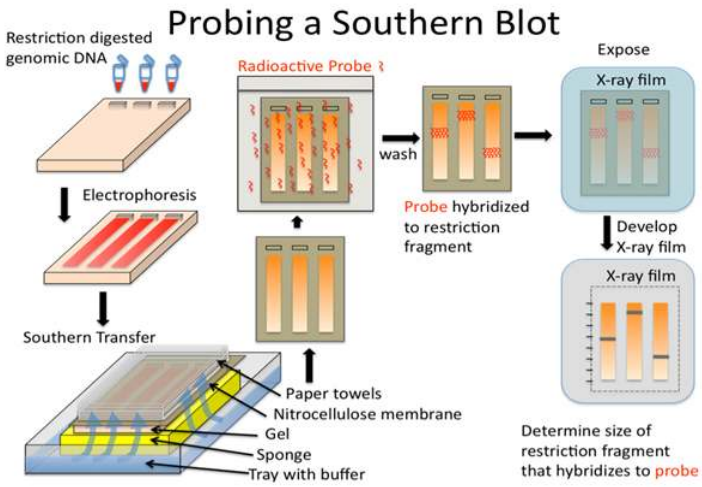


Figure **13.1.1**: A diagram of Southern blotting. Genomic DNA that has been digested with a restriction enzyme is separated on an agarose gel, then the DNA is transferred from the gel to a nylon membrane (grey sheet) by blotting. The DNA is immobilized on the membrane, then probed with a radioactively labeled DNA fragment that is complementary to a target sequence. After stringent washing, the blot is exposed to X-ray film to detect what size frament the probe is bound. In this case, the probe bound to different-sized fragments in lanes 1, 2, and 3. In the last image the orange represent the position of the digested DNA, but it is not actually present on the X-ray film. (Original-J. Locke-CC:AN)

The probe is sequence specific (requires complementarity). However, variation in hybridization temperature and washing solutions can alter the **stringency** of the probe. At maximum stringency (higher temperature) hybridization conditions, probes will only hybridize with the exact target sequences that are perfectly complementary (maximum number of hydrogen bonds). At lower temperatures, probes will be able to hybridize to targets to which they do not match exactly, but only are roughly complementary for part of the sequence.

Southern blotting is useful not only for detecting the presence of a DNA sequence within a mixture of DNA molecules, but also for determining the size of a restriction fragment in a DNA sample. Southern blots are useful for detecting fragments larger than those normally amplified by PCR, and when trying to detect fragments that may be only distantly related to a known sequence. Applications of Southern blotting will be discussed further in the context of molecular markers in a subsequent chapter. Southern

blotting was invented before PCR, but PCR has replaced blotting in many applications because of its simplicity, speed, and convenience. Following the development of the Southern blot, other types of blotting techniques were invented.

Northern Blots

The **Northern blot** involves the size separation of RNA in gels like that of DNA. Because we wish to determine the native size of the RNA transcript (and because RNA is single stranded) no restriction enzymes are ever used. Because most RNA is single stranded and can fold into various conformations thorough intra-molecular base pairing, the electrophoresis separation is more haphazard and the bands are often less sharp, compared to that of double stranded DNA.

Western Blots

In a **Western blot,** protein is size separated on a gel (usually an acrylamide gel) before transferring to a membrane, which is then probed with an antibody that specifically binds to an antigenic site on the target protein. This antibody is then detected by other antibodies with some fluorescent or color production marker system. It will also give bands proportional to the amount and size of the target protein (Figure **13.1.2**).

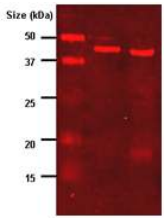


Figure **13.1.2**: Western Blot using a anti-lipoic acid primary antibody and an IR-dye labelled secondary antibody in Leishmania major extracts. (Wikipedia-TimVickers–CC:AS)

A comparison of all three blotting methods is shown in Figure **13.1.3**.

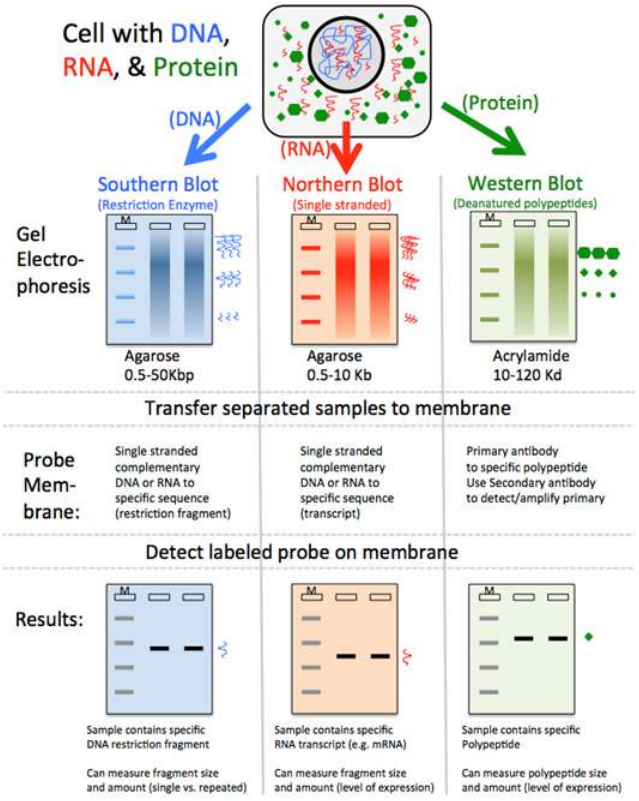


Figure **13.1.3**: Comparison of Southern, Northern, and Western blots. Size and amount of DNA, RNA, and polypeptides can be determined using similar blotting methods. DNA is in blue, RNA in red, and polypeptides in green. A marker lane is shown in the left of each gel to determine size. A eukaryote cell is shown, but the same methods can be applied to prokaryotes, too. (Original-Locke-CC:AN)

This page titled 13.1: DNA Analysis- Blotting and Hybridization is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.7: DNA Analysis- Blotting and Hybridization** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

14: DNA Replication

14.1: DNA Structure

14.2: Genes and Chromatin in Eukaryotes

14.3: Introduction

14.4: DNA Replication

14.5: DNA- The Genetic Material

14: DNA Replication is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

14.1: DNA Structure

A. Early Clues and Ongoing Misconceptions

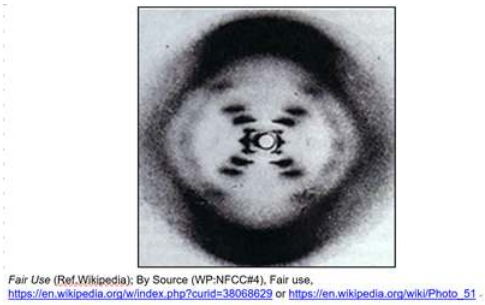
By 1878, a substance in the pus of wounded soldiers derived from cell nuclei (called *nuclein*) was shown to be composed of 5 bases (the familiar ones of DNA and RNA). The four bases known to make up DNA (as part of nucleotides) were thought to be connected through the phosphate groups in short repeating chains of four nucleotides. By the 1940s, we knew that DNA was a long polymer. Nevertheless, it was still considered too simple to account for genes (see above). After the Hershey and Chase experiments, only a few holdouts would not accept DNA as the genetic material. So, the question remaining was how such a "simple" molecule could account for all the genes, even in so simple an organism as a bacterium. The answer to this question was to lie at least in part in an understanding of the physical structure of DNA, made possible by the advent of ***X-Ray Crystallography***.

If a substance can be crystallized, the crystal will diffract X-rays at angles revealing regular (repeating) structures of the crystal. William Astbury demonstrated that high molecular weight DNA had just such a regular structure. His ***crystallographs*** suggested DNA to be a linear polymer of stacked bases (nucleotides), each nucleotide separated from the next by 0.34 nm. Astbury is also remembered for coining the term "*molecular biology*" to describe his studies. The term now covers as all aspects of biomolecular structure, as well as molecular functions (e.g. replication, transcription, translation, gene regulation…).

In an irony of history, the Russian biologist Nikolai Koltsov had already intuited in 1927 that the basis of genetic transfer of traits would be a "giant hereditary molecule" made up of "two mirror strands that would replicate in a semi-conservative fashion using each strand as a template". A pretty fantastic inference if you think about it since it was proposed long before Watson and Crick and their colleagues worked out the structure of the DNA double-helix!

B. Wilkins, Franklin, Watson & Crick

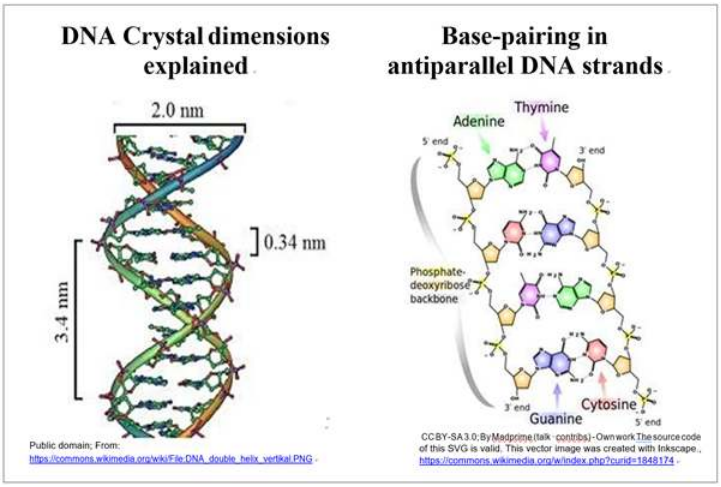
Maurice Wilkins, an English biochemist, was the first to isolate highly pure, high molecular weight DNA. Working in Wilkins laboratory, Rosalind Franklin was able to crystalize this DNA and produce very high-resolution X-Ray diffraction images of the DNA crystals. Franklin's most famous (and definitive) crystallography was "Photo 51", shown below.



Fair Use (Ref.Wikipedia); By Source (WP:NFCC#4), Fair use, https://en.wikipedia.org/w/index.php?curid=38068629 or https://en.wikipedia.org/wiki/Photo_51

This image confirmed Astbury's **0.34 nm** repeat dimension and revealed two more numbers, **3.4 nm** and **2 nm**, reflecting additional repeat structures in the DNA crystal. When James Watson and Francis Crick got hold of these numbers, they used them along with other data to build DNA models out of nuts, bolts and plumbing. Their models eventually revealed DNA to be a pair of **antiparallel complementary** of nucleic acid polymers…, shades of Koltsov's mirror-image macromolecules! Each strand is a string of nucleotides linked by *phosphodiester bonds*, the two strands held together in a *double helix* by complementary H-bond interactions.

Let's look at the evidence for these conclusions and as we do, refer to the two illustrations of the double helix below.



Recalling that Astbury's 0.34 nm dimension was the ***distance between successive nucleotides*** in a DNA strand, Watson and Crick surmised that the 3.4 nm repeat was a structurally meaningful 10-fold multiple of Astbury's number. When they began building their DNA models, they realized from the bond angles connecting the nucleotides that the strand was forming a helix, from which they concluded that the 3.4 nm repeat was the ***pitch*** of the helix, i.e., the distance of one complete turn of the helix. This meant that there were 10 bases per turn of the helix. They further reasoned that the 2.0 nm number might reflect the diameter of helix. When their scale model of a single stranded DNA helix predicted a helical diameter much less than 2.0 nm, they were able to model a ***double helix*** that more nearly met the 2.0 nm diameter requirement. In building their double helix, Watson and Crick realized that bases in opposing strands would come together to form H-bonds, holding the helix together. However, for their double helix to have a constant diameter of 2.0 nm, they also realized that the smaller **pyrimidine** bases, ***Thymine*** (**T**) and ***Cytosine*** (**C**), would have to H-bond to the larger **purine** bases, ***Adenine*** (**A**) and ***Guanosine*** (**G**).

Now to the question of how a "simple" DNA molecule could have the structural diversity needed to encode thousands of different polypeptides and proteins. In early studies, purified *E. coli* DNA was chemically hydrolyzed down to nucleotide monomers. The hydrolysis products contained nearly equal amounts of each base, reinforcing the notion that DNA was that simple molecule that could not encode genes. But Watson and Crick had private access to revealing data from Erwin Chargaff. Chargaff had determined the base composition of DNA isolated from different species, including *E. coli*. He found that the base composition of DNA from different species was not always *equimolar*, meaning that for some species, the DNA was not composed of equal amounts of each of the four bases (see some of this data below).

Base Compositions of DNA from Different Organisms

base		*human*	*yeast*	*fly*
purine	A	27%	21%	15%
pyrimidine	C	23%	32%	35%
purine	G	23%	32%	35%
pyrimidine	T	27%	21%	15%

The mere fact that DNA from some species could have base compositions that deviated from equimolarity put to rest the argument that DNA had to be a very simple sequence. Finally, it was safe to accept that to accept the obvious, namely that DNA was indeed the "stuff of genes".

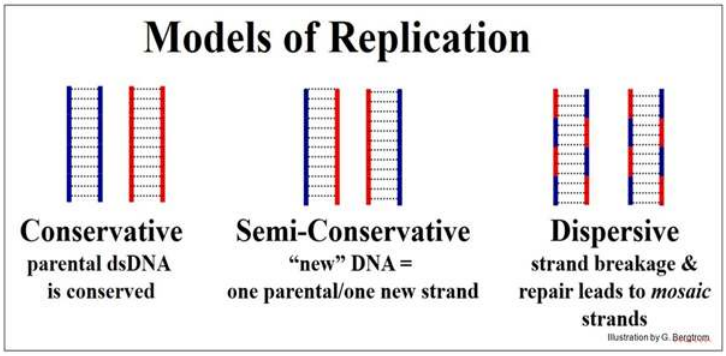
Chargaff's data also showed a unique pattern of base ratios. Although base compositions could vary between species, the **A/T** and **G/C** *ratio* was always one, for every species. Likewise the ratio of **(A+C)/(G+T)** and **(A+G)/(C+T)**. From this information, Watson and Crick inferred that **A** (a purine) would H-bond with **T (a pyrimidine)**, and **G** (a purine) would H-bond with **C** (a pyrimidine) in the double helix. When building their model with this new information, they also found H-bonding between the complementary bases would be maximal only if the two DNA strands were **antiparallel**, leading to the most stable structure of the double helix.

Watson and Crick published their conclusions about the structure of DNA in 1953 (Click here to read their seminal article: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Their article is also famous for predicting a semi-conservative mechanism of replication, something that had been predicted by Koltsov 26 years earlier, albeit based on intuition… and much less evidence!

Watson, Crick and Wilkins shared a Nobel Prize in 1962 for their work on DNA structure. Unfortunately, Franklin died in 1958 and Nobel prizes were not awarded posthumously. There is still controversy about why Franklin did not get appropriate credit for her role in the work. But she has been getting well-deserved, long-delayed recognition, including a university in Chicago named in her honor!

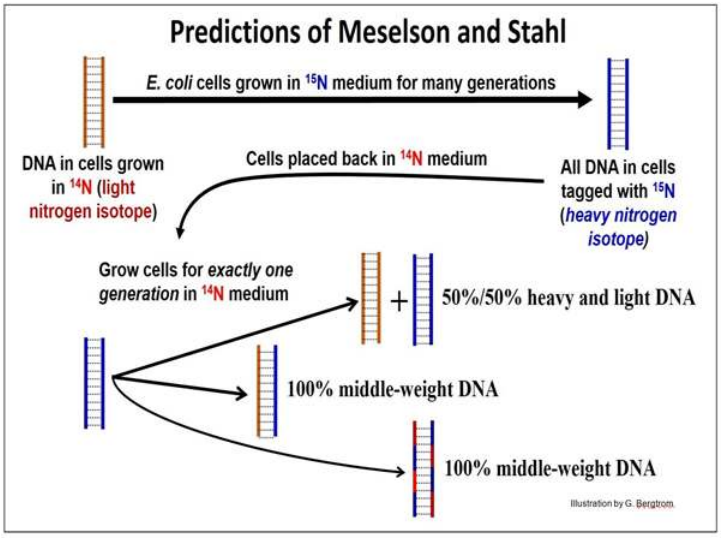
169 Unraveling the Structure of DNA

Confirmation of Watson & Crick's suggestion of semiconservative replication came from Meselson and Stahl's very elegant experiment, which tested the three possible models of replication shown below.



In their experiment, *E. coli* cells were grown for in medium containing 15N, a 'heavy' nitrogen isotope. After many generations, all of the DNA in the cells had become labeled with the heavy isotope. At that point, the 15N-tagged cells were placed back in medium containing the more common, 'light' 14N isotope and allowed to grow for exactly one generation.

Meselson and Stahl's predictions and their experimental design are shown below.



Meselson and Stahl knew that 14N-labeled and 15N-labeled DNA would form separate bands after centrifugation on *CsCl chloride density gradients*. They tested their predictions by purifying and centrifuging the DNA from the 15N-labeled cells grown in 14N medium for one generation. They found that this DNA formed a single band with a density between that of 15N-labeled DNA and 14N-labeled DNA. This result *eliminated* a conservative model of DNA replication (as Watson and Crick also predicted. That left two possibilities: replication was either semiconservative or dispersive. The dispersive model was *eliminated* when DNA isolated from cells grown for a 2nd generation on 14N were shown to contain two bands of DNA on the CsCl density gradients.

Chromosomes

We understood from the start of the 20th century that chromosomes contained genes. Therefore, it becomes necessary to understand the relationship between chromosomes, chromatin, DNA and genes. As noted earlier, chromosomes are a specialized, *condensed* version of chromatin, with key structural features shown below.

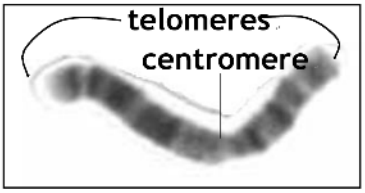
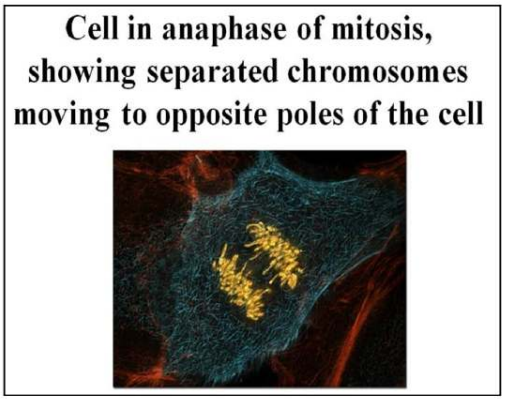


Illustration by G. Bergtrom

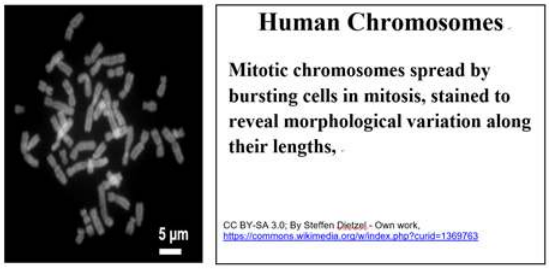
Public Domain. Adapted from National Human Genome Research Institute - Cropped from File:Human male karyotype high resolution.jpg. https://commons.wikimedia.org/w/index.php?curid=40985706 .

We now know that the compact structure of a chromosome prevents damage to the DNA during cell division. This damage can occur when forces on centromeres generated by mitotic or meiotic spindle fibers pull chromatids apart. As the nucleus breaks down during mitosis or meiosis, late 19th century microscopists saw chromosomes condense from the dispersed cytoplasmic background. These chromosomes remained visible as they separated, moving to opposite poles of the cell during cell division. Such observations of chromosome behavior during cell division pointed to their role in heredity. A computer- colorized cell in mitosis is shown below.

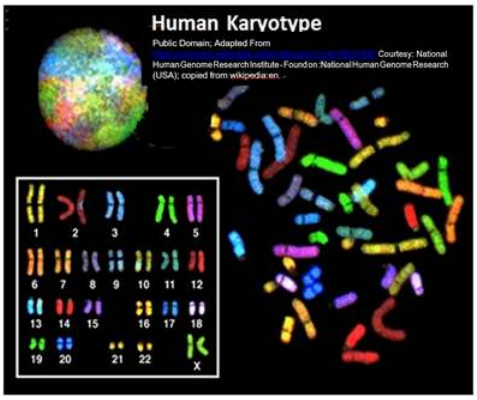


CC-BY-SA-NC; From: https://fascience.wikispaces.com/10+Cell+Growth+and+Division .

It is possible to distinguish one chromosome from another by ***karyotyping***. When cells in metaphase of mitosis are placed under pressure, they burst and the chromosomes spread apart. Such a chromosome spread is shown below.



By the early 1900s, the number, sizes and shapes of chromosomes were shown to be species-specific. What's more, a close look at chromosome spreads revealed that chromosomes came in morphologically matched pairs. This was so reminiscent of Gregor Mendel's paired hereditary factors that chromosomes were then widely accepted as the structural seat of genetic inheritance. Cutting apart micrographs like the one above and pairing the chromosomes by their morphology generates a ***karyotype***. Paired human homologs are easily identified in the colorized micrograph below.



Captured in mitosis, all dividing human cells contain 23 pairs of homologous chromosomes. The ***karyotype*** is from a female; note the pair of homologous sex ("X") chromosomes (lower right of the inset). X and Y chromosomes in males are not truly homologous. Chromosomes in the original spread and in the aligned karyotype stained with fluorescent antibodies to chromosome-specific DNA sequences, 'light up' the different chromosomes.

171 DNA, Chromosomes, Karyotypes & Gene Maps

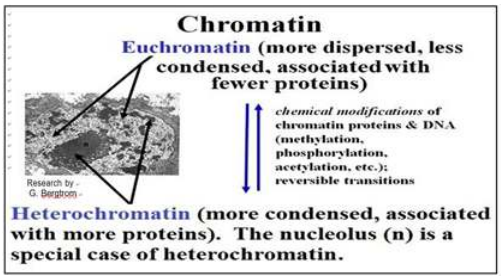
This page titled 14.1: DNA Structure is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **8.3: DNA Structure** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

14.2: Genes and Chromatin in Eukaryotes

Chromosomes and chromatin are a uniquely eukaryotic association of DNA with more or less protein. Bacterial DNA (and prokaryotic DNA generally) is relatively 'naked' – not visibly associated with protein.

The electron micrograph of an *interphase* cell (below) reveals that the chromatin can itself exist in various states of condensation.



Chromatin is maximally condensed during mitosis, forming chromosomes. During interphase, chromatin exists in more or less condensed forms, called ***Heterochromatin*** and ***euchromatin*** respectively. Transition between these chromatin forms involve changes in the amounts and types of proteins bound to the chromatin, and can that can occur during gene regulation, i.e., when genes are turned on or off. Active genes tend to be in the more dispersed euchromatin so that enzymes of replication and transcription have easier access to the DNA. Genes that are inactive in transcription are heterochromatic, obscured by additional chromatin proteins present in heterochromatin. We'll be looking at some experiments that demonstrate this in a later chapter.

We can define three levels of chromatin organization in general terms:

1. DNA wrapped around histone proteins form *nucleosomes* in a "beads on a string" structure.

2. Multiple nucleosomes coil (condense), forming 30 nm fiber (solenoid) structures.

3. Higher-order packing of the 30 nm fiber leads to formation of metaphase chromosomes seen in mitosis & meiosis.

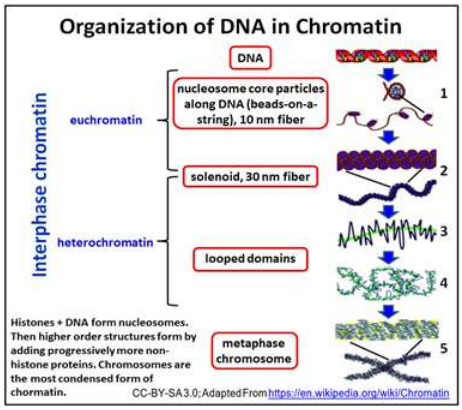
The levels of chromatin structure were determined in part by selective isolation and extraction of interphase cell chromatin, followed by selective chemical extraction of chromatin components. The steps are:

· Nuclei are first isolated from the cells.

· The nuclear envelope gently ruptured so as not to physically disrupt chromatin structure.

· the chromatin can be gently extracted by one of several different chemical treatments (high salt, low salt, acid...).

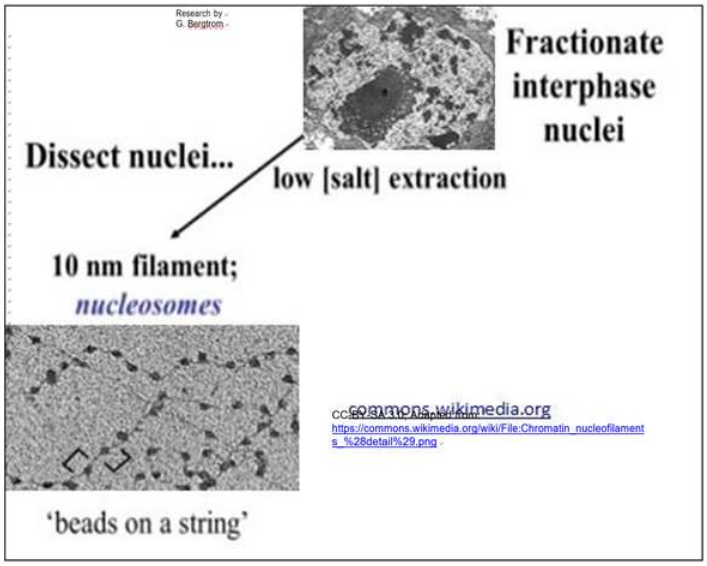
The levels of chromatin structure are illustrated below.



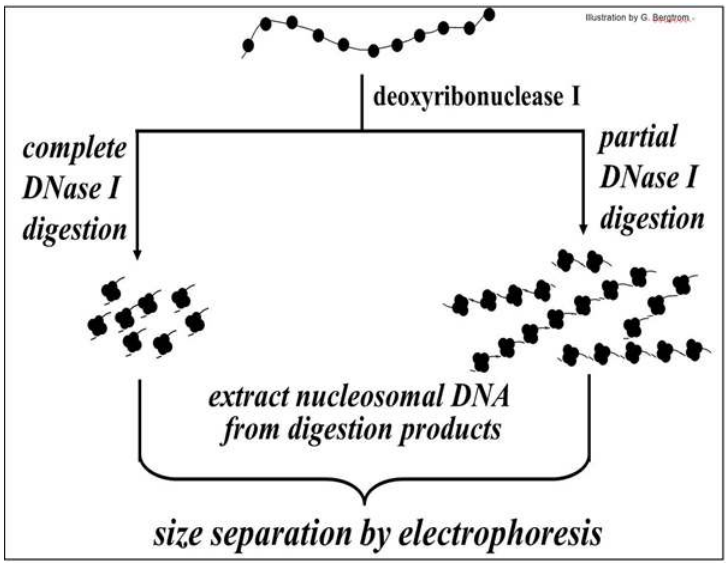
Salt extraction dissociates most of the proteins from the chromatin. When a low salt extract is centrifuged and the pellet resuspended, the remaining chromatin looks like *beads on a string*. DNA-wrapped *nucleosomes* are the beads, which are in turn linked by uniform lengths of metaphorical DNA "string' (**# 1** in the illustration above). A high salt chromatin extract appears as a

coil of nucleosomes, or *30 nm solenoid fiber* (**# 2** above). Other extraction protocols revealed other aspects of chromatin structure shown in #s **3** and **4** above. Chromosomes seen in metaphase of mitosis are the 'highest order', most condensed form of chromatin.

The 10 nm filament of nucleosome 'beads-on-a-string' remaining after a low salt extraction can be seen in an electron microscope as shown below.



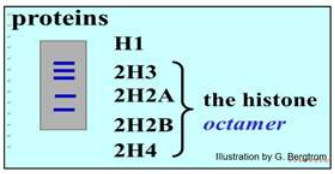
When these nucleosome necklaces were digested with the enzyme *deoxyribonuclease* (**DNAse**), the DNA between the 'beads' was degraded, leaving behind shortened 10nm filaments after a short digest period, or just single beads the beads after a longer digestion (below).



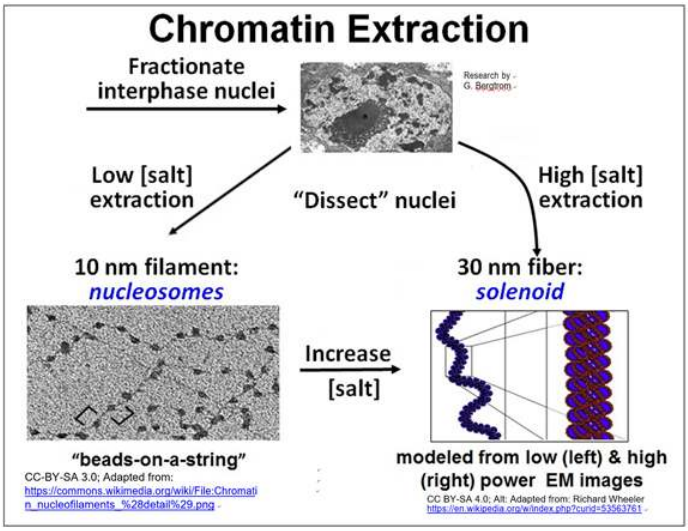
Roger Kornberg (son of Nobel Laureate Arthur Kornberg who discovered the first DNA polymerase enzyme of replication) participated in the discovery and characterization of nucleosomes while he was still a post-doc! Electrophoresis of DNA extracted from these digests revealed *nucleosomes* separated by a "linker" DNA stretch of about 80 base pairs. DNA extracted from the nucleosomes was about 147 base pairs long. This is the DNA that had been wrapped around the proteins of the nucleosome.

172 Nucleosomes-DNA & Protein

After separating all of the proteins from nucleosomal DNA, five proteins were identified (illustrated below).



Histones are basic proteins containing many *lysine* and *arginine* amino acids. Their positively charged side chains enable these amino acids bind the acidic, negatively charged *phosphodiester backbone* of double helical DNA. The DNA wraps around an octamer of histones (2 each of 4 of the histone proteins) to form the *nucleosome*. About a gram of histones is associated with each gram of DNA. After a high salt chromatin extraction, the structure visible in the electron microscope is the 30nm solenoid, the coil of nucleosomes modeled in the figure below.

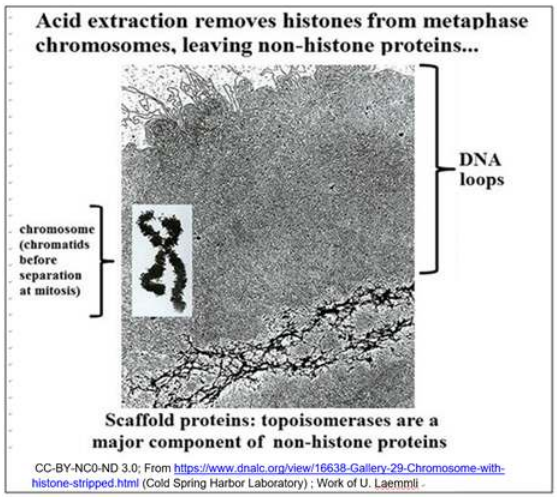


As shown above, simply increasing the salt concentration of an already extracted nucleosome preparation will cause the 'necklace' to fold into the 30nm solenoid structure.

173 Chromatin Structure: Dissecting Chromatin

As you might guess, an acidic extraction of chromatin should selectively remove the basic histone proteins, leaving behind an association of DNA with non-histone proteins. This proved to be the case. An electron micrograph of the chromatin remnant after an acid extraction of metaphase chromosomes is shown on the next page.

DNA freed of the regularly spaced histone-based nucleosomes, loops out, away from the long axis of the chromatin. Dark material along this axis is a protein scaffolding that makes up what's left after histone extraction. Much of this protein is *topoisomerase*, an enzyme that prevents DNA from breaking apart under the strain of replication.



174 Histones and Non-Histone Proteins

This page titled 14.2: Genes and Chromatin in Eukaryotes is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **8.4: Genes and Chromatin in Eukaryotes** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

14.3: Introduction

Replication begins at one or more **origins of replication** along DNA, where **helicase enzymes** catalyze unwinding of the double helix. DNA unwinding creates replicating bubbles, or **replicons**, with **replication forks** at either end. Making a new DNA strand starts with making an RNA **primer** with RNA nucleotides and **primase** enzymes. DNA nucleotides are then added to the 3'-ends of primers by one a **DNA polymerase**. Later, other DNA polymerases catalyze removal of the RNA primers and replacement of the hydrolyzed ribonucleotides with new deoxyribonucleotides. Finally, **DNA ligases** stitch together the fragments of new DNA synthesized at the replication forks. This complex mechanism is common to the replication of 'naked' prokaryotic DNA and of chromatinencased eukaryotic DNA, and must therefore have arisen early in the evolution of replication biochemistry. In this chapter, we look at the details of replication and the differences in detail between prokaryotic and eukaryotic replication that arise because of differences in DNA packing. As with any complex process with many moving parts, replication is error-prone. Therefore, we will also look at how the overall **fidelity** of replication relies mechanisms of **DNA repair** that target specific kinds of replication mistakes, or **mutations**. At the same time, lest we think that uncorrected errors in replication are always a bad thing, they usually *do not* have bad outcomes. Instead, they leave behind the very mutations that allow *natural selection* and the *evolution of diversity*.

Learning Objectives

1. Explain how Cairns interpreted his theta (Θ) images.
2. Compare and contrast the activities of enzymes required for replication.
3. Describe the order of events at an origin of replication and at each *replication fork.*
4. Compare *unidirectional* and *bidirectional* DNA synthesis from an origin of replication.
5. Outline the basic synthesis and proofreading functions of *DNA polymerase.*
6. Identify the major players and their roles in the initiation of replication.
7. Explain how Okazaki's experimental results were not entirely consistent with how both strands of DNA replicate
8. List the major molecular players (enzymes, etc.) that elongate a growing DNA strand.
9. List the non-enzymatic players in replication and describe their functions.
10. Describe how the structure of *telomerase* enables proper replication.
11. Compare the activities of topoisomerases 1 and 2.
12. Explain the reasoning behind the hypothesis of *processive replication.*
13. Compare and contrast the impacts of germline and somatic mutations.
14. Describe common forms of DNA damage.
15. List enzymes of replication that were adapted to tasks of DNA repair.
16. Explain why a DNA glycosylase is useful in DNA repair.
17. Explain why the connection between 'breast cancer genes' and DNA repair.

This page titled 14.3: Introduction is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **9.1: Introduction** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

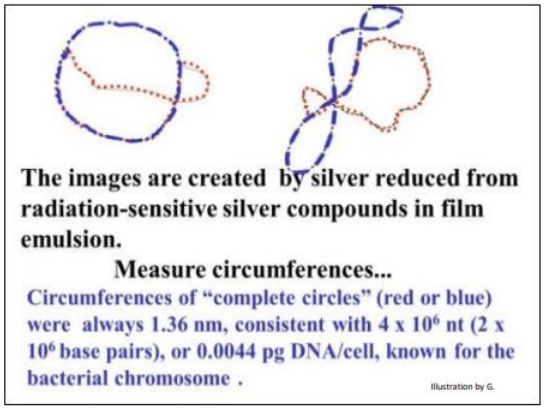
14.4: DNA Replication

As we've seen, DNA strands have directionality, with a 5' nucleotide-phosphate and a 3' deoxyribose hydroxyl end. This is even true for circular bacterial chromosomes…, if the circle is broken! Because the strands of the double helix are *antiparallel*, the 5' end of one strand aligns with the 3'end of the other at both ends of the double helix. The complementary pairing of bases in DNA means that the base sequence of one strand can be used as a template to make a new complementary strand. As we'll see, this structure of DNA created some interesting dilemmas for understanding the biochemistry of replication. The puzzlement surrounding how replication proceeds begins with experiments that visualize replicating DNA.

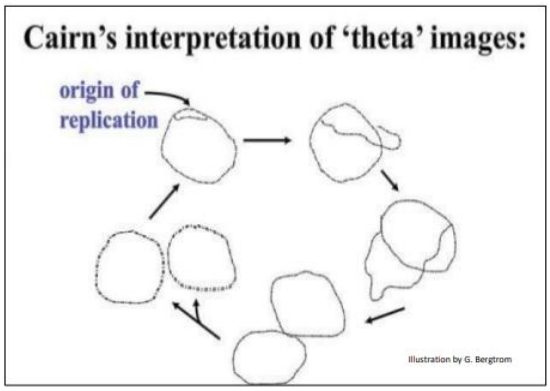
A. Visualizing Replication and Replication Forks

Recall the phenomenon of bacterial conjugation allowed a demonstration bacterial chromosomes were circular. In 1963, John Cairns confirmed this fact by direct visualization of bacterial DNA. He cultured *E. coli* cells for long periods on 3Hthymidine (3H-T) to make all of their cellular DNA radioactive. He then disrupted the cells gently to minimize damage to the DNA. The DNA released was allowed to settle and adhere to membranes. A sensitive film was placed over the membrane and time was allowed for the radiation to expose the film. After Cairns developed the autoradiographs, he examined the results in the electron microscope. He saw tracks of silver grains in the autoradiographs (the same kind of silver grains that create an image on film in old-fashioned photography). Look at the two drawings of his **autoradiographs** on the next page.

Cairns measured the length of the "silver" tracks, which usually consisted of three possible closed loops, or circles. The circumferences of two of these circles were always equal, their length closely predicted by the DNA content of a single, nondividing cell. Cairns therefore interpreted these images to be bacterial DNA in the process of replication. Cairns' autoradiographs and the measurements that led him to conclude that he was looking at images of bacterial circular chromosomes are illustrated below.



He arranged his autoradiograph images in a sequence (below) to make his point.

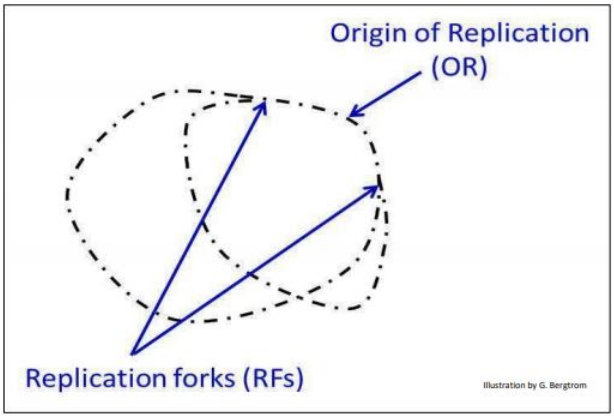


Because the replicating chromosomes looked (vaguely!) like the Greek letter **\(\theta \)**, Cairns called them **theta images**. He inferred that replication starts at a single **origin of replication** on the bacterial chromosome, proceeding around the circle to

completion.

175 Seeing *E.Coli* Chromosomes

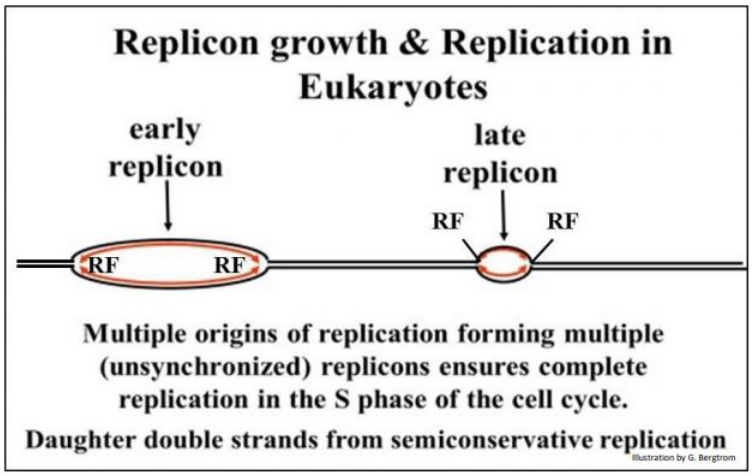
Subsequent experiments by David Prescott demonstrated **bidirectional replication**…, that replication did indeed begin at an origin of replication, after which the double helix was unwound and replicated in both directions, away from the origins, forming **two replication forks** (illustrated below).



176 Semiconservative Bidirectional Replication From Two RFs

Bacterial cells can divide every hour (or even less); the rate of bacterial DNA synthesis is about 2 X 106 base pairs per hour. A typical eukaryotic cell nucleus contains thousands of times as much DNA as a bacterium, and typical eukaryotic cells double every 15-20 hours. Even a small chromosome can contain hundreds or thousands of times as much DNA as a bacterium. It appeared that eukaryotic cells could not afford to double their DNA at a bacterial rate of replication! Eukaryotes solved this problem not by evolving a faster biochemistry of replication, but by using multiple origins of replication from which DNA synthesis proceeds in both directions. This results in the creation of multiple **replicons**.

Each replicon enlarges, eventually meeting other growing replicons on either side to replicate most of each linear chromosome, suggested in the illustration below.



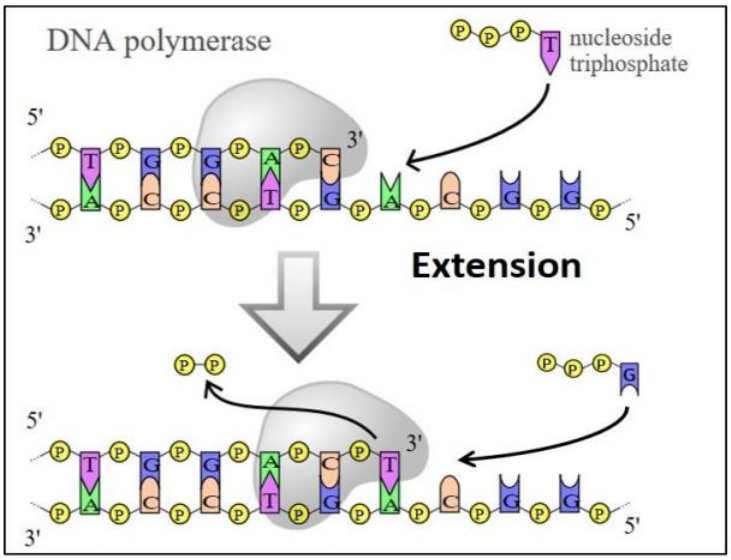
Before we consider the biochemical events at replication forks in detail, let's look at the role of DNA polymerase enzymes in the process.

177 Multiple Replicons in Eukaryotes

B. DNA Polymerases Catalyze Replication

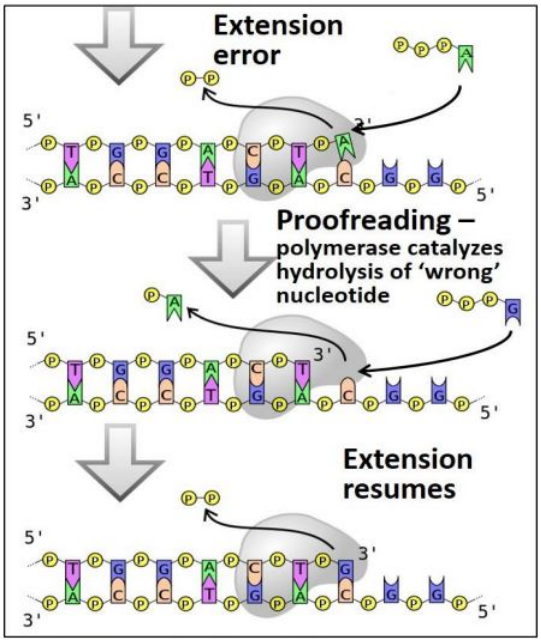
The first of these enzymes was discovered in *E. coli* by Arthur Kornberg, for which he received the 1959 Nobel Prize in Chemistry. Thomas Kornberg, one of Arthur's sons later found two more of DNA polymerases! All DNA polymerases require a template

strand against which to synthesize a new complementary strand. They all grow new DNA by adding to the 3' end of the growing DNA chain in successive condensation reactions. And finally, all DNA polymerases also have the odd property that they can only add to a pre-existing strand of nucleic acid, raising the question of where the 'preexisting' strand comes from! DNA polymerases catalyze the formation of a phosphodiester linkage between the end of a growing strand and the incoming nucleotide complementary to the template strand. The energy for the formation of the phosphodiester linkage comes in part from the hydrolysis of two phosphates (pyrophosphate) from the incoming nucleotide during the reaction. While replication requires the participation of many nuclear proteins in both prokaryotes and eukaryotes, DNA polymerases perform the basic steps of replication, as shown in the illustration below.



178 DNA Polymerases & Their Activities

Although DNA polymerases replicate DNA with high fidelity with as few as one error per 107 nucleotides, mistakes do occur. The proofreading ability of some DNA polymerases corrects many of these mistakes. The polymerase can sense a mismatched base pair, slow down and then catalyze repeated hydrolyses of nucleotides until it reaches the mismatched base pair. This basic proofreading by DNA polymerase is shown below.



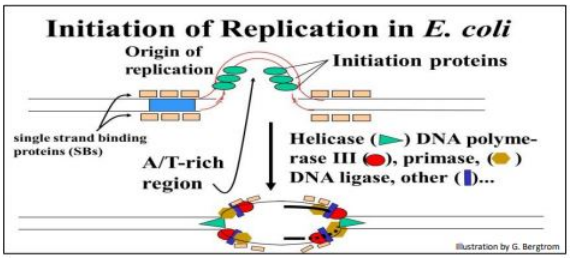
After mismatch repair, DNA polymerase resumes forward movement. Of course, not all mistakes are caught by this or other repair mechanisms (see *DNA Repair*, below). Mutations in the eukaryotic germ line cells that elude correction can cause genetic diseases. However, most are the mutations that fuel evolution. Without mutations in germ line cells (egg and sperm), there would be no mutations and no evolution, and without evolution, life itself would have reached a quick dead end! Other replication mistakes can generate mutations somatic cells. If these somatic mutations escape correction, they can have serious consequences, including the generation of tumors and cancers.

C. The Process of Replication

DNA replication is a sequence of repeated condensation (dehydration synthesis) reactions linking nucleotide monomers into a DNA polymer. Like all biological polymerizations, replication proceeds in three enzymatically catalyzed and coordinated steps: **initiation**, **elongation** and **termination**.

1. Initiation

As we have seen, DNA synthesis starts at one or more origins or replication. These are DNA sequences targeted by *initiator proteins* in *E. coli* (below).



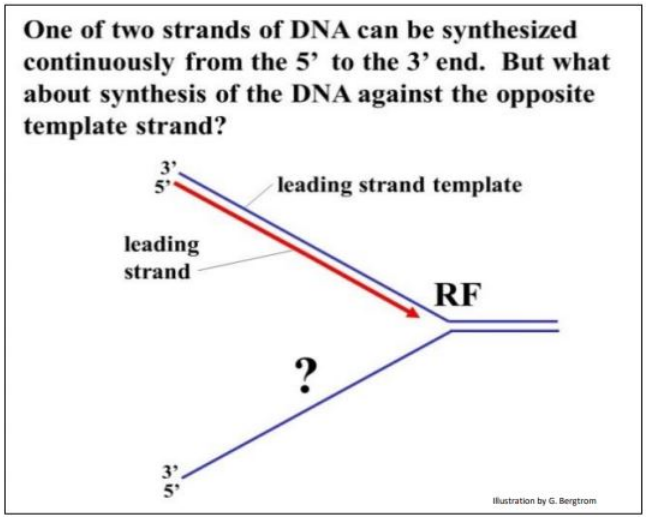
After breaking hydrogen bonds at the origin of replication, the DNA double helix is progressively unzipped in both directions (i.e., by *bidirectional replication*). The separated DNA strands serve as templates for new DNA synthesis. Sequences at replication origins that bind to initiation proteins tend to be rich in adenine and thymine bases. This is because A-T base pairs have two hydrogen (H-) bonds that require less energy to break than the three H-bonds holding G-C pairs together. Once initiation proteins loosen H-bonds at a replication origin, **DNA helicase** uses the energy of ATP hydrolysis to unwind the double helix. DNA polymerase III is the main enzyme that then elongates new DNA. Once initiated, a replication bubble (replicon) forms as repeated cycles of elongation proceed at opposite replication forks.

179 Replication Initiation in *E. coli*

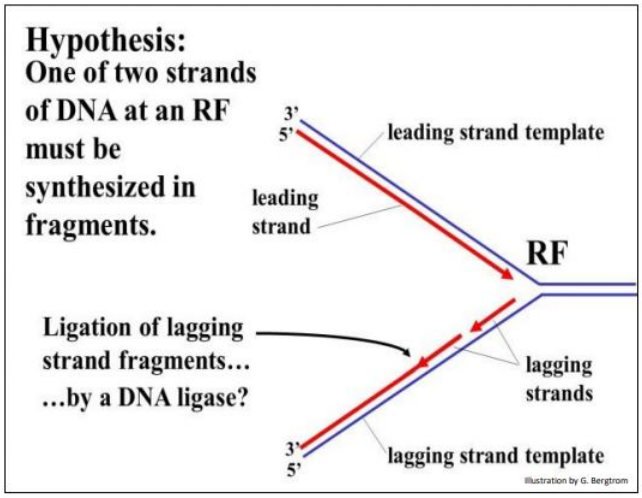
Recalling that new nucleotides can only be added to the free 3' hydroxyl group of a pre-existing nucleic acid strand. Since no known DNA polymerase can start synthesizing new DNA strands from scratch, this is a problem! The action of DNA polymerases therefore requires a **primer**, a nucleic acid strand to which to add nucleotides. The questions were…, what is the primer and where does it come from? Since *RNA polymerases* (enzymes that catalyze RNA synthesis) are the only nucleotide polymerase that can grow a new nucleic acid strand against a DNA template from scratch (i.e., from the first base), it was suggested that RNA might be the primer, After synthesis of a short RNA primer, new *deoxynucleotides* would be added to its 3’ end by DNA polymerase. The discovery of short stretches of RNA nucleotides at the 5’ end of Okazaki fragments confirmed the notion of RNA primers. We now know that cells use **primase**, a special RNA polymerase active during replication, to make those RNA primers against DNA templates before a DNA polymerase can grow the DNA strands at replication forks. As we will see now, the requirement for RNA primers is nowhere more in evidence in events at a replication fork.

2. Elongation

Looking at elongation at one replication fork (below), we see another problem:

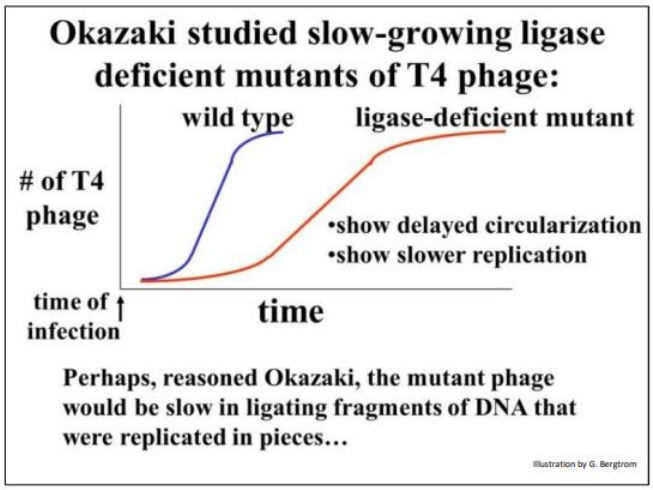


One of the two new DNA strands can grow continuously towards the replication fork as the double helix unwinds. But what about the other strand? Either this other strand must grow in pieces in the opposite direction, or it must wait to begin synthesis until the double helix is fully unwound. If one strand of DNA must be replicated in fragments, then those fragments would have to be stitched (i.e., ligated) together. The problem is illustrated below.



According to this hypothesis, a new **leading strand** of DNA is lengthened continuously by sequential addition of nucleotides to its 3' end against its **leading strand template**. The other strand however, would be made in pieces that would be joined in phosphodiester linkages in a subsequent reaction. Because of the extra step and presumably extra time it takes to make and join these new DNA fragments, this new DNA is called the **lagging strand**, making its template the **lagging strand template**.

Reiji Okazaki and his colleagues were studying mutants of T4 phage that grew slowly in their *E. coli* host cells. They graphed the growth rates of wild-type and mutant T4 phage and demonstrated that slow growth was due to a deficient **DNA ligase** enzyme, already known to catalyze the circularization of linear phage DNA molecules being replicated in infected host cells. The graph below summarizes their results.



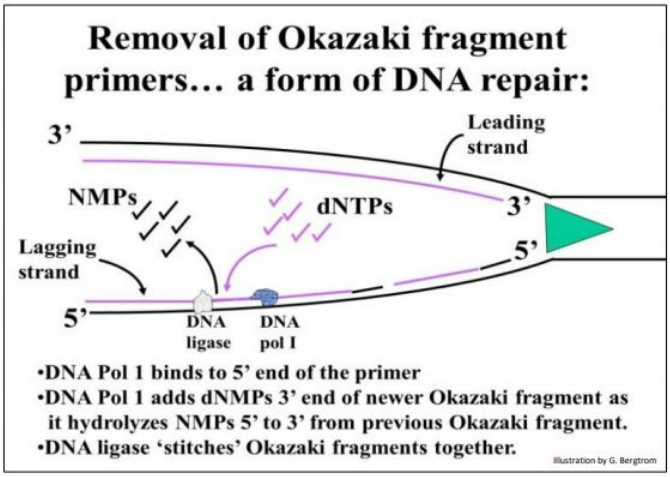
Okazaki's hypothesis was that the deficient DNA ligase in the mutant phage not only slowed down circularization of replicating T4 phage DNA, but would also be slow at joining phage DNA fragments replicated against at least one of the two template DNA strands. When the hypothesis was tested, the Okazakis found that short DNA fragments did indeed accumulate in *E. coli* cells infected with ligasedeficient mutants, but not in cells infected with wild type phage. The lagging strand fragments are now called **Okazaki fragments**.

180 Okazaki Experiments & Fragments - Solving a Problem at an RF

181 Okazaki Fragments are Made Beginning with RNA Primers

You can check out Okazaki's original research at this link.

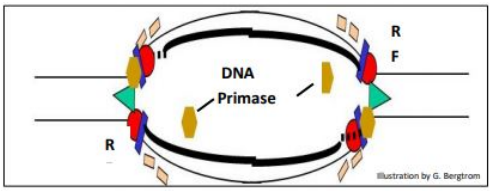
Each Okazaki fragment would have to begin with a 5' RNA primer, creating yet another dilemma! The RNA primer must be replaced with deoxynucleotides before stitching the fragments together. This in fact happens, and the process illustrated below



Removal of RNA primer nucleotides from Okazali fragments requires the action of **DNA polymerase I,** an enzyme that can also catalyze hydrolysis of the phosphodiester bonds between the RNA (or DNA) nucleotides from the 5'-end of a nucleic acid strand. *Flap Endonuclease 1* (**FEN 1**) also plays a role in removing 'flaps' of nucleic acid from the 5' ends of the fragments often displaced by polymerase as it replaces the replication primer. At the same time as the RNA nucleotides are removed, DNA polymerase I catalyzes their replacement by the appropriate deoxynucleotides. Finally, when a fragment is entirely DNA, **DNA ligase** links it to the rest of the already assembled lagging strand DNA. Because of its 5' exonuclease activity (not found in other DNA polymerases), DNA polymerase 1 also plays unique roles in DNA repair (discussed further below). As Cairn's suggested and others demonstrated, replication proceeds in two directions from the origin to form a replicon with its two replication forks (RFs). Each RF has a primase associated with replication of Okazaki fragments along lagging strand templates.

As Cairn's suggested and others demonstrated, replication proceeds in two directions from the origin to form a replicon with its two replication forks (RFs). Each RF has a primase associated with replication of Okazaki fragments along lagging strand templates.

The requirement for primases at replication forks is shown below.

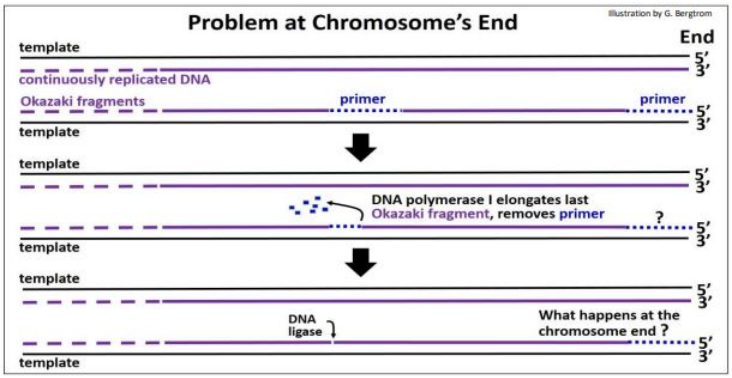


Now we can ask what happens when replicons reach the ends of linear chromosomes in eukaryotes.

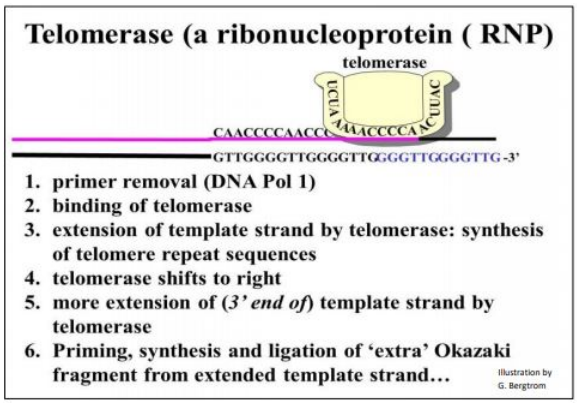
182 Replication Elongation in *E.coli*

3. Termination

In prokaryotes, replication is complete when two replication forks meet after replicating their portion of the circular DNA molecule. In eukaryotes, many replicons fuse to become larger replicons, eventually reaching the ends of the chromosomes…, where there is yet another problem (below)!



When a replicon nears the end of a chromosome (i.e. a double-stranded DNA molecule), the strand synthesized continuously stops when it reaches the 5' end of its template DNA. In theory, synthesis of a last Okazaki fragment can be primed from the 3' end of the lagging template strand. The illustration above implies removal of a primer from the penultimate Okazaki fragment and DNA polymerase catalyzed replacement with DNA nucleotides. But what about the last Okazaki fragment? Would its primer be hydrolyzed? Moreover, without a free 3' end to add to, how are those RNA nucleotides replaced with DNA nucleotides? The problem here is that every time a cell replicates, one strand of new DNA (likely both) would get shorter and shorter. Of course, this would not do…, and does not happen! Eukaryotic replication undergoes a *termination* process involving extending the length of one of the two strands by the enzyme **telomerase**. The action of telomerase is summarized in the illustration below



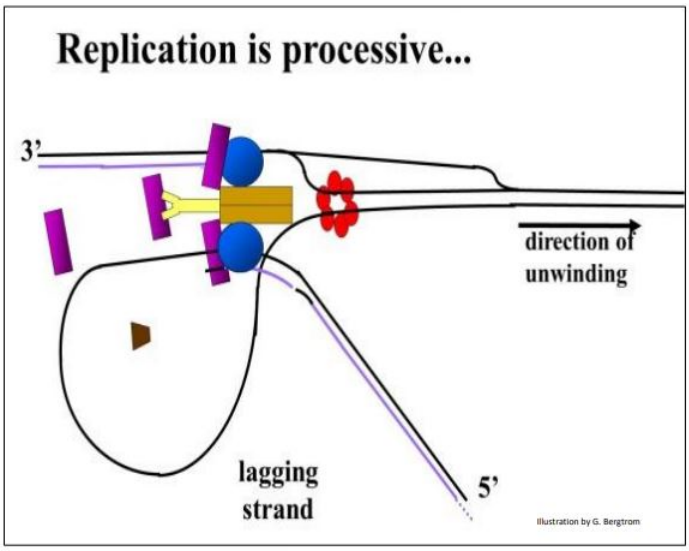
Telomerase consists of several proteins and an RNA. From the drawing,, the RNA component serves as a template for 5'-> 3' extension of the problematic DNA strand. The protein with the requisite reverse transcriptase activity is called ***Te****lomerase* ***R****everse* ***T****ranscriptase*, or **TERT**. The ***Te****lomerase* ***R****NA* ***C****omponent* is called **TERC**. Carol Greider, Jack Szostak and Elizabeth Blackburn shared the 2009 Nobel Prize in Physiology or Medicine for discovering telomerase.

183 Telomerase Replication Prevents Chromosome Shortening https://youtu.be/M4dmfrxGKKU

One of the more interesting recent observations was that differentiated, nondividing cells no longer produce the telomerase enzyme. On the other hand, the telomerase genes are active in normal dividing cells (e.g., stem cells) and cancer cells, which contain abundant telomerase.

4. Is Replication Processive?

Drawings of replicons and replication forks suggest separate events on each DNA strand. Yet events at replication forks seem to be coordinated. Replication may be **processive**, meaning both new DNA strands are replicated in the same direction at the same time, smoothing out the process. How might this be possible? The drawing below shows lagging strand template DNA bending, so that it faces in the same direction as the leading strand at the replication fork.



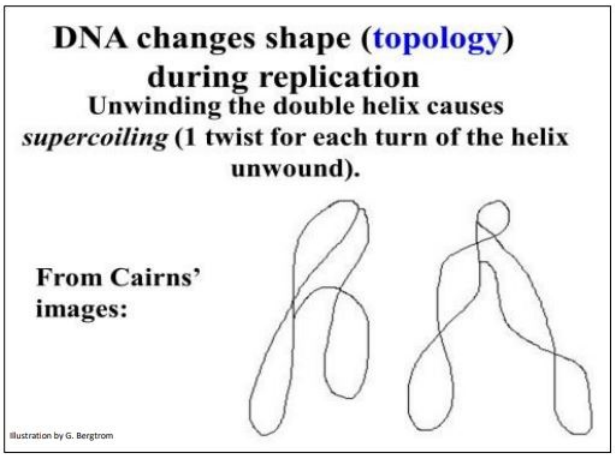
184 Processive Replication

The **replisome** structure cartooned at the replication fork consists of *clamp proteins*, primase, helicase, DNA polymerase and single-stranded binding proteins among others.

Newer techniques of visualizing replication by real-time fluorescence videography call the processive model into question, suggesting that the replication process is anything but smooth! Are lagging and leading strand replication not in fact coordinated? Alternatively, is the jerky movement of DNA elongation in the video an artifact, so that the model of smooth, coordinated replication integrated at a replisome still valid? Or is coordination defined and achieved in some other way? Check out the video yourself in the article here.

5. One More Problem with Replication

Cairns recorded many images of *E.coli* of the sort shown below.



The coiled, twisted appearance of the replicating circles were interpreted to be a natural consequence of trying to pull apart helically intertwined strands of DNA… or intertwined strands of any material! As the strands continued to unwind, the DNA should twist into a **supercoil** of DNA. Increased DNA unwinding would cause the phosphodiester bonds in the DNA to rupture, fragmenting the DNA. Obviously, this does not happen. Experiments were devised to demonstrate supercoiling, and to test hypotheses explaining how cells *relax* the supercoils during replication. Testing these hypotheses revealed the **topoisomerase** enzymes. These enzymes bind and hold on to DNA, catalyze hydrolysis of phosphodiester bonds, control unwinding of the double helix, and finally catalyze the re-formation of the phosphodiester linkages. It is important to note that the topoisomerases are not part of a replisome, but can act far from a replication fork, probably responding to the tensions in overwound DNA. Recall that topoisomerases comprise much of the protein lying along eukaryotic chromatin.

185 Topoisomerases Relieve Supercoiling During Replication

We have considered most of the molecular players in replication. Below is a list of the key replication proteins and their functions (from here).

Enzyme	Function in DNA Replication
DNA Helicase	Also known as helix destabilizing enzyme. Unwinds the DNA double helix at the Replication Fork.
DNA Polymerase	Builds a new duplex DNA strand by adding nucleotides in the 5' to 3' direction. Also performs proof-reading and error correction.
DNA clamp	A protein which prevents DNA polymerase III from dissociating from the DNA parent strand.
Single-Strand Binding (SSB) Proteins	Bind to ssDNA and preven the DNA double helix from re-annealing after DNA helicase unwinds it thus maintaining the strand separation.
Topoisomerase	Relaxes the DNA from its super-coiled nature.
DNA Gyrase	Relieves strain of unwinding by DNA helicase; this is a specific type of topoisomerase
DNA Ligase	Re-anneals the semi-conservative strands and joins Okazaki Fragments of the lagging strand
Primase	Provides a starting point of RNA (or DNA) for DNA polymerase to begin synthesis of the new DNA strand.
Telomerase	Lengthens telomeric DNA by adding repetitive nucleotide sequences to the ends of eukaryotic chromosomes.

This page titled 14.4: DNA Replication is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **9.2: DNA Replication** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

14.5: DNA- The Genetic Material

Learning Objectives

- Detail the molecular nature of a gene.
- Recognize when and where replication of the DNA that is packaged in an organism's chromosome is happening for growth, maintenance, and reproduction.
- Describe how chromosome replication speed can be increased with multiple origins of replication and bidirectional replication.
- Explain the role of each of the primary DNA replicating enzymes in conducting replication of a chromosome. Include the role of these enzyme in ensuring the accuracy of the replication process.
- Predict how the semiconservative replication of a double stranded DNA will happen given the structure of the molecule and the specific function of the DNA replicating enzymes.

Introduction

The genetic (hereditary) material for all living things is composed of DNA (deoxyribonucleic acid). The structure of DNA must enable this substance to store coded information that control the biological function of cells. The genetic material transmits this hereditary information in a stable form for the cell and organism through accurate **replication** of DNA. Although DNA is capable of change (we will discuss how this happens when we talk about DNA mutations), the replication process ensures high accuracy in copying the genetic information so that all progeny cells receive the same information. The DNA from all chromosomes in a human cell would be more than 6.5 feet long! Therefore, to fit inside the nucleus of a cell, DNA is packaged into chromosomes.

Read this article about Watson and Crick for more insight on the experimental facts discovered by chemists and biologist which contributed to the determination of the structure of DNA.

Chemical Structure of DNA Subunits

DNA is a polymer made of nucleotide subunits. A **nucleotide** consists of 3 chemical groups; a sugar, a phosphate and a nitrogenous base (Figure 1). In the case of DNA, the sugar is deoxyribose.

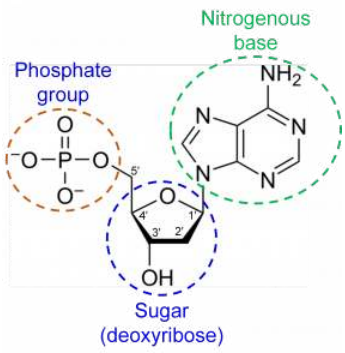


Figure 1. A nucleotide consists of three chemical groups. Image adapted from Wikipedia.

There are 4 different nucleotides in DNA that contain 4 different **bases** (Figure 2). These bases are named adenosine (A), cytosine (C), guanine (G) and thymine (T).

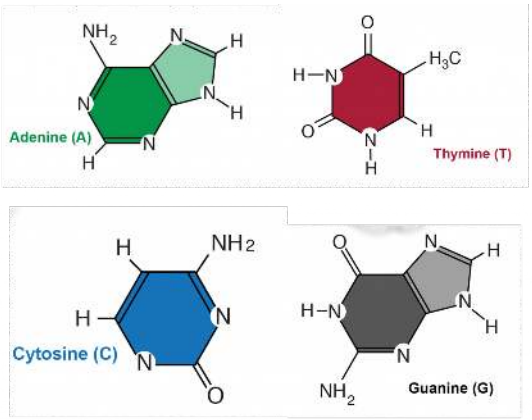


Figure 2. Adenine, Thymine, Cytosine, and Guanine are the nitrogenous bases of DNA. Adapted from NIH-NHGRI.

Complementary, antiparallel DNA strands

DNA codes information by ordering the sequence of the A,T,G and C nucleotides in a long polymer. The nucleotides are connected by phosphate bonds to form a strand. Two strands form a double stranded molecule (Figure 3). Note that in Figure 1, the carbon atoms on the deoxyribose sugar are numbered. The phosphate bond connects carbon #3 (the 3 prime or 3′ carbon) of one sugar to the 5′ carbon of the next. This gives a DNA strand directionality because the 5′ carbon faces one way and the 3′ carbon faces the other.

The bases on one strand form hydrogen bonds with the bases on the other. This is called "**base pairing**" and it is very specific, A only pairs with T and C only pairs with G. Thus, the sequences on the two strands of DNA are "**complementary**"; whenever there is an "A" on one strand, there is a "T" in the corresponding position of the complementary strand.

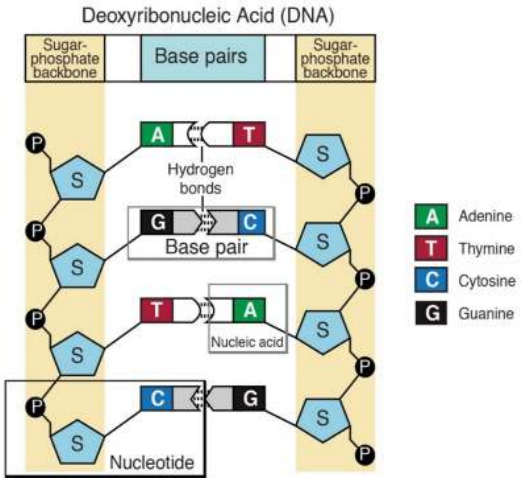


Figure 3. DNA is a double-stranded molecule. Illustration by NIH-NHGRI.

In addition to containing complementary nucleotide sequences, the strands are in opposite orientations. The 5′ end of one strand is oriented toward the 3′ end of the other. For this reason, the strands are referred to as "**antiparallel**".

Double helix

The final feature of the molecular structure is that DNA assumes a helical conformation (Figure 4). This is the most stable configuration that accommodates all the molecular structures and chemical bonds that make up DNA.

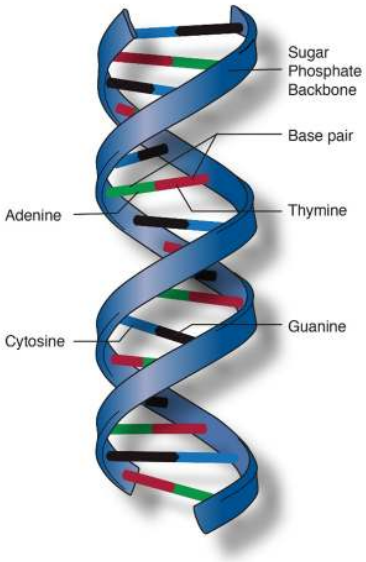


Figure 4. DNA is a double helix. Illustration by NIH-NHGRI.

Activity

Given the following sequence of a DNA strand, predict the sequence and orientation of the complementary strand.

5′-G-A-C-C-G-T-A-A-T-C-G-C-3′

Show Answer

Answer: 3′-C-T-G-G-C-A-T-T-A-G-C-G-5′

Packaging of DNA into chromatin and chromosomes

Chromosomes are long double stranded DNA molecules. Over 150,000,000 nucleotide pairs make up the human X chromosome. The complete replication and orderly transfer of something this large requires the chromosome to be packaged for stability and organization which creates genetic stability in the cells of a species. Every cell in a diploid organism contains two copies (2n) of every chromosome present in that organism (Figure 5). For example, humans have 46 chromosomes in their body, 23 were inherited from the father and 23 from the mother. **Gametes**, the reproductive cells of an organism, (egg or sperm), have only one set (1n) of chromosomes. When the two gametes unite, they form a living embryo with two sets of genetic information. Therefore, we actually have two copies of the genetic information for each trait. Sometimes one copy controls **trait expression**, and other times both copies influence a trait. As a result, the offspring will have characteristics of both the mother and the father.

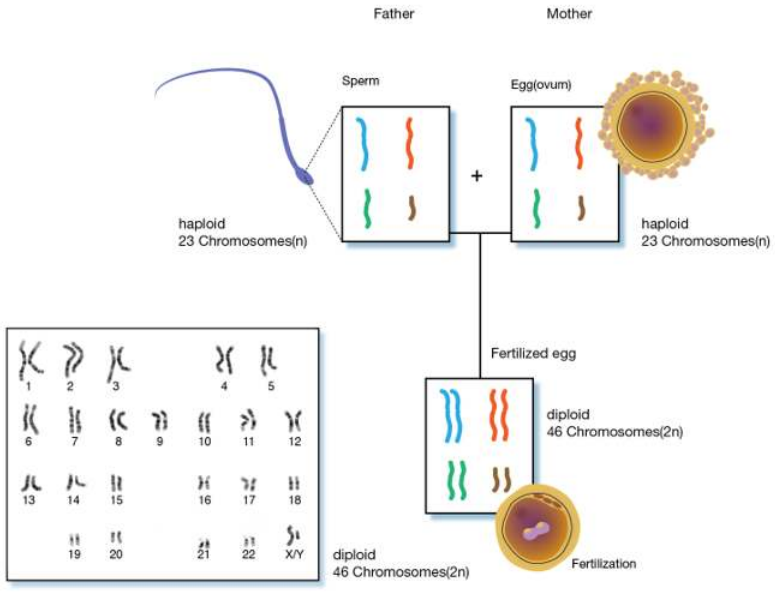


Figure 5. Human non-sex cells are diploid (2n) and contain 46 chromosomes. On the other hand, human sex cells or gametes (sperm and egg) are haploid (n) and contain single set of 23 chromosomes. Image from NIH-NHGRI.

Inside cell nuclei, DNA is assembled, or **packaged**, into material called chromatin, which is composed of both DNA and proteins (Figure 6). Chromatin functions to protect and regulate DNA, as well as to efficiently store the very long DNA molecules that make up chromosomes within the limited space of the nucleus.

DNA from higher eukaryotic organisms is about 109 to 1010 bp. Human and maize chromosomal DNA is about 109 bp, which is equivalent to 1.8 m if all chromosomal DNA were stretched end to end in a linear manner. The diameter of the nucleus is only about 4-6 μm making it a challenge to fit chromosomal DNA inside the cell nucleus. This is accomplished by the ability of DNA to assume a very condensed structure to fit inside the nucleus. Therefore, the organization of eukaryotic DNA into chromatin is an important aspect of DNA packaging.

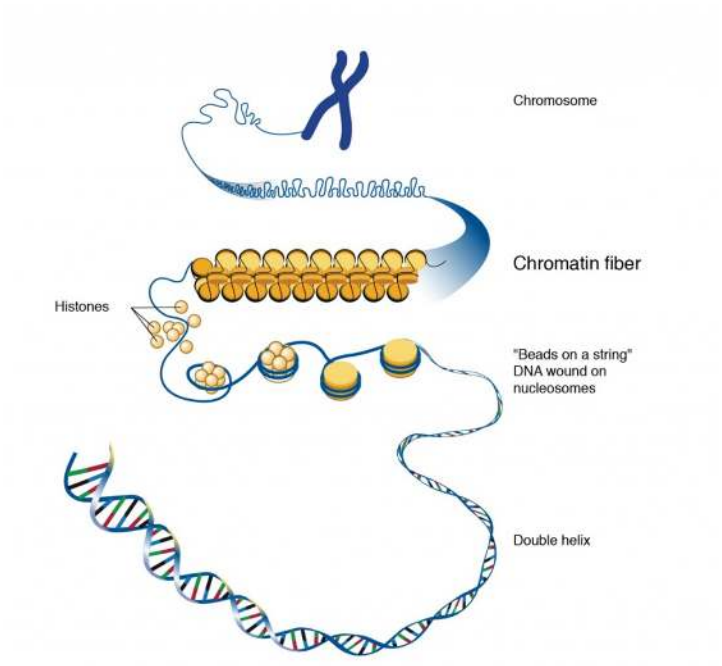


Figure 6. DNA is packed into chromatin and chromosomes. Illustration by NIH-NHGRI.

Nucleosomes and histones

The ordered coiling of chromosomal DNA around a histone protein core forms the **chromatin**. Chromatin is made up of **nucleosomes** (Figure 6) which represent the association of chromosomal DNA with histone proteins. A nucleosome is made up of about 145-147 base pairs of DNA coiled around each histone octamer, for about two complete turns. A histone octamer consists of two copies of each core histone. The total mass of the histones in the nucleus approaches that of the DNA – 2 molecules of each core histone to approximately 200 bp of DNA.

Higher order structures

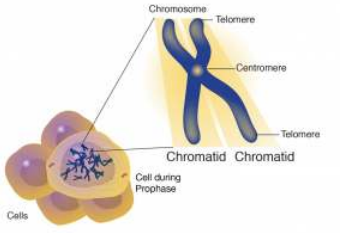


Figure 7. Chromosomes have centromeres and telomeres. Illustration by NIH-NHGRI.

If the higher order chromatin structure is disrupted, electron microscopy reveals the appearance of “beads on a string” with a diameter of about 10 nm. The “beads” represent the DNA wrapped around histones. Nucleosome formation results in a DNA fiber that is about 10 nm and a packaging ratio of about 7. The higher order chromatin structure results when the 10 nm fiber is coiled into a solenoid. The result of nucleosome coiling is a chromatin fiber of 30 nm that is observed by electron microscopy.

Centromeres

The centromere is a chromosomal region that controls chromosome segregation at mitosis and meiosis (Figure 7). Centromeres connect to microtubules of the spindle apparatus, which directs their movement to opposite poles (daughter cell nuclei) during cell division. Scientists were successful in isolating these centromeric sequences from yeast and engineered brewer’s yeast (*Saccharomyces cerevisiae*) plasmids that were able to replicate like the chromosomes. Through these efforts scientists were able to point centromeric function to a DNA stretch of about 120 bp that was resistant to DNAse and bound to a single microtubule.

Telomeres

The telomere lies at the end of the chromosome and confers stability by “sealing” the end of a chromosome (Figure 7). Telomeres consist of long series of short repeated DNA sequence that occurs in tandem arrays and are added to the end of the chromosome during DNA replication by an enzyme called telomerase. In most plant species the sequence **TTTAGGG** constitutes a conserved telomere motif.

The knowledge about the structure and function of centromeres and telomeres has had a profound impact in molecular genetics. For example, yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) have proven useful in the physical mapping of plant genomes requiring the cloning and multiplying large (>100 kb) DNA fragments in yeast or bacterial cells.

DNA Replication

In 1953, J.D. Watson and F.H.C. Crick published a note in NATURE; one of the world’s most widely read research journals. The note cited just six references.

Watson and Crick’s writing had limited chemistry details given the title of their note, “A Structure for Deoxyribonucleic Acid”. A later article would share more of the chemistry behind their proposed double helix structure for Deoxyribonucleic Acid (DNA). This note in Nature was written for a broader audience, including biologists who recognized that the structure of DNA was a prerequisite for understanding the function of the genetic material. They included a diagram to allow readers to visualize this unique double helix structure. To emphasize the structure determines function importance, Watson and Crick included this statement in their note…

> *“It has not escaped our notice that the specific pairing we have postulate immediately suggests a possible copying mechanism for the genetic material.”*

Accurate replication prior to cell division is one of the three key functions of the genetic material. While later experiments would reveal the details of how cells replicate DNA, Watson and Crick were obviously impressed that the proposed double helix structure would fit with the concept that living cells are needed to produce more living cells and the transmission of a complete set of biological instructions. The genetic material must have a structure that lends itself to being copied. Using old to help build new leads to a proposed semiconservative model for replicating the double stranded DNA molecule (Figure 8).

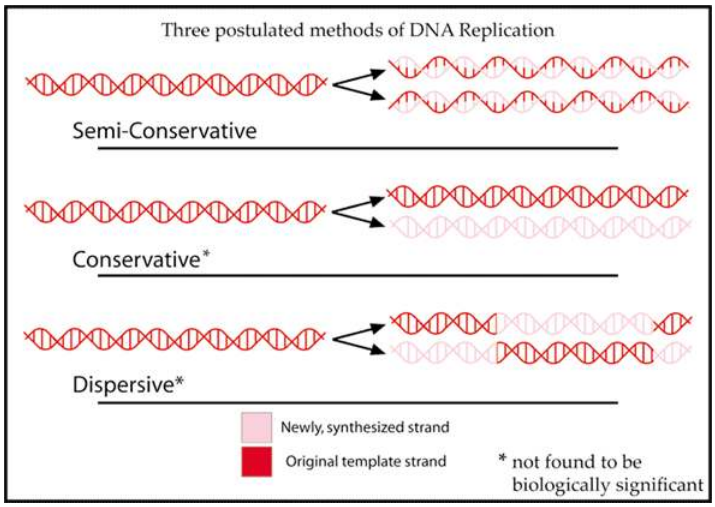


Figure 8. Watson and Crick predicted that the double helix molecule could be replicated in a living cell by the semi-conservative method. Each strand of the old double stranded molecule is read to build a new strand. Early experiments on replication confirmed this prediction. Image Source: Adenosine, CC BY-SA 2.5, via Wikimedia Commons.

Prior to both mitosis and meiosis cell division in multicellular organisms, and cell division in prokaryotes, the DNA inside the cell must be replicated. (see lesson Mitosis and Meiosis). This replication process generates the genetic information needed for either two cells, genetically identical with the original cell or in sexually reproducing organisms, four gamete cells with half of the original cell's genetic information.

When and Where:

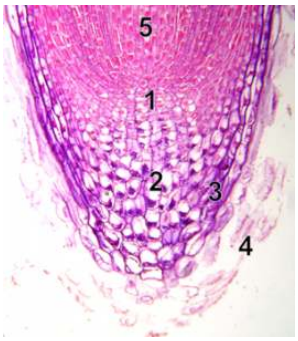


Figure 9. Plant root tip. Image Source: SuperManu, CC BY-SA 2.5, via Wikimedia Commons.

DNA replication will happen before cells divide. In hours old embryos that are developing into multicellular organisms, all the cells are replicating their DNA and dividing into new cells. Hours or days later, multicellular organisms have developed specific meristem regions where new cell production occurs. A visual example of meristem cells where DNA replication is needed is the tip of a plant root (Figure 9). New cells made at the root tip allow the root to grow. The root must also replace the cells that are damaged as the growing root moves through soil. In Figure 8, the meristem region (1) contains meristem cells that are replicating their chromosomes, then dividing into two identical cells by mitosis. Cells on the tip part of the meristem specialize in root cap cells (2,3) with the job of protecting the meristem. The fate of these cells is to die (4), and they need to be replaced by new cells made in the meristem. Cells made on the other half of the meristem (5) can elongate and eventually specialize. These root cells will then function for the plant, all season for annual plants, years for perennial plants. Replication of DNA will only happen in the meristem cells.

Speed of chromosome replication

Fifty-two years after Watson and Crick's Nature note, a group of about 50 molecular geneticists published the complete DNA sequence of the human X chromosome (Figure 5). This discovery revealed that X chromosome is one double-stranded DNA molecule that is about 155 million nucleotide pairs long. Every time a human cell divides, one or two of these 155,000,000 nucleotides must be replicated. This must be done with speed and accuracy.

The main DNA replication enzyme (DNA polII, see below) works fast. Estimates of its replicating speed are around 750 nucleotides per second.

Activity

Let us do the math.

155,000,000 nucleotides / chromosome X 1 second / 750 nucleotides

= 206,666 seconds per X chromosome

Let us convert these seconds to hours.

1 hour / 360 seconds X 206,666 seconds/chromosome = 547 hours per chromosome

That means one replication enzyme would take 547 hours or about 24 days to replicate an X chromosome. What are the biological ramifications of this 24 days? If you cut yourself, it would take 24 days to replicate the chromosomes before the skin cells could divide. The formation of new cells needed for healing cuts in our skin would keep us in a wounded state.

To speed up the chromosome replication process, two tactics are used by living cells. Both tactics are shown in Figure 10. First, replication will not start on just one end of the chromosome. Instead, there are hundreds of origins of replication (ori) along the chromosome. In addition, the replication process moves bidirectionally from the ori. This bidirectional replication creates forks of replication that move toward each other as the chromosome is replicated.

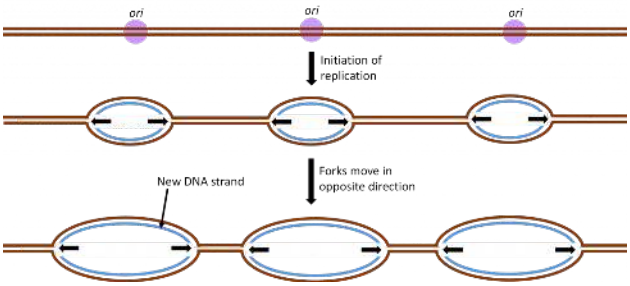


Figure 10. Origin of replication and replication forks. Image by Walter Suza.

As a result of these two tactics, a large chromosome can be replicated in hours rather than weeks and cell division can happen quickly when it is needed for growth or cell replacement.

The Chromosome Replicating Team

Like all processes in living cells, a collection of specific proteins works together to perform functions which control and complete chromosome replication. All the proteins involved in DNA replication make and break bonds, so they are enzymes. The function of the main DNA replicating enzymes is described below.

Five of the main DNA replicating enzymes

- **Helicase**: This is one of the enzymes that unwinds the double stranded DNA molecule. The unwound single strands no longer hydrogen bond to their complementary strand but the sugar-phosphate bonds remain intact.
- **Topoisomerase**: The unwinding work of Helicase creates tension in the double stranded DNA ahead of the replication fork. Topoisomerase relieves this tension by catalyzing a series of sugar phosphate bond breaking and making steps. Without this tension relief, the Double stranded DNA could break and the intact chromosome cannot finish replicating.
- **DNA Polymerase III (DNA pol III)**: This is the main DNA synthesizing enzyme. The enzyme reads the single strand as a template and places in the complementary deoxyribonucleotide. DNA pol III reads the template in the 3′ to 5′ direction and

builds the new strand in the 5′ to 3′ direction, adding the next nucleotide to the 3′ end by catalyzing sugar-phosphate bonding. DNA pol III illustrates the specificity of enzymes several ways; it reads and proofreads the placement of new nucleotides to insure accurate replication. It can only read and build in one direction. DNA pol III can only add nucleotides to a free 3′ end. This last specification means that DNA pol III cannot start the replication process on the single stranded template. Another enzyme needs to be part of the in vivo replication team to prime the work of DNA pol III.

- **Primase**: Biochemists named this enzyme to describe its role in starting or priming the replication process. DNA (or RNA) primase is a special RNA polymerase. The enzyme reads the single stranded DNA template 3′ to 5′ and adds ribonucleic acid (RNA) nucleotides in the 5′ to 3′ direction. Once a few hundred RNA nucleotides are added, primase falls off the template strand and leaves the 3′ end that DNA Pol III needs to continue the process.
- **DNA Polymerase I (DNA pol I)**: The DNA polymerase specializes in the removal of the RNA primers and replacing them with DNA nucleotides. The enzyme works with the same 3′ to 5′ reading and 5′ to 3′ building.
- **DNA Ligase**: After the five enzymes described above have completed their work, some sugar phosphate bonds need to be made to complete the double stranded molecule. DNA ligase has this backbone bond sealing assignment.

Now that we have introduced the replication enzymes, we can describe the step-by-step action of this team. We will focus on the action that happens at one origin of replication.

Step 1: Initiating replication at the Ori

Helicase enzymes bind to the ori sequences and start to unwind the double stranded DNA. This establishes two replication forks (one shown in Figure 10a) and a helicase enzyme will be working to unwind the DNA at each fork as the forks move away from each other. The unwinding exposes single strands of DNA that are the template for building a new strand. Topoisomerase can be seen working to relieve tension ahead of the replication fork in Figure 10a.

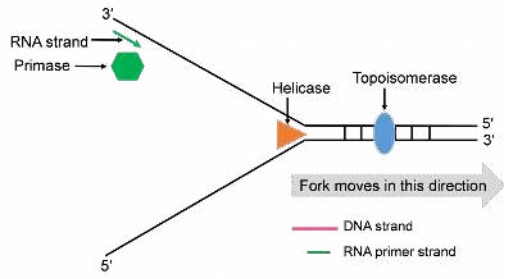


Figure 10a. Helicase and Topoisomerase work together to expose single strand DNA templates. Primase adds RNA nucleotides, reading the old DNA single strand template 3′ to 5′. Image by Walter Suza.

Step 2: Priming DNA with RNA

The process of priming with primase and then replicating with DNA pol III happens at both replication forks (Figure 10a and b). DNA pol III enzymes can replicate new strands as fast as the DNA is unwound at each replication fork. The unwinding creates two template strands at each fork so primase enzyme must work to prime both old strands.

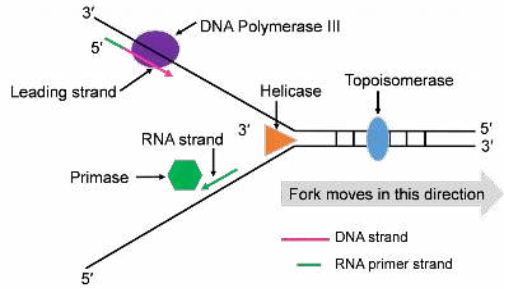


Figure 10b. DNA Polymerase III can add to the 3′ end of the RNA primer and start to read the template 3′ to 5′. Primase can prime the other template strand at this replication fork. Image by Walter Suza.

Step 3: Synthesis of leading and lagging strands

The double strands of DNA are anti-parallel and 'run' in opposite directions and DNA pol III can only work in one direction (Figure 10c). Thus, there is continuous and discontinuous replication happening at each replication fork. The strand that can be replicated as helicase unwinds is the leading strand. The DNA pol III must replicate the other strand in the opposite direction so this strand will lag behind. (Figure 10c).

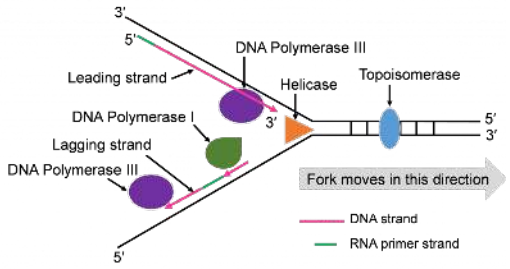


Figure 10c. Both leading and lagging strands are being made at this replication fork. DNA polymerase I removes RNA primer nucleotides and replaces them with DNA. Image by Walter Suza.

There is more priming needed to replicate the lagging strand and the priming leaves RNA-DNA hybrid stretches in the new double stranded molecule. This means there is work for the enzyme DNA pol I. This enzyme will be part of the removal of the RNA nucleotide primers, and replacement with DNA nucleotides (Figure 10c and d).

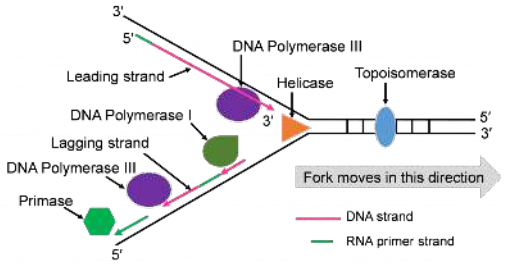


Figure 10d. Lagging strand synthesis must be primed multiple times with primase enzyme creating discontinuous replication. Leading strand just needs to be primed one time. Image by Walter Suza.

Both replication forks have a leading and a lagging strand of new DNA. The DNA nucleotide stretches between the red RNA primers would be the Okazaki fragments.

One of the first scientists to provide evidence to support this model of DNA replication was the team of Reiji and Tsuneko Okazaki in the 1960s. They used radioactive nucleotides to track the newly replicating fragments and found that many of the fragments were short. This fits with the discontinuous model.

As the four enzymes just described attend to their specific roles in replication, two double stranded DNA molecules with an old strand and a new strand are built. If the DNA pol III makes a mistake and adds the wrong, non-complementing nucleotide, the enzyme can proof read its work and replace these replication mistakes. When replication is perfect, the two double stranded DNA molecules will have identical sequences.

Step 4: Ligation of Okazaki fragments

The fifth enzyme that has a final role in completing replication is DNA ligase (not shown in Figure 10d and e). This enzyme will seal the sugar phosphate bond between the last replacement nucleotide added by DNA pol I and the first nucleotide that had been added after priming by DNA pol III. With this bond sealing work, the replication process is complete.

The two double stranded DNA molecules that get made will have identical sequences unless rare mistakes happen as the enzymes conduct their work.

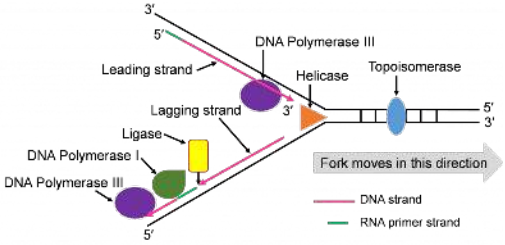


Figure 10e. Ligase must form a sugar-phosphate bond between the last nucleotide placed by DNA pol I and the first nucleotide placed by DNA pol III. Image by Walter Suza.

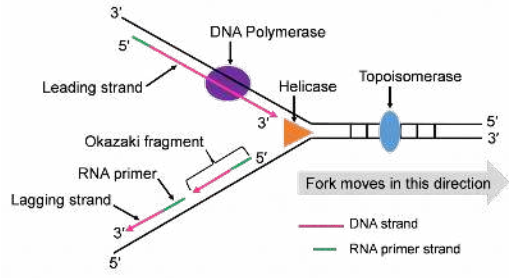


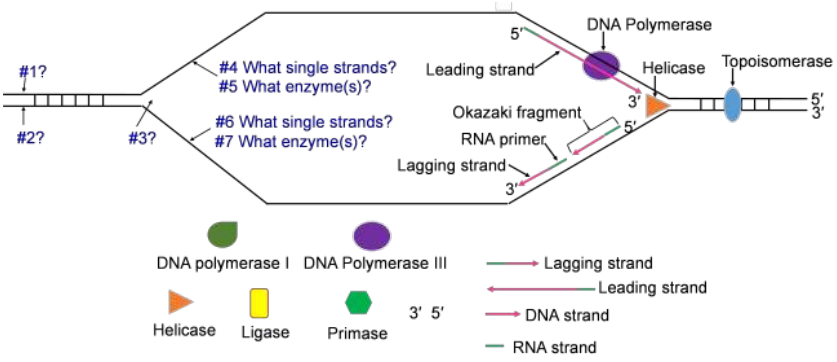
Figure 10f. Short fragments made by discontinuous replication were discovered by the Okazaki team and are called the Okazaki fragments. Image by Walter Suza.

Lesson Summary:

DNA is composed of nucleotides. Nucleotides are connected together by phosphate bonds to form a strand. The bases on one strand form hydrogen bonds with the bases on the other to form a double stranded molecule. The final feature of the molecular structure is that DNA assumes a helical conformation. To fit inside the nucleus, DNA assumes a very condensed structure. Therefore, chromosomal DNA is coiled around a histone protein core to form chromatin. The tight packaging of DNA in chromatin must be modified to allow DNA replication and transcription. In the process of replication, the two DNA strands separate and act as templates for the synthesis of complementary daughter strands.

Learning Activities

Place the single strands, replication enzymes, and 3′ / 5′ labels in the appropriate places (#1-#7)



This page titled 14.5: DNA- The Genetic Material is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Philip Becraft, Patricia Hain, & Patricia Hain (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.2: DNA- The Genetic Material** by Donald Lee, Marjorie Hanneman, Patricia Hain, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

15: Transcription

15.1: Introduction

15.2: Overview of Transcription

15.3: Details of Transcription

Thumbnail: Simplified diagram of mRNA synthesis and processing. (CC BY 3.0 - unported ; Kelvinsong).

15: Transcription is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

15.1: Introduction

Transcription, the synthesis of RNA based on a DNA template, is the central step of the Central Dogma proposed by Crick in 1958. The basic steps of transcription are the same as for replication: initiation, elongation and termination. Differences between transcription in prokaryotes and eukaryotes are in the details.

- *E.coli* uses a single **RNA polymerase** enzyme to transcribe all kinds of RNAs while eukaryotic cells use different RNA polymerases to catalyze **ribosomal RNA** (rRNA), **transfer RNA** (tRNA) and **messenger RNA** (mRNA) synthesis.
- In contrast to eukaryotes, some bacterial genes are part of **operons** whose mRNAs encode multiple polypeptides.
- Bacterial mRNAs are typically translated as they are being transcribed.
- Most RNA transcripts in prokaryotes emerge from transcription ready to use
- Eukaryotic transcripts synthesized as longer precursors undergo **processing** by **trimming**, **splicing**, or both!
- DNA in bacteria is virtually 'naked' in the cytoplasm while eukaryotic DNA is wrapped up in chromatin proteins in a nucleus.
- In bacterial cells the association of ribosomes with mRNA and the translation of a polypeptide can begin even before the transcript is finished. This is because these cells have no nucleus. In our cells, RNAs must exit the nucleus before they encounter *ribosomes* in the cytoplasm.

In this chapter, you will meet bacterial **polycistronic** mRNAs (transcripts of **operons** that encode more than one polypeptide) and the **split genes** of eukaryotes (with their **introns** and **exons**). We will look at some details of transcription of the three major classes of RNA and then at how eukaryotes process precursor transcripts into mature, functional RNAs. Along the way, we will see one example of how protein structure has evolved to interact with DNA.

Learning Objectives

1. Discriminate between the *three steps of transcription* in pro- and eukaryotes, and the factors involved in each.
2. State an hypothesis for why eukaryotes evolved complex *RNA processing* steps.
3. Speculate on why any cell in its right mind would have genes containing *introns* and *exons* so that their transcripts would have to be processed by *splicing*.
4. Articulate the differences between *RNA vs. DNA structure*.
5. Explain the need for *sigma factors* in bacteria.
6. Speculate on *why eukaryotes do not have operons.*
7. List structural features of proteins that bind and *recognize specific DNA sequences.*
8. Explain how proteins that *do not* bind specific DNA sequences *can* still bind to specific regions of the genome.
9. Formulate an hypothesis for why bacteria do not polyadenylate their mRNAs as much as eukaryotes do.
10. Formulate an hypothesis for why bacteria do not cap their mRNAs.

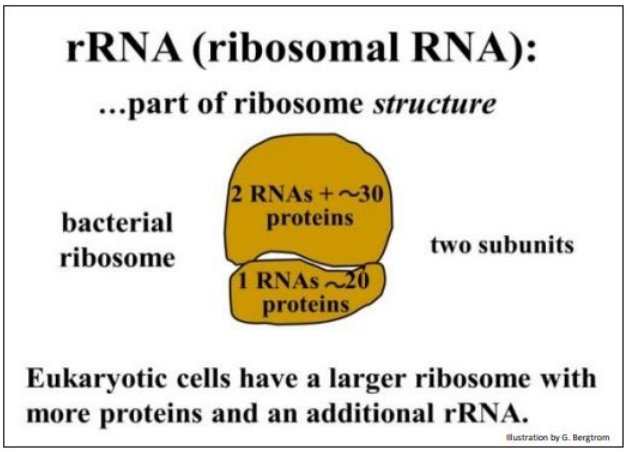
This page titled 15.1: Introduction is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **10.1: Introduction** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

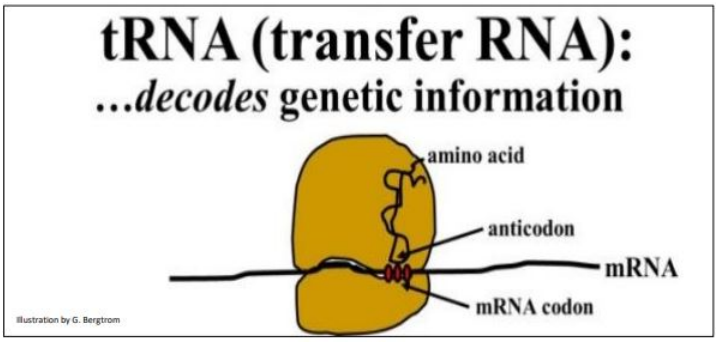
15.2: Overview of Transcription

A. The Major Types of Cellular RNA

All cells make three main kinds of RNA: ribosomal RNA (**rRNA**), transfer RNA (**tRNA**) and messenger RNA (**mRNA**). rRNA is a structural as well as enzymatic component of ribosomes, the protein-synthesizing machine in the cell. Quantitatively, rRNAs are by far the most abundant RNAs in the cell and mRNAs, the least. Three **rRNAs** and about 50 ribosomal proteins make up the two subunits of a bacterial ribosome, as illustrated below.



tRNAs are the *decoding devices* used in protein synthesis (**translation**) to convert nucleic acid sequence information into the amino acid sequences of polypeptides. tRNAs attached to amino acids are positioned on ribosomes based on codonanticodon recognition, as shown below.



During translation, tRNAs decode base sequences in **messenger RNA** (**mRNAs**) into amino acid sequences of polypeptides.

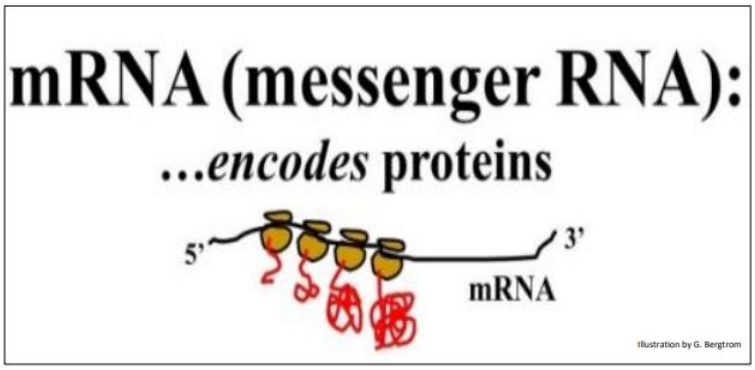
186 Transcription Overview: Ribosomes and Ribosomal RNAs

187 Transcription Overview: Demonstrating the Major RNAs

In 2009, Venkatraman Ramakrishnan, Thomas A. Steitz and ADA Yonath received the Nobel Prize in Chemistry their studies on the structure and molecular biology of the ribosome. Dr. Yonath is one of five women to receive a Nobel Prize – the others were Marie Curie, Irène Joliot-Curie, Dorothy Hodgkin and Barbara Mclintock.

The fact that genes are inside the eukaryotic nucleus and that the synthesis of polypeptides encoded by those genes happens in the cytoplasm led to the proposal that there must be a messenger RNA (mRNA. Sidney Brenner eventually confirmed the existence of mRNAs. Check out his classic experiment in Brenner S (1961, *An unstable intermediate carrying information from genes to ribosomes for protein synthesis*. Nature 190:576-581).

Recall polypeptide synthesis by the formation of polyribosomes (polysomes) along a single mRNA, as illustrated below.

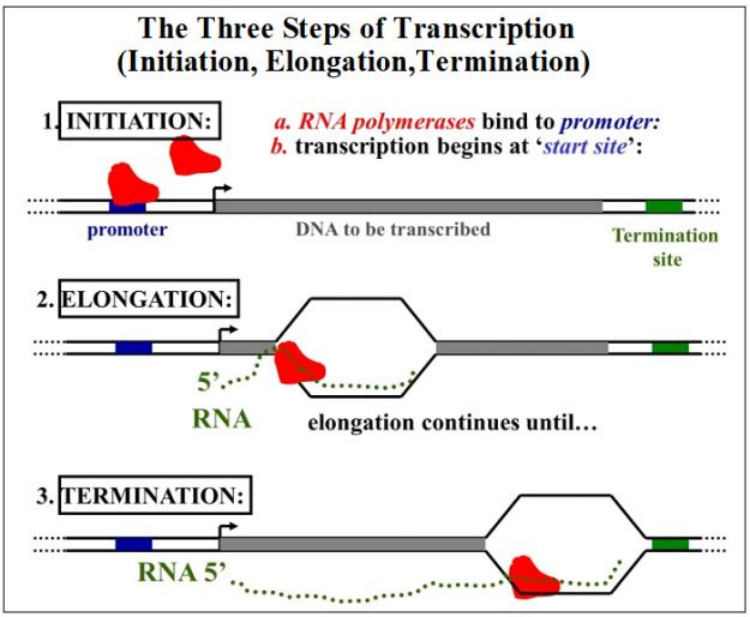


While mRNA is a small fraction of total cellular RNA, there are still smaller amounts of other RNAs such as the transient primers that we saw in DNA replication. We'll encounter still other kinds of low-abundance RNAs later.

B. Key Steps of Transcription

In transcription, an *RNA polymerase* uses the template DNA strand of a gene to catalyze synthesis of a complementary, antiparallel RNA strand. RNA polymerases use ribose nucleotide triphosphate (NTP) precursors, in contrast to DNA polymerases, which use *deoxyribose nucleotide* (dNTP) precursors. In addition, RNAs incorporate *uracil* (U) nucleotides into RNA strands instead of the thymine (T) nucleotides that end up in new DNA. Another contrast to replication - RNA synthesis does not require a primer. With the help of transcription initiation factors, RNA polymerase locates the *transcription start site* of a gene and begins synthesis of a new RNA strand from scratch. Finally, like replication, transcription is error-prone.

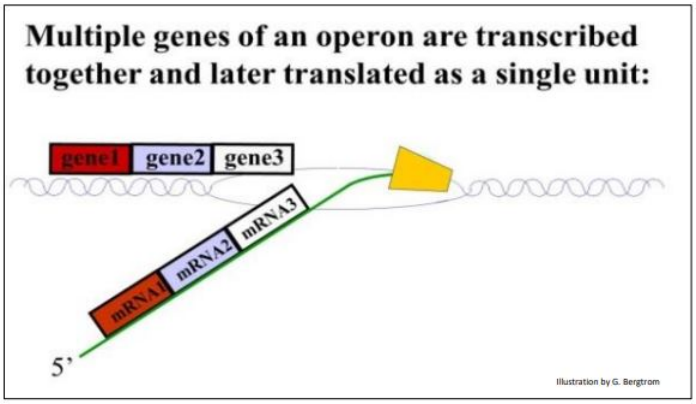
The basic steps of transcription are summarized on the next page. Here we can identify several of the DNA sequences that characterize a gene. The *promoter* is the binding site for RNA polymerase. It usually lies 5' to, or *upstream* of the transcription start site (the bent arrow). Binding of the RNA polymerase positions the enzyme to near the transcription start site, where it will start unwinding the double helix and begin synthesizing new RNA. The transcribed grey DNA region in each of the three panels are the **transcription unit** of the gene. Termination sites are typically 3' to, or downstream from the transcribed region of the gene. By convention, *upstream* refers to DNA 5' to a given reference point on the DNA (e.g., the transcription start-site of a gene). *Downstream* then, refers to DNA 3' to a given reference point on the DNA.



188 Transcription Overview: The Basic Mechanism of RNA Synthesis

In bacteria, some transcription units encode more than one kind of RNA. Bacterial operons are an example of this phenomenon. The resulting mRNAs can be translated into multiple polypeptides at the same time. In the illustration below, RNA polymerase is

transcribing a single mRNA molecule encoding three separate polypeptides.



Bacterial transcription of the different RNAs requires only one RNA polymerase. Different RNA polymerases catalyze rRNA, mRNA and tRNA transcription in eukaryotes. Roger Kornberg received the Nobel Prize in Medicine in 2006 for his discovery of the role of **RNA polymerase II** and other proteins involved in eukaryotic messenger RNA transcription (like father-like son!!).

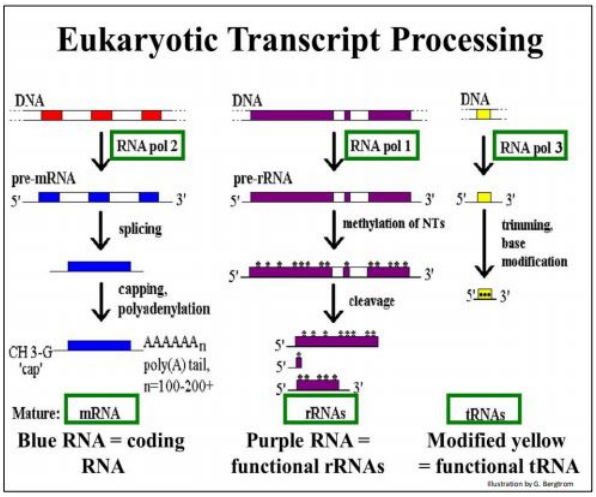
189 RNA Polymerases in Prokaryotes and Eukaryotes

While mRNAs, rRNAs and tRNAs are most of what cells transcribe, a growing number of other RNAs (e.g., *siRNAs*, *miRNAs*, *lncRNAs*…) are also transcribed. Some functions of these transcripts (including control of gene expression or other transcript use) are discussed elsewhere.

C. RNAs are Extensively Processed After Transcription in Eukaryotes

Many eukaryotic RNAs are processed (trimmed, chemically modified) from large precursor RNAs to mature, functional RNAs. These precursor RNAs (pre-RNAs, or primary transcripts) contain in their sequences the information necessary for their function in the cell.

Processing of the three major types of transcripts in eukaryotes is shown below.



To summarize the illustration:

1. Many eukaryotic genes are 'split' into coding regions (exons) and non-coding intervening regions (introns).
2. Transcription of split genes generates a primary mRNA transcript (pre-mRNA).
3. Primary transcripts are spliced to remove the introns from the exons; exons are then ligated into a continuous mRNA. In some cases, the same pre-mRNA is spliced into alternate mRNAs encoding related but not identical polypeptides!
4. Pre-rRNA is cleaved and/or trimmed (not spliced!) to make shorter mature rRNAs.

5. Pre-tRNAs are trimmed, some bases within the transcript are modified and 3 bases (not encoded by the tRNA gene) are enzymatically added to the 3’-end.

190 Post Transcriptional Processing Overview

The details of transcription and processing differ substantially in prokaryotes and eukaryotes. Let’s focus first on details of transcription itself, and then RNA processing.

This page titled 15.2: Overview of Transcription is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

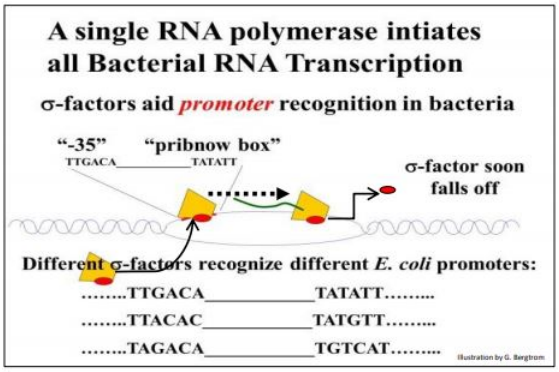
- **10.2: Overview of Transcription** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

15.3: Details of Transcription

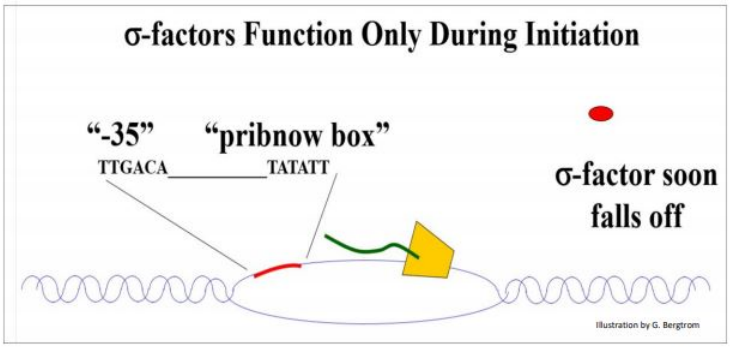
Find a well-written summary of transcription in prokaryotes and eukaryotes at an NIH website (Transcription in Prokaryotes and Eukaryotes). Here (and at this link), you will encounter proteins that bind DNA. Some proteins bind DNA to regulate transcription, inducing or silencing transcription of a gene. We will discuss their role in the regulation of gene expression later. Other proteins interact with DNA simply to allow transcription. These include one or more that, along with RNA polymerase itself, that must bind to the gene promoter to initiate transcription. We will look at bacterial transcription first.

A. Transcription in Prokaryotes

In E. coli, a single RNA polymerase transcribes all kinds of RNA, associating with one of several sigma factor proteins (**σ -factors**) to initiate transcription. It turns out that different promoter sequences and corresponding σ -factors play roles in the transcription of different genes (illustrated below).



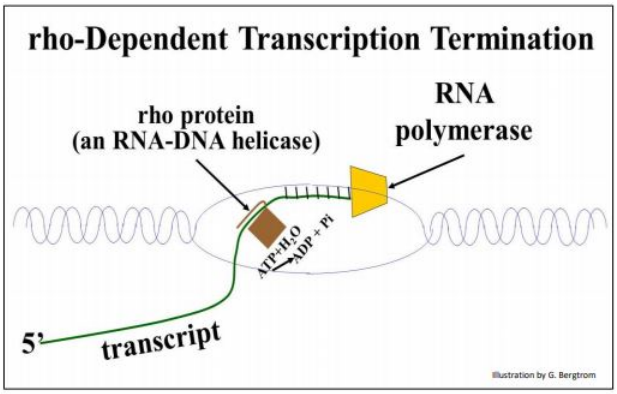
In the absence of the σ-factor, the *E. coli* RNA polymerase can transcribe RNA, but does so at a high rate, and from random sequences in the chromosome. In contrast, when the σ-factor is bound to the RNA polymerase, the complex seems to scan the DNA, recognize and then bind to the promoter sequence of a gene. In this case, the overall transcription rate is slower, but only genes are transcribed, rather than random bits of the bacterial genome! The *Pribnow box*, named for its discoverer, was the first promoter sequence characterized.



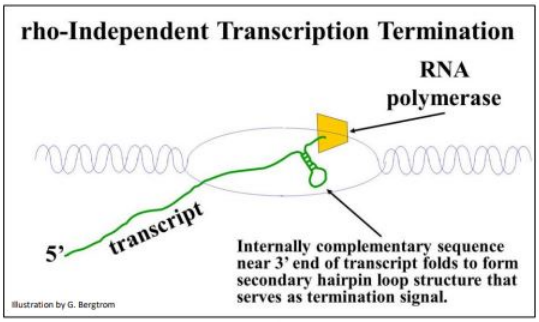
Elongation is the successive addition of nucleotides complementary to their DNA templates, forming phosphodiester linkages. The enzymatic reactions of elongation are similar to the DNA polymerase-catalyzed elongation during replication.

There are two ways that bacterial RNA polymerase 'knows' when it has reached the end of a transcription unit. In one case, as the RNA polymerase nears the 3' end of the nascent transcript, it transcribes a 72 base, C-rich region. At this point, a *termination factor* called the **rho** protein binds to the nascent RNA strand. *Rho* is an ATP-dependent helicase that breaks the H-bonds between the RNA and the template DNA strand, thereby preventing further transcription.

Rho-dependent termination is illustrated below.



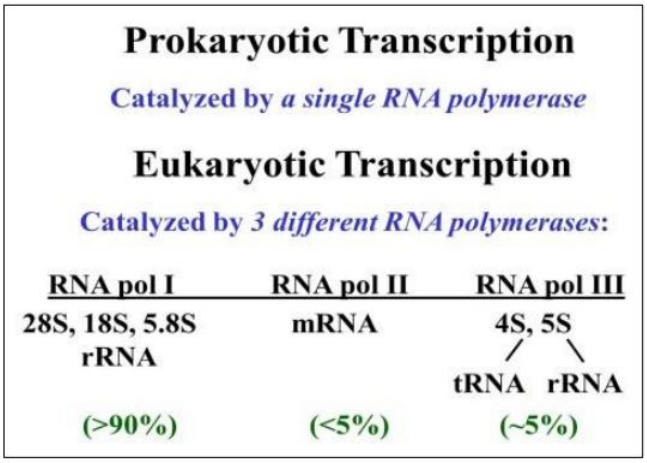
In the other mechanism of termination, the polymerase transcribes RNA whose *termination signal* assumes a secondary *hairpin loop* structure that causes the dissociation of the RNA polymerase, template DNA and the new RNA transcript. The role of the hairpin loop in **rho-independent termination** is illustrated below.



191 Details of Prokaryotic Transcription

B. Transcription in Eukaryotes

Whereas bacteria rely on a single RNA polymerase for their transcription needs, eukaryotes use three different RNA polymerases to synthesize the three major different kinds of RNA, as shown below.

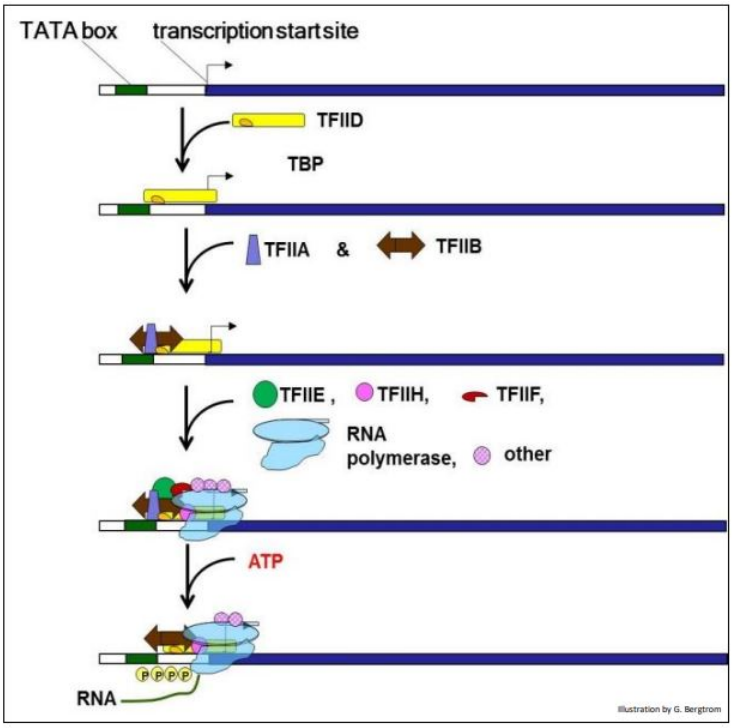


Note that catalysis of the synthesis of most of the RNA in a eukaryotic cell (rRNAs) is by RNA polymerase I. With the help of initiation proteins, each RNA polymerase initiates transcription at a promoter sequence. Once initiated, the RNA polymerases then catalyze the successive formation of phosphodiester bonds to elongate the transcript. Recall that mRNAs are the least abundant in eukaryotes as they are in bacterial cells.

Unfortunately, the details of the termination of transcription in eukaryotes are not as well understood as they are in bacteria. Therefore, we will focus on initiation, and then consider the processing of different eukaryotic RNAs into ready-to-use molecules.

1. Eukaryotic mRNA Transcription

The multiple steps of eukaryotic mRNA transcription are shown on the next page.



Transcription of eukaryotic mRNAs by *RNA polymerase II* begins with the sequential assembly of a eukaryotic *initiation complex* at a gene promoter. The typical eukaryotic promoter for a protein-encoding gene contains a TATA box DNA *sequence motif* as well as additional short upstream sequences. *TATA-binding* protein (TBP) first binds to the TATA box along with TFIID (*transcription initiation factor IID*).

This intermediate recruits *TFIIA* and *TFIIB*. Next, *TFGIIE*, *TFIIF* and *TFIIH*, several other initiation factors and RNA polymerase II bind to form the transcription **initiation complex**. Phosphorylation adds several phosphates to the aminoterminus of the RNA polymerase, after which some of the TF's dissociate from the initiation complex. The remaining RNA polymerase-TF complex can now start making the mRNA.

Unlike prokaryotic RNA polymerase, eukaryotic RNA Polymerase II does not have an inherent helicase activity. For this, eukaryotic gene transcription relies on the multi-subunit TFIIH protein, in which two subunits have helicase activity. Consistent with the closer relationship of *archaea* to eukaryotes (rather to prokaryotes), archaeal mRNA transcription initiation resembles that of eukaryotes, albeit requiring fewer initiation factors during formation of an initiation complex.

192 Eukaryotic mRNA Transcription

A significant difference between prokaryotic and eukaryotic transcription is that RNA polymerase and other proteins involved at a gene promoter do not see naked DNA. Instead, they must recognize specific DNA sequences through chromatin proteins. On the other hand, all proteins that interact with DNA have in common a need to recognize the DNA sequences to which they must bind…, within the double helix. In other words, they must see the bases within the helix, and not on its uniformly electronegative phosphate backbone surface. To this end, they must penetrate the DNA, usually through the **major groove** of the double helix. We will see that DNA regulatory proteins face the same problems in achieving specific shape-based interactions!

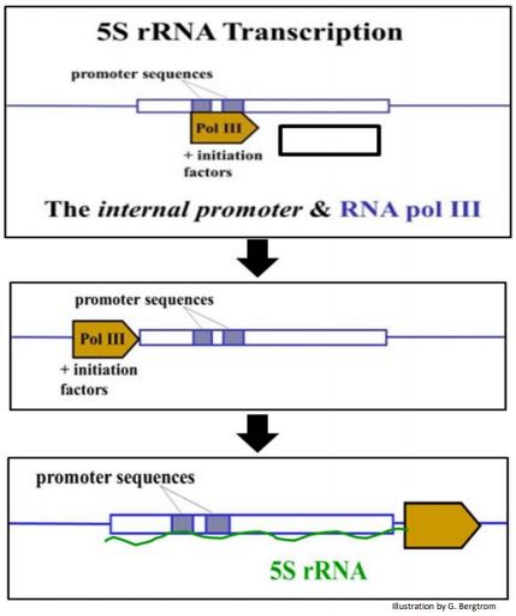
193 Recognition of Transcription factors at Promoters

2. Eukaryotic tRNA and 5SRNA Transcription

Transcription of 5S rRNA and tRNAs by *RNA Polymerase III* is unusual in that the promoter sequence to which it binds (with the help of initiation factors) is not upstream of the transcribed sequence, but lies within the transcribed sequence. After binding to this

internal promoter, the polymerase re-positions itself to transcribe the RNA from the transcription start site so that the final transcript thus contains the promoter sequence!

5S rRNA by RNA polymerase III is shown below.



3. Transcription of the Other Eukaryotic rRNAs

tRNAs are also transcribed by RNA polymerase III in much the same way as the 5S rRNA. The other rRNAs are transcribed by RNA polymerase I, which binds to an upstream promoter along with transcription initiation factors. We know less of the details of this process compared to our understanding of mRNA transcription. We'll explore what we do know next. As already noted, transcription termination is not as well understood in eukaryotes as in prokaryotes. Coupled termination and polyadenylation steps common to most prokaryotic mRNAs are discussed in more detail below, with a useful summary at the NIH-NCBI website here.

This page titled 15.3: Details of Transcription is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **10.3: Details of Transcription** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

16: Transcriptional Regulation (prokaryotes)

16.1: The lac Operon
16.2: The Use of Mutants to Study the lac Operon
16.3: Gene Regulation in Prokaryotes

16: Transcriptional Regulation (prokaryotes) is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

16.1: The lac Operon

Early insights into mechanisms of transcriptional regulation came from studies of *E. coli* by researchers Francois Jacob & Jacques Monod. In *E. coli*, and many other bacteria, genes encoding several different proteins may be located on a single transcription unit called an **operon**. The genes in an operon share the same transcriptional regulation, but are translated individually. Eukaryotes generally do not group genes together as operons (exception is *C. elegans* and a few other species).

Basic *lac* Operon structure

E. coli encounters many different sugars in its environment. These sugars, such as **lactose** and **glucose**, require different enzymes for their metabolism. Three of the enzymes for lactose metabolism are grouped in the ***lac* operon**: ***lacZ***, ***lacY***, and ***lacA*** (Figure **16.1.1**). *LacZ* encodes an enzyme called **β-galactosidase**, which digests lactose into its two constituent sugars: glucose and galactose. *lacY* is a **permease** that helps to transfer lactose into the cell. Finally, *lacA* is a **trans-acetylase**; the relevance of which in lactose metabolism is not entirely clear. Transcription of the *lac* operon normally occurs only when lactose is available for it to digest. Presumably, this avoids wasting energy in the synthesis of enzymes for which no substrate is present. A single mRNA transcript includes all three enzyme-coding sequences and is called polycistronic. A cistron is equivalent to a gene.

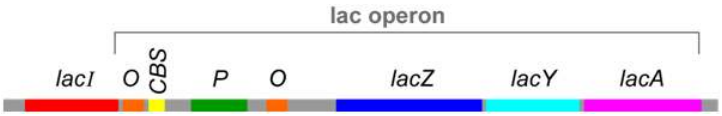


Figure **16.1.1**: Diagram of a segment of an E. coli chromosome containing the lac operon, as well as the lacI coding region. The various genes and cis-elements are not drawn to scale. (Originall-Deyholos-CC:AN)

cis- and *trans*Regulators

In addition to the three protein-coding genes, the *lac* operon contains short DNA sequences that do not encode proteins, but are instead binding sites for proteins involved in transcriptional regulation of the operon. In the *lac* operon, these sequences are called **P (promoter)**, **O (operator)**, and **CBS (CAP-binding site)**. Collectively, sequence elements such as these are called ***cis*-elements** because they must be located on the same piece of DNA as the genes they regulate. On the other hand, the proteins that bind to these *cis*-elements are called ***trans*-regulators** because (as diffusible molecules) they do not necessarily need to be encoded on the same piece of DNA as the genes they regulate.

lacI is an allosterically regulated repressor

One of the major *trans*-regulators of the *lac* operon is encoded by ***lacI***. Four identical molecules of *lacI* proteins assemble together to form a **homotetramer** called a **repressor** (Figure **16.1.2**). This repressor binds to two operator sequences adjacent to the promoter of the *lac* operon. Binding of the repressor prevents RNA polymerase from binding to the promoter (Figure **16.1.3**). Therefore, the operon will not be transcribed when the operator is occupied by a repressor.

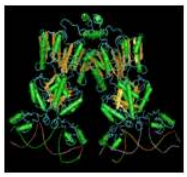


Figure **16.1.2**: Structure of lacI homotetramer bound to DNA (Origianl-Deyholos-CC:AN)

Besides its ability to bind to specific DNA sequences at the operator, another important property of the *lacI* protein is its ability to bind to lactose. When lactose is bound to *lacI*, the shape of the protein changes in a way that prevents it from binding to the operator. Therefore, in the presence of lactose, RNA polymerase is able to bind to the promoter and transcribe the *lac* operon, leading to a moderate level of expression of the *lacZ*, *lacY*, and *lacA* genes. Proteins such as *lacI* that change their shape and functional properties after binding to a ligand are said to be regulated through an **allosteric** mechanism. The role of *lacI* in regulating the *lac* operon is summarized in Figure **16.1.4**.

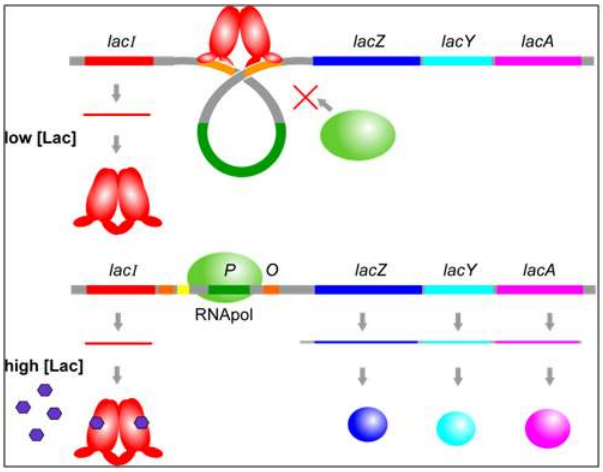


Figure **16.1.3**: When the concentration of lactose [Lac] is low, lacI tetramers bind to operator sequences (O), thereby blocking binding of RNApol (green) to the promoter (P). Alternatively, when [Lac] is high, lactose binds to lacI, preventing the repressor from binding to O, and allowing transcription by RNApol. (Origianl-Deyholos-CC:AN)

CAP is an allosteric activator of the *lac* operon

A second aspect of *lac* operon regulation is conferred by a *trans*-factor called **cAMP binding protein** (**CAP**, Figure **16.1.4**). CAP is another example of an allosterically regulated *trans*-factor. Only when the CAP protein is bound to cAMP can another part of the protein bind to a specific *cis*-element within the *lac* promoter called the **CAP binding sequence (CBS)**. CBS is located very close to the promoter (P). When CAP is bound to at CBS, RNA polymerase is better able to bind to the promoter and initiate transcription. Thus, the presence of cAMP ultimately leads to a further increase in *lac* operon transcription.

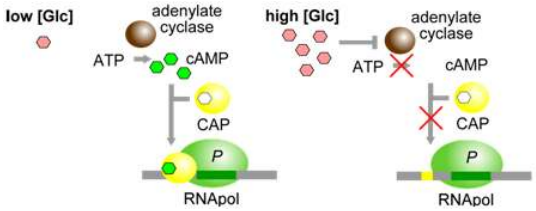


Figure **16.1.4**: CAP, when bound to cAMP, helps RNApol to bind to the lac operon. cAMP is produced only when glucose [Glc] is low. (Origianl-Deyholos-CC:AN)

The physiological significance of regulation by cAMP becomes more obvious in the context of the following information. The concentration of cAMP is inversely proportional to the abundance of glucose: when glucose concentrations are low, an enzyme called **adenylate cyclase** is able to produce cAMP from ATP. Evidently, *E. coli* prefers glucose over lactose, and so expresses the *lac* operon at high levels only when glucose is absent and lactose is present. This provides another layer of logical control of *lac* operon expression: only in the presence of lactose, and in the absence of glucose is the operon expressed at its highest levels.

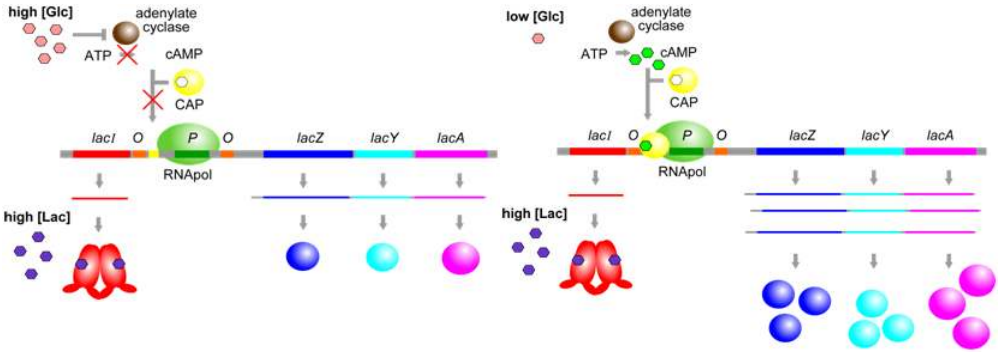


Figure **16.1.6**: When glucose [Glc] and lactose [Lac] are both high, the lac operon is transcribed at a moderate level, because CAP (in the absence of cAMP) is unable to bind to its corresponding cis-element (yellow) and therefore cannot help to stabilize binding of RNApol at the promoter. Alternatively, when [Glc] is low, and [Lac] is high, CAP and cAMP can bind near the promoter and increase further the transcription of the lac operon. (Origianl-Deyholos-CC:AN)

This page titled 16.1: The lac Operon is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **12.1: The lac Operon** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.2: The Use of Mutants to Study the lac Operon

Single mutants of the *lac* operon

The *lac* operon and its regulators were first characterized by studying mutants of *E. coli* that exhibited various abnormalities in lactose metabolism. Some mutants expressed the *lac* operon genes constitutively, meaning the operon was expressed whether or not lactose was present in the medium. Such mutant are called **constitutive** mutants.

The operator locus (*lacO*) - One example is $\boldsymbol{O^c}$, in which a mutation in an operator sequence and reduces or precludes the repressor (the *lacI* gene product) from recognizing and binding to the operator sequence. Thus, in O^c mutants, *lacZ*, *lacY*, and *lacA* are expressed whether or not lactose is present.

The *lacI* locus – One type of mutant allele of *lacI* (callled $\boldsymbol{I^-}$) prevents either the production of a repressor polypeptide or produces a polypeptide that cannot bind to the operator sequence. This is also a constitutive expresser of the *lac* operon because absence of repressor binding permits transcription.

Another type of mutant of *lacI* called $\boldsymbol{I^s}$ prevents the repressor polypeptide from binding lactose, and thus will bind to the operator and be non-inducible.. This mutant constitutively represses the *lac* operon whether lactose is present or not. The lac operon is not expressed and this mutant is called a “super-suppressor”.

The F-factor and two lac operons in a single cell – partial diploid in *E.coli*

More can be learned about the regulation of the *lac* operon when two different copies are present in one cell. This can be accomplished by using the **F-factor** to carry one copy, while the other is on the genomic *E. coli* chromosome. This results in a partial diploid in *E. coli.*

The F-factor is an **episome** that is capable of being either a free plasmid or integrated into the host bacterial chromosome. This switching is accomplished by IS elements where unequal crossing over can recombine the F-factor and adjacent DNA sequences (genes) in and out of the host chromosome. Researchers have used this genetic tool to create partial diploids (merozygotes) that allow them to test the regulation with different combinations of different mutations in one cell. For example, the F-factor copy may have a I^S mutation while the genomic copy might have an O^C mutation. How would this cell respond to the presence/absence of lactose (or glucose)? This partial diploid can be used to determine that I^S is dominant to I^+, which in turn is dominant to I^-. It can also be used to show the O^C mutation only acts in *cis*- while the *lacI* mutation can act in *trans*- .

This page titled 16.2: The Use of Mutants to Study the lac Operon is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

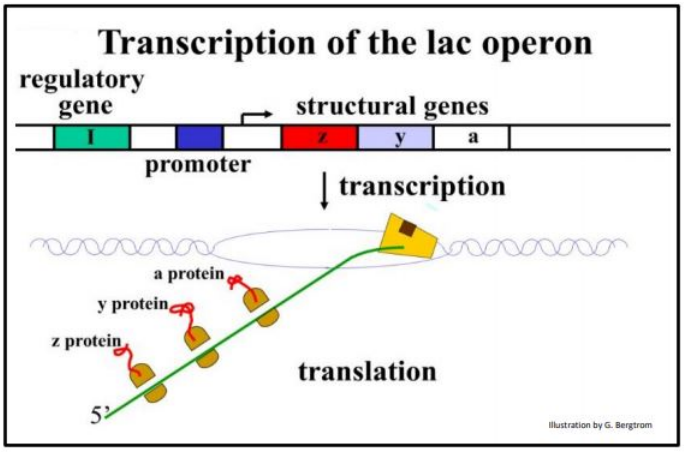
- **12.2: The Use of Mutants to Study the lac Operon** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

16.3: Gene Regulation in Prokaryotes

Many prokaryotic genes are organized in **operons**, linked genes transcribed into a single mRNA encoding two or more proteins. Operons usually encode proteins with related functions. Regulating the activity of an operon (rather than multiple single genes encoding single proteins) allows better coordination of the synthesis of several proteins at once. In *E. coli*, the regulated **lac operon** encodes three enzymes involved in the metabolism of **lactose** (an alternative nutrient to glucose). Regulation of an operon (or of a single gene for that matter) can be by **repression** or by **induction**. When a small metabolite in a cell binds to a regulatory **repressor** or **inducer** protein, the protein undergoes an allosteric change that allows it to bind to a regulatory DNA sequence…, or to un-bind from the DNA. We will see examples of such regulation in the **lac** and **trp** operons. Lac operon gene regulation is an example of **gene repression** as well as **induction**. Trp (tryptophan) operon regulation is by **gene repression**. In both the operons, changes in levels of intracellular metabolites reflect the metabolic status of the cell and elicit appropriate changes in gene transcription. We will look at the regulation of both operons.

216 Overview of Prokaryotic Gene Regulation

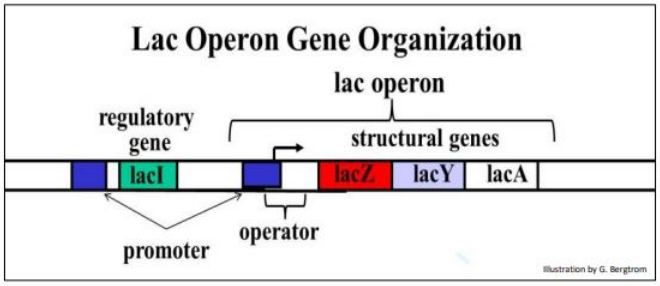
The mRNA transcribed from the **lac operon** is simultaneously translated into those three enzymes, as shown below.



A. Mechanisms of Control of the Lac Operon

In the animal digestive tract (including ours), genes of the *E. coli* **lac operon** regulate the use of *lactose* as an alternative nutrient to glucose. Think cheese instead of chocolate! The operon consists of lacZ, lacY, and lacA genes that were called **structural genes**. By definition, structural genes encode proteins that participate in cell structure and metabolic function. As already noted, the *lac operon* is transcribed into an mRNA encoding the Z, Y and A proteins.

Let's take a closer look at the structure of the lac operon and the function of the Y, Z and A proteins (below).



The lacZ gene encodes **β-galactosidase**, the enzyme that breaks lactose (a disaccharide) into *galactose* and glucose. The lacY gene encodes lactose **permease**, a membrane protein that facilitates lactose entry into the cells. The role of the lacA gene (a **transacetylase**) in lactose energy metabolism is not well understood. The **I gene** to the left of the lac Z gene is a **regulatory gene** (to distinguish it from structural genes). Regulatory genes encode proteins that interact with regulatory DNA sequences associated

with a gene to control transcription. The **operator** sequence separating the I and Z genes is a transcription regulatory DNA sequence.

The *E. coli* lac operon is usually silent (repressed) because these cells prefer glucose as an energy and carbon source. In the presence of sufficient glucose, a **repressor protein** (the I gene product) is bound to the **operator**, blocking transcription of the lac operon. Even if lactose is available, cells will not be use it as an alternative energy and carbon source when glucose levels adequate. However, when glucose levels drop, the lac operon is active and the three enzyme products are translated. We will see how limiting glucose levels induce maximal lac operon transcription by both **derepression** and direct **induction**, leading to maximal transcription of the lac genes only when necessary (i.e., in the presence of lactose and absence of glucose). Let's look at some of the classic experiments that led to our understanding of E. coli gene regulation in general, and of the lac operon in particular.

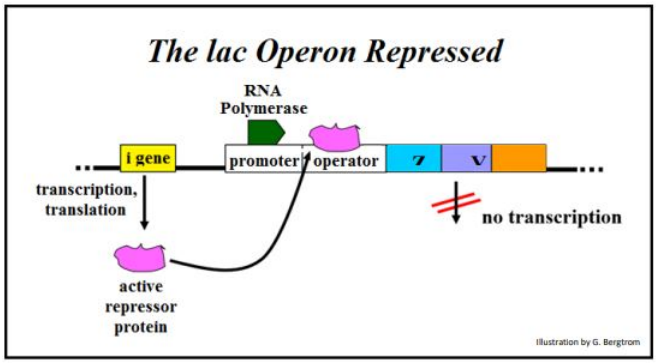
In the late 1950s and early 1960s, Francois Jacob and Jacques Monod were studying the use of different sugars as carbon sources by *E. coli*. They knew that wild type *E. coli* would **not** make the β-galactosidase, β-galactoside permease or β-galactoside transacetylase proteins when grown on glucose. Of course, they also knew that the cells would switch to lactose for growth and reproduction if they were deprived of glucose! They then searched for and isolated different E. coli mutants that could not grow on lactose, even when there was no glucose in the growth medium. Here are some of the mutants they studied:

1. One mutant failed to make active β-galactosidase enzyme but made permease.
2. One mutant failed to make active permease but made normal amounts of β-galactosidase.
3. Another mutant failed to make transacetylase..., but could still metabolize lactose in the absence of glucose. Hence the uncertainty of its role in lactose metabolism.
4. Curiosly, one mutant strain failed to make any of the three enzymes!

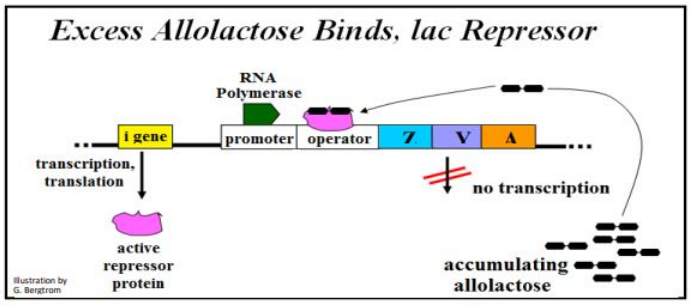
Since double mutants are very rare and triple mutants even rarer, Jacob and Monod inferred that the activation of all three genes in the presence of lactose were controlled together in some way. In fact, it was this discovery that defined the operon as a set of genes transcribed as a single mRNA, whose expression could therefore be effectively coordinated. They later characterized the repressor protein produced by the lacI gene. Jacob, Monod and Andre Lwoff shared the Nobel Prize in Medicine in 1965 for their work on bacterial gene regulation. We now know that *negative* and *positive* regulation of the *lac operon* (described below) depend on two regulatory proteins that together, control the rate of lactose metabolism.

1. Negative Regulation of the Lac Operon by Lactose

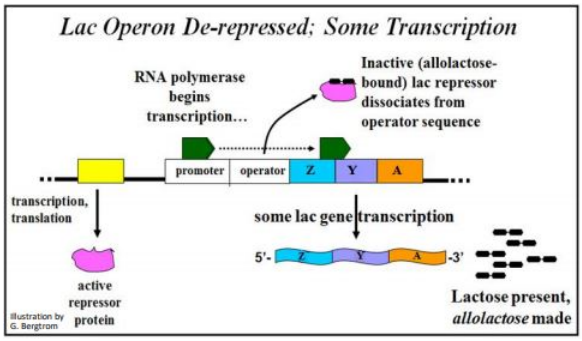
Refer to the illustration below to identify the players in lac operon derepression.



The repressor protein product of the I gene is always made and present in *E. coli* cells. I gene expression is not regulated! In the absence of lactose in the growth medium, the repressor protein binds tightly to the operator DNA. While **RNA polymerase** is bound to the promoter and ready to transcribe the operon, the presence of the repressor bound to the operator sequence close to the Z gene physically blocks its forward movement. Under these conditions, little or no transcript is made. If cells are grown in the presence of lactose, the lactose entering the cells is converted to **allolactose**. Allolactose binds to the repressor sitting on the operator DNA to form a 2-part complex, as shown below.

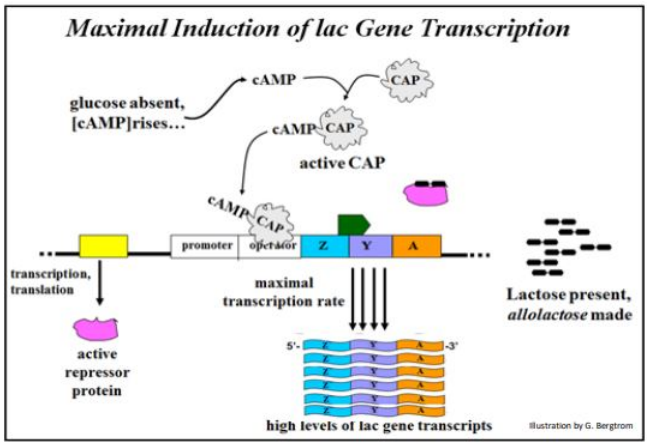


The allosterically altered repressor dissociates from the operator and RNA polymerase can transcribe the *lac* operon genes as illustrated below



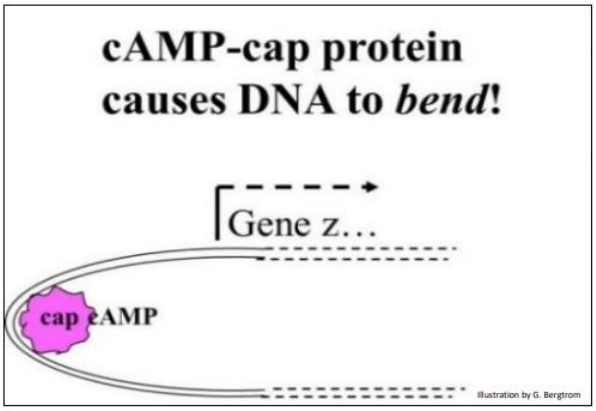
2. Positive Regulation of the Lac Operon; Induction by Catabolite Activation

The second control mechanism regulating lac operon expression is mediated by CAP (cAMP-bound **catabolite activator protein** or cAMP receptor protein). When glucose is available, cellular levels of cAMP are low in the cells and CAP is in an inactive conformation. On the other hand, if glucose levels are low, cAMP levels rise and bind to the CAP, activating it. If lactose levels are also low, the cAMP-bound CAP will have no effect. If lactose is present and glucose levels are low, then allolactose binds the lac repressor causing it to dissociate from the operator region. Under these conditions, the cAMP-bound CAP can bind to the operator in lieu of the repressor protein. In this case, rather than blocking the RNA polymerase, the activated Camp-bound CAP induces even more efficient lac operon transcription. The result is synthesis of higher levels of lac enzymes that facilitate efficient cellular use of lactose as an alternative to glucose as an energy source. Maximal *activation* of the lac operon in high lactose and low glucose is shown below.



217 Regulation of the Lac Operon

cAMP-bound CAP is an **inducer** of transcription. It does this by forcing the DNA in the promoter-operator region to bend. And since bending the double helix loosens H-bonds, it becomes easier for RNA polymerase to find and bind the promoter on the DNA strand to be transcribed…, and for transcription to begin. cAMP-CAPinduced bending of DNA is illustrated below.



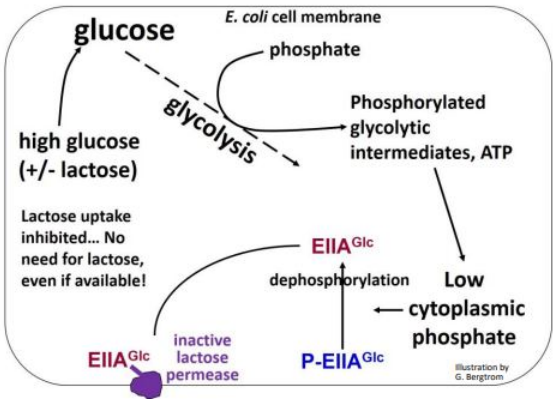
3. Lac Operon Regulation by Inducer Exclusion and Multiple Operators

In recent years, additional layers of lac operon regulation have been uncovered. In one case, the ability of *lac permease* to transport lactose across the cell membrane is regulated. In another, additional operator sequences have been discovered to interact with a multimeric repressor to control lac gene expression.

a) Regulation of Lactose use by Inducer Exclusion

When glucose levels are high (even in the presence of lactose), phosphate is consumed to phosphorylate glycolytic intermediates, keeping cytoplasmic phosphate levels low. Under these conditions, unphosphorylated EIIAGlc binds to the *lactose permease* enzyme in the cell membrane, preventing it from bringing lactose into the cell.

The role of phosphorylated and unphosphorylated $EIIA^{Glc}$ in regulating the lac operon are shown below.

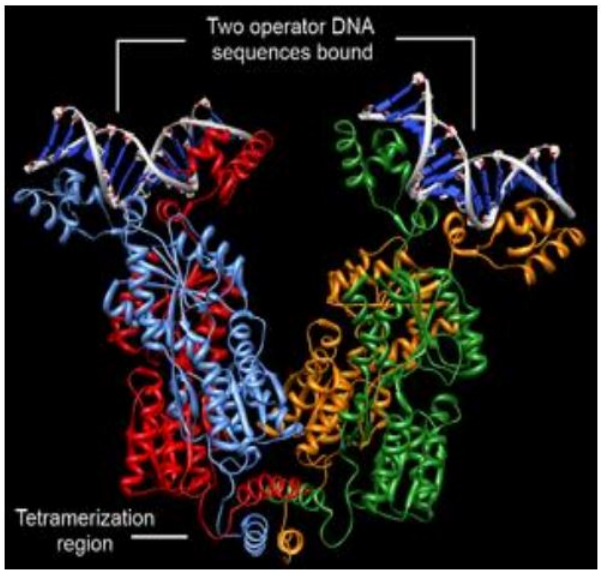


High glucose levels block lactose entry into the cells, effectively preventing allolactose formation and the derepression of the lac operon. Inducer exclusion is thus a logical way for the cells to handle an abundance of glucose, whether or not lactose is present. On the other hand, if glucose levels are low in the growth medium, phosphate concentrations in the cells rise sufficiently for a specific kinase to phosphorylate the EIIAGlc. Phosphorylated EIIAGlc then undergoes an allosteric change and dissociates from the lactose permease, making it active so that more lactose can enter the cell. In other words, the inducer is not “excluded” under these conditions!

The kinase that phosphorylates $EIIA^{Glc}$ is part of a phosphoenolpyruvate (PEP)- *dependent phosphotransferase* system (PTS) cascade. When extracellular glucose levels are low, the cell activates the PTS system in an effort to bring whatever glucose is around into the cell. But the last enzyme in the PTS phosphorylation cascade is the kinase that phosphorylates $EIIA^{Glc}$. Phosphorylated $EIIA^{Glc}$ dissociates from the lactose permease, re-activating it, bringing available lactose into the cell from the medium.

b) Repressor Protein Structure and Additional Operator Sequences

The lac repressor is a tetramer of identical subunits (below).



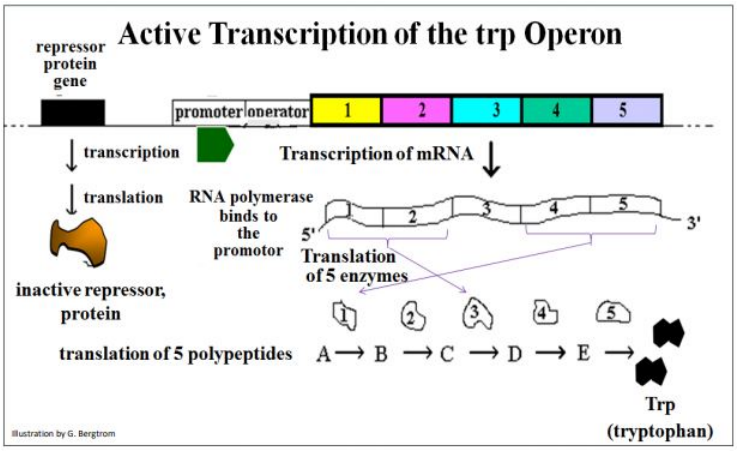
Each subunit contains a *helix-turn-helix* motif capable of binding to DNA. However, the operator DNA sequence downstream of the promoter in the operon consists of a pair of *inverted repeats* spaced apart in such a way that they can only interact two of the repressor subunits, leaving the function of the other two subunits unknown… that is, until recently!

Two more operator regions were recently characterized in the lac operon. One, called $\mathbf{O_2}$, is within the *lac z gene* itself and the other, called $\mathbf{O_3}$, lies near the end of, but within the *lac I* gene. Apart from their unusual location within actual genes, these operators, which interact with the remaining two repressor subunits, went undetected at first because mutations in the O2 or the O3 region individually do not contribute substantially to the effect of lactose in derepressing the lac operon. Only mutating both regions at the same time results in a substantial reduction in binding of the repressor to the operon.

B. Mechanism of Control of the Tryptophan Operon

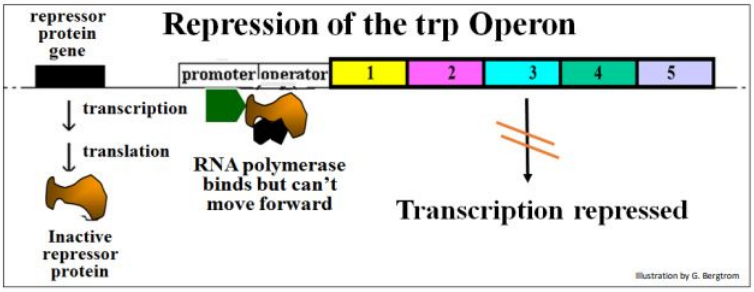
If ample tryptophan (**trp**) is available, the tryptophan synthesis pathway can be inhibited in two ways. First, recall how feedback inhibition by excess trp can allosterically inhibit the trp synthesis pathway. A rapid response occurs when tryptophan is present in excess, resulting in rapid feedback inhibition by blocking the first of five enzymes in the trp synthesis pathway. The *trp operon* encodes polypeptides that make up two of these enzymes.

Enzyme 1 is a *multimeric* protein, made from polypeptides encoded by the *trp5* and *trp4* genes. The trp1 and trp2 gene products make up **Enzyme 3**. If cellular tryptophan levels drop because the amino acid is rapidly consumed (e.g., due to demands for proteins during rapid growth), E. coli cells will continue to synthesize the amino acid, as illustrated below.



On the other hand, if tryptophan consumption slows down, tryptophan accumulates in the cytoplasm. Excess tryptophan will bind to the trp repressor. The trp-bound repressor then binds to the trp operator, blocking RNA polymerase from transcribing the operon.

The repression of the trp operon by trp is shown below.



In this scenario, tryptophan is a **co-repressor**. The function of a co-repressor is to bind to a repressor protein and change its conformation so that it can bind to the operator.

219 Repression of the Tryptophan (TRP) Operon

This page titled 16.3: Gene Regulation in Prokaryotes is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **12.2: Gene Regulation in Prokaryotes** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

17: RNA processing

17.1: Details of Eukaryotic mRNA Processing
17.2: Gene Expression- Transcription

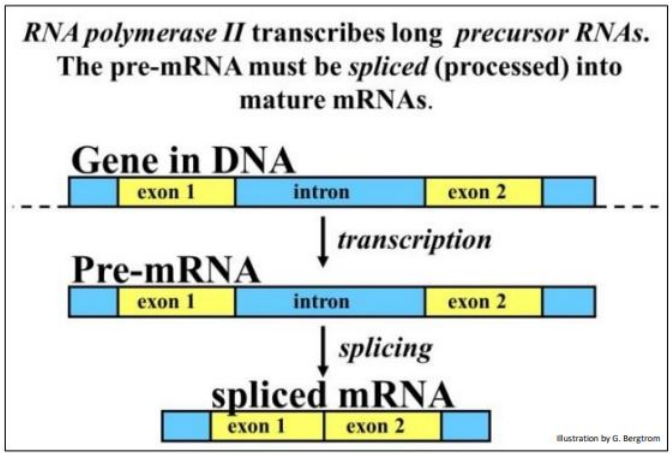
17: RNA processing is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

17.1: Details of Eukaryotic mRNA Processing

Eukaryotic mRNA primary transcripts undergo extensive processing, including *splicing*, *capping* and, *polyadenylation*. The steps described here are considered in order of (sometimes overlapping!) occurrence. We begin with splicing, an mRNA phenomenon.

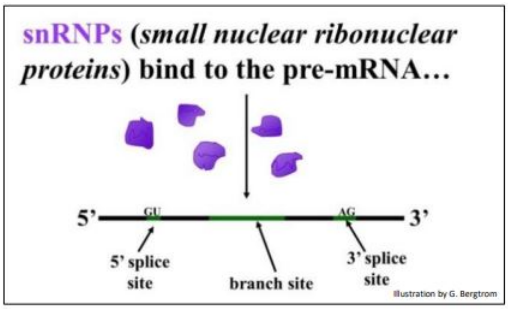
A. Spliceosomal Introns

Bacterial gene coding regions are continuous. The discovery of eukaryotic split genes with introns and exons came as quite a surprise. Not only did it seem incongruous for evolution to have stuck irrelevant DNA in the middle of coding DNA, no one could have dreamt up such a thing! For their discovery of split genes, by Richard J. Roberts and Phillip A. Sharp shared the Nobel Prize for Physiology in 1993. In fact, all but a few eukaryotic genes are split, and some have one, two (or more than 30-50!) **introns** separating bits of coding DNA, the **exons**. **Splicing** is summarized below.

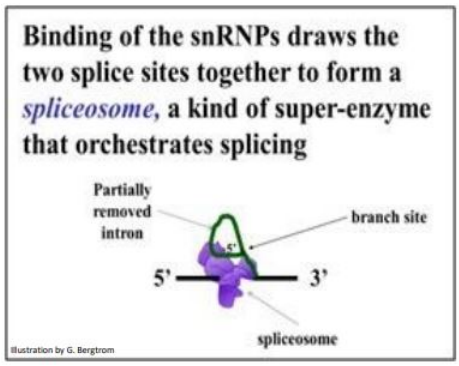


Splicing involves a number of *small ribonuclear proteins* (**snRNPs**). snRNPs are particles composed of RNA and proteins. They bind to specific sites in an mRNA and then direct a sequential series of cuts and ligations (the *splicing*) necessary to process the mRNAs.

The role of snRNPs in splicing pre-mRNAs is illustrated below.



snRNP binding to a pair of splice sites flanking an intron in a pre-mRNA forms the **spliceosome** that completes the splicing, including removal of the **lariat** (the intermediate structure of the intron). The last step is to ligate exons into a continuous mRNA with all its codons intact and ready for translation. Spliceosome action is summarized below.

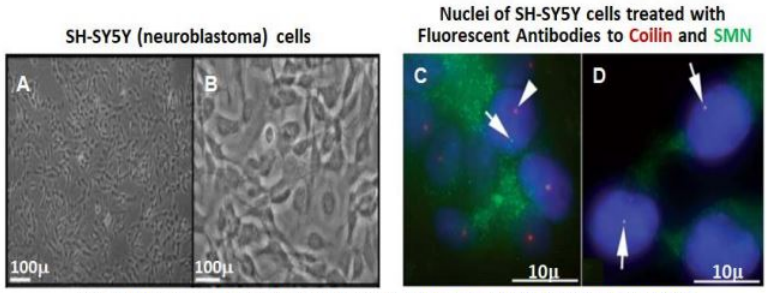


194 The Discovery of Split Genes

195 mRNA Splicing

B. Specific Nuclear bodies and their associated proteins facilitate the assembly and function of the SnRNPs

Recall the organization of nuclei facilitated by nuclear bodies. **Cajal bodies** (**CBs**) and **Gems** are nuclear bodies that are similar in size and have related functions in assembling spliceosomal SnRNPs. Some splicing defects correlate with mutations in the *coil* protein that associate with Cajal bodies; others correlate with mutations in SMN proteins normally associated with Gems. An hypothesis was that CBs and Gems interact in SnRNP and spliceosome assembly…, but how? Consider the results of an experiment in which antibodies to **coilin** and the **SMN** protein were localized in undifferentiated and differentiated neuroblastoma cells.



Creative Commons Deed, NC, Attribution 2.5 Adapted from: https://www.ncbi.nlm.nih.gov/pubmed/19735367

A and C are undifferentiated cells in culture; B and D are cells that were stimulated to differentiate. In the fluorescence micrographs at the right, arrows point to fluorescent nuclear bodies. The coilin protein is associated with CBs and SMN is found in Gems. Therefore, we expect that fluorescent antibodies to coilin (green) will localize to CBs and antibodies to SMN protein (red) will bind to Gems. This is what happens in the nuclei of undifferentiated cells (panel C). But in panel D, the two antibodies colocalize, suggesting that the CBs and Gems aggregate in the differentiated cells. This would explain the need for both functional coilin and SMN protein to produce functional SnRNPs. The CBs and Gems may be aggregating in differentiated cells due to an observed increase in expression of the SMN protein. This could lead to more active Gems more able to associate with the CBs.

This and similar experiments demonstrate that different nuclear bodies do have specific functions. They are not random structural artifacts, have evolved to organize nuclear activities in time and space in ways that are essential to the cell.

C. Group I and Group II *Self-Splicing* Introns

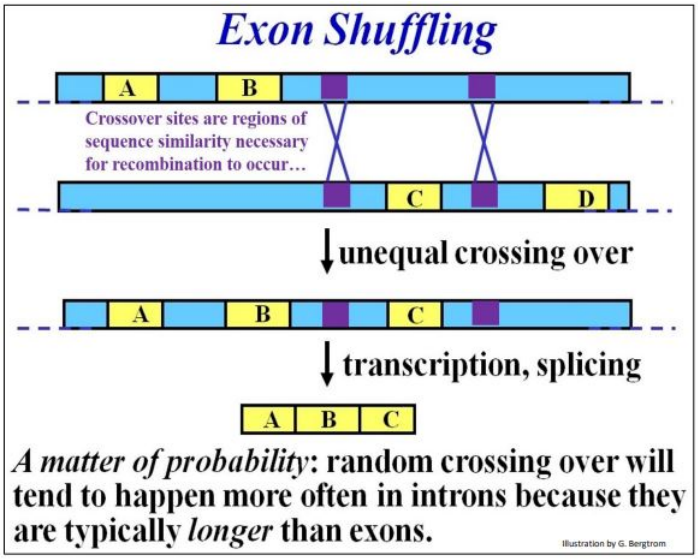
While Eukaryotic *Spliceosomal introns* are spliced using snRNPs as described above, *Group I* or *Group II* introns are removed by different mechanisms. **Group I introns** interrupt mRNA and tRNA genes in bacteria and in mitochondrial and chloroplast genes. They are occasionally found in bacteriophage genes, but rarely in nuclear genes, and then only in lower eukaryotes. Group I introns are *self-splicing*! Thus, they are *ribozymes* that do not require snRNPs or other proteins. Instead, they fold into a secondary stem-loop structure that positions catalytic nucleotides at appropriate splice sites, excise themselves, and re-ligate the exons. **Group II introns** in chloroplast and mitochondrial rRNA, mRNA, tRNA and some bacterial mRNAs can be quite long, form complex stem-

loop tertiary structures, and self-splice, at least in a test tube! However, *Group II introns* encode proteins required for their own splicing *in vivo*. Like spliceosomal introns, they form a lariat structure at an A residue branch site. All this suggests that the mechanism of spliceosomal intron splicing evolved from that of *Group II* introns.

D. So, Why Splicing?

The puzzle implied by the question of course is why higher organisms have split genes in the first place. While the following discussion can apply to all splicing, it will reference mainly spliceosomal introns. Here are some answers to the question "Why splicing?"

- Introns in nuclear genes are typically longer (often much longer!) than exons. Since they are non-coding, they are large targets for mutation. In effect, noncoding DNA, including **introns can buffer the ill effects of random mutations**.
- You may recall that gene duplication on one chromosome (and loss of a copy from its homolog) arise from unequal recombination (non-homologous crossing over). It occurs when similar DNA sequences align during synapsis of meiosis. In an organism that inherits a chromosome with both gene copies, the duplicate can accumulate mutations as long as the other retains original function. The diverging gene then becomes part of a pool of selectable DNA, the grist of evolution, in the descendants of organisms that inherit the duplicated genes, increasing species diversity. Unequal recombination can also occur between similar sequences (e.g., in introns) in the same or different genes. Introns can also enable the sharing of exons between genes. After unequal recombination between introns flanking an exon, one gene will acquire another exon while the other will lose it. Once again, as long as an organism retains a copy of the participating genes with original function, the organism can make the required protein and survive. Meanwhile, the gene with the extra exon may produce the same protein, but one with a new structural domain and function. Like a complete duplicate gene, one with a new exon and added function is in the pool of selectable DNA. Thus, this phenomenon of **exon shuffling increases species diversity**! The evidence indicates that exon shuffling has occurred, creating proteins with different overall functions that nonetheless share at least one domain and one common function. An example discussed earlier involves calcium-binding proteins that regulate many cellular processes. Structurally related calcium (Ca++) binding domains are common to many otherwise structurally and functionally unrelated proteins. Consider exon shuffling in the unequal crossover (non-homologous recombination) illustrated below.



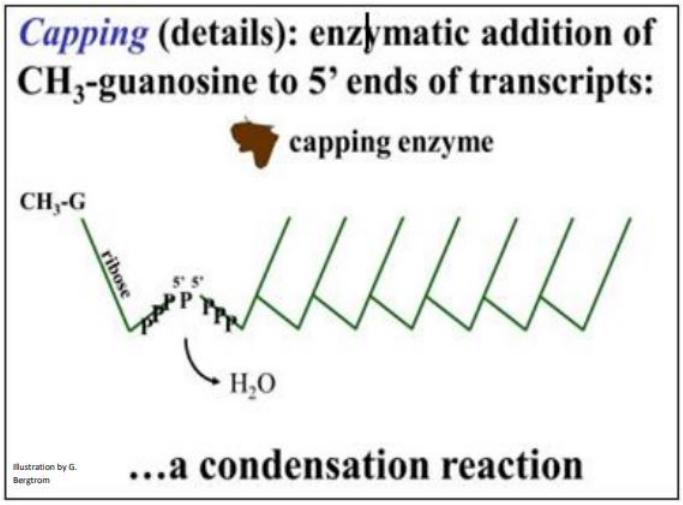
In this example, regions of strong similarity in different (non-homologous) introns in the same gene align during synapsis of meiosis. Unequal crossing over between the genes inserts exon C in one of the genes. The other gene loses the exon (not shown in the illustration). In sum, introns are *buffers against deleterious mutations*, and equally valuable, are potential targets for gene duplication and exon shuffling. This makes introns key players in creating *genetic diversity, the hallmark of evolution*.

196 Origin of Introns

197 Intron Evolution-What was selected here?

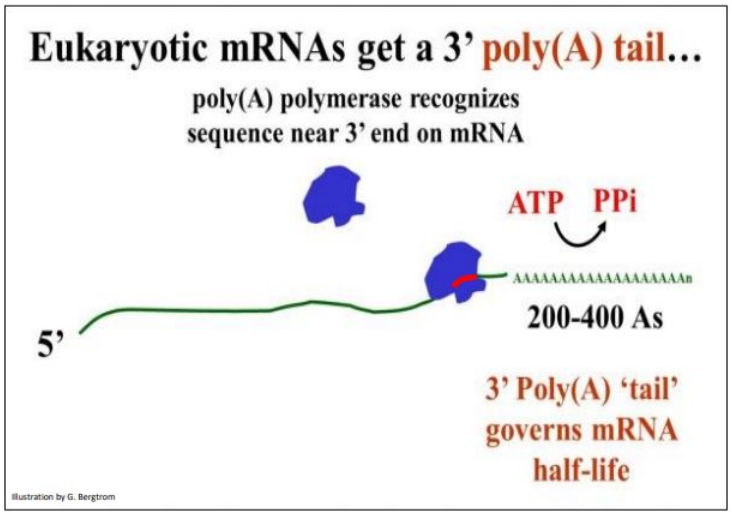
E. Capping

A **methyl guanosine cap** added 5'-to-5' to an mRNA functions in part to help mRNAs leave the nucleus and associate with ribosomes. The cap is added to an exposed 5' end, even as transcription and splicing are still in progress. A *capping enzyme* places a methylated guanosine residue at the 5'-end of the mature mRNA. The 5' cap structure is shown below (check marks are 5'-3' linked nucleotides).



F. Polyadenylation

After transcription termination, **poly(A) polymerase** catalyzes the addition of multiple AMP residues (several hundred in some cases) to the 3' terminus by the enzyme. The enzyme binds to an **AAUAA** sequence near the 3' end of an mRNA and begins to catalyze the addition of the adenosine monophosphates. The AAUAA *poly(A)* recognition site is indicated in red in the illustration of polyadenylation shown below.



The result of *polyadenylation* is a 3' **poly (A) tail** whose functions include assisting in the transit of mRNAs from the nucleus and regulating the half-life of mRNAs in the cytoplasm. The poly (A) tail shortens each time a ribosome completes translating the mRNA.

198 mRNA 5' Capping and 3' Polyadenylation

This page titled 17.1: Details of Eukaryotic mRNA Processing is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **10.4: Details of Eukaryotic mRNA Processing** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

17.2: Gene Expression- Transcription

Learning Objectives

- Describe the roles that the promoter, coding region and untranslated regions of a gene play in gene expression.
- Describe mRNA processing steps.
- Draw the process of transcription and include the following in your drawing. DNA template and non-template strands, RNA polymerase, new RNA strand, and direction of RNA synthesis.
- Draw the process of mRNA processing and include the following in your diagram, Gene (DNA), promoter, coding region, introns, exons, pre-mRNA, mature mRNA, poly A tail, cap.

Introduction

Genes are DNA sequences that control traits in an organism by coding for proteins (Figure 1). Organisms such as plants and animals have tens of thousands of genes. The impact that a single gene's information can have on an organism, however, is tremendous. Furthermore, organisms have all their genes in each of their cells, but they only need to use the information from a subset of these genes, depending on the type of cell and the cell's stage of development. Therefore, the key to gene function is controlling its expression.

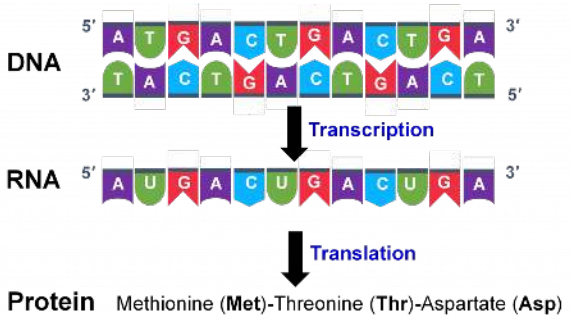


Figure 1. The central dogma of molecular genetics. Image by Marjorie Hanneman and Walter Suza.

Gene structure and transcription

The DNA sequence contains the information to control all biological functions, including the manifestation of traits important to agriculture (yield, drought tolerance, disease resistance, etc., etc.) How is the information contained in DNA sequences converted into the cellular activities necessary for plants and other organisms to function? DNA sequences are used to direct the synthesis of other molecules that actually perform these cellular functions.

Most typical genes encode proteins. The production of a protein from a gene involves several different processes (Figure 1). Transcription involves the copying of the DNA nucleotide sequence into an intermediate nucleotide molecule called RNA (ribonucleic acid). The primary RNA molecule is processed into a mature messenger RNA (mRNA) which then provides the information for the synthesis of a protein through the process of translation. Proteins are composed of amino acids connected by peptide bonds. The sequence of amino acids is determined by the sequence of nucleotide bases in the mRNA. The 20 amino acids have different chemical and physical properties and the sequence of amino acids determine the structure and function of the protein.

Gene structure

Only particular regions of chromosomal DNA are transcribed. A **gene** can be considered as the region of transcribed DNA, along with associated regions of DNA important for the regulation of transcription (Figure 1). A gene has several parts that are each important to the function of the gene.

The regulatory region (also known as the promoter) contains DNA sequence involved in the control of where and when the genes will be turned on to produce mRNA. The coding region is the part of the gene that is used as template to produce RNA molecules

in a process called **transcription**. Some RNA molecules perform cellular functions directly while many others (messenger RNAs) are used to direct the synthesis of proteins in a process called **translation**.

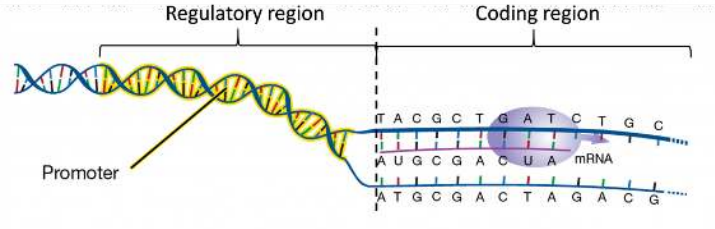


Figure 2. Every gene has a promoter and a coding region. Adapted from NIH-NHGRI.

a. Gene Promoter

The signals for starting and stopping transcription are located within DNA sequences. Specific nucleotide segments called promoters are recognized by RNA polymerase to start RNA synthesis. After the transcription of full-length RNA strand is completed, a second segment of DNA called terminator invokes termination of RNA synthesis and the detachment of RNA polymerases from the DNA template.

b. Protein-coding region

The protein-coding region of a gene is composed of the sequence of nucleotides that codes for **amino acids**. As described further in the section on **translation**, the coding region begins with an ATG **start codon** (AUG in RNA) and then ends with one of three **stop codons**. These sequences include only exons, but not all exonic sequences are protein coding as they may include **untranslated regions**.



Figure 3. The protein-coding region of a gene contains nucleotides that codes for amino acids. The 5′ and a 3′ UTR sequences do not code for amino acids but contain regulatory sequences that influence gene expression. In mRNA, translation start is AUG and translation stop can be UAA, UAG, and UGA. Image by Walter Suza.

c. Untranslated regions (UTRs)

Mature transcripts contain some sequences that do not code for amino acid sequences in proteins. These are referred to as untranslated regions or UTRs. Most mRNA transcripts contain a 5′ and a 3′ UTR. The 5′ UTR contains sequences toward the 5′ end of the mRNA sequence, before the **start codon**. These sequences can often be important for translational regulation, and sometimes other functions. The sequences following the **stop codon** are the 3′ UTR. The 3′ UTR may also have important functions regulating transcript stability or directing transcript localization within cells, or sometimes even transport (trafficking) between cells.

UTRs and introns are often useful in genetic studies. Protein coding regions are under strong selective pressure to produce functional proteins and so sequence variation is relatively rare. UTRs and introns on the other hand are under less stringent selection and are therefore sources of sequence variants that can be used to develop genetic markers.

Transcription

The genetic information of DNA is transferred to an intermediate molecule called RNA that is often translated to amino acid sequences used to build proteins. RNA is a nucleic acid, like DNA, but with some important differences. RNA contains a ribose sugar group instead of the deoxyribose found in DNA. RNA molecules are single stranded, instead of being double stranded. RNA contains a uridine (U) base and does not contain a thymidine base. The other bases (A, C, G) are contained in both RNA and DNA. U has the property of base-pairing with A.

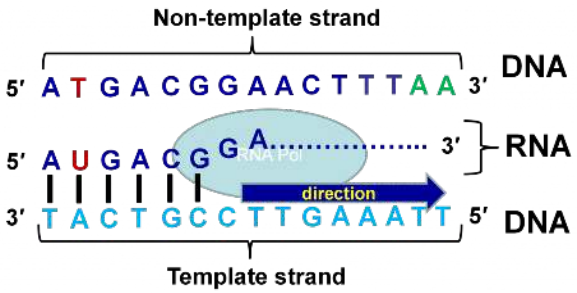


Figure 4. The enzyme RNA polymerase (RNA Pol) uses the template strand to synthesize RNA in the 5′ to 3′ prime direction. The base T is replaced with U in RNA. Image by Walter Suza.

RNA synthesis is directed by a DNA **template** in a process called **transcription**. A protein complex containing the enzyme **RNA polymerase** synthesizes an RNA molecule by adding nucleosides to the 3′ end of a growing chain. The principle of base pairing is used again and each nucleotide base added is complementary to the corresponding base on the DNA template. Thus, the RNA is complementary in sequence to the **template strand** of DNA, which is also referred to as "antisense" or "negative" strand (Figure 4). The RNA is identical in sequence (except U replaces T) to the other strand, which is called the "sense" or "positive" strand. Because RNA molecules are produced by the process of transcription, they are often referred to as **transcripts**.

RNA processing

Coding (transcribed) region

This is the region that is transcribed by RNA polymerase, also known as the RNA coding region. As described below, it may include **introns**, sequences that are removed from the mature RNA molecule during **RNA processing**. The transcribed region is demarcated by promoter and terminator sequences.

Introns and exons

As mentioned, and described in detail below, **introns** are sequences that are removed from transcripts during RNA processing. Sequences that are retained in mature transcripts are called **exons**. The corresponding stretches of DNA are typically referred to with the same terms. Introns are commonly found in genes of eukaryotes but are rare in prokaryotic organisms.

Intron splicing

The process of transcription produces pre-mRNA that contains both introns and exons. The process of splicing involves removal of introns from pre-mRNA and joining together the exons. A complex group of proteins that form a spliceosome perform the splicing reaction.

Introns sometimes serve as boundaries for sequences encoding functional protein domains, leading to possibility for new and variant proteins by exon shuffling. Also, introns can provide possibility for productions of variant RNA forms through alternate splicing allowing more than one gene product from a single gene. Some introns result from the insertion of transposable elements and may be spliced perfectly of imperfectly, offering more possibility for new genetic diversity.

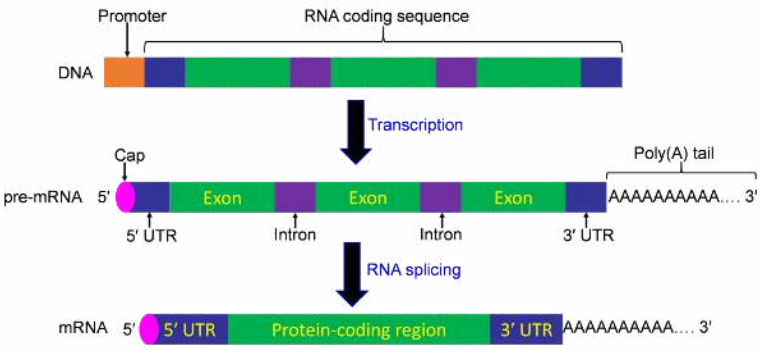


Figure 5. DNA (gene) transcription produces precursor-mRNA (pre-mRNA) that contains both introns and exons. The 5′ cap is 7-methyl guanidine. The enzyme poly(A) polymerase adds the poly(A) tail. The process of splicing involves removal of introns from pre-mRNA and joining together the exons to form mature mRNA. Image by Walter Suza

5′ Capping

The 5’ capping is the addition of a 7-methyl guanidine to the first nucleotide of mRNA molecule, usually and adenine or guanidine. The phosphodiester linkage between 7-methyl guanidine and the target nucleotide is 5′-5′ instead of 5′-3′, and 3 phosphates rather than 1 are retained in the linkage. The cap stabilizes the 5′ end of the mRNA and plays a role in translation initiation.

Poly adenylation

The transcription of a gene may proceed beyond what ends up as 3′ end of mature mRNA. Thus the 3′ end of mRNA is formed after transcription. The enzyme poly(A) polymerase adds numerous adenosines to the 3′ end to result in what is called the poly(A) tail. The poly(A) tail is necessary for proper processing and transport of mRNA to the cytoplasm. The poly(A) tail is also important for the stability of mRNA, and initiation of translation in eukaryotic organisms.

Summary

The genetic information of DNA is transferred through transcription to an intermediate molecule called RNA. The signals for starting and stopping transcription are located within the DNA sequence and referred to as promoter and terminator sequences. The coding region of a gene is composed of a sequence of nucleotides that are transcribed into RNA. These sequences include exons and introns. Exons are the sequences that code for proteins. The coding region of a gene contains exons and introns. Also, pre-mRNA contains both introns and exons. The introns in pre-mRNA are removed through a process called intron splicing. The mRNA is processed by 5′ capping and addition of a poly(A) tail.

Learning Activities

Learning Activity 1

Given the following sequence of double-stranded DNA, predict the sequence of the RNA strand.

Learning Activity 2

Watch This Video: *Transcription Detail*

A video element has been excluded from this version of the text. You can watch it online here: https://iastate.pressbooks.pub/genagbiotech/?p=24

This page titled 17.2: Gene Expression- Transcription is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Philip Becraft, & Philip Becraft (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.5: Gene Expression- Transcription** by Donald Lee, Marjorie Hanneman, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

18: Transcriptional Regulation (eukaryotes)

18.1: Eukaryotic Gene Regulation

18.2: Regulatory Elements in Evolution

18.3: Additional Levels of Regulating Transcription

18.4: Gene Regulation in Eukaryotes

18.5: Transgenic organisms

18: Transcriptional Regulation (eukaryotes) is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

18.1: Eukaryotic Gene Regulation

Like prokaryotes, transcriptional regulation in eukaryotes involves both *cis*-elements and *trans*-factors, only there are more of them and they interact in a more complex way. A diagram of a typical eukaryotic gene, including several types of *cis*-elements, is shown in Figure 18.1.7.

Proximal Regulatory Sequences

As in prokaryotes the RNA polymerase binds to the gene at its promoter to begin transcription. In eukaryotes, however, RNApol is part of a large protein complex that includes additional proteins that bind to one or more specific *cis*-elements in the promoter region, including **GC-rich boxes**, **CAAT boxes**, **and TATA boxes**. High levels of transcription require both the presence of this protein complex at the promoter, as well as their interaction with other *trans*-factors described below. The approximate position of these elements relative to the **transcription start site** ($^{+}1$) is shown in Figure 18.1.7, but it should be emphasized that the distance between any of these elements and the transcription start site can vary, but are typically within ~200 base pairs of the start of transcription. This contrasts the next set of elements.

Distal Regulatory elements

Even more variation is observed in the position and orientation of the second major type of *cis*-regulatory element in eukaryotes, which are called **enhancer elements**. Regulatory *trans*-factor proteins called **transcription factors** bind to enhancer sequences, then, while still bound to DNA, these proteins interact with RNApol and other proteins at the promoter to enhance the rate of transcription. There are a wide variety of different transcription factors and each recognizes a specific DNA sequence (enhancer element) to promote gene expression in the adjacent gene under specific circumstances. Because DNA is a flexible molecule, enhancers can be located near (~100s of bp) or far (~10K of bp), and either upstream or downstream, from the promoter (Figures 18.1.7 and 18.1.8).

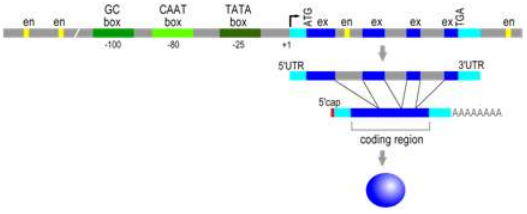


Figure 18.1.7: Structure of a typical eukaryotic gene. RNA polymerase binding may involve one or more cis-elements within the proximal region of a promoter (green boxes). Enhancers (yellow boxes) may be located any distance upstream or downstream of the promoter and are also involved in regulating gene expression. The processing of a primary transcript to a mature mRNA is also shown. (Origianl-Deyholos-CC:AN)

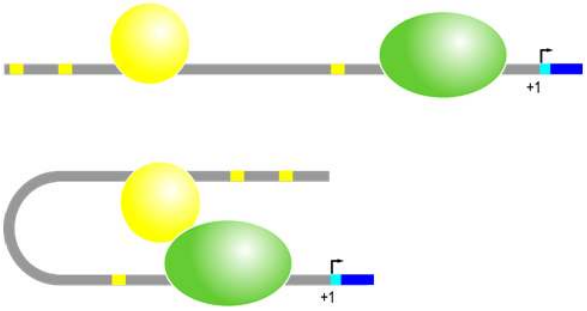


Figure 18.1.8: A transcription factor (yellow) bound to an enhancer that is located far from a promoter. Because of the flexibility of the DNA molecule, the transcription factor and RNApol are able to interact physically, even though the cis-elements to which they are bound are located far apart. In eukaryotic cells, RNApol is actually part of a large complex of proteins (not shown here) that assembles at the promoter. (Origianl-Deyholos-CC:AN)

Example 1: Drosophila yellow gene

The *yellow* gene of Drosophila provides an example of the modular nature of enhancers. This gene encodes an enzyme in the pathway that produces a dark pigment in the insect's exoskeleton. Mutants have a yellow cuticle rather than the wild type darker pigmented cuticle. (Why call the gene "yellow": recall that genes are often named after their mutant phenotype.) Figure 18.1.9

shows three enhancer elements (left side - wing, body, mouth), each binds a different tissue specific transcription factor to enhancer transcription of *yellow*$^{+}$ in that tissue and makes the pigment. So, the wing cells will have a transcription factor that binds to the wing enhancer to drive expression; likewise in the body and mouth cells. Thus, specific combinations of *cis*-elements and *trans*-factors control the differential, tissue-specific expression of genes. This type of combinatorial action of enhancers is typical of the transcriptional activation of most eukaryotic genes: specific transcription factors activate the transcription of target genes under specific conditions.

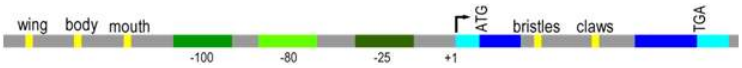


Figure 18.1.9: Tissue-specific cis-regulatory elements within a simplified representation of the yellow gene of Drosophila. (Origianl-Deyholos-CC:AN)

While enhancer sequences promote expression, there is an oppositely acting type of element, called **silencers**. These elements function in much the same manner, with transcription factors that bind to DNA sequences, but they act to silence or reduce transcription from the adjacent gene.

Again, a gene's expression profile (transcription level, tissue specific, temporal specific) is a combination of various enhancer and silencer elements.

Example 2: Gal4-UAS system from yeast – a genetic tool

Genetic researchers have taken advantage of a yeast distal enhancer sequence to make the **GAL4-UAS system**, a powerful technique for studying the expression of genes in other eukaryotes. It relies on two parts: a "**driver**" and a "**responder**" (Figure 18.1.10). The driver part is a gene encoding a yeast transcriptional activator protein called Gal4. It is separate from the responder part, which contains the enhancer sequence, or upstream activation sequence (UAS, as it is called in yeast) to which the Gal4 protein specifically binds. This UAS is placed upstream (using genetic engineering) from a promoter transcribing a reporter gene, or other gene of interest, such as GFP (green fluorescent protein).

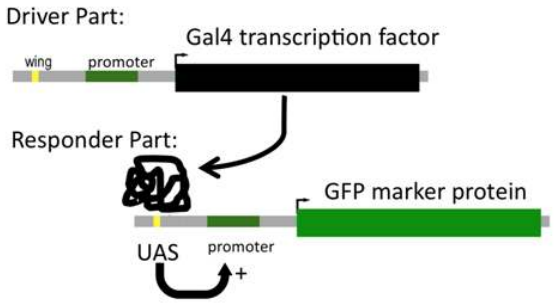


Figure 18.1.10: The GAL4-UAS system. The driver, with a wing enhancers, expresses the Gal4 protein that then binds to the UAS element upstream of a marker gene, GFP. This would express the GFP in the wing tissues. The modular aspect of this system would let the wing enhancer be replaced by any other enhancer and the GFP marker replaced with any other gene. (Origianl-Locke-CC:AN)

Both parts must be present in the same cell for the system to express the responder gene. If the driver is absent, the responder product will not be expressed. However, both are in the same cell (or organism) the pattern of expression of the driver part will induce the responder part's expression in the same pattern.

This system works is a variety of eukaryotes, including humans. It has been especially well exploited in Drosophila where 100's (1,000's ?) of differently expressing driver lines are available. These lines permit the tissue specific expression of any responder gene to examine its effect on development.

This page titled 18.1: Eukaryotic Gene Regulation is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **12.3: Eukaryotic Gene Regulation** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

18.2: Regulatory Elements in Evolution

Mutations can occur in both *cis*-elements and *trans*-factors; both can result in altered patterns of gene expression. If an altered pattern of gene expression results in a selective advantage (or at least do not produce a major disadvantage), they may be selected and maintained in future populations. They may even contribute to the evolution of new species. An example of a sequence change in an enhancer is found in the *Pitx* gene.

Example: *Pitx* expression in Stickleback

The three-spined **stickleback** (Figure 18.2.11) provides an example of natural selection of a mutation in a *cis*-regulatory element. This fish occurs in two forms: (1) populations that inhabit deep, open water and have a spiny pelvic fin that deters larger predator fish from feeding on them; (2) populations from shallow water environments and lack this spiny pelvic fin. In shallow water, it appears that a long, spiny pelvic fin would be a disadvantage because it frequently contacts the sediment at the bottom of the pond and allows parasitic insects in the sediment to invade the stickleback. Researchers compared gene sequences of individuals from both deep and shallow water environments as shown in Figure 18.2.11. They observed that in embryos from the deep-water population, a gene called *Pitx* was expressed in several groups of cells, including those that developed into the pelvic fin. Embryos from the shallow-water population expressed *Pitx* in the same groups of cells as the other population, with an important exception: *Pitx* was not expressed in the pelvic fin **primordium** in the shallow-water population. Further genetic analysis showed that the absence of *Pitx* gene expression from the developing pelvic fin of shallow-water stickleback was due to the absence (mutation) of a particular enhancer element upstream of *Pitx*.

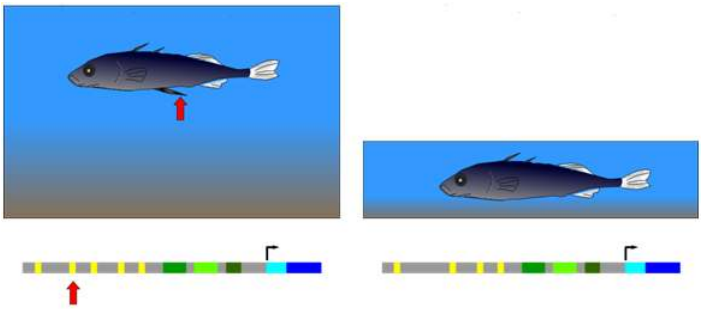


Figure 18.2.11: Development of a large, spiny pelvic fin in deep-water stickleback (left) depends on the presence of a particular enhancer element upstream of a gene called Pitx. Mutants lacking this element, and therefore the large pelvic fin (right), have been selected for in shallow-water environments. (Wikipedia-Richard Wheeler-GFDL)

Example: Hemoglobin expression in placental mammals.

Hemoglobin is the oxygen-carrying component of red blood cells (erythrocytes). Hemoglobin usually exists as tetramers of four non-covalently bound hemoglobin molecules (Fig 12.12). Each hemoglobin molecule consists of a **globin** polypeptide with a covalently attached heme molecule. Heme is made through a specialized metabolic pathway and is then bound to globin polypeptide through **post-translational modification**.

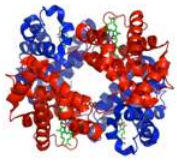


Figure 18.2.12: A tetramer of human hemoglobin, type $\alpha 2\beta 2$. The α chains are labeled red, and the β chains are labeled blue. Heme groups are green. (Wikipedia-from Wood, W.G. 1976 Br. Med. Bull. 32, 282-CC:AS)

The composition of the tetramers changes during development (Figure 18.2.13). From early childhood onward, most tetramers are of the type $\alpha_2\beta_2$, which means they contain of two copies of each of two slightly different globin proteins named α and β. A small amount of adult hemoglobin is $\alpha_2\delta_2$, which has δ globin instead of the more common β globin. Other tetrameric combinations predominate before birth: $\zeta_2\varepsilon_2$ is most abundant in embryos, and $\alpha_2\gamma_2$ is most abundant in fetuses. Although the six globin proteins (α = alpha, β = beta , γ = gamma, δ =delta, ε =epsilon , ζ = zeta) are very similar to each other, they do have slightly different

functional properties. For example, fetal hemoglobin has a higher oxygen affinity than adult hemoglobin, allowing the fetus to more effectively extract oxygen from maternal blood. The specialized γ globin genes that are characteristic of fetal hemoglobin are found only in placental mammals.

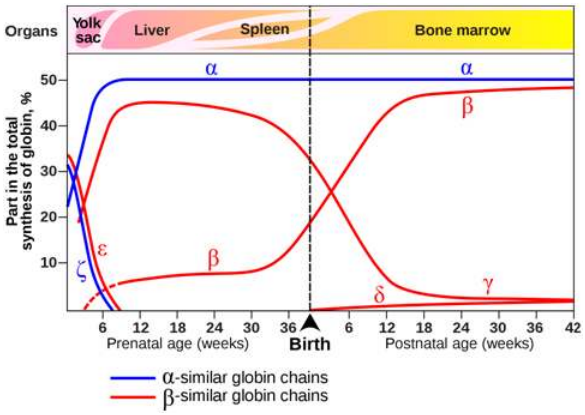


Figure **18.2.13**: Expression of globin genes during prenatal and postnatal development in humans. The organs in which globin genes are primarily expressed at each developmental stage are also indicated. (Origianl-Deyholos-CC:AN)

Each of these globin polypeptides is encoded by a different gene. In humans, globin genes are located in clusters on two chromosomes (Figure **18.2.14**). We can infer that these clusters arose through a series of duplications of an ancestral globin gene. **Gene duplication** events can occur through rare errors in processes such as DNA replication, meiosis, or transposition. The duplicated genes can accumulate mutations independently of each other. Mutations can occur in either the regulatory regions (e.g. promoter regions), or in the coding regions, or both. In this way, the promoters of globin genes have evolved to be expressed at different phases of development, and to produce proteins optimized for the prenatal environment.

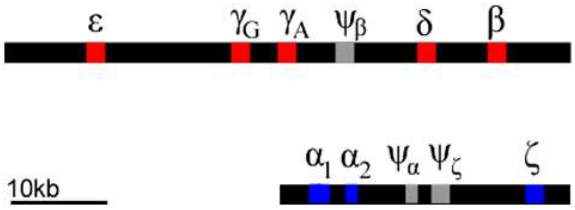


Figure **18.2.14**: Fragments of human chromosome 11 and human chromosome 16 on which are located clusters of β-like and α-like goblin genes, respectively. Additional globin genes (θ, μ) have also been described by some researchers, but are not shown here. (Origianl-Deyholos-CC:AN)

Of course, not all mutations are beneficial: some mutations can lead to inactivation of one or more of the products of a gene duplication. This can produce what is called a **pseudogene**. Examples of pseudogenes (ψ) are also found in the globin clusters. Pseudogenes have mutations that prevent them from being expressed at all. The globin genes provide an example of how gene duplication and mutation, followed by selection, allows genes to evolve specialized expression patterns and functions. Many genes have evolved as **gene families** in this way, although they are not always clustered together as are the globins.

This page titled 18.2: Regulatory Elements in Evolution is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **12.4: Regulatory Elements in Evolution** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

18.3: Additional Levels of Regulating Transcription

Eukaryotes regulate transcription via promoter sequences close to the transcription unit (as in prokaryotes) and also use more distant enhancer sequences to provide more variation in the timing, level, and location of transcription, however, there are still additional levels of genetic control. This consists of two major mechanism: (1) large-scale changes in chromatin structure, and (2) modification of bases in the DNA sequence. These two are often inter-connected.

Chromatin Dynamics

Despite the simplified way in which we often represent DNA in figures such as those in this chapter, DNA is almost always associated with various chromatin proteins. For example, histones remain associated with the DNA even during transcription. Thus the rate of transcription is also controlled by the accessibility of DNA to RNApol and regulatory proteins. So, in regions were the chromatin is highly compacted, it is unlikely that any gene will be transcribed, even if all the necessary *cis*- and *trans*- factors are present in the nucleus. The extent of chromatin compaction in various regions is regulated through the action of **chromatin remodeling** proteins. These protein complexes include enzymes that add or remove chemical tags, such as methyl or acetyl groups, to various DNA bound proteins. These modifications alter the local chromatin density and thus the availability for transcription. **Acetylated** histones, for example, tend to be associated with actively transcribed genes, whereas **deacetylated** histone are associated with genes that are silenced (Figure 18.3.15).

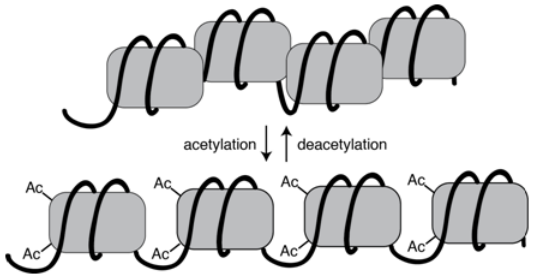


Figure 18.3.15: Acetylation of histone proteins is associated with more a more open chromatin configuration. Acetylation is a reversible process. (Origianl-Deyholos-CC:AN)

Likewise, **methylation** of DNA itself is also associated with transcription regulation. Cytosine bases, particularly when followed by a guanine (**CpG sites**) are important targets for DNA methylation (Figure 18.3.16). Methylated cytosine within clusters of CpG sites is often associated with transcriptionally inactive DNA.

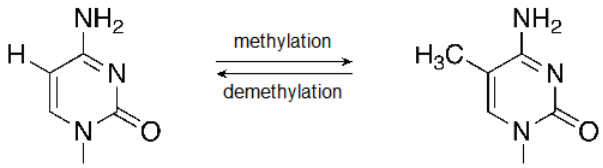


Figure 18.3.16: The methylation reaction shown here produces 5-methylcytosine (5mC). Methyl groups may also be removed by various processes. (flickr-Beardy Git-CC:AND)

The modification of DNA and its associated proteins is enzymatically reversible (acetylation/deacetylation; methylation/demethylation) and thus a cyclical activity. Regulation of this provides another layer through which eukaryotic cells control the transcription of specific genes.

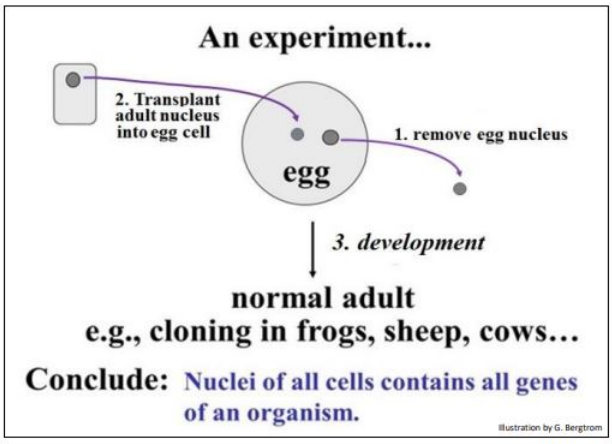
This page titled 18.3: Additional Levels of Regulating Transcription is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **12.5: Additional Levels of Regulating Transcription** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

18.4: Gene Regulation in Eukaryotes

A. The Difference between Eukaryotic and Prokaryotic Gene Regulation

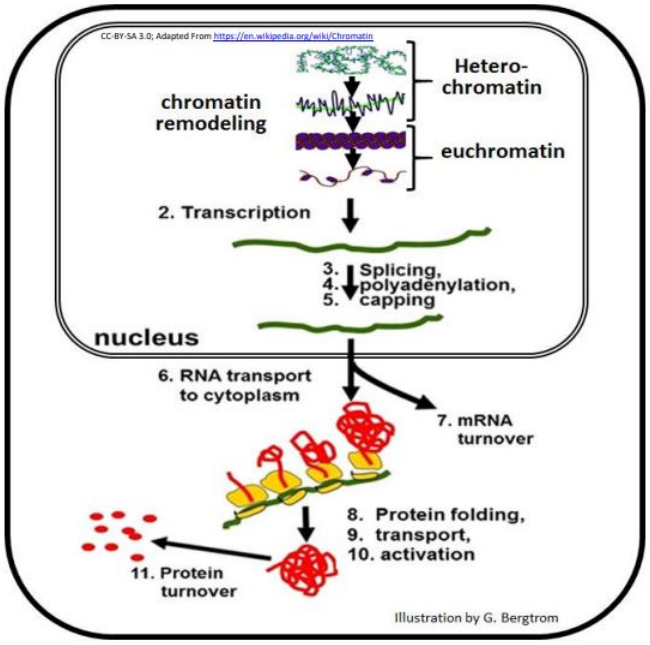
Let's Recall an experiment described earlier and illustrated below.



Results of this experiment provided the evidence that even very different cells of an organism contain the same genes. In fact, in any multicellular eukaryotic organism, every cell contains the same DNA (genes). Therefore, the different cell types in an organism differ not in which genes they contain, but which sets of genes they express! Looked at another way, cells differentiate when they turn on new genes and turn off old ones. Thus, gene regulation produces different sets of gene products during differentiation, leading to cells that look and function differently in the organism.

220 An Experiment: All of an Organism's Cells Have the Same Genome

Compared to prokaryotes, many steps in eukaryotes lie between transcription of an mRNA and the accumulation of a polypeptide end product. Eleven of these steps are shown in the pathway from gene to protein below.

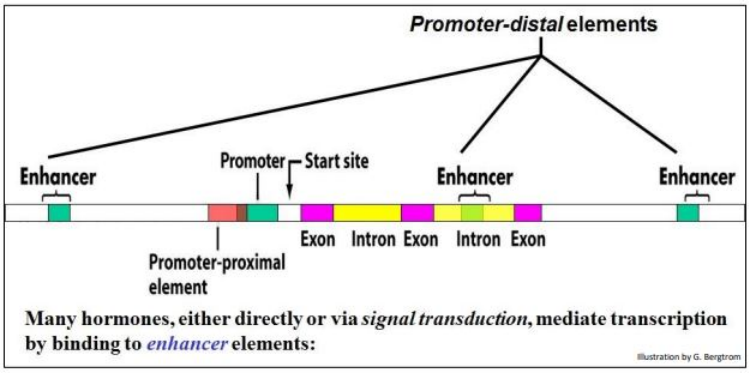


Theoretically, cells could turn on, turn off, speed up or slow down any of the steps in this pathway, changing the steady state concentration of a polypeptide in the cells. While regulation of any of these steps is possible, the expression of a single gene is typically controlled at only one or a few steps. A common form of gene regulation is at the level of transcription initiation, similar to transcriptional control in bacteria, in principle if not in detail.

B. Complexities of Eukaryotic Gene Regulation

Gene regulation in eukaryotes is more complex than in prokaryotes. This is in part because their genomes are larger and because they encode more genes. For example, the E. coli genome houses about 5,000 genes, compared to around 25,000 genes in humans. Furthermore, eukaryotes can produce even more than 25,000 proteins by alternative splicing of mRNAs and in at least a few cases, by initiating transcription from alternative start sites in the same gene. And of course, the activity of many more genes must be coordinated without the benefit of multigene operons! Finally, eukaryotic gene regulation is made more complicated because all nuclear DNA is wrapped in protein in the form of chromatin.

All organisms control gene activity with **transcription factors** that bind to specific DNA sequences (**cis regulatory elements**). In eukaryotes, these elements can be **proximal to** (near) the promoter of a gene, or **distal to** (quite far from) the gene they regulate. A eukaryotic map showing the components of a typical gene and its associated cis-acting regulatory elements is shown below.

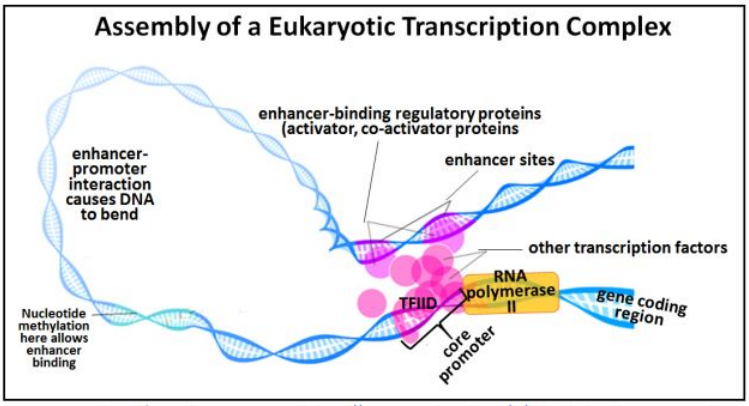


Enhancers are typical *distal* cis elements that recognize and bind transcription factors to increase the rate of transcription of a gene. Oddly enough, these short DNA elements can be in the 5' or 3' non-translated region of the gene, or even within introns, and can lie thousands of base pairs away from the genes they control. Note that enhancer elements are even in introns can also be very far from the start-site of transcription of a gene.

Upstream regulatory regions of eukaryotic genes (to the left of a gene promoter as shown above) often have distal binding sites for more than a few transcription factors, some with positive (*enhancing*) and others with negative (*silencing*) effects. Of course, which of these DNA regions are active in controlling a gene depends on which transcription factor(s) are present in the nucleus. Sets of positive regulators will work together to coordinate and maximize gene expression when needed, and sets of negative regulators will bind negative regulatory elements to silence a gene.

222 Transcription Factors Bind DNA Near and Far

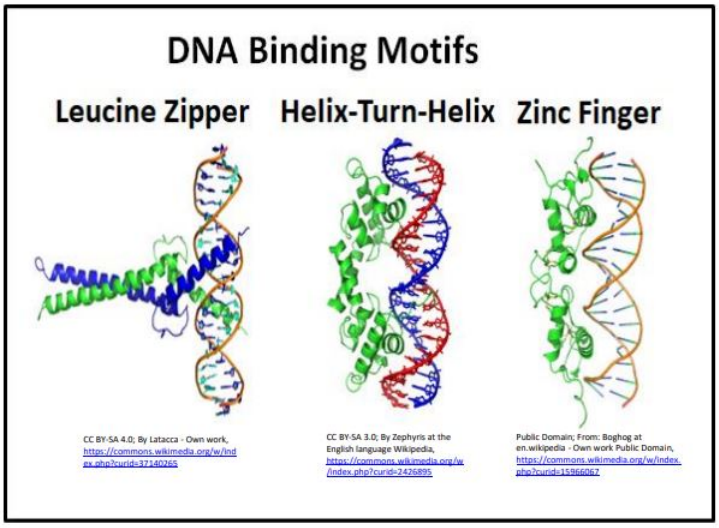
We saw that in eukaryotes, the initiation of transcription involves many transcription factors and RNA polymerase II acting at a gene promoter to form a **transcription preinitiation complex**. **TFIID**, or **TATA binding protein** is one of the first factors to bend, causing the DNA in the promoter region to bend, much like the CAP protein in bacteria. TFIID also recruits other transcription factors to the promoter. As in bacteria, bending the DNA loosens H-bonds between bases, facilitating unwinding the double helix near the gene. Bending eukaryotic DNA also brings distal regulatory proteins bound to enhancer sequences far from the promoter together with the proteins bound to more proximal regulatory elements, as shown in the drawing below.



Nucleotide methyation sites may facilitate regulatory protein-enhancer binding. When such regulatory proteins, here called *activators* (i.e., of transcription), bind to their enhancers, they acquire an affinity for protein cofactors that enable recognition and binding to other proteins in the transcription initiation complex. This attraction stabilizes the bend in the DNA that then makes it easier for RNA polymerase II to initiate transcription

223 Assembling a Eukaryotic Transcription Initiation Complex

It is worth reminding ourselves that it is shape and allosteric change that allow DNAprotein interactions (in fact, any interactions of macromolecules). The lac repressor we saw earlier is a transcription factor with *helix-turn-helix* DNA binding motifs. This motif and two others (*zinc finger*, and *leucine zipper*) characterize DNA binding proteins are illustrated below.



DNA-binding motifs in each regulatory protein shown here bind one or more regulatory elements ‘visible’ to the transcription factor in the major groove of the double helix.

224 Transcription Factor Domains/Motifs Bind Specific DNA Sequences

We will look next at some common ways in which eukaryotic cells are signaled to turn genes on or off, or to increase or decrease their rates of transcription. As we describe these models, remember that eukaryotic cells regulate gene expression in response to changes in extracellular environments. These can be unscheduled, unpredictable changes in blood or extracellular fluid composition (ions, small metabolites), or dictated by changes in a long-term genetic program of differentiation and development. Changes in gene expression even obey circadian (daily) rhythms, the ticking of a clock. In eukaryotes, changes in gene expression, expected or not, are usually mediated by the timely release of chemical signals from specialized cells (e.g., hormones, cytokines, growth factors, etc.). We will focus on some betterunderstood models of gene regulation by these chemical signals.

C. Regulation of Gene Expression by Hormones that enter Cells and Those That Don't

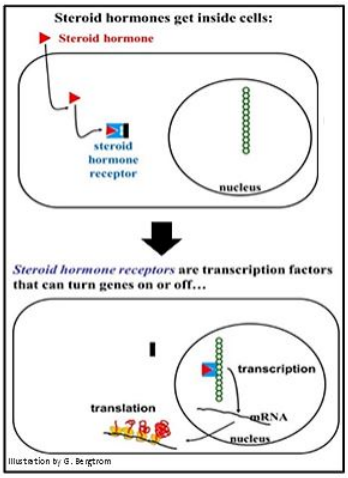
Gene-regulatory (cis) elements in DNA and the transcription factors that bind to them have co-evolved. But not only that! Eukaryotic organisms have evolved complete pathways that respond to environmental or programmed developmental cues and lead to an appropriate cellular response. Chemicals that regulate genes in prokaryotes are not usually signals communicated by other cells. In eukaryotes, chemicals released by some cells signal other cells to respond, thus coordinating the activity of the whole organism. Hormones released by cells in endocrine glands are well-understood signal molecules; hormones affect *target cells* elsewhere in the body.

225 Chemicals That Control Gene Expression

1. How Steroid Hormones Regulate Transcription

Steroid hormones cross the cell membranes to have their effects. Common steroid hormones include testosterone, estrogens, progesterone, glucocorticoids and mineral corticoids. Once in target cell, such hormones bind to a **steroid hormone** receptor protein to form a **steroid hormone-receptor complex**. The receptor may be in the cytoplasm or in the nucleus, but in the end, the hormone-receptor complex must bind to DNA regulatory elements of a gene to either enhance or silence transcription. Therefore, a steroid hormone must cross the plasma membrane, and may also need to cross the nuclear envelope.

Follow the Binding of a steroid hormone to a cytoplasmic receptor below.



Here the hormone (the triangle) enters the cell. An allosteric change in the receptor releases a protein subunit called Hsp90 (the black rectangle in the illustration). The remaining hormone-bound receptor enters the nucleus.

The fascinating thing about Hsp90 is that it was first discovered in cells subjected to heat stress. When the temperature gets high enough, cells shut down most transcription and instead transcribe Hsp 90 and/or other special *heat shock* genes. The resulting **heat shock proteins** seem to protect the cells against metabolic damage until temperatures return to normal. Since most cells never experience such high temperatures, the evolutionary significance of this protective mechanism is unclear. As we now know, heat shock proteins have critical cellular functions, in this case blocking the DNA-binding site of a hormone receptor until a specific steroid hormone binds to it.

Back to hormone action! No longer associated with the Hsp90 protein, the receptor bound to its hormone cofactor binds to a cis-acting transcription control element in the DNA, turning transcription of a gene on or off. The hormone receptors for some steroid hormones are already in the nucleus of the cell, so the hormone must cross not only the plasma membrane, but also the nuclear envelope in order to access the receptor.

As for steroid hormone functions, we already saw that glucocorticoids turn on the genes of gluconeogenesis. Steroid hormones also control sexual development and reproductive cycling in females, salt and mineral homeostasis in the blood, metamorphosis in arthropods, etc., all by regulating gene expression.

226 Steroid Hormones Regulate Gene Transcription

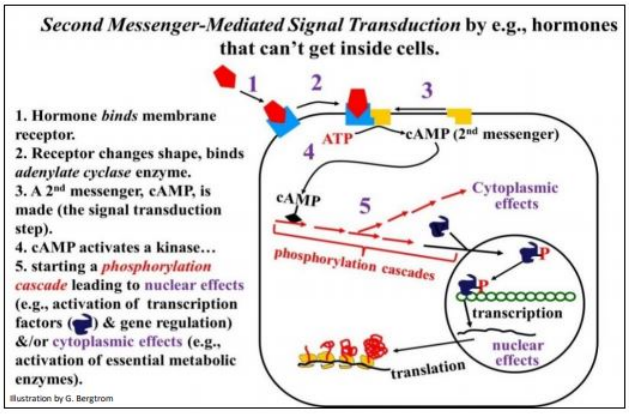
2. How Protein Hormones Regulate Transcription

Protein hormones are of course large and soluble, with highly charged surfaces. Other hormones might be relatively small (e.g., adrenalin), but are charged. Large or highly charged signal molecules cannot get across the phospholipid barrier of the plasma membrane. To have any effect at all, they must bind to receptors on the surface of cells. These receptors are typically membrane glycoproteins.

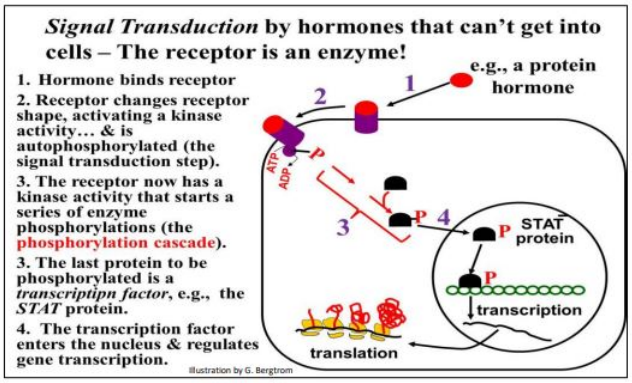
The information (signals) carried by protein hormones must be conveyed into the cell indirectly, by a process called *signal transduction*. There are two well-known pathways of signal transduction, each of which involves activating pathways of protein phosphorylation in cytoplasm. The **phosphorylation cascade** that results activates a transcription factor that binds to regulatory DNA, turning a gene on or off.

Binding of a hormone to a cell surface receptor leads to an allosteric change in the receptor. This in turn activates other proteins either in the plasma membrane or in the cytoplasm, leading to the synthesis of a cytoplasmic **second messenger**. The second messenger typically binds to a protein kinase in the cytoplasm, launching a series of protein phosphorylations, or a phosphorylation cascade. The last in the series of proteins to be phosphorylated is an activated transcription factor that will bind to a cis-regulatory DNA sequence.

cAMP was the first second messenger metabolite to be discovered. It mediates many hormonal responses, controlling both gene activity and enzyme activity. cAMP forms when the hormone-receptor in the membrane binds to and activates a membrane-bound adenylate cyclase enzyme. The cAMP produced then binds to a protein kinase, the first of several in a phosphorylation cascade. Signal transduction mediated by cAMP is summarized in the illustration below.



A different kind of signal transduction involves a hormone receptor that is itself the protein kinase. The role of **enzyme-linked hormone receptors** in signal transduction is summarized below.

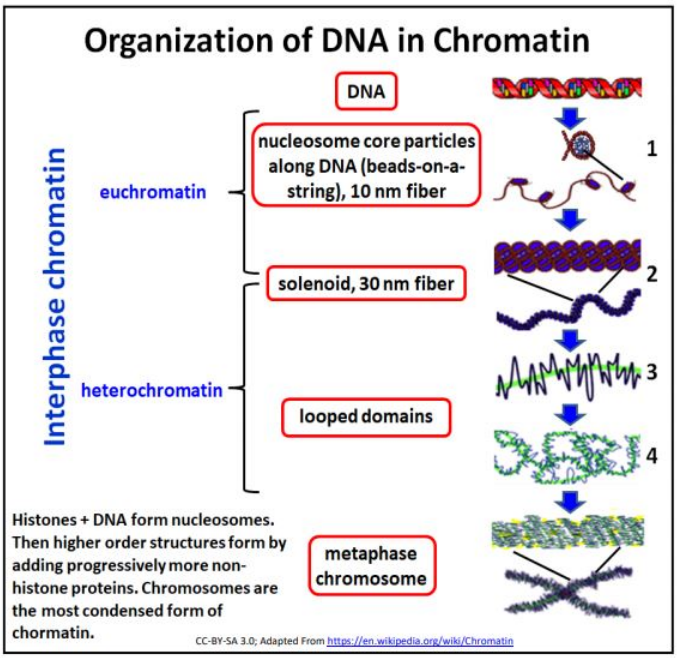


Binding of the signal protein (e.g. hormone) to the enzyme-linked receptor causes an allosteric change that activates the receptor kinase, starting phosphorylation cascade resulting in an active transcription factor. We look at signal transduction in more detail in another chapter.

227 Signal Transduction Can Lead to Gene Regulation

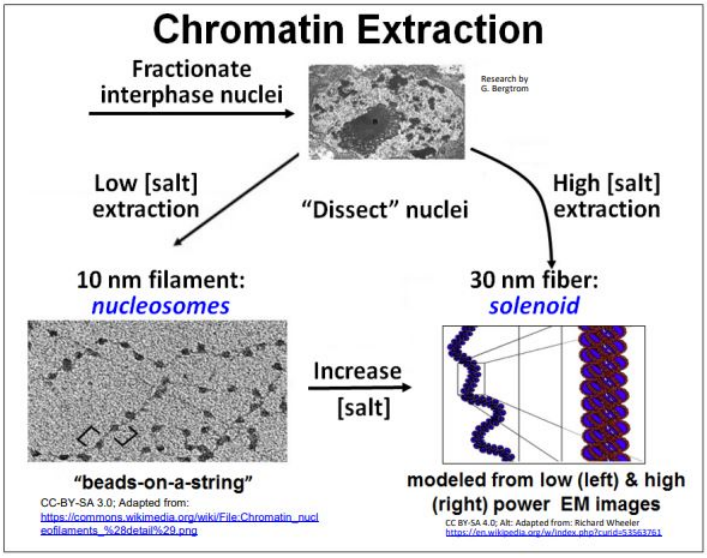
D. Regulating Eukaryotic Genes Means Contending with Chromatin

Consider again the illustration of the different levels of chromatin structure (below).



Transcription factors bind specific DNA sequences by detecting them through the grooves (mainly the major groove) in the double helix. The drawing above reminds us however, that unlike the nearly naked DNA of bacteria, eukaryotic (nuclear) DNA is coated with proteins that, in aggregate are by mass, greater than the mass of DNA that they cover. The protein-DNA complex of the genome is of course, chromatin.

Again, as a reminder, DNA coated with histone proteins forms the 9 nm diameter beads-on-a-string structure in which the beads are the **nucleosomes**. The association of specific non-histone proteins causes the nucleosomes to fold over on themselves to form the **30 nm solenoid**. As we saw earlier, it is possible to selectively extract chromatin. Take a second look at the results of typical extractions of chromatin from isolated nuclei below.



Further accretion of non-histone proteins leads to more folding and the formation of **euchromatin** and **heterochromatin** characteristic of non-dividing cells. In dividing cells, the chromatin further condenses to form the **chromosomes** that separate during either *mitosis* or *meiosis*.

Recall that biochemical analysis of the 10 nm filament extract revealed that the DNA wraps around histone protein octamers, the nucleosomes or beads in this beads-on-astring structure. Histone proteins are highly conserved in the eukaryotic evolution (they are not found in prokaryotes). They are also very basic (many *lysine* and *arginine* residues) and therefore very positively charged. This explains why they are able to arrange themselves uniformly along DNA, binding to the negatively charged *phosphodiester* backbone of DNA in the double helix.

Since the DNA in euchromatin is less tightly packed than it is in heterochromatin, perhaps active genes are to be found in euchromatin and not in heterochromatin. Experiments in which total nuclear chromatin extracts were isolated and treated with the enzyme deoxyribonuclease (DNAse) revealed that the DNA in active genes was degraded more rapidly than non-transcribed DNA. More detail on these experiments can be found in the two links below.

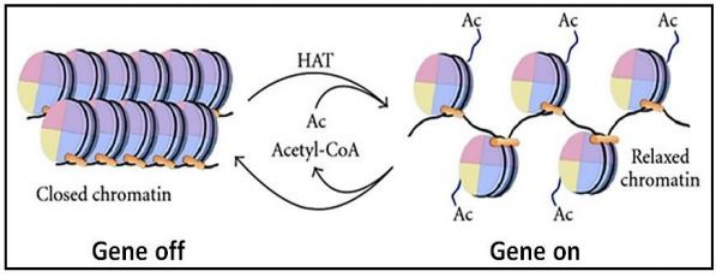
228 Question: Is Euchromatic DNA Transcribed?

229 Experiment and Answer: Euchromatin *is* Transcribed

The results of such experiments are consistent with the suggestion that active genes are more accessible to DNAse because they are in less coiled, or less condensed chromatin. DNA in more condensed chromatin is surrounded by more proteins, and thus is less accessible to, and protected from DNAse attack. When packed up in chromosomes during mitosis or meiosis, all genes are largely inactive.

Regulating gene transcription must occur in non-dividing cells or during the interphase of cells, where changing the shape of chromatin (**chromatin remodeling**) in order to silence some and activate other genes is possible. Changing chromatin conformation involves chemical modification of chromatin proteins and DNA. For example, chromatin can be modified by histone acetylation, de-acetylation, methylation and phosphorylation, reactions catalyzed by *histone acetyltransferases* (HAT enzymes), de-acetylases, methyl transferases and kinases, respectively. For example, acetylation of lysines near the amino end of histones H2B and H4 tends to unwind nucleosomes and open the underlying DNA for transcription. De-acetylation then, promotes condensation of the chromatin in the affected regions of DNA. Likewise, methylation of lysines or arginines (the basic amino acids that characterize histones!) of H3 and H4 can open DNA for transcription, while demethylation has the opposite effect. In one case, di-methylation of a lysine in H3 can suppress transcription. These chemical modifications affect recruitment of other proteins that alter chromatin conformation and ultimately activate or block transcription.

This reversible and its effect on chromatin are illustrated below.



Nucleosomes themselves can be moved, slid and otherwise repositioned by complexes that hydrolyze ATP for energy to accomplish the physical shifts. Some cancers are associated with mutations in genes for proteins involved in chromatin remodeling. This is no doubt, because failures of normal remodeling could adversely affect normal cell cycling and normal replication. In fact, a single, specific pattern of methylation may mark DNA in multiple cancer types (check out Five Cancers with the Same Genomic Signature - Implications).

E. Regulating all the Genes on a Chromosome at Once

Recall that X chromosomes in human female somatic cells is inactivated, visible in the nucleus as a Barr body. One of the two X chromosomes in female fruit flies is also inactivates. However, both males and females of Drosophila (presumably also us!) require X chromosome gene expression during embryogenesis. Given the difference in X chromosome **gene dosage** between males and females, do males get by with fewer X chromosome gene products than females?

Experiments looking at the expression of X chromosome gene in male and female flies revealed similar levels of gene products. It turns out that the activity of a nuclear body called **HLB** (**H**istone **L**ocus **B**ody) is required for increase in X chromosome gene transcription. A protein, called **CLAMP** (**C**hromatin-**L**inked **A**daptor for **M**ale-specific lethal (MSL) **P**rotein), was shown to bind to **GAGA** nucleotide repeats lying between the genes for histones 3 and 4. As there are about 100 repeats of the fivegene histone locus on X chromosomes, and thus about 100 repeats of the GAGA repeats. Therefore, many CLAMP proteins bind to the HLBs, where they recruit many MSL proteins. The MSL protein complexes that form then globally increase male X chromosome gene expression, compensating for the lower X gene dosage in males. Read the original research here (L.E. Reider et al. (2018) Genes & Development 31:1-15). And finally, there is emerging evidence that the HLB action may also be involved in inactivation of an entire female X chromosome later in embryogenesis in females!

This page titled 18.4: Gene Regulation in Eukaryotes is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **12.4: Gene Regulation in Eukaryotes** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

18.5: Transgenic organisms

General principles of transgenesis

Transgenic organisms contain foreign DNA that has been introduced using biotechnology. Foreign DNA (the **transgene**) is defined here as DNA from another species, or else recombinant DNA from the same species that has been manipulated in the laboratory then reintroduced. The terms transgenic organism and genetically modified organism (**GMO**) are generally synonymous. The process of creating transgenic organisms or cells to be come whole organisms with a permanent change to their germline has been called either **transformation** or **transfection**. (Unfortunately, both words have alternate meanings. Transformation also refers to the process of mammalian cell becoming cancerous, while transfection also refers to the process of introducing DNA into cells in culture, either bacterial or eukaryote, for a temporary use, not germ line changes.) Transgenic organisms are important research tools, and are often used when exploring a gene's function. Transgenesis is also related to the medical practice of gene therapy, in which DNA is transferred into a patient's cells to treat disease. Transgenic organisms are widespread in agriculture. Approximately 90% of canola, cotton, corn, soybean, and sugar beets grown in North America are transgenic. No other transgenic livestock or crops (except some squash, papaya, and alfalfa) are currently produced in North America.

To make a transgenic cell, DNA must first be transferred across the cell membrane, (and, if present, across the cell wall), without destroying the cell. In some cases, **naked DNA** (meaning plasmid or linear DNA that is not bound to any type of **carrier**) may be transferred into the cell by adding DNA to the medium and temporarily increasing the porosity of the membrane, for example by **electroporation**. When working with larger cells, naked DNA can also be **microinjected** into a cell using a specialized needle. Other methods use **vectors** to transport DNA across the membrane. Note that the word "vector" as used here refers to any type of carrier, and not just plasmid vectors. Vectors for transformation/transfection include **vesicles** made of lipids or other polymers that surround DNA; various types of particles that carry DNA on their surface; and infectious viruses and bacteria that naturally transfer their own DNA into a host cell, but which have been engineered to transfer any DNA molecule of interest. Usually the foreign DNA is a complete expression unit that includes its own cis-regulators (e.g. promoter) as well as the gene that is to be transcribed.

When the objective of an experiment is to produce a **stable** (i.e. heritable) transgenic eukaryote, the foreign DNA must be incorporated into the host's chromosomes. For this to occur, the foreign DNA must enter the host's nucleus, and recombine with one of the host's chromatids. In some species, the foreign DNA is inserted at a random location in a chromatid, probably wherever strand breakage and non-homologous end joining happen to occur. In other species, the foreign DNA can be targeted to a particular locus, by flanking the foreign DNA with DNA that is homologous to the host's DNA at that locus. The foreign DNA is then incorporated into the host's chromosomes through homologous recombination.

Furthermore, to produce multicellular organisms in which all cells are transgenic and the transgene is stably inherited, the cell that was originally transformed must be either a gamete or must develop into tissues that produce gametes. Transgenic gametes can eventually be mated to produce homozygous, transgenic offspring. The presence of the transgene in the offspring is typically confirmed using PCR or Southern blotting, and the expression of the transgene can be measured using reverse-transcription PCR (RT-PCR), RNA blotting, and Western (protein blotting).

The rate of transcription of a transgene is highly dependent on the state of the chromatin into which it is inserted (i.e. **position effects**), as well as other factors. Therefore, researchers often generate several independently transformed/transfected lines with the same transgene, and then screen for the lines with the highest expression. It is also good practice to clone and sequence the transgenic locus from a newly generated transgenic organism, since errors (truncations, rearrangements, and other mutations) can be introduced during transformation/transfection.

Producing a transgenic plant

The most common method for producing transgenic plants is **Agrobacterium-mediated transformation** (Figure 18.5.1). *Agrobacterium tumifaciens* is a soil bacterium that, as part of its natural pathogenesis, injects its own tumor-inducing (**T_i**) plasmid into cells of a host plant. The natural T_i plasmid encodes growth-promoting genes that cause a gall (i.e. tumor) to form on the plant, which also provides an environment for the pathogen to proliferate. Molecular biologists have engineered the T_i plasmid by removing the tumor-inducing genes and adding restriction sites that make it convenient to insert any DNA of interest. This engineered version is called a **T-DNA** (transfer-DNA) plasmid; the bacterium transfers a linear fragment of this plasmid that

includes the conserved “left-border (LB)”, and right-border (RB)” DNA sequences, and anything in between them (up to about 10 kb). The linear T-DNA fragment is transported into the nucleus, where it recombines with the host-DNA, probably wherever random breakages occur in the host’s chromosomes.

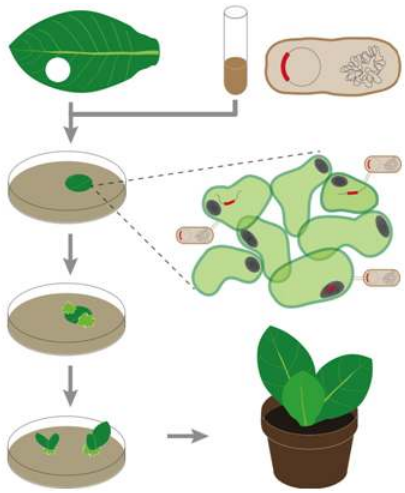


Figure **18.5.1**: Production of a transgenic plant using Agrobacterium-mediated transformation. The bacterium has been transformed with a T-DNA plasmid that contains the transgene and a selectable marker that confers resistance to a herbicide or antibiotic. A bacterial culture and plant tissue (e.g. a leaf punch) are co-cultured on growth medium in a Petri dish. Some of the plant cells will become infected by the bacterium, which will transfer the T-DNA into the plant cytoplasm. In some cases the transgene will become integrated into the chromosomal DNA of a plant cell. In the presence of certain combinations of hormones, the plant cells will dedifferentiate into a mass of cells called callus. The presence of a selective agent (e.g. herbicide or antibiotic) in the growth medium prevents untransformed cells from dividing. Therefore, each callus ideally consists only of transgenic plant cells. The resistant calli are transferred to media with other combinations of hormones that promote organogenesis, i.e. differentiation of callus cells into shoots and then roots. The regenerated transgenic plants are transferred to soil. Their seeds can be harvested and tested to ensure that the transgene is stably inherited. (Original-Deyholos-CC:AN)

In Arabidopsis and a few other species, flowers can simply be dipped in a suspension of Agrobacterium, and ~1% of the resulting seeds will be transformed. In most other plant species, cells are induced by hormones to form a mass of undifferentiated tissues called a callus. The Agrobacterium is applied to a callus and a few cells are transformed, which can then be induced by other hormones to regenerate whole plants (Figure **18.5.2**). Some plant species are resistant (i.e. “**recalcitrant**”) to transformation by Agrobacterium. In these situations, other techniques must be used such as **particle bombardment**, whereby DNA is non-covalently attached to small metallic particles, which are accelerated by compressed air into callus tissue, from which complete transgenic plants can sometimes be regenerated. In all transformation methods, the presence of a selectable marker (e.g. a gene that confers antibiotic resistance or herbicide resistance) is useful for distinguishing transgenic cells from non-transgenic cells at an early stage of the transformation process.

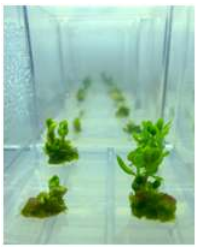


Figure **18.5.2**: Organogenesis of flax shoots from calli. (Original-J. McDill-CC:AN)

Producing a transgenic mouse

In a commonly used method for producing a transgenic mouse, **stem cells** are removed from a mouse embryo, and a transgenic DNA construct is transferred into the stem cells using electroporation, and some of this transgenic DNA enters the nucleus, where it may undergo homologous recombination (Figure **18.5.3**). The transgenic DNA construct contains DNA homologous to either side of a locus that is to be targeted for replacement. If the objective of the experiment is simply to delete (“**knock-out**”) the targeted locus, the host’s DNA can simply be replaced by selectable marker, as shown. It is also possible to replace the host’s DNA at this

locus with a different version of the same gene, or a completely different gene, depending on how the transgenic construct is made. Cells that have been transfected and express the selectable marker (i.e. resistance to the antibiotic neomycin resistance, neoR, in this example) are distinguished from unsuccessfully transfected cells by their ability to survive in the presence of the selective agent (e.g. an antibiotic). Transfected cells are then injected into early stage embryos, and then are transferred to a foster mother. The resulting pups are chimeras, meaning that only some of their cells are transgenic. Some of the chimeras will produce gametes that are transgenic, which when mated with a wild-type gamete, will produce mice that are hemizygous for the transgene. Unlike the chimeras, these hemizygotes carry the transgene in all of their cells. Through further breeding, mice that are homozygous for the transgene can be obtained.

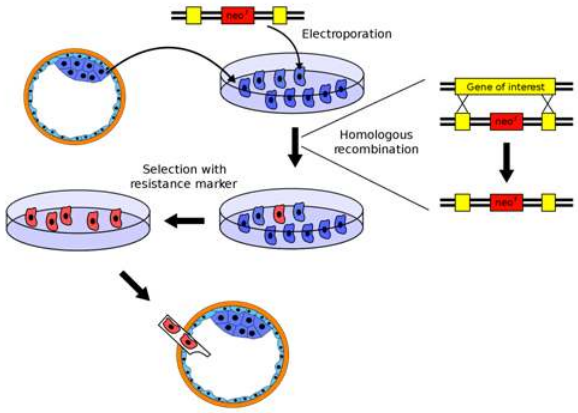


Figure **18.5.3**: Production of a transgenic mouse. Stem cells are removed from an embryo, and are transfected (using electroporation) with a transgenic construct that bears a neomycin resistance gene (neo^r) flanked by two segments of DNA homologous to a gene of interest. In the nucleus of a transgenic cell, some of the foreign DNA will recombine with the targeted gene, disrupting the targeted gene and introducing the selectable marker. Only cells in which neo^r has been incorporated will survive selection. These neomycin resistant cells are then transplanted into another embryo, which will grow into a chimera within a foster mother. (Wikipedia-Kiaergaard-CC:AN)

Human gene therapy

Many different strategies for human gene therapy are under development. In theory, either the **germline** or **somatic** cells may be targeted for transfection, but most research has focused on somatic cell transfection, because of risks and ethical issues associated with germline transformation. Gene therapy approaches may be further classified as either ***ex vivo*** or ***in vivo***, with the former meaning that cells (e.g. stem cells) are transfected in isolation before being introduced to the body, where they replace defective cells. *Ex vivo* gene therapies for several blood disorders (e.g. immunodeficiencies, thalassemias) are undergoing clinical trials. For *in vivo* therapies, the transfection occurs within the patient. The objective may be either stable integration, or **non-integrative** transfection. As described above, stable transfection involves integration into the host genome. In the clinical context, stable integration may not be necessary, and carries with it higher risk of inducing mutations in either the transgene or host genome). In contrast, transient transfection does not involve integration into the host genome and the transgene may therefore be delivered to the cell as either RNA or DNA. Advantages of RNA delivery include that no promoter is needed to drive expression of the transgene. Besides mRNA transgenes, which could provide a functional version of a mutant protein, there is great interest in delivery of **siRNA** (small-inhibitory RNAs), which can be used to silence specific genes in the host cell's genome.

Vectors for *in vivo* gene therapy must be capable of delivering DNA or RNA to a large proportion of the targeted cells, without inducing a significant immune response, or having any toxic effects. Ideally, the vectors should also have high specificity for the targeted cell type. Vectors based on viruses (e.g. **lentiviruses**) are being developed for in both *in vivo* and *ex vivo* gene therapies. Other, non-viral vectors (e.g. vesicles and nanoparticles) are also being developed for gene therapy as well.

This page titled 18.5: Transgenic organisms is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **8.8: Transgenic organisms** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

19: Small RNAs

19.1: Post-transcriptional Control of Gene Expression

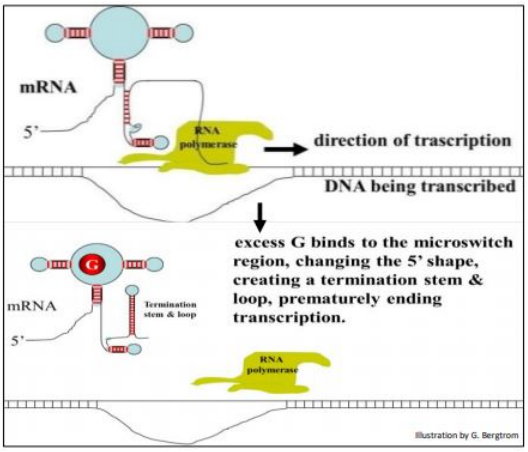
19: Small RNAs is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

19.1: Post-transcriptional Control of Gene Expression

Not too long ago we thought that very little of the eukaryotic genome was ever transcribed. We also thought that the only non-coding RNAs were tRNAs and rRNAs. Now we know that other RNAs play roles in gene regulation and the degradation of spent cellular DNA or unwanted foreign DNA. These are discussed in detail below.

A. Ribosomes

The *riboswitches* is a bacterial transcription mechanism for regulating gene expression. While this mechanism is not specifically post-transcriptional, it is included here because the action occurs after transcription initiation and aborts completion of an mRNA. When the mRNA for an enzyme in the guanine synthesis pathway is transcribed, it folds into stem-&-loop structures. Enzyme synthesis will continue for as long as the cell needs to make guanine. But if guanine accumulates in the cell, excess guanine will bind stem-loop elements near the 5' end of the mRNA, causing the RNA polymerase and the partially completed mRNA dissociate from the DNA, prematurely ending transcription. The basis of guanine riboswitch regulation of expression of a guanine synthesis pathway enzyme is shown below.



232 Riboswitches Interrupt Bacterial Transcription

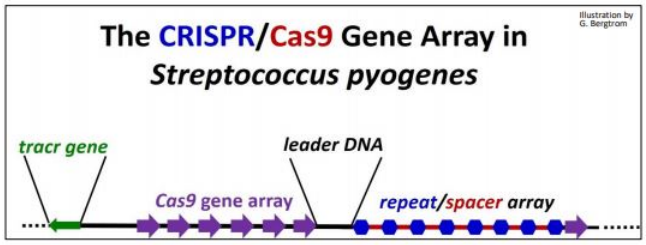
The ability to form folded, stem-loop structures at the 5' ends of bacterial mRNAs seems to have allowed the evolution of translation regulation strategies. Whereas guanine interaction with the stem-loop structure of an emerging 5' mRNA can abort its own transcription, similar small metabolite/mRNA and even protein/mRNA interactions can also regulate (in this case prevent) translation. As we will see shortly, 5' mRNA folded structures also play a role in eukaryotic translation regulation.

233 Small Metabolites Also Regulate Bacterial mRNA translation

B. CRISPR/Cas: RNA-Protein Complex of a Prokaryotic Adaptive Immune System

In higher organisms, the *immune system* is adaptive. It remembers prior exposure to a pathogen, and can thus mount a response to a second exposure to the same pathogen. The discovery of an '*adaptive immune system*' in many prokaryotes (bacteria, archaebacteria) was therefore something of a surprise.

CRISPR (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeat) RNAs are derived from phage transcripts that have interacted with CRISPR-Associated (Cas) proteins. They make up the **CRISPR/Cas** system that seems to have evolved to fight of viral infection by targeting phage DNA for destruction. When viral DNA gets into a cell during a phage infection, it can generate a CRISPR/Cas gene array in the bacterial genome, with **spacer** DNA sequences separating repeats of the CRISPR genes. These remnants of a phage infection are the *memory* of this p*rokaryotic immune system*. When a phage attempts to re-infect a previously exposed cell, spacer RNAs and Cas genes are transcribed. After Cas mRNA translation, the Cas protein and spacer RNAs will engage and target the incoming phage DNA for destruction to prevent infection. Thus, the CRISPR/Cas systems (there is more than one!) *remember* prior phage attacks, and transmit that memory to progeny cells. The CRISPR/Cas9 system in *Streptococcus pyogenes* is one of the simplest of these immune defense systems (illustrated below).



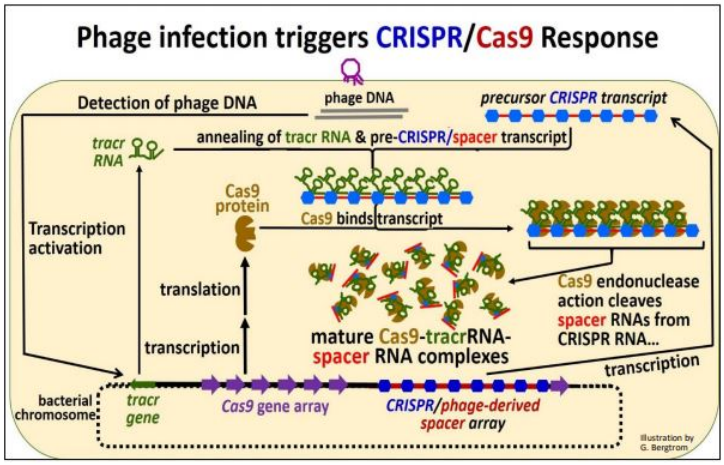
The CRISPR/Cas gene array consists of the following components:

- Cas: Genes native to host cells
- CRISPR: 24-48 bp repeats native to host cells
- Spacer DNA: DNA between CRISPR repeats: typically, phage DNA from prior phage infection or plasmid transformation
- leader DNA: Contains promoter for CRISPR/spacer RNA transcription
- tracr gene: Encodes transcription activator (tracr) RNA (not all systems)

Let's look at CRISPR/Cas in action.

1. The CRISPR/Cas Immune Response

Consider the mechanism of action of this prokaryotic immune system. The action begins when infectious phage DNA gets into the cell, as drawn below.



Let’s summarize what has happened here:

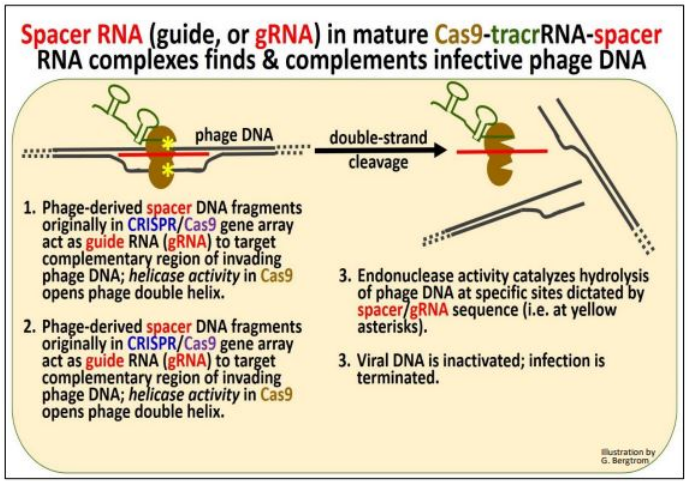
a) Incoming phage DNA was detected after phage infection.

b) Then the tracr and Cas genes are transcribed along with the CRISPR/spacer region. Cas mRNAs are translated to make the Cas protein. Remember, the spacer DNAs in the CRISPR region are the legacy of a prior phage infection.

c) CRISPR/spacer RNA forms hydrogen bonds with a complementary region of the tracr RNA as the two RNAs associate with Cas proteins.

d) Cas protein endonucelases hydrolyze spacer RNA from CRISPR RNA sequences. The spacer RNAs remain associated with the complex while the actual, imperfectly palindromic CRISPR sequences (shown in blue in the illustration above) fall off.

In the next steps, *phage-derived spacer* RNAs, now called **guide RNAs** (or **gRNAs**) ‘guide’ mature *Cas9/tracrRNA/spacer* RNA complexes to new incoming phage DNA resulting from a phage attack. The association of the complex with the incoming phage DNA and subsequent events are illustrated below.



Once again, let's summarize:

a) Spacer (i.e., gRNA) in the complex targets incoming phage DNA.

b) Cas helicase unwinds incoming phage DNA at complementary regions.

c) gRNA H-bonds to incoming phage DNA.

d) Cas endonucleases create a double-stranded break (hydrolytic cleavage) at specific sites in incoming phage DNA. Because precise site DNA strand cleavage is guided by RNA molecules, CRISPR/Cas endonucleases are classified as type V restriction enzymes.

e) The incoming phage DNA is destroyed and a new phage infection is aborted.

Check out here to learn more about how bacteria acquire spacer DNAs, and therefore how this primitive adaptive immune system 'remembers') in the first place

2. Using CRISPR/Cas to Edit/Engineer Genes

Early studies demonstrated the reproducible cleavage of incoming phage DNA at specific nucleotides. Several labs quickly realized that it might be possible to adapt the system to cut DNA at virtually any specific nucleotide in a target DNA! It has turned out that the system works both in vivo and in vitro, allowing virtually unlimited potential for editing genes and RNAs in a test tube… or in any cell. Here is the basic process:

a) Engineer gDNA with a Cas-specific DNA sequence that targets a desired target in genomic DNA.

b) Fuse the gDNA to tracr DNA to make a single guide DNA (sgDNA) so that it can be made as a single guide transcript (sgRNA).

c) Engineer a CRISPR/Cas9 gene array that substitutes this sgDNA for its original spacer DNAs.

d) Place engineered array in a plasmid next to regulated promoters.

e) Transform cells by 'electroporation' (works for almost any cell type!)

f) Activate the promoter to transcribe the CRISPR/Cas9 genes…

The applications are powerful… and controversial!

3. The Power and the Controversy

The application of gene editing with CRISPR/Cas systems has already facilitated studies of gene function in vitro, in cells and in whole organisms. Click here for a description of CRISPR/Cas applications already on the market! The efficiency of specific gene editing using CRISPR/Cas systems holds great promise for understanding basic gene structure and function, for determining the genetic basis of disease, and for accelerating the search for gene therapies. Here are just a few examples of how CRISPR/Cas approaches are being applied.

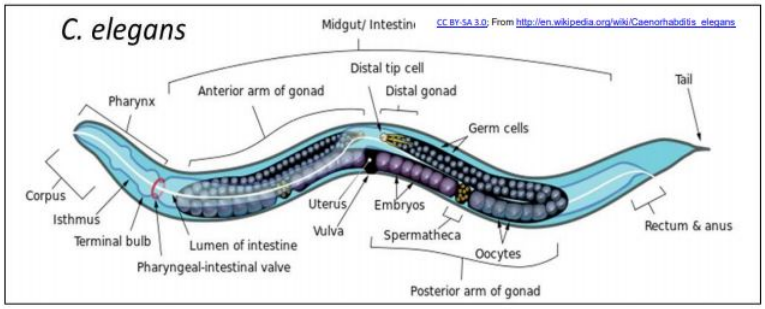
- One can engineer an sgRNA with desired mutations targeting specific sites in chromosomal DNA. Then clone sgRNA into the CRISPR/Cas9 array on a plasmid. After transformation of appropriate cells, the engineered CRISPR/Cas9 forms a complex with target DNA sequences. Following nicking of both strands of the target DNA, DNA repair can insert the mutated guide sequences into the target DNA. The result is loss or acquisition of DNA sequences at *specific, exact sites*, or **Precision Gene Editing**. It is the ability to do this in living cells that has excited the basic and clinical research communities.
- Before transforming cells, engineer the **CRISPR/Cas9** gene array on the plasmid to eliminate both **endonuclease** activities from the Cas protein. Upon transcription of the array in transformed cells, the *CRISPR/Cas9-sgRNA* still finds an *sgRNA*-targeted gene. However, lacking CAS protein endonuclease activities, the complex that forms just sits there **blocking transcription**. This technique is sometimes referred to as **CRISPRi** (*CRISPER interference*), by analogy to *RNAi*. Applied to organisms (and not just in vitro or to cells), it mimics the much more difficult **knockout** mutation experiments that have been used in studies of behavior of cells or organisms rendered unable to express a specific protein.
- There are now several working CRISPR/Cas systems capable of *Precision Gene Editing*. They are exciting for their speed, precision, their prospects for rapid, targeted gene therapies to fight disease, and their possibilities to alter entire populations (called **Gene Drive**). By inserting modified genes into the germline cells of target organisms, gene drive can render harmless entire malarial mosquito populations, to eliminate pesticide resistance in e.g. insects, eliminate herbicide resistance in undesirable plants, or genetically eliminate invasive species. For more information, click Gene drive; for an easy read about this process and the controversies surrounding applications of CRISPR technologies to mosquitoes in particular, check out J. Adler, (2016) *A World Without Mosquitoes*. Smithsonian, 47(3) 36-42, 84.
- It is even possible to delete an entire chromosome from cells. This bit of global genetic engineering relies on identifying multiple unique sequences on a single chromosome and then targeting these sites for CRISPR/Cas. When the system is activated, the chromosome is cut at those sites, fragmenting it beyond the capacity of DNA repair mechanisms to fix the situation. Click here to learn more.

If for no other reason than its efficiency and simplicity, precision gene editing with CRISPR/Cas techniques has raised ethical issues. Clearly, the potential exists for abuse, or even for use with no beneficial purpose at all. It is significant that, as in all discussions of biological ethics, scientists are very much engaged in the conversation. Despite the controversy, we will no doubt continue to edit genes with CRISPR/Cas, and we can look for a near future Nobel Prize for its discovery and application! If you still have qualms, maybe RNA editing will be the answer. Check out the link at Why edit RNA? for an overview of the possibilities!

Finally, “mice and men” (and women and babies too) have antibodies to Cas9 proteins, suggesting prior exposure to microbial CRISPR/Cas9 antigens. This observation may limit clinical applications of the technology! See Uncertain Future of CRISPR-Cas9 Technology.

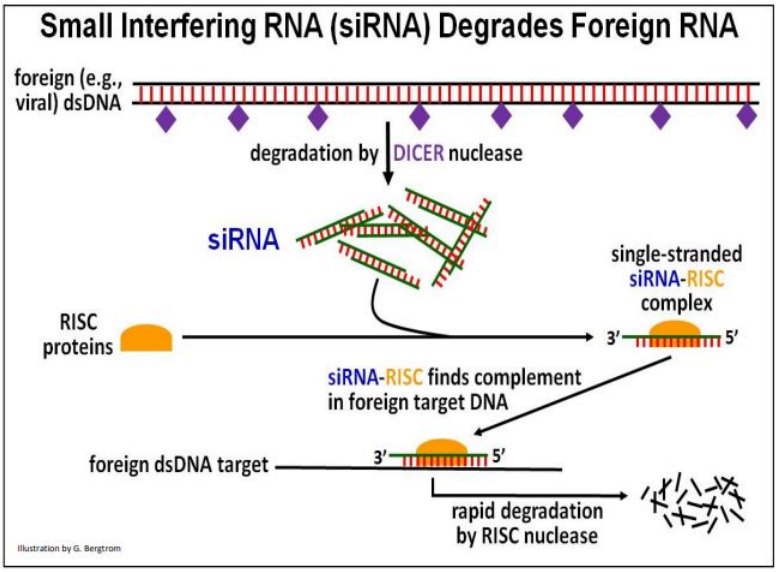
C. The Small RNAs: miRNA and siRNA in Eukaryotes

Micro RNAs (miRNAs) and small interfering RNAs (siRNAs) are found in *C. elegans*, a small nematode (roundworm) that quickly became a model for studies of cell and molecular biology and development. The particular attractions *C. elegans* are that (a) its genome has ~21,700 genes, comparable to the ~25,000 genes in a human genome!; (b) it uses the products of these genes to produce an adult worm consisting of just 1031 cells organized into all of the major organs found in higher organisms; (c) It is possible to trace the embryonic origins of every single cell in its body! *C. elegans* is illustrated below.



1. Small Interfering RNA (siRNA)

siRNA was first found in plants as well as in *C. elegans*. However, siRNAs (and miRNAs) are common in many higher organisms. siRNAs were so-named because they *interfere* with the function of other RNAs foreign to the cell or organism. Their action was dubbed *RNA interference* (**RNAi**). For their discovery of siRNAs, A. Z. Fire and C. C. Mello shared the 2006 Nobel Prize in Physiology or Medicine. The action of siRNA targeting foreign DNA is illustrated below.



When cells recognize foreign double-stranded RNAs (e.g., some viral RNA genomes) as alien, the **DICER** a *nuclease* called hydrolyzes them. The resulting short double-stranded hydrolysis products (the **siRNAs**) combine with **R**NAi **I**nduced **S**ilencing **C**omplex, or **RISC** proteins. The *antisense* siRNA strand in the resulting **siRNA-RISC** complex binds to complementary regions of foreign RNAs, targeting them for degradation. Cellular use of RISC to control gene expression in this way may have derived from the use of RISC proteins by miRNAs as part of a cellular defense mechanism, to be discussed next.

Custom-designed siRNAs have been used to disable expression of specific genes in order to study their function *in vivo* and *in vitro*. Both siRNAs and miRNAs are being investigated as possible therapeutic tools to interfere with RNAs whose expression leads to cancer or other diseases.

234 siRNA Post Transcriptional Regulation

235 Did siRNA Coopt RISC strategy to Trash Corrupt or Worn out RNA?

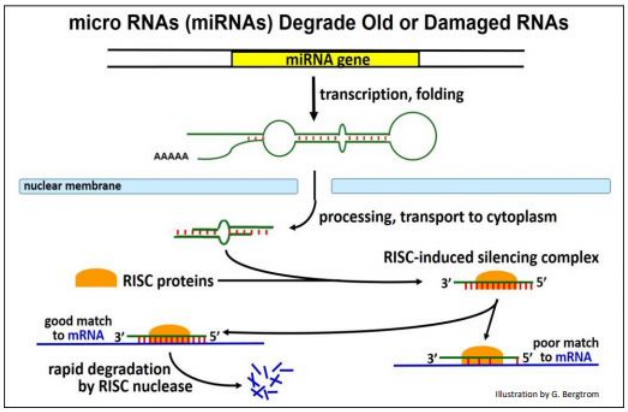
For an example check out a Youtube video of unexpected results of an RNAi experiment at this link. In the experiment described, RNAi was used to block embryonic expression of the **orthodenticle** (**odt**) gene that is normally required for the growth of horns in a dung beetle. The effect of this *knock-out mutation* was, as expected, to prevent horn growth. What was unexpected however, was the development of an eye in the middle of the beetle's head ('third eye' in the micrograph).

The 3rd eye not only looks like an eye, but is a functional one. This was demonstrated by preventing normal eye development in *odt*-knockout mutants. The 3rd eye appeared…, and was responsive to light! Keep in mind that this was a beetle with a 3rd eye, not *Drosophila*! To quote Justin Kumar from Indiana University, who though not involved in the research, stated that "…lessons learned from *Drosophila* may not be as generally applicable as I or other Drosophilists, would like to believe … The ability to use RNAi in non-traditional model systems is a huge advance that will probably lead to a more balanced view of development."

2. Micro RNAs (miRNA)

miRNAs target unwanted *endogenous* cellular RNAs for degradation. They are transcribed from genes now known to be widely distributed in eukaryotes. The pathway from pre-miRNA transcription through processing and target mRNA degradation is illustrated on the next page.

As they are transcribed, pre-miRNAs fold into a stem-loop structure that is lost during cytoplasmic processing. Like SiRNAs, mature miRNAs combine with RISC proteins. The **RISC** protein-miRNA complex targets old or no-longer needed mRNAs or mRNAs damaged during transcription.



An estimated 250 miRNAs in humans may be sufficient to H-bond to diverse target RNAs; only targets with strong complementarity to a RISC protein-miRNA complex will be degraded.

236 miRNA Post-Transcriptional Regulation

D. Long Non-Coding RNAs

Long non-coding RNAs (**lncRNAs**) are a yet another class of eukaryotic RNAs. They include transcripts of antisense, intronic, intergenic, pseudogene and retroposon DNA. Retroposons are one kind of transposon, or mobile DNA element; pseudogenes are recognizable genes with mutations that make them non-functional. While some lncRNAs might turn out to be incidental transcripts that the cell simply destroys, others have a role in regulating gene expression.

A recently discovered lncRNA is XistAR that, along with the Xist gene product, is required to form *Barr bodies*. Barr bodies form in human females when one of the X chromosomes in somatic cells is inactivated. For a review of lncRNAs, see Lee, J.T. (2012. *Epigenetic Regulation by Long Noncoding RNAs*; Science 338, 1435-1439).

An even more recent article (at lncRNAs and smORFs) summarizes the discovery that some long non-coding RNAs contain short open reading frames (**smORFs**) that are actually translated into short peptides of 30+ amino acids! Who knows? The human genome may indeed contain more than 21,000-25,000 protein-coding genes!

E. Circular RNAs (circRNA)

Though discovered more than 20 years ago, circular RNAs (circRNAs) are made in different eukaryotic cell types. Click Circular RNAs (circRNA) to learn more about this peculiar result of *alternative* splicing. At first circRNAs were hard to isolate. When they were isolated, circRNAs contained “scrambled” exonic sequences and were therefore thought to be nonfunctional errors of mRNA splicing.

In fact, circRNAs are fairly stable. Their levels can rise and fall in patterns suggesting that they are functional molecules. Levels of one circRNA, called **circRims1**, rise specifically during neural development. In mice, other circRNAs accumulate during synapse formation, likely influencing how these neurons will ultimately develop and function. Thus, circRNAs do not seem to be ‘molecular mistakes’. In fact, *errors in their own synthesis* may be correlated with disease! Speculation on the functions of circRNAs also includes roles in gene regulation, particularly the genes or mRNAs from which they themselves are derived.

F. "Junk DNA" in Perspective

Not long ago, we thought that less than 5% of a eukaryotic genome was transcribed (i.e., into mRNA, rRNA and tRNA), and that much of the non-transcribed genome served a structural function… or no function at all. The latter, labeled junk DNA, included non-descript intergenic sequences, pseudogenes, ‘dead’ transposons, long stretches of intronic DNA, etc. Thus, junk DNA was DNA we could do without. Junk DNAs were thought to be accidental riders in our genomes, hitchhikers picked up on the evolutionary road.

While miRNA genes are a small proportion of a eukaryotic genome, their discovery, and that of more abundant lnc RNAs suggest a far greater amount of functional DNA in the genome. Might there be in fact, no such thing as “junk DNA”? The debate about how much of our genomic DNA is a relic of past evolutionary experiments and without genetic purpose continues. Read all about it at Junk DNA - not so useless after all and Only 8.2% of human DNA is functional.

Perhaps we need to re-think what it means for DNA to be “junk” or to be without “genetic purpose”. Maintenance of more than 90% of our own DNA with no known genetic purpose surely comes at an energy cost. At the same time, all of that DNA is grist for future selection, a source of the diversity required for long-term survival. The same natural selection that picks up ‘hitchhiker’ DNA sequences, as we have seen, can at some point, put them to work!

G. The RNA Methylome

Call this an *RNA epi-transcriptome* if you like! Recall that methyl groups direct cleavage of ribosomal RNAs from eukaryotic 45S pre-RNA transcripts. tRNAs among other transcripts, are also post-transcriptionally modified. Known since the 1970s, such modifications were thought to be non-functional. But are they?

This page titled 19.1: Post-transcriptional Control of Gene Expression is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **13.2: Post-transcriptional Control of Gene Expression** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

20: The Genetic Code

20.1: The Function of Genes

20.2: An Overview of the Genetic Code

20.3: Gene and Protein Colinearity and Triplet Codons

20.4: Translation

20.5: Gene Expression- Translation

20.6: Gene Expression- Applied Example (Part 1)

20: The Genetic Code is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

20.1: The Function of Genes

Beadle and Tatum: one gene, one enzyme hypothesis

Life depends on (bio)chemistry to supply energy and to produce the molecules to construct and regulate cells. In 1908, A. Garrod described "in born errors of metabolism" in humans using the congenital disorder, alkaptonuria (black urine disease), as an example of how "genetic defects" led to the lack of an enzyme in a biochemical pathway and caused a disease (phenotype). Over 40 years later, in 1941, **Beadle and Tatum** built on this connection between genes and metabolic pathways. Their research led to the "**one gene, one enzyme (or protein)**" hypothesis, which states that each of the enzymes that act in a biochemical pathway is encoded by a different gene. Although we now know of many exceptions to the "one gene, one enzyme (or protein)" principle, it is generally true that each different gene produces a protein that has a distinct catalytic, regulatory, or structural function.

Beadle and Tatum used the fungus ***Neurospora crassa*** (a mold) for their studies because it had practical advantages as a laboratory organism. They knew that Neurospora was **prototrophic**, meaning that it could synthesize its own amino acids when grown on **minimal medium**, which lacked most nutrients except for a few minerals, simple sugars, and one vitamin (biotin). They also knew that by exposing Neurospora spores to X-rays, they could randomly damage its DNA to create mutations in genes. Each different spore exposed to X-rays potentially contained a mutation in a different gene. After genetically screening many, many spores for growth, most appeared to still be prototrophic and still able to grow on minimal medium. However, some spores had mutations that changed them into **auxotrophic** strains that could no longer grow on minimal medium, but did grow on **complete medium** supplemented with nutrients (Figure **20.1.12**). In fact, some auxotrophic mutations could grow on minimal medium with only one, single nutrient supplied, such as arginine.

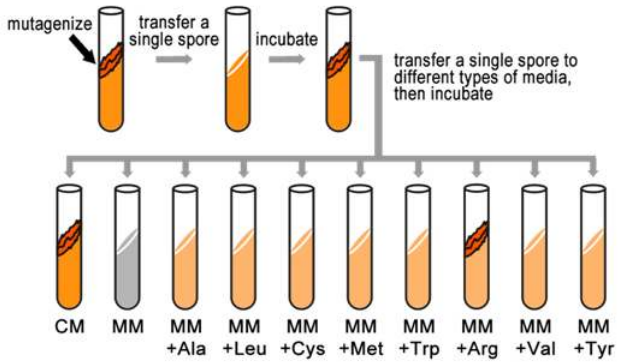


Figure **20.1.12**: A single mutagenized spore is used to establish a colony of genetically identical fungi, from which spores are tested for their ability to grow on different types of media. Because spores of this particular colony are able to grown only on complete medium (CM), or on miminal medium supplemented with arginine (MM+Arg), they are considered Arg auxotrophs and we infer that they have a mutation in a gene in the Arg biosynthetic pathway. This type of screen is repeated many times to identify other mutants in the Arg pathway and in other pathways. (Original-Deyholos-CC:AN)

B&T's 1 gene: 1 enzyme hypothesis led to Biochemical Pathway dissection using genetic screens and mutations

Beadle and Tatum's experiments are important not only for its conceptual advances in understanding genes, but also because they demonstrate the utility of **screening for genetic mutants** to investigate a biological process – **genetic analysis**. Beadle and Tatum's results were useful to investigate biological processes, specifically the metabolic pathways that produce amino acids. For example, Srb and Horowitz in 1944 tested the ability of the amino acids to **rescue** auxotrophic strains. They added one of each of the amino acids to minimal medium and recorded which of these restored growth to independent mutants. For example, if the progeny of a mutagenized spore could grow on minimal medium only when it was supplemented with arginine (Arg), then the auxotroph must bear a mutation in the Arg biosynthetic pathway and was called an "arginineless" strain (arg-).

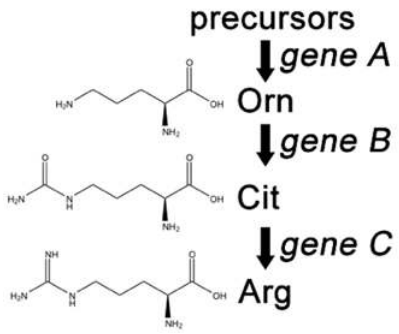


Figure **20.1.13**: A simplified version of the Arg biosynthetic pathway, showing citrulline (Cit) and ornithine (Orn) as intermediates in Arg metabolism. These chemical reactions depend on enzymes represented here as the products of three different genes. (Original-Deyholos-CC:AN)

Synthesis of even a relatively simple molecule such as arginine requires many steps, each with a different enzyme. Each enzyme works sequentially on a different intermediate in the pathway (Figure **20.1.13**). For arginine (Arg), two of the intermediates are ornithine (Orn) and citrulline (Cit). Thus, mutation of any one of the enzymes in this pathway could turn Neurospora into an Arg auxotroph (arg-). Srb and Horowitz extended their analysis of Arg auxotrophs by testing the intermediates of amino acid biosynthesis for the ability to restore growth of the mutants (Figure **20.1.14**).

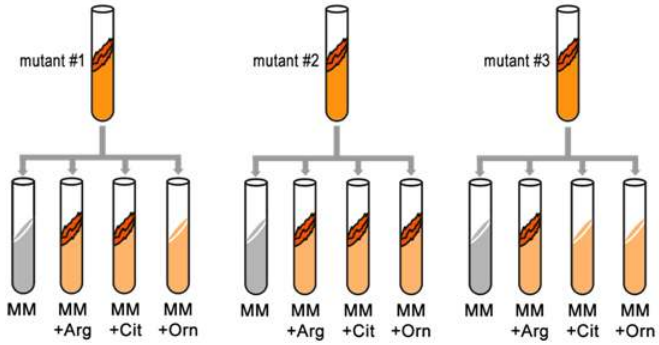


Figure **20.1.14**: Testing different Arg auxotrophs for their ability to grow on media supplemented with intermediates in the Arg biosynthetic pathway. (Original-Deyholos-CC:AN)

They found that some of the Arg auxotrophs could be rescued only by Arg, while others could be rescued by either Arg or Cit, and still other mutants could be rescued by Arg, Cit, or Orn (Table **20.1.1**). Based on these results, they deduced the location of each mutation in the Arg biochemical pathway, (i.e. which gene was responsible for the metabolism of which intermediate).

Table **20.1.1**: Ability of auxotrophic mutants of each of the three enzymes of the Arg biosynthetic pathways to grow on minimal medium (MM) supplemented with Arg or either of its precursors, Orn and Cit. Gene names refer to the labels used in Figure **20.1.14**

	MM + Orn	MM + Cit	MM + Arg
gene A mutants	**Yes**	**Yes**	**Yes**
gene B mutants	**No**	**Yes**	**Yes**
gene C mutants	**No**	**No**	**Yes**

This page titled 20.1: The Function of Genes is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **1.5: The Function of Genes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

20.2: An Overview of the Genetic Code

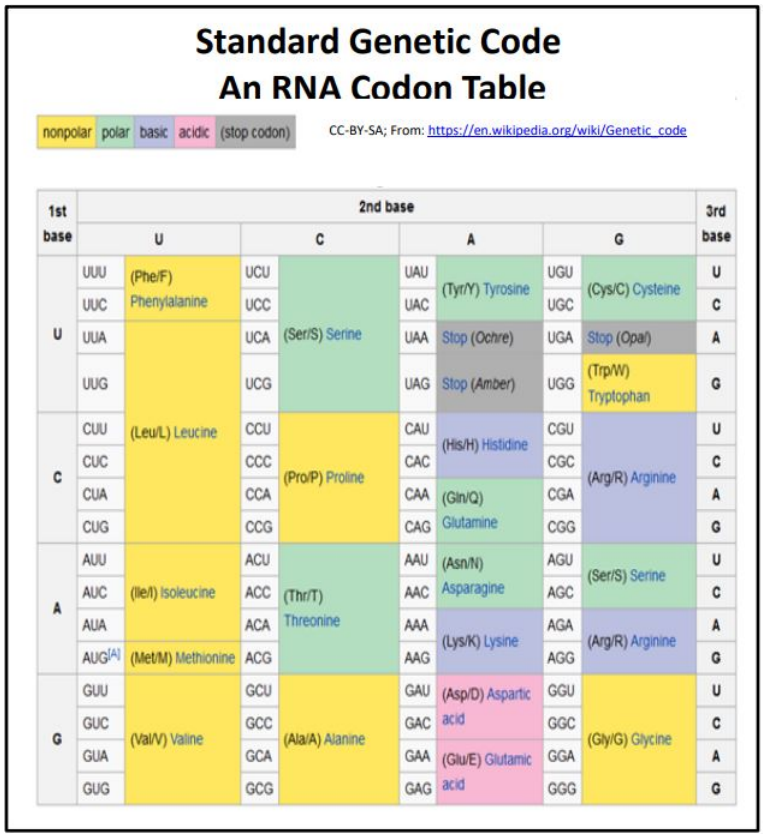
A. The (Nearly) Universal, Degenerate Genetic Code

The **genetic code** is the information for linking amino acids into polypeptides in an order based on the base sequence of 3-base code words (**codons**) in a gene and its messenger RNA (mRNA). With a few exceptions (some prokaryotes, mitochondria, chloroplasts), **the genetic code is universal** – it's the same in all organisms from viruses and bacteria to humans. The table of the **Standard Universal Genetic Code** on the next page shows the RNA version of triplet codons and their corresponding amino acids. There is a single codon for two amino acids (methionine and tryptophan), but two or more codons for each of the other 18 amino acids. For the latter reason, we say that the genetic code is **degenerate**. The three **stop codons** in the *Standard Genetic Code* 'tell' ribosomes the location of the last amino acid to add to a polypeptide. The last amino acid itself can be any amino acid consistent with the function of the polypeptide being synthesized. However, evolution has selected AUG as the **start codon** for all polypeptides, regardless of function, as well as for the placement of methionine within a polypeptide. Thus, all polypeptides begin life with a methionine at their amino-terminal end. As we will see in more detail, the mRNA translation machine is the **ribosome** and the *decoding device* is tRNA. Each amino acid attaches to a tRNA whose short sequence contains a 3-base *anticodon* that is complementary to an mRNA codon. Enzymatic reactions catalyze the **dehydration synthesis** (**condensation**) reactions that link amino acids in peptide bonds in the order specified by codons in the mRNA.

201 The Genetic Code Dictionary

B. Comments on the Nature and Evolution of Genetic Information

The near-universality of the genetic code from bacteria to humans implies that the code originated early in evolution. It is probable that portions of the code were in place even before life began. Once in place however, the genetic code was highly constrained against evolutionary change. The degeneracy of the genetic code enabled and contributed to this constraint by permitting base many base changes that do not affect the amino acid encoded in a codon.



The near universality of the genetic code and its resistance to change are features of our genomes that allow us to compare gene and other DNA sequences to establish evolutionary relationships between organisms (species), groups of organisms (genus, family, order, etc.) and even individuals within a species.

In addition to constraints imposed by a universal genetic code, some organisms show *codon bias*, a recent constraint on which universal codons an organism uses. Codon bias is seen in organisms preferably use A-T rich codons, or in organisms that favor codons richer in G and C. Interestingly, codon bias in genes often accompanies corresponding *genomic nucleotide bias*. An organism with an AT codon bias may also have an AT-rich genome (likewise GC-rich codons in GC-rich genomes). You can recognize genome nucleotide bias in Chargaff's base ratios!

Finally, we often think of genetic information as genes for proteins. Obvious examples of *non-coding genetic information* include the genes for rRNAs and tRNAs, common to all organisms. The amount of these kinds of informational DNA (i.e., genes for polypeptides, tRNAs and rRNAs) as a proportion of total DNA can range across species, although it is higher in eukaryotes prokaryotes. For example, ~88% of the E. coli circular chromosome encodes polypeptides, while that figure is less ~1.5% for humans. Some less obvious informative DNA sequences in higher organisms are transcribed (e.g., introns). Other informative DNA in the genome is never transcribed. The latter include regulatory DNA sequences, DNA sequences that support chromosome structure and other DNAs that contribute to development and phenotype. As for that amount of truly non-informative (useless) DNA in a eukaryotic genome, that amount is steadily shrinking as we sequence entire genomes, identify novel DNA sequences and discover novel RNAs (topics covered elsewhere in this text).

This page titled 20.2: An Overview of the Genetic Code is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **11.2: An Overview of the Genetic Code** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

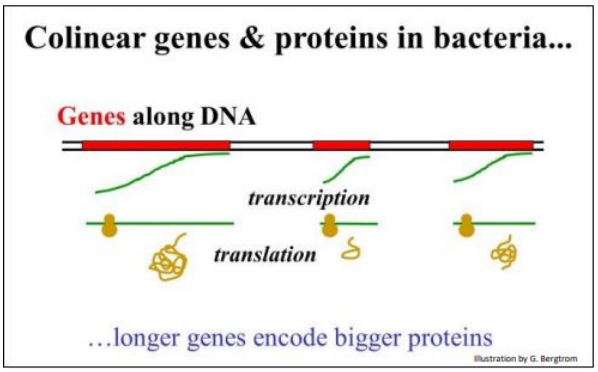
20.3: Gene and Protein Colinearity and Triplet Codons

Serious efforts to understand how proteins are encoded began after Watson and Crick used the experimental evidence of Maurice Wilkins and Rosalind Franklin (among others) to determine the structure of DNA. Most hypotheses about the genetic code assumed that DNA (i.e., genes) and polypeptides were **colinear**.

A. Colinearity

For genes and proteins, colinearity just means that the length of a DNA sequence in a gene is proportional to the length of the polypeptide encoded by the gene. The gene mapping experiments in *E. coli* already discussed certainly supported this hypothesis.

The concept of colinearity is illustrated below.

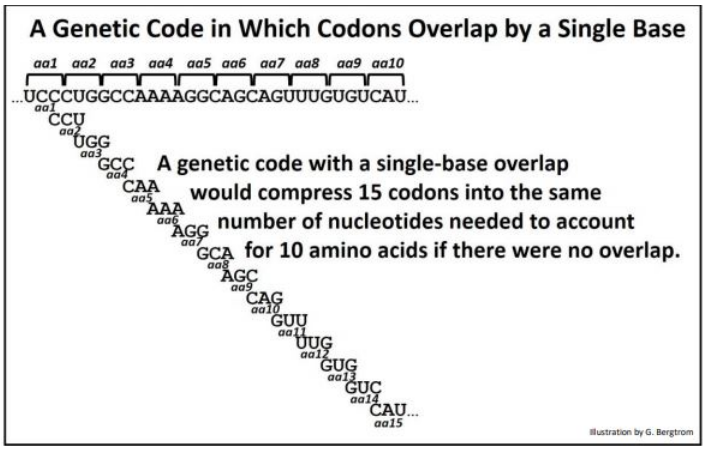


If the genetic code is collinear with the polypeptides it encodes, then a one-base **codon** obviously does not work because such a code would only account for four amino acids. A two-base genetic code also doesn't work because it could only account for 16 (4 2) of the twenty amino acids found in proteins. However, threenucleotide codons could code for a maximum of 43 or 64 amino acids, more than enough to encode the 20 amino acids. And of course, a 4-base code also works; it satisfies the expectation that genes and proteins are collinear, with the' advantage' that there would be 256 possible codons to choose from (i.e., 44 possibilities).

B. How is the Genetic Code 'Read' to Account for All of an Organisms' Gene?

George Gamow (a Russian Physicist working at George Washington University) was the first to propose **triplet codons** to encode the twenty amino acids, the simplest hypothesis to account for the colinearity of gene and protein, and for encoding 20 amino acids. One concern that was raised was whether there is enough DNA in an organism's genome to fit the all codons it needs to make all of its proteins? Assuming genomes did not have a lot of extra DNA laying around, how might genetic information be compressed into short DNA sequences in a way that is consistent with the colinearity of gene and polypeptide. One idea assumed 44 meaningless and 20 *meaningful* 3-base codons (one for each amino acid) and 44 *meaningless* codons, and that the *meaningful* codons in a gene (i.e., an mRNA) would be read and translated in an overlapping manner.

A code where codons overlap by one base is shown below.



You can figure out how compressed a gene could get with codons that overlapped by two bases. However, as attractive as an overlapping codon hypothesis was in achieving genomic economies, it sank of its own weight almost as soon as it was floated! If you look carefully at the example above, you can see that each succeeding amino acid would have to start with a specific base. A look back at the table of 64 triplet codons quickly shows that only one of 16 amino acids, those that begin with a C can follow the first one in the illustration. Based on amino acid sequences accumulating in the literature, virtually any amino acid could follow another in a polypeptide. Therefore, overlapping genetic codes are untenable. The genetic code must be non-overlapping!

Sidney Brenner and Frances Crick performed elegant experiments that directly demonstrated the non-overlapping genetic code. They showed that bacteria with a single base deletion in the coding region of a gene failed to make the expected protein. Likewise, deleting two bases from the gene. On the other hand, bacteria containing a mutant version of the gene in which three bases were deleted were able to make the protein. The protein it made was slightly less active than bacteria with genes with no deletions.

The next issue was whether there were only 20 **meaningful** codons and 44 **meaningless** ones. If only 20 triplets actually encoded amino acids, how would the translation machinery recognize the correct 20 codons to translate? What would prevent the translational machinery from 'reading the wrong' triplets, i.e., reading an mRNA *out of phase*? If for example, if the translation machinery began reading an MRNA from the second or third bases of a codon, it would likely encounter a meaningless 3-base sequence in short order.

One speculation was that the code was *punctuated*. That is, perhaps there were the chemical equivalent of commas between the meaningful triplets. The commas would be of course, additional nucleotides. In such a punctuated code, the translation machinery would recognize the 'commas' and would not translate *any* meaningless 3- base triplet, avoiding out-of-phase translation attempts. Of course, a code with nucleotide 'commas' would increase the amount of DNA needed to specify a polypeptide by a third!

Then, Crick proposed the *Commaless Genetic Code*. He divided the 64 triplets into 20 *meaningful* codons that encoded the amino acids, and 44 *meaningless* ones that did not. The result was such that when the 20 *meaningful* codons are placed in any order, any of the triplets read in overlap would be among the 44 *meaningless* codons. In fact, he could arrange several different sets of 20 and 44 triplets with this property! Crick had cleverly demonstrated how to read the triplets in correct sequence without nucleotide 'commas'.

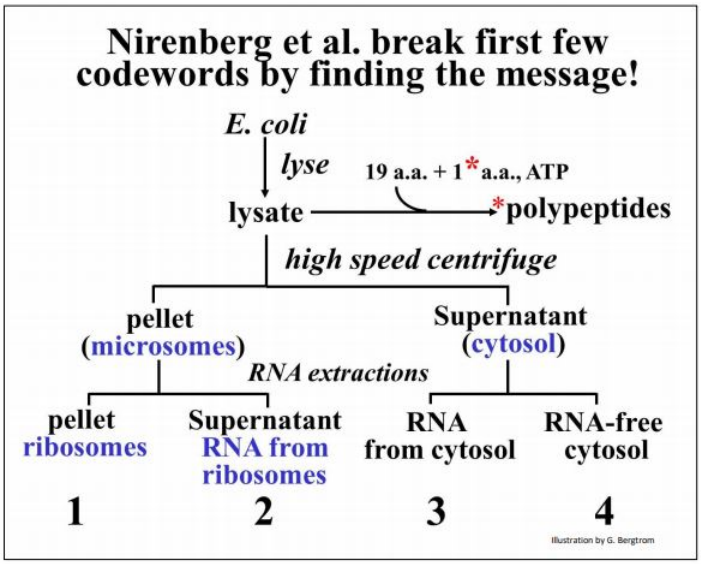
202 Speculations About a Triplet Code

As we know now, the genetic code is indeed 'commaless'… but not in the sense that Crick had envisioned. What's more, Thanks to the experiments described next, we know that ribosomes read the correct codons in the right order because they know exactly where to start!

C. Breaking the Genetic Code

When the genetic code was actually broken, it was found that 61 of the codons specify amino acids and therefore, that the code is **degenerate**. Breaking the code began when Marshall Nirenberg and Heinrich J. Matthaei decoded the first triplet. They fractionated *E. coli* and identified which fractions had to be added back together in order to get polypeptide synthesis in a test tube (*in vitro* translation).

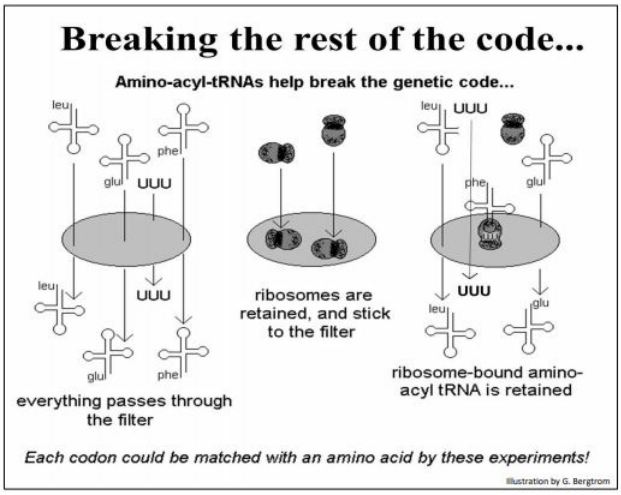
The cell fractionation is summarized below.



Check out the original work in the classic paper by Nirenberg MW and Matthaei JH [(1961) *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribo-nucleotides.* Proc. Natl. Acad. Sci. USA 47:1588-1602]. The various cell fractions isolated by this protocol were added back together along with amino acids (one of which was radioactive) and ATP as an energy source. After a short incubation, Nirenberg and his coworkers looked for the presence of high molecular weight radioactive proteins as evidence of cell-free protein synthesis.

They found that all four final sub-fractions (1-4 above) must be added together to make radioactive proteins in the test tube. One of the essential cell fractions consisted of RNA that had been gently extracted from ribosome (fraction 2 in the illustration). Reasoning that this RNA might be mRNA, they substituted a synthetic poly(U) preparation for this fraction in their cell-free protein synthesizing mix, expecting poly(U) to encode a simple repeating amino acid.

They set up 20 reaction tubes, with a different amino acid in each…, and made only poly-phenylalanine. The experiment is illustrated below.



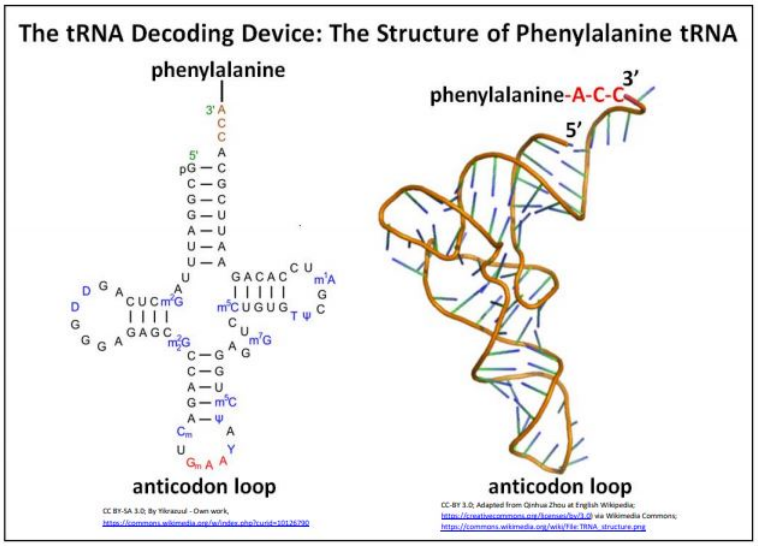
So, the triplet codon UUU means *phenylalanine*. Other polynucleotides were synthesized by G. Khorana, and in quick succession, poly(A) and poly(C) were shown to make poly-lysine and poly-proline in this experimental protocol. Thus AAA and CCC must encode lysine and proline respectively. With a bit more difficulty and ingenuity, poly di- and tri-nucleotides were also used in the cell free system to decipher several additional codons.

203 Deciphering the First Codon

M. W. Nirenberg, H. G. Khorana and R. W. Holley shared the 1968 Nobel Prize in Physiology or Medicine for their contributions to our understanding of protein synthesis. Deciphering the rest of the genetic code was based on Crick's realization that chemically, amino acids have no attraction for either DNA or RNA (or triplets thereof). Instead, he predicted the existence of an **adaptor molecule** that would contain nucleic acid and amino acid information *on the same molecule*. Today we recognize this molecule as tRNA, the genetic *decoding device*.

Nirenberg and Philip Leder designed the experiment that pretty much *broke* the rest of the genetic code. They did this by adding individual amino acids to separate test tubes containing tRNAs, in effect causing the synthesis of specific aminoacyl-tRNAs.

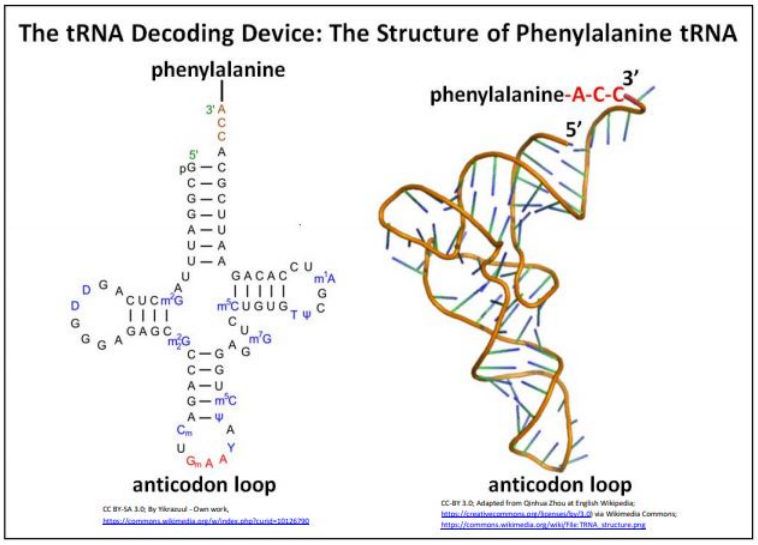
They then mixed their amino acid-bound tRNAs with isolated ribosomes and synthetic triplets. Since they had already shown that synthetic three-nucleotide fragments would bind to ribosomes, they hypothesized that triplet-bound ribosomes would in turn, bind appropriate amino acid-bound tRNAs. The experiment is shown below.



Various combinations of tRNA, ribosomes and aminoacyl-tRNAs were placed over a filter. Nirenberg and Leder knew that aminoacyl-tRNAs alone passed through the filter and that ribosomes did not. They predicted then, that triplets would associate with the ribosomes, and further, that this complex would bind the tRNA with the amino acid encoded by the bound triplet. This 3-part complex would also be retained by the filter, allowing the identification of the amino acid retained on the filter, and therefore the triplet code-word that had enabled binding the amino acid to the ribosome.

204 Deciphering all 64 Triplet Codons

After the code was largely deciphered, Robert Holley actually sequenced a yeast tRNA, and from regions of internal complementarity, predicted the folded structure of the tRNA. This first successful sequencing of a nucleic acid was possible because the tRNA was short, and contained several modified bases that facilitated the sequencing chemistry. Holley found the amino acid alanine at one end of the tRNA and he found one of the anticodons for an alanine codon roughly in the middle of the tRNA sequence. Holley predicted that this (and other) tRNAs would fold and assume a *stem-loop*, or *cloverleaf* structure with a central **anticodon loop**. The illustration below shows this structure for a phenylalanine tRNA along with subsequent computer-generated structures (below right) showing a now familiar "L"-shaped molecule with an **amino acid attachment site** at the 3'-end at the top of the molecule, and the *anticodon loop* at the other, bottom 'end'



205 tRNA Structure and Base Modifications

After a brief overview of translation, we'll break translation down into its 3 steps and see how aminoacyl-tRNAs function in the initiation and elongation steps of translation, as well as the special role of an *initiator* tRNA.

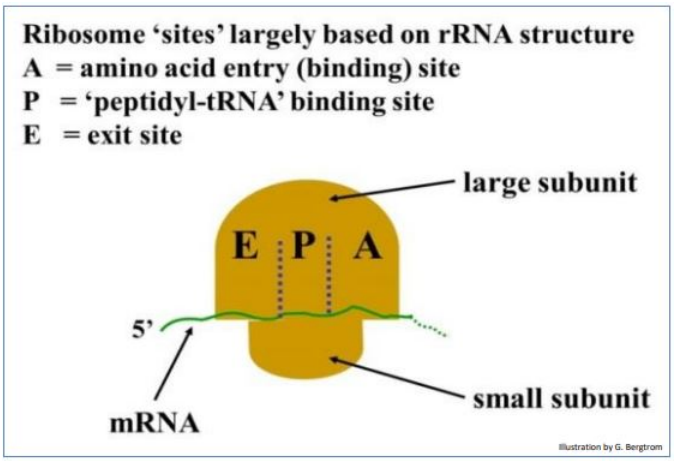
This page titled 20.3: Gene and Protein Colinearity and Triplet Codons is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **11.3: Gene and Protein Colinearity and Triplet Codons** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

20.4: Translation

A. Overview of Translation (Synthesizing Proteins)

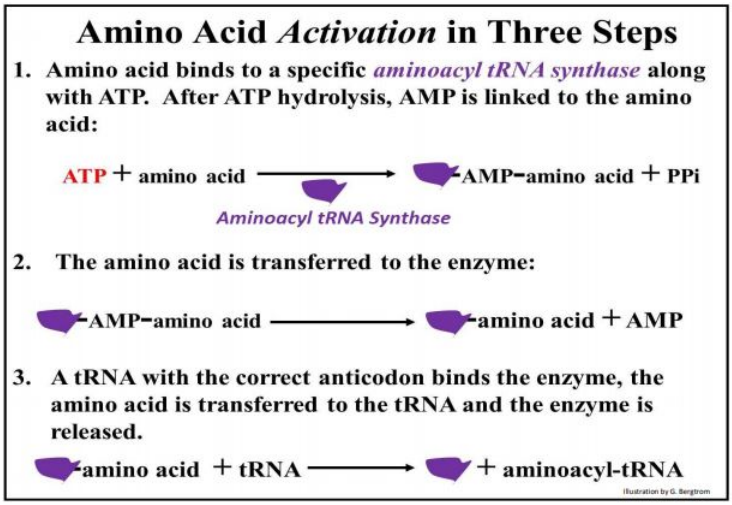
Like any polymerization in a cell, translation occurs in three steps: **initiation** brings a ribosome, mRNA and an *initiator* tRNA together to form an initiation complex. **Elongation** is the successive addition of amino acids to a growing polypeptide. **Termination** is signaled by sequences (one of the stop codons) in the mRNA and protein **termination factors** that interrupt elongation and release a finished polypeptide. The events of translation occur at specific **A**, **P** and **E** sites on the ribosome (see drawing below).



B. Translation - First Steps

1. Making Aminoacyl-tRNAs

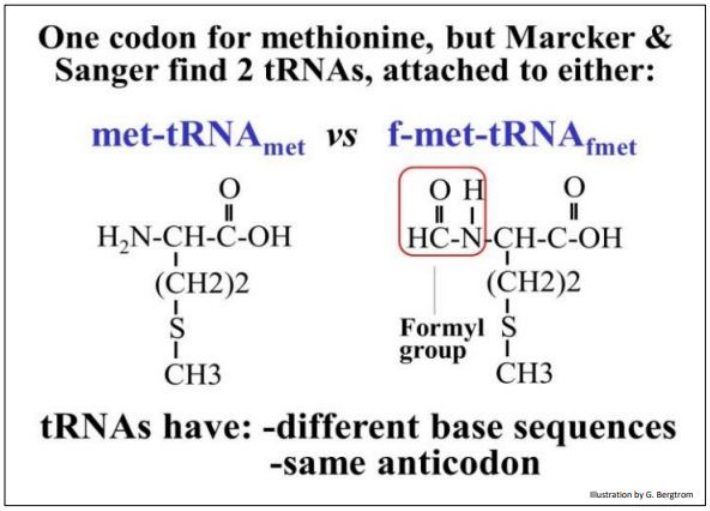
Translation is perhaps the most energy-intensive job a cell must do, beginning with the attachment of amino acids to their tRNAs. The basic amino-acylation reaction is the same for all amino acids. A specific **aminoacyl-tRNA** synthase attaches each tRNA to (**charges**) an appropriate amino acid. Charging tRNAs requires ATP and proceeds in three steps (shown below).



In the first step, ATP and an appropriate amino acid bind to the aminoacyl-tRNA synthase. ATP is hydrolyzed releasing a pyrophosphate (PPi) and leaving an enzyme-AMP-amino acid complex. Next, the amino acid is transferred to the enzyme, releasing the AMP. Finally, the tRNA binds to the enzyme, the amino acid is transferred to the tRNA and the intact enzyme is regenerated and released. The charged tRNA is ready to use in translation.

Several studies had already established that polypeptides are synthesized from their amino (N-) terminal end to their carboxyl (C-) terminal end. When it became possible to determine the amino acid sequences of polypeptides, it turned out that around 40% of *E. coli* proteins had an N-terminal methionine, suggesting that *all* proteins began with a methionine. It also turned out that, even though there is only one codon for methionine, two different tRNAs for methionine could be isolated. One of the tRNAs was bound to a methionine modified by *formylation*, called **formylmethionine-tRNA$_{fmet}$** (or **fmet-tRNAf** for short). The other was **methionine-tRNA$_{met}$** (**met-tRNA$_{met}$** for short), charged with an unmodified methionine.

Methionine and formylated methionine are shown below.



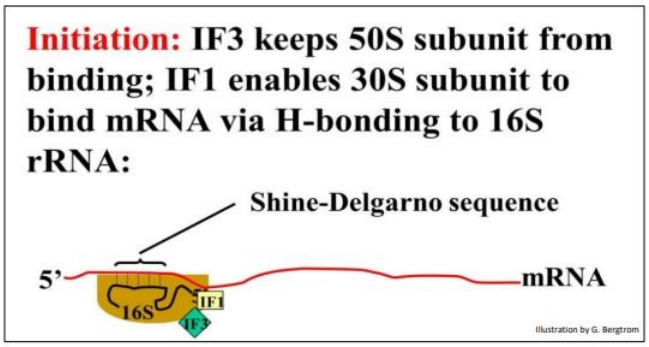
tRNA$_{met}$ and **tRNAf** each have an anticodon to AUG, the only codon for methionine, but have different base sequences encoded by different tRNA genes. tRNA$_{met}$ is used to insert methionine in the middle of a polypeptide. tRNAf is the initiator tRNA, and is only used to start new polypeptides with formylmethionine. In prokaryotes, methionine on met-tRNAf is formylated at its amino group to make the fmet-tRNAf. The formylating enzyme that does this does not recognize methionine on met-tRNA$_{met}$.

In *E. coli*, a *formylase* enzyme removes the formyl group from all N-terminal formyl methionines at some point after translation has begun. As we noted, the methionine itself (and sometimes more N-terminal amino acids) are also removed from about 60% of E. coli polypeptides. Eukaryotes have inherited both the initiator tRNAf and the tRNAmet, using only met-tRNAf during initiation. However, methionine on the eukaryotic initiator met-tRNAf is never formylated in the first place. What's more, methionine is absent from virtually all mature eukaryotic polypeptides.

Early in evolution, the need for an initiator tRNA must have ensured a correct starting point for translation on an mRNA and therefore growth of a polypeptide from one end to the other, that is, from its N- to its C-terminus. At one time, formylation of the N-terminal methionine may have served to block accidental addition of amino acids or other modifications at the N-terminus of a polypeptide. Today, formylation seems to be a kind of molecular appendix in bacteria. Since then, evolution (in eukaryotes at least) has selected other features to replace actual formylation as the protector of the N-terminus of polypeptides.

2. Initiation

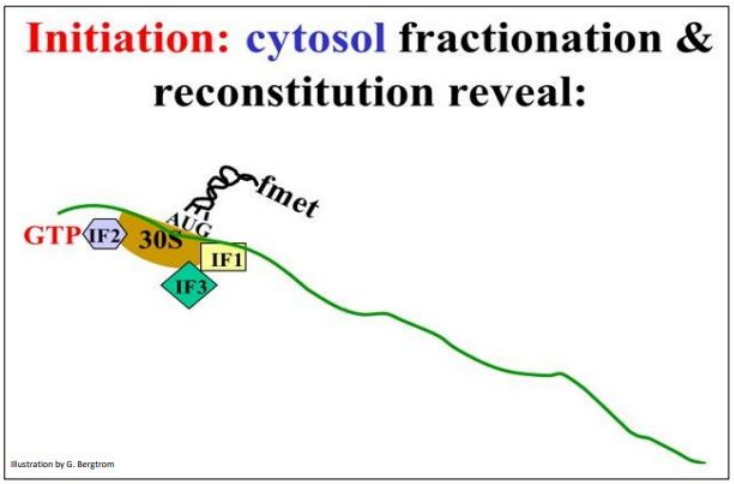
Now that we have charged the tRNAs, we can look more closely at the three steps of translation. Understanding translation initiation began with a molecular dissection of the components of E. coli cells required for cell-free (in vitro) protein synthesis, including cell fractionation, protein purification and reconstitution experiments. Initiation starts with when the **Shine-Delgarno** sequence forms Hbonds with a complementary sequence in the **16S rRNA** bound to 30S ribosomal subunit. The Shine-Delgarno sequence is a short nucleotide sequence in the 5' **untranslated region** (**5'-UTR**) of the messenger RNA, just upstream of the initiator AUG codon. This requires the participation of initiation factors **IF1** and **IF3**. In this event, IF1 and IF3 as well as the mRNA are bound to the 30S ribosomal subunit (below).



Demonstration of the binding of an mRNA to a ribosomal subunit required isolation and separation of the 30S ribosomal subunit, an RNA fraction of the cell, and the purification of initiation factor proteins from the bacterial cells. This was followed by reconstitution (adding the separated fractions back together) in the correct order show that mRNA would only bind to the 30S subunit in the presence of the two specific initiation factor proteins.

206 Translation Initiation: mRNA Associates with 30S Ribsomal Subunit

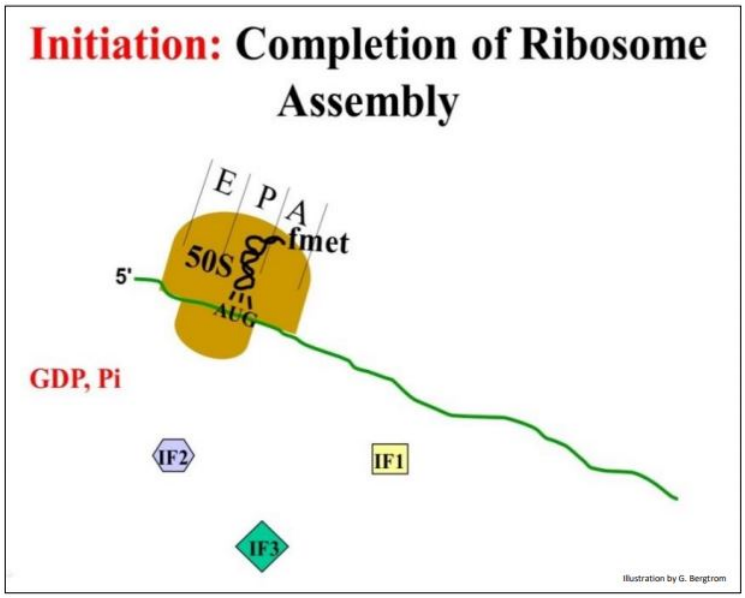
Next, with the help of GTP and another initiation factor (**IF2**), the initiator fmettRNAf recognizes and binds to the AUG start codon found in all mRNAs. Some call the resulting structure (shown below) the **Initiation Complex**, which includes the 30S ribosomal subunit, Ifs 1, 2 and 3, and the fmet-tRNAf.



207 Initiation Complex Formation

In the last step of initiation, the large ribosomal subunit binds to this complex. IFs 1, 2 and 3 disassociate from the ribosome and the initiator fmet-tRNA$_{fmet}$ ends up in the P site of the ribosome.

Some prefer to call the structure formed at this point the initiation complex (below).



208 Adding the Large Ribosomal Subunit

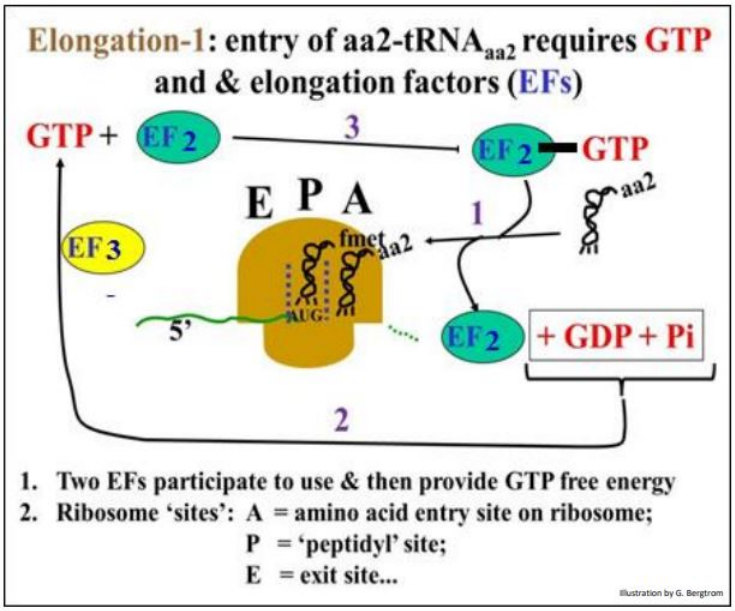
Initiation can happen multiple times on a single mRNA, forming the polyribosome, or polysome described in Chapter 1. Each of the complexes formed above will engage in the elongation of a polypeptide described next.

3. Elongation

Elongation is a sequence of protein factor-mediated condensation reactions and ribosome movements along an mRNA. As you will see, polypeptide elongation requires a considerable input of free energy.

a) Elongation 1

The first step in elongation is the entry of the next aminoacyl-tRNA (**aa2- tRNAaa2**), which requires the free energy of GTP hydrolysis. The energy is supplied by the hydrolysis of GTP bound elongation factor 2 (**EF2-GTP**). The aa2-tRNAaa2 enters the ribosome based on codon-anticodon interaction at the **A** site as shown below.

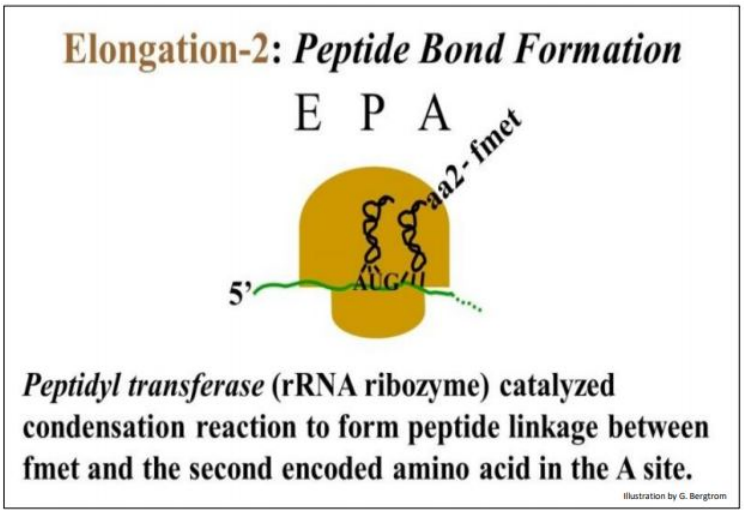


The GDP dissociates from EF2 as aa2-tRNAaa2 binds the anticodon in the A site. To keep elongation moving along, elongation factor (EF3) rephosphorylates the GDP to GTP, which can re-associate with free EF2.

209 Elongation: Elongation Factors and GTP

b) Elongation 2

Peptidyl transferase, a **ribozyme** component of the ribosome itself, links the incoming amino acid to a growing chain in a condensation reaction.



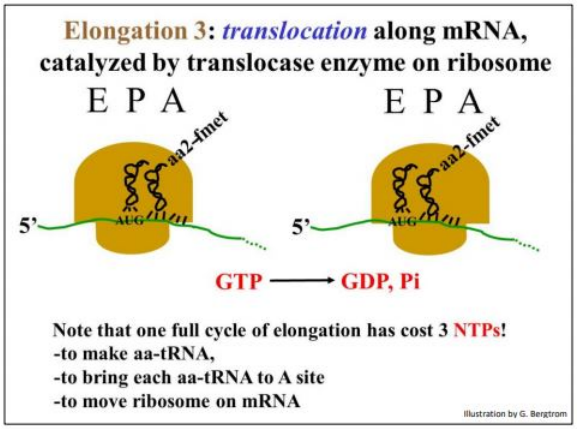
In this reaction, the fmet is transferred from the initiator tRNAf in the P site to aa2-tRNAaa2 in the A site, forming a peptide linkage with aa2.

210 Elongation: A Ribozyme Catalyzes Peptide Linkage Formation

c) Elongation 3

Translocase catalyzes GTP hydrolysis as the ribosome moves (translocates) along the mRNA. After translocation, the next mRNA codon shows up in the A site of the ribosome and the first tRNA (in this example, tRNAf) ends up on the E site of the ribosome.

The movement of the ribosome along the mRNA is illustrated below.



The tRNAf, no longer attached to an amino acid, will exit the E site as the next (3rd) aa-tRNA enters the empty A site, based on a specific codon-anticodon interaction (assisted by elongation factors and powered by GTP hydrolysis) to begin another cycle of elongation. Note that in each cycle of elongation, an ATP is consumed to attach each amino acid to its tRNA, and two GTPs are hydrolyzed in the cycle itself. In other words, at the cost of three NTPs, protein synthesis is the most expensive polymer synthesis reaction in cells!

211 Elongation: Translocase Moves Ribosomes along mRNA

212 Adding the Third Amino Acid

213 Big Translation Energy Costs

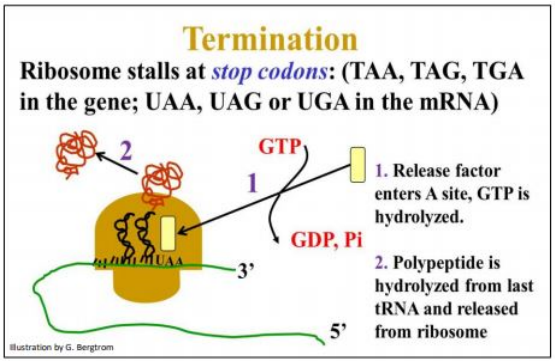
As polypeptides elongate, they eventually emerge from a groove in the large ribosomal subunit. As noted, a formylase enzyme in *E. coli* cytoplasm removes the formyl group from the exposed initiation fmet from all growing polypeptides. While about 40% of *E. coli* polypeptides still begin with methionine, specific proteases catalyze the hydrolytic removal of the amino-terminal methionine (and sometimes even more amino acids) from the other 60% of polypeptides. The removal of the formyl group and one or more N-terminal amino acids from new polypeptides are examples of post-translational processing.

214 The Fates of fMet and Met: Cases of Post-Translational Processing

4. Termination

Translation of an mRNA by a ribosome ends when translocation exposes one of the three stop codons in the A site of the ribosome. Stop codons are not situated some distance from the 3' end of an mRNA. The region between a stop codon to the end of the mRNA is called the **3' untranslated region** of the messenger RNA (**3'UTR**).

Since there is no aminoacyl-tRNA with an anticodon to the stop codons (UAA, UAG or UGA), the ribosome actually stalls and the translation slow-down is just long enough for a protein **termination factor** to enter the A site. This interaction causes release of the new polypeptide and the disassembly of the ribosomal subunits from the mRNA. The process requires energy from yet another GTP hydrolysis. After dissociation, ribosomal subunits can be reassembled with an mRNA for another round of protein synthesis. Translation termination is illustrated below.



215 Translation Termination

We have seen some examples of post-translational processing (removal of formyl groups in *E. coli*, removal of the N-terminal methionine from most polypeptides, etc.) Most proteins, especially in eukaryotes, undergo one or more additional steps of posttranslational processing before becoming biologically active. We will see examples in upcoming chapters.

Let's conclude this chapter with a "we thought we knew everything" moment! A recent study reports that ribosomes can sometimes **re-initiate** translation in the 3' UTR of an mRNA using AUG codons upstream of the normal start codon of the mRNA. There is evidence that the resulting short polypeptides may be functional! Click here to read more: here.

This page titled 20.4: Translation is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **11.4: Translation** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

20.5: Gene Expression- Translation

Learning Objectives

- Determine the amino acid sequence of a protein given the **nucleotide** sequence of a gene.
- Recognize the biological connection among mRNA, tRNA, amino acids, and proteins.
- Understand the concept of protein structure and its relation to protein function.

Introduction

The importance of gene expression is evident when you observe the changes plants go through during their lifecycle or during a season. Trees and bushes, for example have dormant buds through the winter. Environmental signals affiliated with the coming of spring induce genes in the buds to turn on and drive the dramatic changes of leaf development and flowering. The genes were always present in those bud cells but were controlled to turn on at the proper time. Understanding gene expression thus requires an examination of two processes, the activation of **gene expression** to make a "message" and the reading of this message to build a specific protein.

Amino acids are used to make proteins

Amino acids are the building blocks of proteins. Except for the amino acid proline, they all consist of a central carbon atom covalently bonded to an amino group, a hydrogen atom, a carboxyl group, and a variable R group (Figure 1). In the physiological range, both the carboxylic acid and amino groups are completely ionized allowing amino acids to act as either an acid or a base. Thus, amino acids do not assume a neutral form in an aqueous environment.

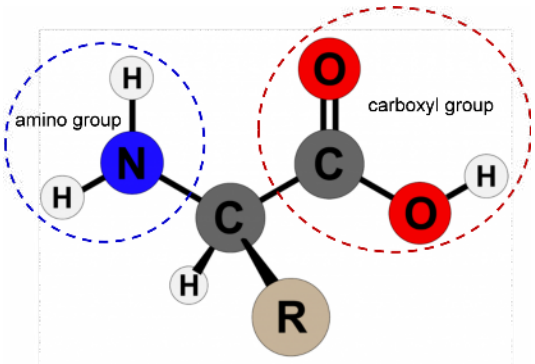


Figure 1. The structure of an amino acid. The amino (-NH2) and carboxyl (-COOH) are the same for all amino acids but the side chain (R group) is specific for each amino acid. Image Source: Techguy78, CC BY-SA 4.0, via Wikimedia Commons.

The chemical and physical properties of the side chains provide the functionally important properties to amino acids. Some of the important properties of side chains include acidic, basic, or neutral, hydrophobic vs. hydrophilic, as well as the size of side chains. They are categorized according to their side groups as nonpolar, polar, positively, or negatively charged or aromatic. All these properties and others affect how each amino acid contributes to the structure and function of a mature protein.

There are 20 different amino acids making up the subunits of **proteins** (Table 1). All proteins are made from differing combinations of the amino acids (Figure 2). Therefore, if you are going to make proteins you need to either make amino acids first (Figure 3) or consume amino acids in your diet. Animals consume some of their amino acids (the essential amino acids), but plants and bacteria are able make all their own amino acids.

Table 1. Amino acids and their abbreviations[1]

Full Name	Abbreviation (3 Letter)	Abbreviation (1 Letter)
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

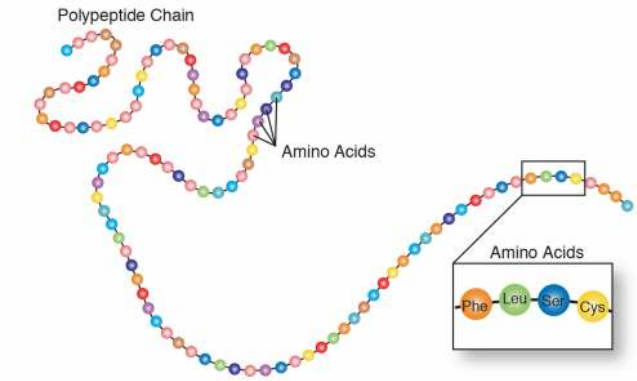


Figure 2. Proteins are polypeptides of different amino acids. Adapted from NIH-NHGRI.

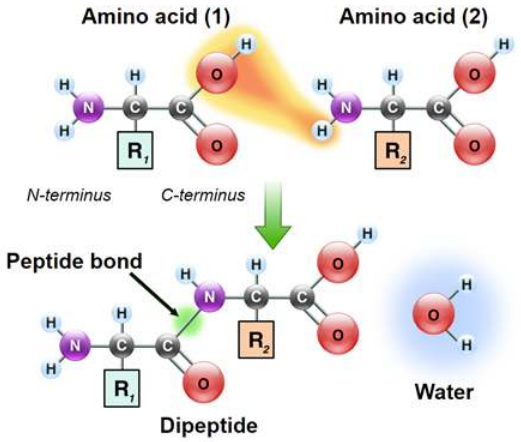


Figure 3. Proteins or polypeptides are made by connecting amino acids together with peptide bonds. The formation of peptide bond is a dehydration reaction that releases water with each new bond that is formed.

Codons and the Genetic Code

The search for the genetic code (Table 2) revealed that genetic information is stored in nucleotide triplets referred to as **codons**. The genetic code is degenerate because many amino acids are specified by more than one codon. Sixty one of the 64 possible combinations of the three bases in a codon are used to code for specific amino acids. Three **stop codons**, UAA, UAG, and UGA do not code for any amino acids but specify the termination of peptide chain synthesis during translation. The AUG **start codon** is used to initiate polypeptide synthesis and codes for the amino acid methionine.

Table 2. The genetic code. The 20 common amino acids are listed in their three and one-letter formats Stop codons are UAA, AUG and UGA.

First Position	Second Position				Third Position
	U	C	A	G	
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
					C
	Leu (L)		Stop	Stop	A
			Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
					C
			Gln (Q)		A
					G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
					C
			Lys (K)	Arg (R)	A
	Met (M)				G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
					C
			Glu (E)		A
					G

Why a Triplet Code?

Prior to understanding the details of **transcription** and **translation**, geneticists predicted that DNA could encode **amino acids** only if a code of at least three nucleotides was used. The logic is that the **nucleotide** code must be able to specify the placement of 20 amino acids. Since there are only four nucleotides, a code of single nucleotides would only represent four amino acids, such that A, C, G and U could be translated to encode amino acids. A doublet code could code for 16 amino acids (4 x 4). A triplet code could make a genetic code for 64 different combinations (4 X 4 X 4) genetic code and provide plenty of information in the DNA molecule to specify the placement of all 20 amino acids. When experiments were performed to crack the genetic code, it was found to be a code that was triplet. These three letter codes of nucleotides (AUG, AAA, etc.) are called codons.

The genetic code only needed to be cracked once because it is universal (with some rare exceptions). That means all organisms use the same codons to specify the placement of each of the 20 amino acids in **protein** formation. A codon table can therefore be constructed and any **coding region** of nucleotides read to determine the amino acid sequence of the protein encoded. A look at the genetic code in the codon table below reveals that the code is redundant meaning many of the amino acids can be coded by four or six possible codons. The amino acid sequence of **proteins** from all types of organisms is usually determined by sequencing the gene that encodes the protein and then reading the genetic code from the DNA sequence

Ribosomes, tRNA and Anti-codons

The ribosomes are a large molecular complexes containing an assembly of several ribosomal RNAs (rRNA) and many proteins. Ribosomes are the machinery of protein synthesis, facilitating the ordered addition of amino acids to a nascent polypeptide chain under the guidance of an mRNA template (Figure 5A). Prior to the initiation of protein synthesis the ribosome occurs in two separate subunits of size 60s and 40s ("s" stands for Svedberg units and is a measure of a particles sedimentation rate during centrifugation).

The meaning of a codon for a specific amino acid is determined by the tRNA (transfer RNA). Each tRNA (Figure 5B) contains a triplet of nucleotides referred to as an anticodon, which is complementary to a specific codon. For each of the 20 amino acids, a specific enzyme (aminoacyl-tRNA synthetase) catalyzes its linkage to the 3′ end of its specific tRNA adapter. The function of aminoacyl-tRNA synthetase is critical as it provides a check for accuracy in protein synthesis by adding a specific amino acid to a specific tRNA molecule. In this way, one particular amino acid is targeted to each codon triplet of mRNAs.

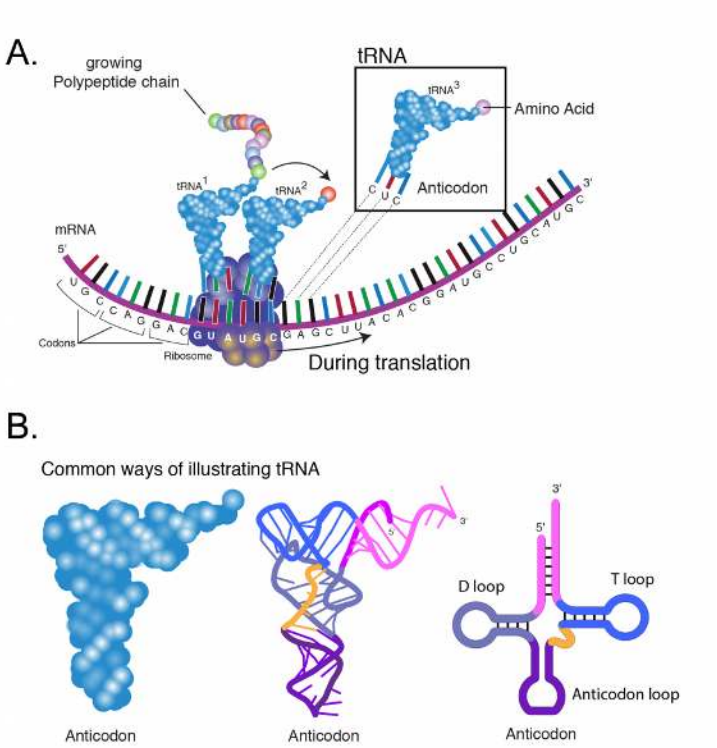


Figure 5: The key molecules needed for translation are the mRNA, ribosome, tRNA, and amino acids. Illustration by NIH-NHGRI.

Peptide chain formation

Protein synthesis is initiated by the 40s subunit through attachment it's to the mRNA, and recognition of the AUG codon. This is followed by the attachment of the 60s subunit to the 40s-mRNA complex to provide the structure necessary to align each successive aminoacyl-charged tRNA for transfer of its amino acid to the growing polypeptide. Successive amino acids are attached to the growing polypeptide by the formation of **peptide bonds** that form between the carboxyl group of one amino acid and the amino group of the next.

Like nucleic acids, polypeptides also have molecular directionality. One end of a polypeptide chain will have a free amino group and the opposite end will have a free carboxy group. As such, these ends are referred to as the amino-terminus or carboxy-terminus, respectively. During translation, mRNA templates are read from the 5' end toward the 3' end. Proteins are synthesized beginning at the amino terminal end going toward the carboxy terminus.

Protein structure and function

Proteins assume complex 3-dimensional structures that are essential for their various functions as enzymes, regulatory factors, or structural proteins. Overall structure is determined by complex physico-chemical interactions among amino acid side groups.

Several levels of structure are considered (Figure 6). The primary structure of a protein is the amino acid sequence of its polypeptide chains. The secondary structure is derived through the interactions between neighboring amino acids to form local structural elements such as beta sheets or alpha helices. The tertiary structure refers to the three-dimensional structure of an entire polypeptide and is determined by the diverse properties of amino acid side groups. For example, in an aqueous environment, hydrophobic amino acids will often interact at the core of a protein structure while hydrophilic groups may be exposed on the protein surface.

Many proteins are multimeric, composed of several polypeptide chains called “subunits”, which associate non-covalently or in some cases via disulphide bonds. Such higher order spatial associations among polypeptides are also the result of interactions among amino acids and result in a quaternary structure. Some multimeric proteins consist of multiple copies of the same polypeptide.

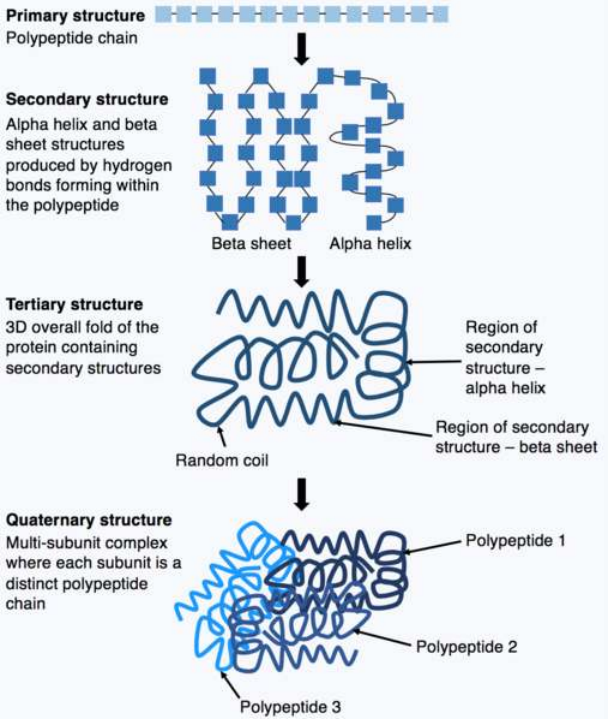


Figure 6. Levels of protein structure. Image Source: Kep17, CC BY-SA 4.0, via Wikimedia Commons.

Summary

The genetic information of DNA is transferred to through transcription to an intermediate molecule called RNA. The signals for starting and stopping transcription are located within the DNA sequence, and referred to as promoter and terminator sequences. The coding region of a gene is composed of a sequence of nucleotides that are transcribed into RNA. These sequences include exons and introns. Exons are the sequences that code for proteins. The coding region of a gene contains exons and introns. Also, pre-mRNA contains both introns and exons. The introns in pre-mRNA are removed through a process called intron splicing. The mRNA is processed by 5’ capping and addition of a poly(A) tail. Mature RNA is then translated to amino acids used to build proteins.

1. Table Source: Donald Lee, University of Nebraska-Lincoln ↵

This page titled 20.5: Gene Expression- Translation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Philip Becraft, & Philip Becraft (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.6: Gene Expression- Translation** by Donald Lee, Marjorie Hanneman, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

20.6: Gene Expression- Applied Example (Part 1)

Learning Objectives

At the completion of this lesson you should be able to:

1. Define the roles of DNA and proteins in cell development and metabolism
2. Determine the amino acid sequence of a protein given the **nucleotide** sequence of a gene.
3. Describe the roles that the **promoter**, **coding region** and **termination sequence** of a gene play in gene expression.
4. Recognize the differences between the structure of proteins, **amino acids**, **genes**, and **nucleotides**.
5. Draw the process of gene expression and include the following in your drawing. Gene, **RNA polymerase**, promoter, coding region, termination sequence, cell, **nucleus**, cytoplasm, **mRNA**, tRNA, ribosome, anticodon, codon, amino acid, protein, peptide bond.

You have learned this fundamental biology fact: genes provide the information to instruct cells how to build specific proteins and these proteins control cell development and metabolism. Some biologists refer to this fact as the "Central Dogma". We prefer to refer to this process as gene expression. This lesson will integrate this biology fundamental with some gene expression examples. These examples emphasize an additional fundamental biology idea that life is a continuous cycle. In order for living cells to live, they require the genes, RNAs, and proteins that came from prior existing living cells.

Example #1: Gene Expression of the Plant Acetolactate Synthase Enzyme Gene

We have selected this gene because the protein the gene encodes is the target site for several types of herbicides and thus is an important example in agriculture. The gene also illustrates an important circle of life concept. All living things need amino acids to build proteins and some living things need proteins to build amino acids.

The ALS **gene** is not found in animals but is found in plants and bacteria. Plants and bacteria are able make all their own amino acids. Animals must consume a subset of the 20 amino acids (see Lesson on Translation) in their diet. One gene all plants and bacteria must have to live encodes the enzyme acetolactate synthase (ALS). ALS enzyme allows plants to make their own supply of three of the 20 amino acids; leucine, isoleucine, and valine. The ALS protein is an **enzyme** and is needed to catalyze a reaction in the chemical pathway that makes these three **amino acids**. Let us detail the information transfer steps needed for gene expression: use the genetic code of DNA sequences in the gene to encode the amino acid sequence of the ALS enzyme.

Three parts of the Gene and a Two-Step Process

The ALS gene consists of several thousand nucleotides of DNA and like all genes has three main parts, the **promoter**, **coding region**, and termination sequences (Figure 1, gene sequence abbreviated for demonstration).

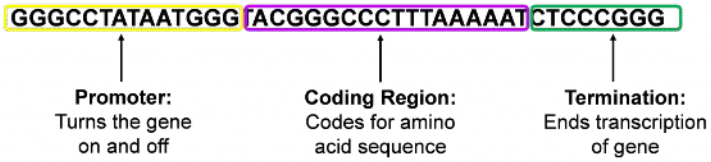


Figure 1. The parts of the ALS gene; the promoter and termination sequences control transcription while the coding region encodes the protein. Image by Marjorie Hanneman.

Each part of the ALS gene plays a role in controlling gene expression and this is a two-part process. The promoter is the on/off switch of the gene, the coding region determines the amino acids sequence of the **protein** that will get made and the termination sequence signals where the gene information ends. The DNA sequence that makes up a gene is a part of a much larger chromosome of continuous DNA. The gene sequence is organized to make an RNA (ribonucleic acid) copy of the gene's nucleotide sequence to serve as a mobile message for protein building. RNA building is called transcription because DNA nucleotides encode RNA nucleotides.

Step 1: Transcription: control from the promoter and termination sequence

The **promoter's** role is to selectively turn on the ALS gene in cells that will need the ALS enzyme. The promoter accomplishes this by its specific nucleotide sequences that react with regulatory molecules in the cell. We will discuss the regulation of gene expression in more detail later. What cells in a plant will need the ALS **enzyme**? All cells that are making proteins. This will be all cells in the plant at some stage of development. Therefore, the ALS gene has a promoter sequence that allows the gene to be turned on in all types of cells. When the ALS **gene** is turned on, transcription can take place. Like most chemistry in the cell, enzymes will be needed for **gene expression**. **RNA polymerase** is the *transcription* enzyme (Figure 2) which will bind to the DNA sequences in the ALS promoter and interact with one strand of the gene (the coding strand).

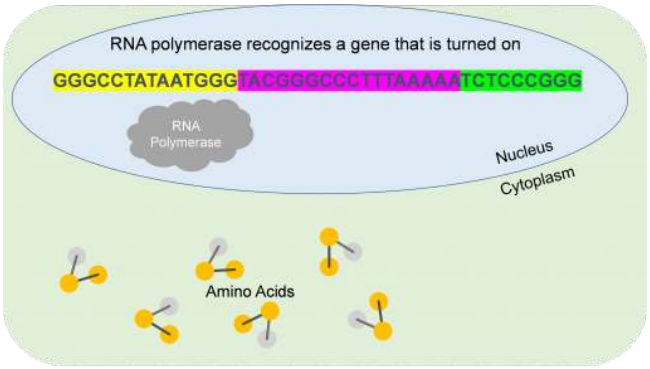


Figure 2. RNA Polymerase binds to the promoter of the ALS gene to initiate transcription. Image by Marjorie Hanneman.

RNA polymerase works by moving along the gene promoter until it encounters a series of A and T nucleotides called the TATA box (Figure 3). The TATA box signals RNA polymerase that it has reached the end of the promoter. Now the enzyme has the "green light" to read the coding strand **DNA** and make RNA. The procedure works much like DNA replication. RNA polymerase will read the DNA in a 3' to 5' direction and build the RNA 5' to 3'. Unlike DNA replication though, only one DNA strand is read and the transcription process must end once the RNA polymerase reaches the end of the ALS gene. If RNA polymerase kept going along that strand of the DNA making up the chromosome, other genes that are on the same **chromosome** might be expressed in the wrong cells or at the wrong times.

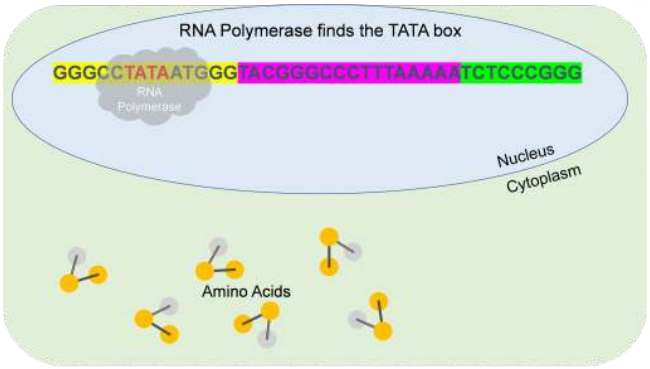


Figure 3. RNA polymerase encounters the TATA box at the end of the promoter. This signals where to begin the reading of the coding region. Image by Marjorie Hanneman.

How is the RNA polymerase signaled that it has reached the end of the gene? That is the role of a **termination sequence** in the gene (Figure 4). The termination sequences signal the end of the gene. One transcription termination strategy will be described here. The **nucleotides** that make up the termination sequence of the ALS gene could be ordered to form a palindrome. This means that the nucleotides being placed in the RNA can fold back on themselves and form a hairpin loop (Figure 5).

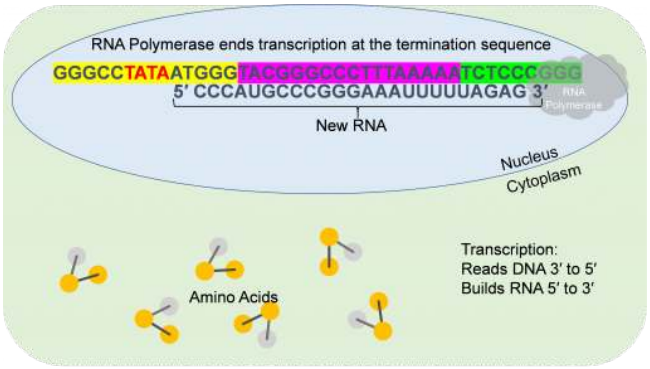


Figure 4. When RNA polymerase reads the termination sequence it is signaled to quit reading the gene's coding strand. Image by Marjorie Hanneman.

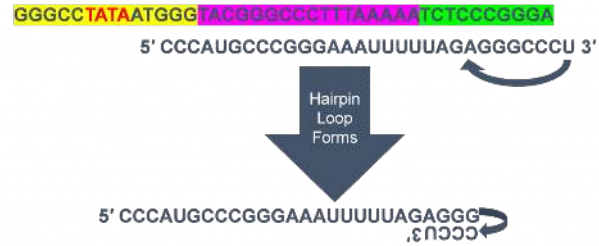


Figure 5. The termination sequence is palindromic. The RNA being made has a sequence that can form a hairpin loop. The formation of the hairpin in the newly made RNA disrupts the RNA polymerase, halting transcription. Image by Marjorie Hanneman.

While no geneticist has actually viewed this hairpin loop formation in action, it is believed that the hairpin formation snaps the RNA polymerase off the DNA coding strand and releases the RNA message. Transcription of an RNA that has the **coding region** information is now completed. Each time an RNA polymerase goes through the process, one copy of the RNA is made by reading the DNA template. The type of RNA made from transcription of the ALS gene is called a messenger RNA or mRNA for short. Some **genes** can encode transfer RNA (tRNA) or ribosomal RNA (rRNA). These RNA molecules have special roles in the next part of gene expression called translation. Once an **mRNA** is transcribed, some modifications will occur in the nucleus. We will describe some of the post-transcriptional processes later. For now, let's follow the ALS mRNA onto the **translation** step of gene expression.

Video Examples

Watch this video from UNL for more information about the Transcription process

Step 2: Translation: reading the code to make proteins

The **translation** process requires a conversion of information from a chain of **nucleotides** in RNA to a chain of **amino acids** in **protein**. The translation term was coined based on the idea that we are converting to a new molecular language at this stage of **gene** expression. The conversion of information in translation is a little more complex than **transcription** and requires several biomolecules to interact and work together (Figure 7). We will describe the ribosomes and tRNAs molecules first and then describe how they work together to translate a **mRNA**.

- **tRNA:** The 't' prefix means transfer. The tRNA is a small RNA (70 to 80 nucleotides) and has a sequence that allows hairpins to form which gives a secondary structure to the molecule. Their small size makes them mobile to provide the cells means to transfer amino acids to the site of protein building. Their structure also makes them specific in how they deliver their cargo of a single, specific amino acid. A tRNA is about 50 times larger than an amino acid.
- **Ribosomes:** Ribosomes are macromolecules that are made of both RNAs (rRNAs) and proteins. These components assemble into a workbench for the translation process to occur. The ribosome is a big macromolecule: it can hold three tRNAs in position for the translation work and can catalyze bond breaking and bond formation.

We will use a simple representation of the mRNA, tRNA, ribosome, and amino acid to show how they work together (Figure 6).

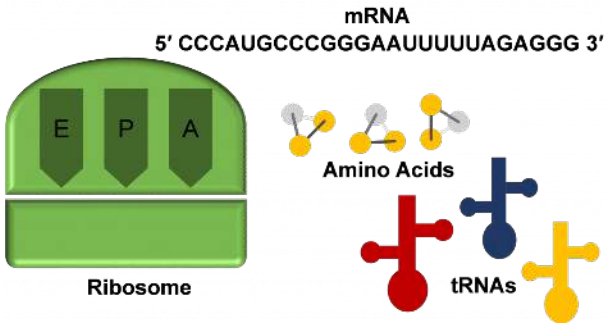


Figure 6. The key molecules needed for translation are the mRNA, ribosome, tRNA, and amino acids. Image by Marjorie Hanneman.

The tRNAs shown in Figure 6 are not ready for translation. The structure of the tRNA is recognized by special enzymes in the cell that attach the proper amino acid to the tRNAs. The tRNAs that are bound to their designated amino acid are called charged tRNAs (Figure 7). The near the middle of the 70-80 nucleotides in a tRNA are three nucleotides called the anticodon. When the tRNA folds, the anticodon becomes one end of the molecule and the other end will be attached to a specific amino acid. In the figures for this lesson, the tRNA is shown with its secondary structure and only the anticodon sequence is shown (Figure 7).

Let's describe how the key players get protein building started, how they keep it going, and how they end it. Active ribosomes are assembled from small and large subunits when the ribosome "workbench" is ready to translate a mRNA (Figure 7). When mRNAs are present in the cytoplasm, the small and large ribosome subunits assemble at the 5' end of these messages and are ready to read mRNAs like blueprints to build specific proteins.

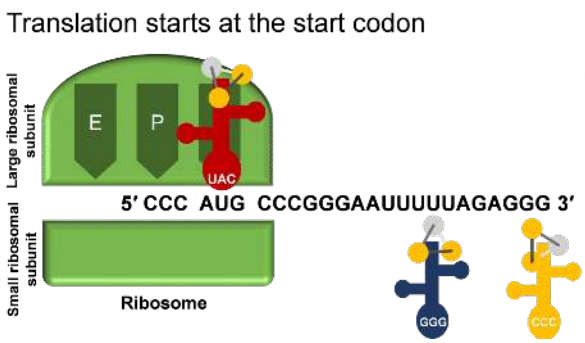


Figure 7. The ribosome translates the mRNA 5' to 3' and starts by finding an AUG sequence. The ribosome allows a tRNA with the UAC anticodon to bind and bring in an amino acid. The tRNA is bound to the A site of the ribosome. Image by Marjorie Hanneman.

The ribosome reads the mRNA quickly and accurately, but it must determine where to start. As the ribosome moves in the 5' to 3' direction along the mRNA it "looks" for the sequence AUG (Figure 7). This sequence is called the start codon and signals the ribosome that this is where to begin building the protein sequence.

Video Examples

Watch this UNL video on Translation for more information about these processes.

The Reading Frame, Codons and Anticodons

The mRNAs, tRNAs, ribosomes, and **amino acids** are now in position to combine their specific functions into the building of a protein. The AUG start codon establishes the beginning of what is called a reading frame on a mRNA. The nucleotides in the mRNA must have the specificity to code for 20 different amino acids to build the protein. The genetic code is read as a three-letter code which provides this specificity. The ribosome will move along and read the sequence in the mRNA one codon (three nucleotides) at a time to follow the reading frame and build the correct protein (Figure 8).

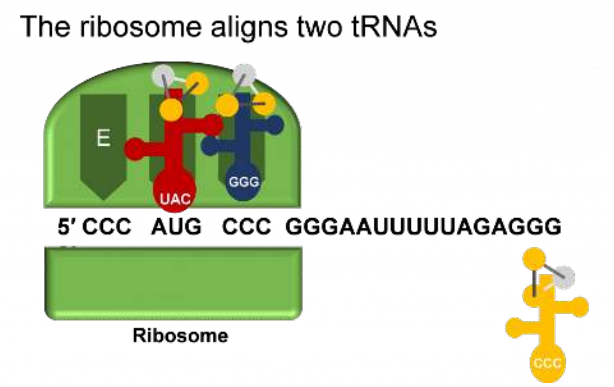


Figure 8. The ribosome allows two tRNAs with complementing anticodons to enter the 'A' and 'P' sites and bring in their amino acids. Image by Marjorie Hanneman.

Getting Translation Started with the Start Codon

The AUG start codon signals the ribosome to place in the amino acid methionine. The A, P and E sites on the ribosome site works somewhat like a vice on the workbench, holding stuff in place so it can be worked on. In Figure 9, the ribosome has moved in the 5' to 3' direction on the mRNA .mRNA. This allowed a tRNA with the anticodon GGG to move into the A site and placed two amino acids very close to each other on the ribosome. The ribosome will now move one more codon in the 5' to 3' direction which moves the two tRNAs now to the E and P sites (Figure 9). Now the ribosome is ready to start binding together amino acids with peptide bonds to start to form a protein. A third tRNA (with the CCC anticodon) can now arrive at the open A site, bringing the amino acid designated by the GGG codon.

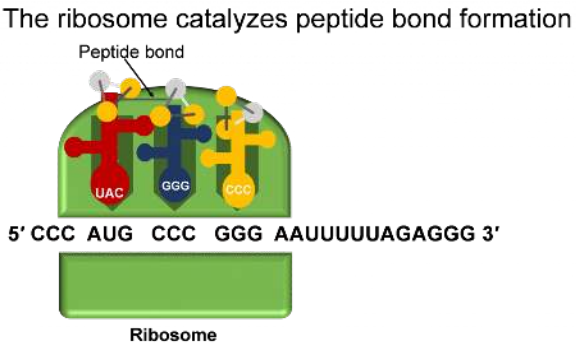


Figure 9. The ribosome can link two amino acids together with peptide bonds because the tRNAs holding them are next to each other at the 'A' and 'P' sites of the ribosome. Image by Marjorie Hanneman.

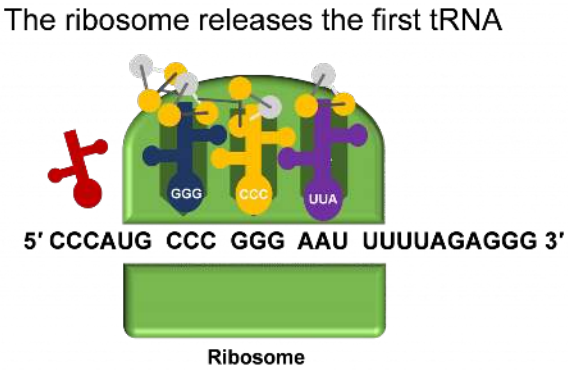


Figure 10: The ribosome moves over to the 'GGG' codon, releases the tRNA with the 'UAC' anticodon and is ready for the next tRNA. Image by Marjorie Hanneman.

Peptide Bond Formation and Protein Building

As you can see, a ribosome “workbench” has catalytic tools to go along with its ‘P’ and ‘A’ “vice sites”. The ribosome has enzymatic functions that allow it to break and form bonds. The ribosome will break the bond that binds the amino acid to the tRNA at the ‘E’ site. Simultaneously the ribosome forms a peptide bond between the second and the third **amino acids** that were brought in by the tRNAs. The close proximity of these amino acids at the ribosome sites provides the opportunity for peptide bond formation and then release of the tRNA in a smooth continuous process (Figure 10 and 11). Peptide bonds always bind the acid end of one amino acid with the amino end of the next amino acid. Peptide bonds are strong covalent bonds which keep the amino acids connected during and after **translation**. The tRNA at the ‘E’ site has now accomplished its task by bringing in the first amino acid. This tRNA will then move off the ribosome (Figure 11). The tRNA released from a ribosome is recycled by the cell. It can be charged again by binding another amino acid in the cytoplasm and contribute to the synthesis of another protein. The protein. The first three amino acids in the **protein** have now been put together.

The synthesis of our protein is far from over. The ribosome continues to shift one codon in the 5’ to 3’ direction as a tRNA leaves the E site .site. A tRNA can now come into the A site if it has the anticodon that is complementary to the next codon on the **mRNA**. **Proteins** are hundreds of **amino acids** long. In our example, the steps described continue until the complete mRNA information is read to build the 667 amino acid long ALS enzyme.

The End of Translation: stop codons looking for something they cannot find

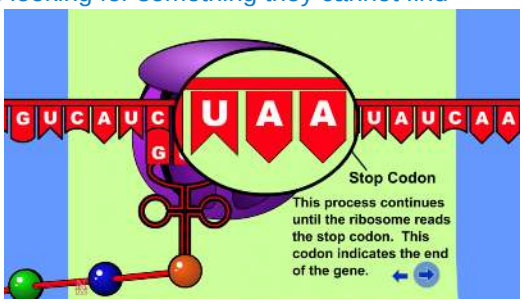


Figure 11: Ribosome encounters one of the three sequences in the genetic code that function as stop codons. Image by Donald Lee.

Start codons start **translation** so it is logical that stop codons stop translation (Figure 11) How do stop codons do this? There are 61 tRNAs with different anticodons (see Appendix 1: Codon Table). That means there are three codons that do not have corresponding tRNAs with complementary anticodons. These three codons serve as stop codons. When a ribosome encounters a stop codon on a **mRNA** it will wait for a tRNA with the right anticodon to come over. It will not skip the codon or shift over one **nucleotide** to form a new reading frame. The ribosome waits for the right tRNA, but it does not wait for long. A stalled ribosome will quickly cleave off the bound tRNA with the newly built **protein** chain and then move on to translate another mRNA. No tRNAs in the cell have anticodons that complement any of the three possible stop codons. Therefore, stop codons are able to end the translation process when the completed protein is made.

Sequence, structure, and function

In the case of our ALS protein, the assembly of all 667 amino acids into the polypeptide or protein chain in the correct order is important because the sequence of the amino acids determines how the protein folds. The folded structure of the protein determines its function. If some amino acids are changed and the structure changes, the ALS will not function properly as an enzyme in plant or microbe cells.

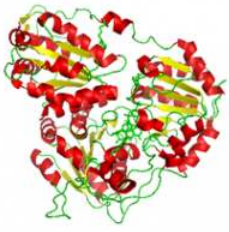


Figure 12. ALS enzyme structure. The red and green stretches depict different kinds of protein structure features that provide the enzyme its ability to function as a specific enzyme. Image Source: Rj Joplin, CC BY-SA 3.0, via Wikimedia Commons.

Summary

Organisms use the process of **transcription** to make **mRNA** as a temporary molecule to transport **protein** coding information from a **gene**. The **translation** process reads the mRNA and makes the intended protein, one amino acid at a time. This process is controlled by the **promoter** of the gene to ensure that **proteins** are made in the right cells at the right time to drive growth, development, metabolism, stress response and other processes needed to complete the organism's life cycle.

Video Examples

Watch this video from UNL for an overview of Transcription and Translation.

Circle of Life and Gene Expression

Let us revisit the idea that gene expression works because living cells give rise to more living cells. Imagine yourself when you were just a single cell zygote with a copy of all the genes in the human genome from your biological mother and a copy from your biological father. What else did you need to get from your parents before you could express your first gene and build your first protein? You needed to already have an RNA polymerase in the nucleus of the cell to transcribe your first gene and you needed to have ribosomes, tRNAs and amino acids ready for translating that mRNA in the cytoplasm. Luckily, the egg cell from your mother had the 23 human chromosomes plus all of these gene expression biomolecules. Later, you could turn on your own genes for RNA polymerase, tRNAs and ribosomes but your mother's egg cell was a living environment that allowed you to start your life.

A plant treated with a herbicide which binds to and blocks the action of the ALS enzyme likewise demonstrates that gene expression has a cycle of dependency on living cells. Small plants in an environment for rapid growth will have their growth halted by the ALS binding herbicides. The plants do not die from the herbicide, they just stop growing. The reason for this is that plants need to build new proteins to grow and to build new proteins they need the ALS enzyme to be working and building new leucine, isoleucine, and valine amino acids. ALS herbicide treated plants stop making these three amino acids and translation stops for every protein in the cell that requires at least one of these amino acids. Living things need amino acids to build proteins and plants and microbes need the ALS enzyme to build three of the amino acids. The next lesson (**Gene expression part 2**) will apply the gene expression principles from this lesson to the observation that some plants can be resistant to ALS inhibitor herbicides.

This page titled 20.6: Gene Expression- Applied Example (Part 1) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Donald Lee, Patricia Hain, & Patricia Hain (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.7: Gene Expression- Applied Example (Part 1)** by Donald Lee, Marjorie Hanneman, Patricia Hain, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

21: Alleles and Gene Function

21.1: Types of Mutations
21.2: DNA Mutations

21: Alleles and Gene Function is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

21.1: Types of Mutations

Mutations (changes in a gene sequence) can result in mutant alleles that no longer produce the same level or type of active product as the wild-type allele. Any mutant allele can be classified into one of five types: (1) amorph, (2) hypomorph, (3) hypermorph, (4) neomorph, and (5) antimorph.

- **Amorph** alleles are complete loss-of-function. They make no active product – zero function. The absence of function can be due to a lack of transcription (gene regulation mutation) or due to the production of a malfunctioning (protein coding mutation) product. These are also sometimes referred to as a **Null** allele.
- **Hypomorph** alleles are only a partial loss-of-function. They make an incompletely functioning product. This could occur via reduced transcription or via the production of a product that lacks complete activity. These alleles are sometimes referred to as **Leaky** mutations, because they provide some function, but not complete function.

Both amorphs and hypomorphs tend to be recessive to wild type because the wild type allele is usually able to supply sufficient product to produce a wild type phenotype (called haplo-sufficient - see Chapter 6). If the mutant allele is not haplo-sufficient, then it will be dominant to the wild type.

While the first two classes involve a **loss-of-function**, the next two involve a **gain-of-function** – quantity or quality. Gain-of-function alleles are almost always dominant to the wild type allele.

- **Hypermorph** alleles produce more of the same, active product. This can occur via increased transcription or by changing the product to make it more efficient/effective at its function.
- **Neomorph** alleles produce an active product with a new, different function, something that the wild type allele doesn't do. It can be either new expression (new tissue or time) or a mutation in the product to create a new function (additional substrate or new binding site), not present in the wild type product.

Antimorph alleles are relatively rare, and have an activity that is dominant and opposite to the wild-type function. These alleles usually have no normal function of their own and they interfere with the function from the wild type allele. Thus, when an antimorph allele is heterozygous with wild type, the wild type allele function is reduced. While at the molecular level there are many ways this can happen, the simplest model to explain antimorph effect is that the product acts as a dimer (or any multimer) and one mutant subunit poisons the whole complex. Antimorphs are also known as dominant negative mutations.

Identifying Muller's Morphs - All mutations can be sorted into one of the five morphs base on how they behave when heterozygous with other alleles – deletion alleles (zero function), wild type alleles (normal function), and duplication alleles (double normal function).

This page titled 21.1: Types of Mutations is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **4.4: Types of Mutations** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

21.2: DNA Mutations

Learning Objectives

- List the types of mutation.
- Predict the effects of specific mutations in gene coding regions on the protein encoded by the gene.
- Describe ways in which mutations arise, naturally and experimentally.

Introduction

Crop improvement relies on genetic variation, which is synonymous with DNA variation. Mutations are the source of genetic variation that is the driver for trait improvement. In general, mutations are rare events, and most are deleterious and get selected against during the course of evolution. However, occasionally mutations create genetic variation that is beneficial. As such, it is often of interest for plant improvement or for genetic studies to induce genetic mutations by exposing seed or plant tissue to certain agents, called mutagens.

By definition, a mutation is a heritable change in DNA sequence. This can happen in several ways: substitution of a DNA base, insertion or deletion of one or more DNA bases, or by large-scale chromosomal rearrangements, the latter of which will not be considered here. It has also become clear in recent years that chromatin carries additional information to the DNA sequence in the form of base modifications such as cytosine methylation. This is known as **epigenetic** modification and can impact gene function and plant traits.

A. DNA base substitutions

The simplest type of mutation is a substitution of one base for another in the DNA sequence. Substitutions most often arise as errors during DNA replication or repair. The most common type is the **transition**, where one pyrimidine may be substituted by the other, or one purine by the other. The less common type is the **transversion**, in which a purine may be substituted by a pyrimidine or vice versa. As described in the next section, these can have various effects on gene function depending on where they occur in a gene.

B. Insertions & deletions



Figure 1. Example of DNA base substitution. In this example, the base G is substituted for T. Image by NIH-NHGRI.

Insertions or deletions of DNA are other types of mutations that occur quite frequently. They can vary in size from one to thousands or more nucleotide bases. Their effects on gene function depend on the size of the insertion or deletion, and the location relative to the gene. As described below, insertions and deletions that are in multiples of 3 are referred to as in-frame insertions and deletions and result in the addition or deletion of amino acids from the protein sequence. When not in multiples of three, insertions and deletions lead to **frameshift mutations**, altering the amino acid sequence of the protein encoded by the mutated gene.



Figure 2. Example of deletions. In this example, the base A is deleted. Image by NIH-NHGRI.

A common source of inserted DNA is from **transposons**, sequences of DNA that are able to move from one genomic site and insert into another. Transposons were discovered in maize by Barbara McClintock. For this discovery, she was awarded the Nobel Prize in Physiology and Medicine in 1983.

Some transposons copy themselves and multiply in number within a genome. Transposons can cause wide scale chromosome rearrangements and gene mutations. A feature of mutations caused by many plant transposons is that they are unstable. Excision of

a transposon from the mutated gene can often restore it to wild type or create a new stable mutant allele. Transposons are found in every living organism studied and are a major driving force in genome evolution.



Figure 3. Example of a DNA insertion. In this example, the base A is inserted between T and C. Image by NIH-NHGRI.

Effects of mutations on gene function

A. Coding region mutations

Silent mutation

A silent mutation is a mutation that results in the change of a codon without a change in the amino acid represented by the codon. Because of the redundancy of the genetic code, DNA base substitutions may not lead to the incorporation of an incorrect amino acid in the protein.

Missense mutation (amino acid substitution)

A missense mutation is a single nucleotide base substitution that alters a codon such that it codes for a different amino acid that is incorporated into the encoded protein. The amino acid substitution may affect the function of the protein if it occurs in a critical portion of the protein.

Frameshift mutation

Due to the triplet nature of the genetic code, an insertion or deletion can change the **reading frame** for the entire subsequent sequence. For example, if a particular sequence is read sequentially (Figure 4).

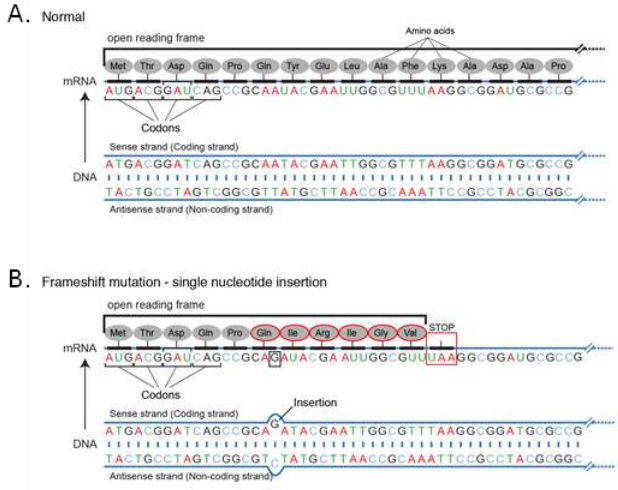


Figure 4. Impact of frameshift mutation on gene function. The normal gene sequence is shown in part A. The insertion of the base G shown in part B, results in a new protein. Image by NIH-NHGRI.

Because the new sequence of codons is different from the original, the entire amino acid sequence is changed from the point of insertion. A similar effect is seen for deletions. This effect is seen whenever the number of nucleotide bases inserted or deleted is not a multiple of 3 and it usually will result in the loss of the function of the protein.

Amino acid additions or deletions

On the other hand, if insertions or deletions are in multiples of 3, amino acid additions or deletions can arise. Consider the following 3-base insertion (XXX):

UUU CCC XXX AAA GGG (codons)

aa1 aa2 aa5 aa3 aa4 (amino acids)

If such insertions or deletions occur in important regions of a protein, they might impair function.

Nonsense mutation

A nonsense mutation arises when a functional codon is changed to a stop codon. Nonsense mutations may cause premature termination of translation.

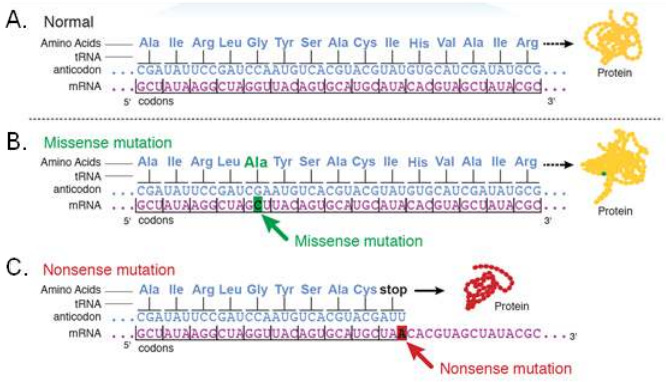


Figure 5. Impact of missense and nonsense mutations on gene function. The normal gene sequence is shown in part A. In part B, a missense mutation leads to replacement of the Gly by Ala. In part C, a nonsense mutation results in premature stop codon and shortening of the protein. Adapted from NIH-NHGRI.

B. Regulatory region mutations

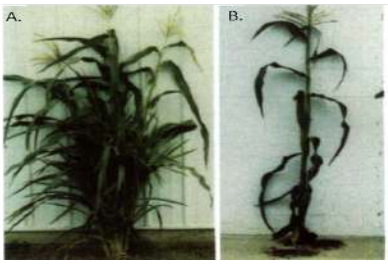


Figure 6. Regulatory region mutation of the tb1 gene converted teosinte (A) to modern maize (B). Image courtesy of John Doebley.

Promoters and **enhancers** control temporal, spatial and quantitative aspects of gene expression. As such, mutations in these regulatory regions of a gene can alter the regulation of the gene's expression. Genes can become expressed at inappropriate times, places or levels, which can affect a plant's phenotype.

Teosinte, the progenitor of modern maize, produces multiple branches (Figure 6A) due to reduced apical dominance. In teosinte and modern maize, apical dominance is under the control of a gene called *teosinte branched1*, ***tb1*** (Doebley et al. 1997). An insertion of a transposon in the regulatory region of the ***tb1*** gene in modern maize causes the gene to be strongly expressed (Doebley et al. 1997; Studer et al. 2011), resulting in enhanced apical dominance, and suppression of excessive branching (Figure 6B).

C. Recessive vs. dominant mutations

Recessive mutations

Recessive mutations most commonly result from a loss of function. A loss of function mutation causes a decrease in the function of a gene. A complete loss of function is known as a **null mutation** while partial loss of function mutants are often referred to as **hypomorphs** or "leaky". Loss of function mutations can affect gene expression or the function of the resultant protein. Most loss of function mutations are recessive because in a heterozygote, the normal allele can provide the necessary function.

Dominant mutations

Dominant mutations can affect genes in several different ways. One general class is gain of function mutations. These encompass several different types.

1. **Overexpression mutants**. As the name implies, the affected gene is expressed at levels that are inappropriately high, causing too much of a gene product to be produced.
2. **Neomorphic mutations** arise when a mutation causes a qualitatively new effect not seen with normal alleles. This can occur when a mutation affects the regulation of a gene such that it becomes expressed in a new time or place in the plant. It can also happen when a mutation alters the properties of the encoded protein. For example, a transcription factor could be mutated to recognize a different promoter sequence, or an enzyme could recognize a new substrate
3. **Dominant negative mutations** alter a gene product such that the mutant gene product interferes with the function of the normal one. This can occur at the protein level. For example, a protein might function as a multimeric complex and the presence of one mutant subunit could compromise the function of the entire complex. Dominant negatives can also occur at the RNA level for example if an insertion, deletion or rearrangement causes the wrong strand to be transcribed, creating an antisense or RNAi suppression effect.

Effects of mutations on traits

Example of a loss of function mutation – gene function is required for a trait

Sweetness is an important quality trait in maize (corn). One of the loss of function mutants used to improve sweetness in corn is ***sugary1*** (***su1***). The Su1 gene is expressed in the endosperm (Lertrat and Pulam, 2007), and it encodes and enzyme that is required for normal starch biosynthesis (Rahman et al. 1998). In the ***su1*** mutant, with defective starch biosynthesis, the concentration of sugar is 3 times higher than wild type making this mutant valuable for sweet corn (Lertrat and Pulam, 2007). In addition, ***su1*** kernels accumulate a highly branched polysaccharide called phytoglycogen, which gives the kernels a creamy texture. New commercial sweet corn hybrids are developed using a combination of starch biosynthesis mutants (Nelson and Pan, 1995; Lertrat and Pulam, 2007).

Example of a gain of function mutation – gene function is sufficient for a trait

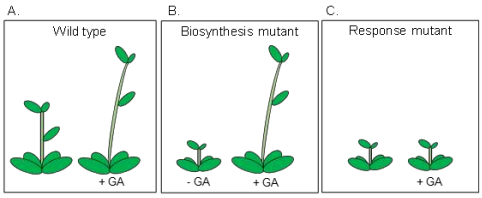


Figure 7. GA promotes plant growth. (A) When applied to wild type plants, GA induces height. (B) A mutation that blocks GA biosynthesis will result in short plants, and exogenous addition of GA to a biosynthetic mutant restores growth. (C) However, a mutation that affects GA signaling leads to failure to sense GA. Thus, exogenous addition of GA does not restore growth of a GA response mutant (C). (Image informed by Peng et al. 1999).

The plant growth hormone, gibberellin (GA) regulates plant height, thus, a plant that fails to produce GA, or does not respond to GA will have a dwarfed phenotype (Figure 7). At the molecular level, the cell response to GA involves several processes, including the destruction of repressor proteins, referred to as DELLAs.

The “green revolution gene” ***Reduced height-1 (Rht-1)*** from wheat encodes a DELLA repressor of GA signaling (Peng et al. 1999). A stretch of five amino acids referred to as the DELLA domain is required for the destruction of the repressor in response to GA. Mutations in the DELLA domain of ***Rht-1*** genes result in a protein product that is resistant to degradation, leading to failure to signal a GA response. The failure to signal the presence of GA in the ***rht-1*** mutants is responsible for their short stature. When the mutant ***rht-1*** gene is transformed into rice, the result is also short plants, which is proof that the new function of the DELLA protein is sufficient to induce dwarfism (Peng et al. 1999).

Generation of mutations

Because mutations are the ultimate source of genetic variation for breeding and genetic studies, it is sometimes of interest to generate new mutations. New DNA mutations (Figure 8) can arise spontaneously or they can be induced by experimental methods.

Spontaneous mutations arise “naturally”

That is, no particular effort was made on the part of the scientist to increase the mutation rate. Spontaneous mutations generally occur at very low rates, on the order of 1 per gene per million gametes.

Mistakes in DNA replication

Very rarely, incorrect bases are incorporated or bases omitted from a DNA strand during synthesis. Such mistakes can lead to spontaneous substitutions, insertions or deletions. For example, strand slippage due to the formation of a loop.

Environmental mutagens

There are several chemical agents present in the environment with potential to damage DNA and create mutations. These include naturally occurring compounds as well as man-made pollutants. Some modify DNA nucleotide bases through deamination (e.g., nitrous acid), while others promote oxidative reactions that may damage DNA (e.g., ozone).

Radiation is another type of environmental mutagen. Ionizing radiation (e.g., X-rays) can shatter DNA sequences and promote chromosome rearrangements. The less powerful ultraviolet rays can penetrate the cell and promote the formation of pyrimidine dimers which may inhibit replication and transcription.

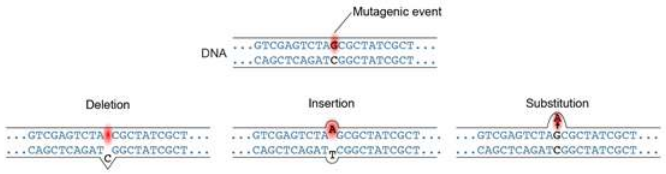


Figure 8. Mutations can arise spontaneously or they can be induced by experimental methods. Illustration by NIH-NHGRI.

Experimental mutagenesis involves applying a mutagenic agent to a plant or plant part to generate a mutation. For the mutation to be useful, it must be heritable, which imposes limitations on what tissues can be treated. Following the mutagenic treatment, a subsequent breeding scheme must be implemented to generate pure breeding lines of appropriate genotypes to display mutant phenotypes. These will depend on the species and the design of the mutagenesis.

Chemical mutagenesis

A commonly used chemical mutagen for experimental plant biology and mutation breeding is ethyl methanesulfonate (EMS). The compound is known as an alkylating agent and induces a CG to TA transition.

Seed Mutagenesis

This is the most common method of chemical mutagenesis in most species. Seeds are soaked in EMS solution, planted and allowed to flower and set seed. During EMS mutagenesis, every cell in the embryo will be independently mutagenized and the resultant seed and subsequent plant will be chimeric. These seeds can be sown and plants screened for traits of interest. For self-pollinating species, the required breeding is very simple, whereas for dioecious species it would be much more elaborate and labor-intensive.

Pollen Mutagenesis

For a few species such as maize, pollen mutagenesis is the method of choice. Pollen are treated with EMS and applied to silks of an acceptor plant. The benefit is that the resultant seed is not chimeric but rather all the cells of the embryo carry the same mutagenized paternal genome inherited from the pollen grain.

Insertional mutagenesis

The idea behind insertional mutagenesis is that a mobile or introduced piece of DNA can sometimes insert into a gene, thereby inactivating or modifying the function of the gene. The approach is useful in gene cloning because the inserted DNA serves as a "tag" that can be used to locate the gene of interest, and subsequently isolate that gene.

Another strategy used by researchers is mutagenesis using T-DNA from the ***Agrobacterium tumefaciens*** Ti plasmid. A larger number of plants (up to 10,000 for Arabidopsis) are transformed with T-DNA to generate mutants at random.

Similar to the example of transposon mutagenesis above, a gene can be "tagged" by a T-DNA and the tag used to isolate the gene. However, a T-DNA insertions are stable and thus disrupt the gene permanently.

In recent years, insertional mutagens have been engineered in many innovative ways. For example, strong promoters have been engineered into transposons such that if the transposon inserts near a gene, it can cause strong transcriptional expression of that

gene. Such “activation tagging” approaches have been successful in identifying many genes that were not amenable to other mutational approaches.

Summary

Mutations alter A-T and G-C base pairs in DNA. A mutation in a coding sequence may alter the sequence and function of the protein product. A frameshift mutation changes the reading frame through insertions or deletions to produce an entirely novel product. A point mutation on the other hand alters only the amino acid represented by the codon in which the mutation exists. Mutations in regulatory regions might alter the expression of a gene. Mutations have been powerful tools for genetic studies as well as to study biological functions, for example, growth and development. The natural occurrence of mutations can be enhanced experimentally by applying agents referred to as mutagens. Researchers have used certain mutagens to induce mutations to study gene function and apply new traits for crop improvement.

Activity 1

For this activity, it may be helpful to refer to the genetic code chart.

For the following DNA sequence:

5′ ATG TTG GAG AAG GTT GAA ACT TTC 3′

Write the coding strand

3′ TAC AAG CTC TTC CAA CTT TGA AAG 5′

Write the mRNA sequence from the given DNA sequence

AUG UUC GAG AAG GUU GAA ACU UUG

Write the resulting peptide sequence

Met-Phe-Glu-Lys-Val-Glu-Thr-Leu

If the **third** triplet of the original DNA sequence is mutated to GA<u>A</u> to result in the following sequence

5′ ATG TTG GAA AAG GTT GAA ACT TTC 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAA AAG GUU GAA ACU UUG

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val-Glu-Thr-Leu

What type of coding region mutation is this? Give a reason for your answer.

1. Silent
2. Amino acid substitution
3. Frameshift
4. Nonsense

If the **sixth** triplet of the original DNA sequence is mutated to <u>T</u>AA to result in the following sequence

5′ ATG TTG GAG AAG GTT TAA ACT TTC 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAG AAG GUU UAA ACU UUG

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val

What type of coding region mutation is this? Give reason for your answer.

1. Silent
2. Amino acid substitution

3. Frameshift
4. Nonsense

The original DNA sequence is mutated to result in the following sequence

5′ ATG TTG GAG AAG GTT CGA AAC TTT C 3′

What will be the mRNA sequence from the mutated DNA sequence?

AUG UUC GAG AAG GUU CGA AAC UUU C

What will be the peptide sequence from the mutated DNA sequence?

Met-Phe-Glu-Lys-Val-Arg-Asn-Phe

What type of coding region mutation is this? Give a reason for your answer.

1. Silent
2. Amino acid substitution
3. Frameshift
4. Nonsense

Activity 2

1. A hypothetical plant gene encodes a protein with the following amino acid sequence:

Met-Leu-Lys-Thr-Phe-Val-Glu

A mutation of a single nucleotide alters the amino acid sequence to:

Met-Tyr-Asp-Pro-Gly-Ala-Lys-Ile-Arg

Describe the type of the mutation and the position of the codon that is affected.

2. For the following DNA sequence

3′ **A**TG GCC GGC AAT CAA CTA TAT TG**A** 5′

1 24 (nucleotide number)

a. Write the coding strand

b. Write the mRNA strand

c. Write the protein strand

d. Give the altered amino acid sequence of the protein for each of the following mutations:

i. A transition at nucleotide 11

ii. A transition at nucleotide 13

iii. A single nucleotide deletion at nucleotide 7

iv. A **T** to **A** change at nucleotide 15

v. An addition of ACC after nucleotide 6

vi. A transition at nucleotide 9

References

Doebley et al. 1997. The evolution of apical dominance in maize. Nature 386: 485-488.

Lertrat, K., Pulam, T. 2007. Breeding for increased sweetness in sweet corn. Int. J. Plant. Breed. 1: 27-30.

Nelson, O., Pan, D. 1995. Starch synthesis in maize endosperms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 475-496.

Peng, J., Richards, D.E., Hartley, N.M., Murphy, G.P., Devos, K.M., Flintman, J.E., Beales, J., Fish, L.J., Worland, A.J., Pelica, F., Sudhakar, D., Christou, P., Snape, J.W., Gale, M.D., Harberd, N.P. 1999. “Green revolution” genes encode mutant gibberellin

response modulators. Nature 400: 256-261.

This page titled 21.2: DNA Mutations is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Marjorie Hanneman, Walter Suza, Philip Becraft, do, & do (Iowa State University Digital Press) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.3: DNA Mutations** by do, Marjorie Hanneman, Philip Becraft, Walter Suza is licensed CC BY-NC-SA 4.0. Original source: https://iastate.pressbooks.pub/genagbiotech.

CHAPTER OVERVIEW

22: Cancer Genetics

22.1: Classification of Cancers
22.2: Cancer Cell Biology
22.3: Hallmarks of Cancer
22.4: Mutagens and Carcinogens
22.5: Oncogenes
22.6: Tumor Suppressor Genes
22.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

22: Cancer Genetics is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

22.1: Classification of Cancers

Cancers can be classified based on the tissues in which they originate. **Sarcomas** are cancers that originate in mesoderm tissues, such as bone or muscle, and cancers arising in glandular tissues (e.g. breast, prostate) are classified as **adenocarcinomas**. **Carcinomas** originate in epithelial cells (both inside the body and on its surface) and are the most common types of cancer (~85%). Each of these classifications may be further sub--divided. For example, **squamous cell carcinoma (SCC)**, **basal cell carcinoma (BCC)**, and **melanoma** are all types of skin cancers originating respectively in the squamous cells, basal cells, or melanocytes of the skin.

This page titled 22.1: Classification of Cancers is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.1: Classification of Cancers** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.2: Cancer Cell Biology

Cancer is a progressive disease that usually begins with increased frequency of cell division (Figure **22.2.2**). Under the microscope, this may be detectable as increased cellular and nuclear size, and an increased proportion of cells undergoing mitosis. As the disease progresses, cells typically lose their normal shape and tissue organization. Tissues with increased cell division and abnormal tissue organization exhibit **dysplasia**. Eventually a tumor develops, which can grow rapidly and expand into adjacent tissues.

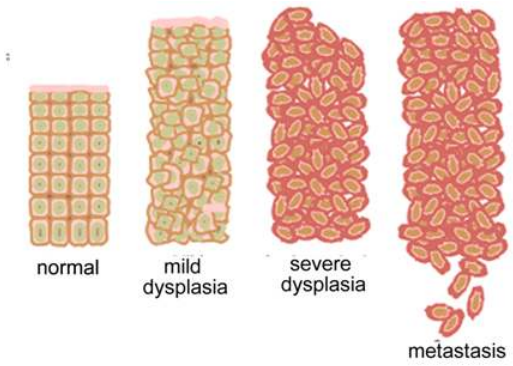


Figure **22.2.2**: Progressive increases in cell division and abnormal cell morphology associated with cancer. (Wikipedia-NIH-PD)

As cellular damage accumulates and additional control mechanisms are lost, some cells may break free of the primary tumor, pass into the blood or lymph system, and be transported to another organ, where they develop into new tumors (Figure **22.2.3**). The early detection of tumors is important so that they can be treated or removed before the onset of metastasis, but note that not all tumors will lead to cancer. Tumors that do not metastasize are classified as **benign,** and are not usually considered life threatening. In contrast, **malignant** tumors become invasive, and ultimately result in cancer.



Figure **22.2.3**: Secondary tumors (white) develop in the liver from cells of a metastatic pancreatic cancer. (Wikipedia-J. Hayman-PD)

This page titled 22.2: Cancer Cell Biology is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.2: Cancer Cell Biology** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.3: Hallmarks of Cancer

Researchers have identified six molecular and cellular traits that characterize most cancers. These six hallmarks of cancer are summarized in Table **22.3.1**. In this chapter, we will focus on the first two hallmarks, namely growth signal autonomy and insensitivity to anti--growth signals.

***Table* 22.3.1** *Ten Hallmarks of Cancer (Hanahan and Weinberg, 2000; Hanahan 2011)*

1. Growth signal autonomy
Cancer cells can divide without the external signals normally required to stimulate division.
2. Insensitivity to growth inhibitory signals
Cancer cells are unaffected by external signals that inhibit division of normal cells.
3. Evasion of apoptosis
When excessive DNA damage and other abnormalities are detected, apoptosis (a type of programmed cell death) is induced in normal cells, but not in cancer cells.
4. Reproductive potential not limited by telomeres
Each division of a normal cell reduces the length of its telomeres. Normal cells arrest further division once telomeres reach a certain length. Cancer cells avoid this arrest and/or maintain the length of their telomeres.
5. Sustained angiogenesis
Most cancers require the growth of new blood vessels into the tumor. Normal angiogenesis is regulated by both inhibitory and stimulatory signals not required in cancer cells.
6. Tissue invasion and metastasis
Normal cells generally do not migrate (except in embryo development). Cancer cells invade other tissues including vital organs.
7. Deregulated metabolic pathways
Cancer cells use an abnormal metabolism to satisfy a high demand for energy and nutrients.
8. Evasion of the immune system
Cancer cells are able to evade the immune system.
9. Chromosomal instability
Severe chromosomal abnormalities are found in most cancers.
10. Inflammation
Local chronic inflammation is associated with many types of cancer.

This page titled 22.3: Hallmarks of Cancer is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.3: Hallmarks of Cancer** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.4: Mutagens and Carcinogens

A **carcinogen** is any agent that directly increases the incidence of cancer. Most, but not all carcinogens are mutagens. Carcinogens that do not directly damage DNA include substances that accelerate cell division, thereby leaving less opportunity for cell to repair induced mutations, or errors in replication. Carcinogens that act as mutagens may be biological, physical, or chemical in nature, although the term is most often used in relation to chemical substances.

Human Papilloma Virus (**HPV**, Figure 22.4.4) is an example of a biological carcinogen. Almost all cervical cancers begin with infection by HPV, which contains genes that disrupt the normal pattern of cell division within the host cell. Any gene that leads to an uncontrolled increase in cell division is called an **oncogene**. The HPV E6 and E7 genes are considered oncogenes because they inhibit the host cell's natural tumor suppressing proteins (include p53, described below). The product of the E5 gene mimics the host's own signals for cell division, and these and other viral gene products may contribute to dysplasia, which is detected during a Pap smear (Figure 22.4.5). Detection of abnormal cell morphology in a Pap smear is not necessarily evidence of cancer. It must be emphasized again that cells have many regulatory mechanisms to limit division and growth, and for cancer to occur, each of these mechanisms must be disrupted. This is one reason why only a minority of individuals with HPV infections ultimately develop cancer. Although most HPV-related cancers are cervical, HPV infection can also lead to cancer in other tissues, in both women and men.

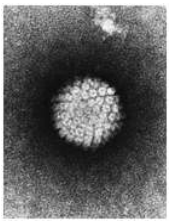


Figure* 22.4.4*: *Electron micrograph of HPV. (Wikipedia-Unknown-PD)*

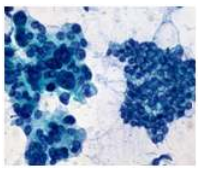


Figure* 22.4.5*: *Dysplastic (left) and normal (right) cells from a Pap smear. (Flickr-Ed Uthman-CC:AS)*

Radiation is a well-known physical carcinogen, because of its potential to induce DNA damage within the body. The most damaging type of radiation is **ionizing**, meaning waves or particles with sufficient energy to strip electrons from the molecules they encounter, including DNA or molecules that can subsequently react with DNA. Ionizing radiation, which includes x-rays, gamma rays, and some wavelengths of ultraviolet rays, is distinct from the non-ionizing radiation of microwave ovens, cell phones, and radios. As with other carcinogens, mutation of multiple, independent genes that normally regulate cell division is required before cancer develops.

Chemical carcinogens (Table 22.4.2) can be either natural or synthetic compounds that, based on animal feeding trials or **epidemiological** (i.e. human population) studies, increase the incidence of cancer. The definition of a chemical as a carcinogen is problematic for several reasons. Some chemicals become carcinogenic only after they are metabolized into another compound in the body; not all species or individuals may metabolize chemicals in the same way. Also, the carcinogenic properties of a compound are usually dependent on its dose. It can be difficult to define a relevant dose for both lab animals and humans.

Nevertheless, when a correlation between cancer incidence and chemical exposure is observed, it is usually possible to find ways to reduce exposure to that chemical.

***Table* 22.4.2:** *Some classes of chemical carcinogens (Pecorino 2008)*

1. PAHs (polycyclic aromatic hydrocarbons)
e.g. benzo[a]pyrene and several other components of the smoke of cigarettes, wood, and fossil fuels
2. Aromatic amines
e.g. formed in food when meat (including fish, poultry) are cooked at high temperature
3. Nitrosamines and nitrosamides
e.g. found in tobacco and in some smoked meat and fish
4. Azo dyes
e.g. various dyes and pigments used in textiles, leather, paints.
5. Carbamates
e.g. ethyl carbamate (urethane) found in some distilled beverages and fermented foods
6. Halogenated compounds
e.g. pentachlorophenol used in some wood preservatives and pesticides.
7. Inorganic compounds
e.g. asbestos; may induce chronic inflammation and reactive oxygen species
8. Miscellaneous compounds
e.g. alkylating agents, phenolics

This page titled 22.4: Mutagens and Carcinogens is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.4: Mutagens and Carcinogens** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.5: Oncogenes

The control of cell division involves many different genes. Some of these genes act as signaling molecules to activate normal progression through the cell cycle. One of the pre-requisites for cancer occurs when one or more of these activators of cell division become mutated.

The mutation may involve a change in the coding sequence of the protein, so that it is more active than normal, or a change in the regulation of its expression, so that it is produced at higher levels than normal, or persists in the cell longer than normal. Genes that are a part of the normal regulation of cell division, but which after mutation contribute to cancer, are called **proto-oncogenes**. Once a proto-oncogene has been abnormally activated by mutation, it is called an oncogene. More than 100 genes have been defined as proto-oncogenes. These include genes at almost every step of the signaling pathways that normally induce cell to divide, including growth factors, **receptors**, **signal transducers**, and transcription factors.

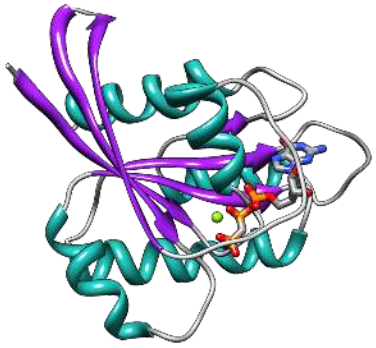


Figure 22.5.6: HRas structure PDB 121p, ribbon showing strands in purple, helices in aqua, loops in gray. Also shown are the bound GTP analog and magnesium ion. (CC-BY-SA ElaineMeng)

ras is an example of a proto-oncogene. *ras* acts as a switch within signal transduction pathways, including the regulation of cell division. When a receptor protein receives a signal for cell division, the receptor activates *ras*, which in turn activates other signaling components, ultimately leading to activation of genes involved in cell division. Certain mutations of the *ras* sequence causes it to be in a permanently active form, which can lead to constitutive activation of the cell cycle. This mutation is dominant as are most oncogenes. An example of the role of ras in relaying a signal for cell division in the EGF pathway is shown in Figure 22.5.7.

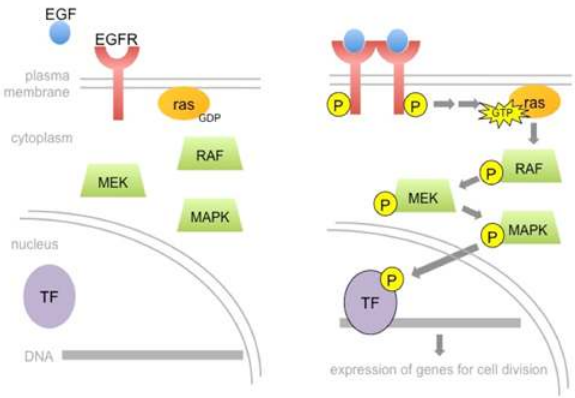


Figure 22.5.7: Simplified representation of the epidermal growth factor (EGF) signaling pathway. In the panel on the left, the components are shown in their inactive forms, prior to stimulation of the pathway. The components include the soluble ligand, EGF, its receptor (EGFR, a tyrosine kinase), ras (a G protein), several kinases (RAF, MEK, MAPK), and a transcription factor (TF). In the right panel, the activate pathway is shown. Binding of the ligand to its receptor leads to autophosphorylation of the receptor. Through a series of proteins not shown here, the phosphorpylated simulates conversion of ras to its active, GTP-bound form. The activated ras then stimulates phosphorylation of a series of kinases, which ultimately activate transcription factors and the expression of genes required for cell proliferation. (Original-Deyholos-CC:AN)

This page titled 22.5: Oncogenes is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.5: Oncogenes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.6: Tumor Suppressor Genes

More than 30 genes are classified as **tumor suppressors**. The normal functions of these genes include repair of DNA, induction of programmed cell death (**apoptosis**) and prevention of abnormal cell division. In contrast to proto-oncogenes, in tumor suppressors it is loss-of-function mutations that contribute to the progression of cancer. This means that tumor suppressor mutations tend to be recessive, and thus both alleles must be mutated in order to allow abnormal growth to proceed. It is perhaps not surprising that mutations in tumor suppressor genes, are more likely than oncogenes to be inherited. An example is the tumor suppressor gene, ***BRCA1***, which is involved in DNA-repair. Inherited mutations in *BRCA1* increase a woman's lifetime risk of breast cancer by up to seven times, although these heritable mutations account for only about 10% of breast cancer. Thus, sporadic rather than inherited mutations are the most common sources of both oncogenes and disabled tumor suppressor genes.

An important tumor suppressor gene is a transcription factor named **p53**. Other proteins in the cell sense DNA damage, or abnormalities in the cell cycle and activate p53 through several mechanisms including **phosphorylation** (attachment of phosphate to specific site on the protein) and transport into the nucleus. In its active form, p53 induces the transcription of genes with several different types of tumor suppressing functions, including DNA repair, cell cycle arrest, and apoptosis. Over 50% of human tumors contain mutations in p53. People who inherit only one function copy of p53 have a greatly increased incidence of early onset cancer. However, as with the other cancer related genes we have discussed, most mutations in p53 are sporadic, rather than inherited. Mutation of p53, through formation of pyrimidine dimers in the genes following exposure to UV light, has been causally linked to squamous cell and basal cell carcinomas (but not melanomas, highlighting the variety and complexities of mechanisms that can cause cancer).



Figure **22.6.8**: p53 bound to its target site on a DNA molecule. (Wikipedia-Thomas Spettstoesser from Cho etal, Science 265PP346, 1994-CC:AS)

This page titled 22.6: Tumor Suppressor Genes is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.6: Tumor Suppressor Genes** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

22.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)

Chronic myelogenous leukemia (CML)

Chronic myelogenous leukemia (**CML**) is a type of cancer of white blood cells, myeloid cells, that are mutated and proliferate uncontrollably through three stages (chronic, accelerated, and blast crisis) and lead eventually to death. Cytogenetics showed the myeloid cells of CML patients usually also have a consistent chromosome translocation (the mutant event) between the long arms of chromosomes 9 and 22, t(9:22)(q34;q11). It is also known as the **Philadelphia chromosome** (Ph^+). This translocation involves breaks in two genes, *c-abl* and *bcr*, on chromosomes 9 and 22, respectively. The fusion of the translocation breaks result in a chimeric gene, called ***bcr-abl***, that contains exons 1 and/or 2 from *bcr* (this varies from patient to patient) and 2-11 from *abl* and it produces a chimeric protein (BCR-ABL or $p185^{bcr\text{-}abl}$) that is transcribed like *bcr* and contains *abl* enzyme sequences. This chimeric protein has a tyrosine-kinase from the *abl* gene sequences that is unique to the CML mutant cell. The consistent, unregulated expression of this gene and its kinase product causes activation of a variety of intracellular signaling pathways, promoting the uncontrolled proliferative and survival properties of CML cells (the cancer). Thus the BCR-ABL tyrosine kinase enzyme exists only in cancer cells (and not in healthy cells) and a drug that inhibits this activity could be used to target and prevent the uncontrolled growth of the cancerous CML cells.

Inhibiting the Bcr-Abl tyrosine kinase activity

Knowing that the kinase activity was the key to treatment, pharmaceutical companies screened chemical libraries of potential kinase inhibitory compounds. After initially finding low potency inhibitors, a relationship between structure and activity suggested other compounds that were optimized to inhibit the BCR-ABL tyrosine kinase activity. The lead compound was STI571, now called **Gleevec**™ or imatinib (Figure **22.7.9**). This drug was shown to inhibit the BCR-ABL tyrosine kinase activity and to inhibit CML cell proliferation *in vitro* and *in vivo*. Gleevec™ works via targeted therapy—only the kinase activity in cancer cells was targeted and thereby killed through the drug's action. In this regard, Gleevec™ was one of the first cancer therapies to show the potential for this type of targeted action. It was dependent upon the genetic identification of the cause and protein target and is often cited as a paradigm for genetic research in cancer therapeutics.

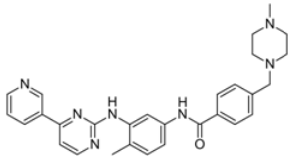


Figure* 22.7.9*: *Biochemical structure of Gleevec™ or Imatinib. . (Wikipedia-Fuse809-CC:AN)*

Caution

This is a simplified presentation of the CML/cancer targeting by the drug Gleevec™. There are many more details than could be presented here. It is represents as a model of finding a drug for each type of cancer, rather than the one, single “magic bullet” that kills all cancers. Remember, there are always complexities in this type of research to treatment process, such as patient genetic and environmental variation that leads to differences in drug metabolism, uptake, and binding. Also, changes in drug dose, mutation of the *bcr-abl* gene, and other events can affect the effectiveness of the treatment and the relapse rate. Biological systems are extremely complex and difficult to modulate in the specific, targeted manner necessary to treat cancer ideally.

Remember, the drug, Gleevec™, is not a cure, but only a treatment. It prevents the uncontrolled proliferation of the CML cells, but doesn’t kill them directly. The arrested cells will die eventually, but there is always a small pool of CML cells that will proliferate if the drug is discontinued. While sustained use of this expensive drug is beneficial to the pharmaceutical companies, it is certainly not the ideal situation for the patient.

This page titled 22.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib) is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **13.7: The “Poster Boy” of Genetic Research Leading to a Cancer Treatment – Gleevec™ (Imatinib)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

23: Genomics

23.1: ‘Omics Technologies

23.2: DNA Sequencing

23.3: Whole Genome Sequencing

Thumbnail: Fluorescently labeled products can be separated electrophoretically based on their length. (Original-Deyholos-CC:AN)

23: Genomics is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

23.1: ‘Omics Technologies

The complete set of DNA within an organism is called its **genome**. **Genomics** is therefore the large-scale description, using techniques of molecular biology of many genes or even whole genomes at once. This type of research is facilitated by technologies that increase throughput (i.e. rate of analysis), and decrease cost. The –**omics** suffix has been used to indicate high-throughput analysis of many types of molecules, including transcripts (**transcriptomics**), proteins (**proteomics**), and the products of enzymatic reactions, or metabolites (**metabolomics**; Figure **23.1.1**). Interpretation of the large data sets generated by –omics research depends on a combination of computational, biological, and statistical knowledge provided by experts in **bioinformatics**. Attempts to combine information from different types of –‘omics studies is sometimes called **systems biology**.



Figure **23.1.1**: A word cloud listing some of the different –omics technologies. Terms that are more widely used are written in the largest characters. There is no significance to the color of each word. (Original-Deyholos-CC:AN)

This page titled 23.1: ‘Omics Technologies is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **11.1: ‘Omics Technologies** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

23.2: DNA Sequencing

DNA sequencing determines the order of nucleotide bases within a given fragment of DNA. This information can be used to infer the RNA or protein sequence encoded by the gene, from which further inferences may be made about the gene's function and its relationship to other genes and gene products. DNA sequence information is also useful in studying the regulation of gene expression. If DNA sequencing is applied to the study of many genes, or even a whole genome, it is considered an example of genomics.

Dideoxy sequencing

Recall that DNA polymerases incorporate nucleotides (dNTPs) into a growing strand of DNA, based on the sequence of a template strand. DNA polymerases add a new base only to the 3'-OH group of an existing strand of DNA; this is why primers are required in natural DNA synthesis and in techniques such as PCR. Most of the currently used DNA sequencing techniques rely on the random incorporation of modified nucleotides called terminators. Examples of terminators are the **dideoxy nucleotides** (**ddNTPs**), which lack a 3'-OH group and therefore cannot serve as an attachment site for the addition of new bases to a growing strand of DNA (Figure **23.2.1**). After a ddNTP is incorporated into a strand of DNA, no further elongation can occur. Terminators are labeled with one of four fluorescent dyes, each specific for one the four nucleotide bases.

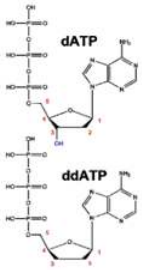


Figure **23.2.1**: *ddNTPs (Original-Deyholos-CC:AN)*

To sequence a DNA fragment, you need many copies of that fragment (Figure **23.2.2**). Unlike PCR, DNA sequencing does not amplify the target sequence and only one primer is used. This primer is hybridized to the denatured template DNA, and determines where on the template strand the sequencing reaction will begin. A mixture of dNTPs, fluorescently labeled terminators, and DNA polymerase is added to a tube containing the primer-template hybrid. The DNA polymerase will then synthesize a new strand of DNA until a fluorescently labeled nucleotide is incorporated, at which point extension is terminated. Because the reaction contains millions of template molecules, a corresponding number of shorter molecules is synthesized, each ending in a fluorescent label that corresponds to the last base incorporated.

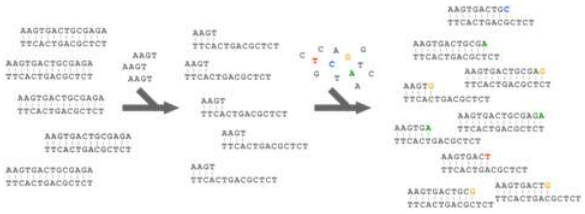


Figure **23.2.2**: A sequencing reactions begins with many identical copies of a template DNA fragment. The template is denatured, then primers are annealed to the template. Following the addition of polymerase, regular dNTPS, and fluorescently labeled terminators, extension begins at the primer site. Elongation proceeds until a fluorescently labeled terminator (shown here in color) is incorporated. (Original-Deyholos-CC:AN)

The newly synthesized strands can be denatured from the template, and then separated electrophoretically based on their length (Figure **23.2.3**). Since each band differs in length by one nucleotide, and the identity of that nucleotide is known from its fluorescence, the DNA sequence can be read simply from the order of the colors in successive bands. In practice, the maximum length of sequence that can be read from a single sequencing reaction is about 700 bp.

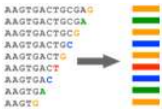


Figure **23.2.3**: Fluorescently labeled products can be separated electrophoretically based on their length. (Original-Deyholos-CC:AN)

A particularly sensitive electrophoresis method used in the analysis of DNA sequencing reactions is called capillary electrophoresis (Figure **23.2.6**). In this method, a current pulls the sequencing products through a gel-like matrix that is encased in a fine tube of clear plastic. As in conventional electrophoresis, the smallest fragments move through the capillary the fastest. As they pass through a point near the end of the capillary, the fluorescent intensity of each dye is read. This produces a graph called a chromatogram. The sequence is determined by identifying the highest peak (i.e. the dye with the most intense fluorescent signal) at each position.

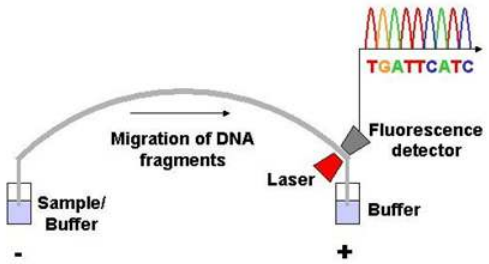


Figure **23.2.4**: Fluorescently labeled products can be separated by capillary electrophoresis, generating a chromatogram from which the sequence can be read.(Wikipedia-Abizar Lakdawalla-PD)

Next-generation sequencing

Advances in technology over the past two decades have increased the speed and quality of sequencing, while decreasing the cost. This has become especially true with the most recently developed methods called next-generation sequencing. Not all of these new methods rely on terminators, but one that does is a method used in instruments sold by a company called **Illumina**. Illumina sequencers use a special variant of PCR called bridge PCR to make many thousands of copies of a short (45bp) template fragment. Each of these short template fragments is attached in a cluster in a small spot on a reaction surface. Millions of other clusters, each made by different template fragment, are located at other positions on the reaction surface. DNA synthesis at each template strand then proceeds using dye-labeled terminators that are used are reversible. Synthesis is therefore terminated (temporarily) after the incorporation of each nucleotide. Thus, after the first nucleotide is incorporated in each strand, a camera records the color of fluorescence emitted from each cluster. The terminators are then modified, and a second nucleotide is incorporated in each strand, and again the reaction surface is photographed. This cycle is repeated a total of 45 times. Because millions of 45 bp templates are sequenced in parallel in a single process, Illumina sequencing is very efficient compared to other sequencing techniques. However, the short length of the templates currently limits the application of this technology.

This page titled 23.2: DNA Sequencing is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **11.2: DNA Sequencing** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

23.3: Whole Genome Sequencing

The need for assembly

Given that the length of a single, individual sequencing read is somewhere between 45bp and 700bp, we are faced with a problem determining the sequence of longer fragments, such as the chromosomes in an entire genome of humans (3×10^9 bp). Obviously, we need to break the genome into smaller fragments. There are two different strategies for doing this:

1. clone-by-clone sequencing, which relies on the creation of a physical map first then sequencing, and
2. whole genome shotgun sequencing, which sequences first and does not require a physical map.

Physical mapping

A physical map is a representation of a genome, comprised of cloned fragments of DNA. The map is therefore made from physical entities (pieces of DNA) rather than abstract concepts such as the linkage frequencies and genes that make up a genetic map (Figure **23.3.1**). It is usually possible to correlate genetic and physical maps, for example by identifying the clone that contains a particular molecular marker. The connection between physical and genetic maps allows the genes underlying particular mutations to be identified through a process call map-based cloning.

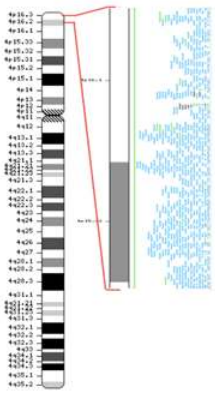


Figure **23.3.1**: A portion of the physical map for human chromosome 4. The entire chromosome is shown at left. The physical map is made up of small blue lines, each of which represents a cloned piece of DNA approximately 100kb in length. (NCBI-unknown-PD)

To create a physical map, large fragments of the genome are cloned into plasmid vectors, or into larger vectors called bacterial artificial chromosomes (BACs). BACs can contain approximately 100kb fragments. The set of BACs produced in a cloning reaction will be redundant, meaning that different clones will contain DNA from the same part of the genome. Because of this redundancy, it is useful to select the minimum set of clones that represent the entire genome, and to order these clones respective to the sequence of the original chromosome. Note that this is all to be done without knowing the complete sequence of each BAC. Making a physical map may therefore rely on techniques related to Southern blotting: DNA from the ends of one BAC is used as a probe to find clones that contain the same sequence. These clones are then assumed to overlap each other. A set of overlapping clones is called a **contig**.

Clone-by-clone sequencing

Physical mapping of cloned sequences was once considered a pre-requisite for genome sequencing. The process would begin by breaking the genome into BAC-sized pieces, arranging these BACs into a map, then breaking each BAC up into a series of smaller clones, which were usually then also mapped. Eventually, a minimum set of smaller clones would be identified, each of which was small enough to be sequenced (Figure **23.3.8**). Because the order of clones relative to the complete chromosome was known prior to sequencing, the resulting sequence information could be easily assembled into one complete chromosome at the end of the project. Clone-by-clone sequencing therefore minimizes the number of sequencing reactions that must be performed, and makes sequence assembly straightforward and reliable. However, a drawback of this strategy is the tedious process of building physical map prior to any sequencing.

Whole genome shotgun sequencing

This strategy breaks the genome into fragments that are small enough to be sequenced, then reassembles them simply by looking for overlaps in the sequence of each fragment. It avoids the laborious process of making a physical map (Figure **23.3.2**). However, it requires many more sequencing reactions than the clone-by-clone method, because, in the shotgun approach, there is no way to avoid sequencing redundant fragments. There is also a question of the feasibility of assembling complete chromosomes based simply on the sequence overlaps of many small fragments. This is particularly a problem when the size of the fragments is smaller than the length of a repetitive region of DNA. Nevertheless, this method has now been successfully demonstrated in the nearly complete sequencing of many large genomes (rice, human, and many others). It is the current standard methodology.

However, shotgun assemblies are rarely able to complete entire genomes. The human genome, for example, relied on a combination of shotgun sequence and physical mapping to produce contiguous sequence for the length of each arm of each chromosome. Note that because of the highly repetitive nature of centromeric and telomeric DNA, sequencing projects rarely include these heterochromatic, gene poor regions.

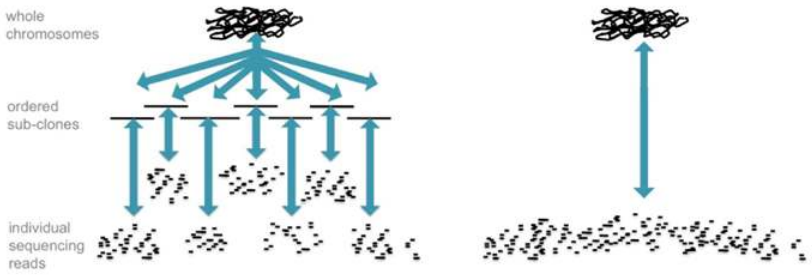


Figure **23.3.2**: Genome sequencing strategies. A clone-by-clone strategy (left) in which the genome is divided into progressively smaller units (clones) before sequencing. Whole genome shotgun strategy (right) assembles the sequence from all the smaller reads. (Origianl-Deyholos-CC:AN)

Genome analysis

An assembled genome is a string of millions of A's,C's,G's,T's. Which of these represent nucleotides that encode proteins, and which of these represent other features of genes and their regulatory elements? The process of **genome annotation** relies on computers to define features such a start and stop codons, introns, exons, and splice sites. However, few of the predictions made by these programs is entirely accurate, and most must be verified experimentally for any gene of particular importance or interest.

This page titled 23.3: Whole Genome Sequencing is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **11.3: Whole Genome Sequencing** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

24: Functional Genomics

24.1: Functional Genomics – Determining Function(s)

24.2: Genomic Approaches- The DNA Microarray

Thumbnail: A DNA microarray (CC BY-SA 3.0 Unported; Guillaume Paumier via Wikipedia)

24: Functional Genomics is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

24.1: Functional Genomics – Determining Function(s)

Having identified putative genes within a genome sequence, how do we determine their function? Techniques of functional genomics are an experimental approach to address this question. One widely used technique in functional genomics is called **microarray analysis** (Figure **24.1.1**). Microarrays can measure the abundance of mRNA for hundreds or thousands of genes at once. The abundance of mRNA of a particular gene is usually correlated with the activity of that gene. For example, genes that are involved in neuronal development likely produce more mRNA in brain tissue than in heart tissue. We can therefore learn about the relationship between particular genes and particular processes by comparing transcript abundance under different conditions. This can identify tissue specific expression (e.g. the nerve/heart example above), as well as differences in temporal expression (development), or exposure to external agents (eg, disease, hormones, drugs, etc.).

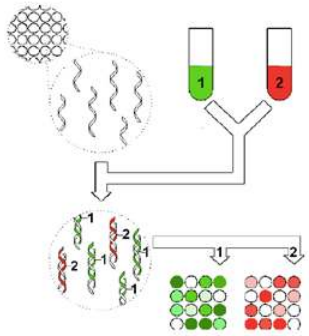


Figure **24.1.1**: An example of a type of DNA microarray. Fluorescently labeled molecules derived from the transcripts of two tissue samples are hybridized to immobilized DNA molecules on the surface of an array. The labeled molecules bind in proportion to their abundance in the original tissue samples. The amount of green or red fluorescence at each spot can be measured using filters for the appropriate wavelengths. Thus, spots on the micoarray that are more green than red represent genes that are more abundant in the tissue sample from which green-labeled molecules were derived. (Original-Deyholos-CC:AN)

This page titled 24.1: Functional Genomics – Determining Function(s) is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

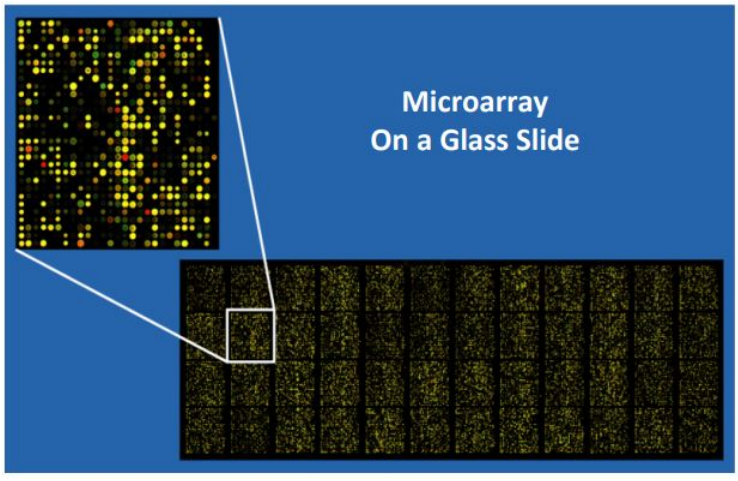
- **11.4: Functional Genomics – Determining Function(s)** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

24.2: Genomic Approaches- The DNA Microarray

Traditionally, when cellular levels of a protein were known to change in response to a chemical effector, molecular studies focused on control of the transcription of its gene. These studies often revealed that the control of gene expression was at the level of transcription, turning a gene on or off through interactions of transcription factors with DNA. However, protein levels are also controlled post-transcriptionally, by regulating the rate of mRNA translation or degradation. Studies of transcriptional and posttranscriptional regulation mechanisms are seminal to our understanding of how the correct protein is made in the right amounts at the right time.

We may have suspected, but now know that control of gene expression and cellular responses can be more complex than increasing or decreasing the transcription of a single gene or translation of a single protein. Whole genome sequences and new techniques make possible the study of the expression of virtually all genes in a cell at the same time, a field of investigation called **genomics**. Genomic studies reveal networks of regulated genes that must be understood to more fully explain the developmental and physiological changes in an organism. When you can 'see' all of the RNAs being transcribed from active genes in a cell, you are looking at a cell's **transcriptome**. By analogy to genomics, **transcriptomics** defines studies of 'webs' of interactive RNAs. Again, by analogy to genomics and transcriptomics, the broad study of active and inactive proteins in cells or tissues, how they are modified (processed) before use and how they interact is called **proteomics**. The technologies applied to proteomic studies include protein microarrays, immunochemical techniques and others uniquely suited to protein analysis (click Proteomics Techniques-Wikipedia for more information). *Protein Microarrays* are increasingly being used to identify protein-protein interactions, as well as the different states of proteins under different cellular conditions. Read even more about these exciting developments and their impact on basic and clinical research at Protein Microarrays from ncbi.

Finally think about this: creating a proteomic library analogous to a genomic library would seem a daunting prospect. But efforts are underway. Check out A stab at mapping the Human Proteome for original research leading to the sampling of a tissue-specific human proteome, and click Strategies for Approaching the Proteome for more general information. Let's look at some uses of DNA microarrays. This technology involves 'spotting' DNA (e.g., cloned DNA from a genomic or cDNA library, PCR products, oligonucleotides…) on a glass slide, or chip. In the language of microarray analysis, the slides are the probes. Spotting a chip is a robotic process. Because the DNA spots are microscopic, a cellspecific transcriptome (cDNA library) can fit on a single chip. A small genome microarray might also fit on a single chip, while larger genomes might need several slides. A primary use of DNA microarrays is **transcriptional profiling**. A genomic microarray can probe a mixture of fluorescently tagged target cDNAs made from mRNAs, in order to identify many (if not all) of the genes expressed in the cells at a given moment (i.e., its transcriptome). cDNA microarray probes can also probe quantitative differences in gene expression in cells or tissues during normal differentiation or in response to chemical signals. They are also valuable for genotyping, (i.e. characterizing the genes in an organism). Microarrays are so sensitive that they can even distinguish between two genes or regions of DNA that differ by a single nucleotide. Click Single Nucleotide Polymorphisms, or SNPs to learn more. In the microarray below, each colored spot (red, yellow, green) is a different fluorescently tagged molecule hybridizing to target sequences on the microarray. In the fluorescence microscope, the spots fluoresce different colors in response to UV light.



With quantitative microarray methods, the brightness (intensity) of the signal from each probe can be measured. In this way, we can compare the relative amounts of cDNA (and thus, different RNAs) in the *transcriptome* of different tissues or resulting from different tissue treatments. A table of different applications of microarrays (adapted from Wikipedia) is shown on the next page.

Application of Technology	Synopsis
Gene Expression Profiling	In a transcription (mRNA or gene expression) profiling experiment the expression levels of thousands of genes are simultaneously monitored to study the effects of certain treatments, diseases, and developmental stages on gene expression.
Comparative genomic hybridization	Assessing genome content in different cells or closely related organisms, where one organism's genome is the probe for a target genome from a different species.
GeneID	Small microarrays to check IDs of organisms in food and feed for genetically modified organisms (GMOs), mycoplasmas in cell culture, or pathogens for disease detection. These detection protocols often combine PCR and microarray technology.
CHIP; Chromatin immunoprecipitation	DNA sequences bound to a particular protein can be isolated by immunoprecipitating the protein. The fragments can be hybridized to a microarray (such as a tiling array) allowing the determination of protein binding site occupancy throughout the genome.
DamID	Analogously to ChIP, genomic regions bound by a protein of interest can be isolated and used to probe a microarray to determine binding site occupancy. Unlike ChIP, DamID does not require antibodies but makes use of adenine methylation near the protein's binding sites to selectively amplify those regions, introduced by expressing minute amounts of protein of interest fused to bacterial DNA adenine methyltransferase.
SNP detection	Identifying single nucleotide polymorphism among alleles within or between populations. Some microarray applications make use of SNP detection, including Genotyping, forensic analysis, measuring predisposition to disease, identifying drug-candidates, evaluating **germline** mutations in individuals or **somatic** mutations in cancers, assessing loss of heterozygosity, or genetic linkage analysis.
Alternative splicing protection	An exon junction array design uses probes specific to the expected or potential splice sites of predicted exons for a gene. It is of intermediate density, or coverage, to a typical gene expression array (with 1-3 probes per gene) and a genomic tiling array (with hundreds or thousands of probes per gene). It is used to assay the expression of alternative splice forms of a gene. Exon arrays have a different design, employing probes designed to detect each individual exon for known or predicted genes, and can be used for detecting different splicing isoforms
Tiling array	Genome tiling arrays consist of overlapping probes designed to densely represent a genomic region of interest, sometimes as large as an entire human chromosome. The purpose is to empirically detect expression of transcripts or alternatively spliced forms which may not have been previously known or predicted.



The Power of Microarrays. https://youtu.be/88rzbpclscM

If you like world records, check out the salamander with the largest genome, 10X bigger than our own: The HUGE Axolotl Genome. What do they do with all that DNA? And can our current technologies figure it out? For the original report, click on the following link: here.

This page titled 24.2: Genomic Approaches- The DNA Microarray is shared under a CC BY license and was authored, remixed, and/or curated by Gerald Bergtrom.

- **15.6: Genomic Approaches- The DNA Microarray** by Gerald Bergtrom is licensed CC BY 4.0. Original source: https://dc.uwm.edu/biosci_facbooks_bergtrom/10/.

CHAPTER OVERVIEW

25: Genetic Linkage

25.1: Linkage
25.2: Recombination
25.3: Linkage Reduces Recombination Frequency
25.4: Crossovers Allow Recombination of Linked Loci
25.5: Inferring Recombination From Genetic Data
25.6: Genetic Mapping

25: Genetic Linkage is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

25.1: Linkage

As we learned in Chapter 6, Mendel reported that the pairs of loci he observed behaved independently of each other; for example, the segregation of seed color alleles was independent from the segregation of alleles for seed shape. This observation was the basis for his Second Law (Independent Assortment), and contributed greatly to our understanding of heredity. However, further research showed that Mendel's Second Law did not apply to every pair of genes that could be studied. In fact, we now know that alleles of loci that are located close together on the same chromosome tend to be inherited together. This phenomenon is called **linkage**, and is a major exception to Mendel's **Second Law of Independent Assortment.** Researchers use linkage to determine the location of genes along chromosomes in a process called genetic mapping. The concept of gene linkage is important to the natural processes of heredity and evolution.

This page titled 25.1: Linkage is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.1: Linkage** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

25.2: Recombination

The term “recombination” is used in several different contexts in genetics. In reference to heredity, **recombination** is defined as any process that results in gametes with combinations of alleles that were not present in the gametes of a previous generation (see Figure **25.2.2**). **Interchromosomal recombination** occurs either through **independent assortment** of alleles whose loci are on different chromosomes (Chapter 6). **Intrachromosomal recombination** occurs through **crossovers** between loci on the same chromosomes (as described below). It is important to remember that in both of these cases, recombination is a process that occurs during meiosis (mitotic recombination may also occur in some species, but it is relatively rare). If meiosis results in recombination, the products are said to have a **recombinant genotype**. On the other hand, if no recombination occurs during meiosis, the products have their original combinations and are said to have a non-recombinant, or **parental genotype**. Recombination is important because it contributes to the genetic variation that may be observed between individuals within a population and acted upon by selection to produce evolution.

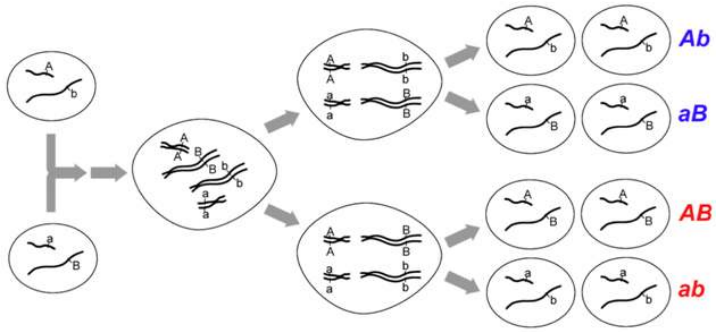


Figure **25.2.2**: When two loci are on non-homologous chromosomes, their alleles will segregate in combinations identical to those present in the parental gametes (Ab, aB), and in recombinant genotypes (AB, ab) that are different from the parental gametes. (Original-Deyholos-CC:AN)

As an example of interchromosomal recombination, consider loci on two different chromosomes as shown in Figure **25.2.2**. We know that if these loci are on different chromosomes, there are no physical connections between them, so they are **unlinked** and will segregate independently as did Mendel’s traits. The segregation depends on the relative orientation of each pair of chromosomes at metaphase. Since the orientation is random and independent of other chromosomes, each of the arrangements (and their meiotic products) is equally possible for two unlinked loci as shown in Figure **25.2.2**. More precisely, there is a 50% probability for recombinant genotypes, and a 50% probability for parental genotypes within the gametes produced by a meiocyte with unlinked loci. Indeed, if we examined all of the gametes that could be produced by this individual (which are the products of multiple independent meioses), we would note that approximately 50% of the gametes would be recombinant, and 50% would be parental. **Recombination frequency** (RF) is simply the number of recombinant gametes, divided by the total number of gametes. A frequency of approximately 50% recombination is therefore a defining characteristic of unlinked loci. Thus the greatest recombinant frequency expected is ~50%.

This page titled 25.2: Recombination is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: Recombination** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

25.3: Linkage Reduces Recombination Frequency

Having considered unlinked loci above, let us turn to the opposite situation, in which two loci are so close together on a chromosome that the parental combinations of alleles always segregate together (Figure **25.3.3**). This is because during meiosis they are so close that there are no crossover events between the two loci and the alleles at the two loci are physically attached on the same chromatid and so they always segregate together into the same gamete. In this case, no recombinants will be present following meiosis, and the recombination frequency will be 0%. This is **complete** (or **absolute**) **linkage** and is rare, as the loci must be so close together that crossovers are never detected between them.

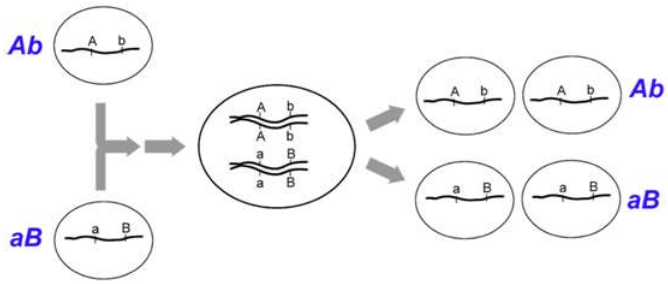


Figure **25.3.3**: If two loci are completely linked, their alleles will segregate in combinations identical to those present in the parental gametes (Ab, aB). No recombinants will be observed. (Original-Deyholos-CC:AN)

This page titled 25.3: Linkage Reduces Recombination Frequency is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.3: Linkage Reduces Recombination Frequency** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

25.4: Crossovers Allow Recombination of Linked Loci

Thus far, we have only considered situations with either no linkage (50% recombination) or complete linkage (0% recombination). It is also possible to obtain recombination frequencies between 0% and 50%, which is a situation we call **incomplete** (or **partial**) **linkage**. Incomplete linkage occurs when two loci are located on the same chromosome but the loci are far enough apart so that crossovers occur between them during some, but not all, meioses. Genes that are on the same chromosome are said to be **syntenic** regardless of whether they are completely or incompletely linked. All linked genes are syntenic, but not all syntenic genes are linked, as we will learn later.

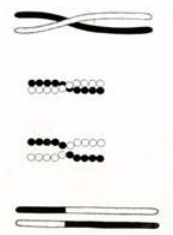


Figure **25.4.4**: A depiction of a crossover from Morgan's 1916 manuscript. Only one pair of non-sister chromatids is shown, but remember there are 4 chromatids for each homologous pair in meiosis. (Wikipedia-Morgan-PD)

Crossovers occur during prophase I of meiosis, when pairs of homologous chromosomes have aligned with each other in a process called **synapsis**. Crossing over begins with the breakage of DNA of a pair of non-sister chromatids. The breaks occur at corresponding positions on two non-sister chromatids, and then the ends of non-sister chromatids are connected to each other resulting in a reciprocal exchange of double-stranded DNA (Figure **25.4.4**). Generally every pair of chromosomes has at least one (and often more) crossovers during meioses (Figure **25.4.5**).

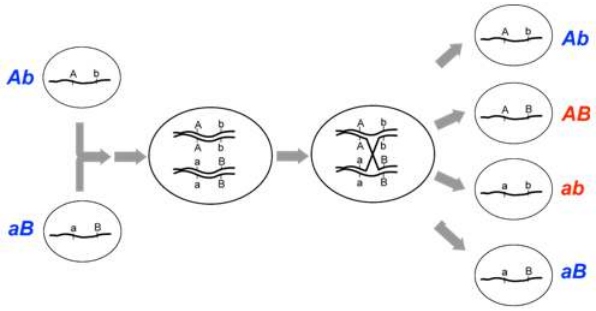


Figure* 25.4.5: A crossover between two linked loci can generate recombinant genotypes (AB, ab), from the chromatids involved in the crossover. Remember that multiple, independent meioses occur in each organism, so this particular pattern of recombination will not be observed among all the meioses from this individual. (Original-Deyholos-CC:AN)*

Because the location of crossovers is essentially random along the chromosome, the greater the distance between two loci, the more likely a crossover will occur between them. Furthermore, loci that are on the same chromosome, but are sufficiently far apart from each other, will on average have multiple crossovers between them and they will behave as though they are completely unlinked. A recombination frequency of 50% is therefore the maximum recombination frequency that can be observed, and is indicative of loci that are either on separate chromosomes, or are located very far apart on the same chromosome.

This page titled 25.4: Crossovers Allow Recombination of Linked Loci is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Crossovers Allow Recombination of Linked Loci** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

25.5: Inferring Recombination From Genetic Data

In the preceding examples, we had the advantage of knowing the approximate chromosomal positions of each allele involved, before we calculated the recombination frequencies. Knowing this information beforehand made it relatively easy to define the parental and recombinant genotypes, and to calculate recombination frequencies. However, in most experiments, we cannot directly examine the chromosomes, or even the gametes, so we must infer the arrangement of alleles from the phenotypes over two or more generations. Importantly, it is generally not sufficient to know the genotype of individuals in just one generation; for example, given an individual with the genotype *AaBb*, we do not know from the genotype alone whether the loci are located on the same chromosome, and if so, whether the arrangement of alleles on each chromosome is *AB* and *ab* or *Ab* and *aB* (Figure **25.5.6**). The top cell has the two dominant alleles together and the two recessive alleles together and is said to have the genes in the **coupling** (or **cis**) **configuration**. The alternative shown in the cell below is that the genes are in the **repulsion** (or **trans**) **configuration**.

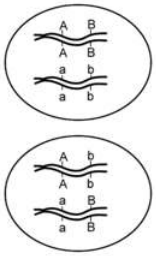


Figure **25.5.6**: Alleles in coupling configuration (top) or repulsion configuration (bottom). (Original-Deyholos-CC:AN)

Fortunately for geneticists, the arrangement of alleles can sometimes be inferred if the genotypes of a previous generation are known. For example, if the parents of *AaBb* had genotypes *AA*BB and *aabb* respectively, then the parental gametes that fused to produce *AaBb* would have been genotype *AB* and genotype *ab*. Therefore, prior to meiosis in the dihybrid, the arrangement of alleles would likewise be *AB* and *ab* (Figure **25.5.7**). Conversely, if the parents of *AaBb* had genotypes *aaBB* and *AAbb*, then the arrangement of alleles on the chromosomes of the dihybrid would be *aB* and *Ab*. Thus, the genotype of the previous generation can determine which of an individual's gametes are considered recombinant, and which are considered parental.

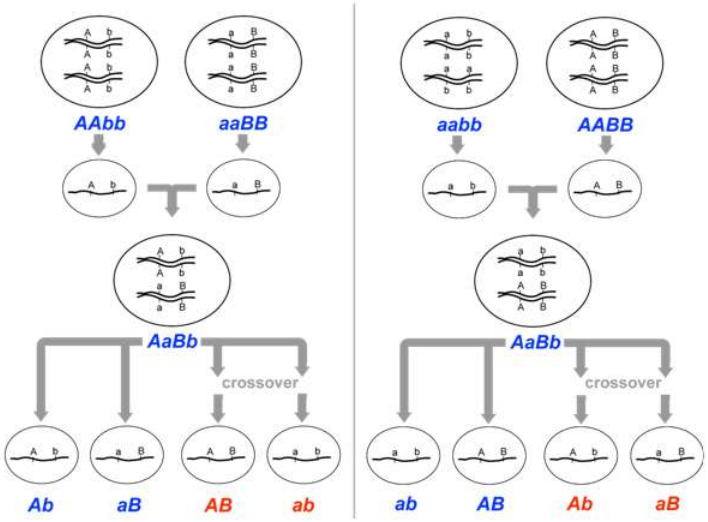


Figure **25.5.7**: The genotype of gametes can be inferred unambiguously if the gametes are produced by homozygotes. However, recombination frequencies can only be measured among the progeny of heterozygotes (i.e. dihybrids). Note that the dihybrid on the left contains a different configuration of alleles than the dihybrid on the rightdue to differences in the genotypes of their respective parents. Therefore, different gametes are defined as recombinant and parental among the progeny of the two dihybrids. In the cross at left, the recombinant gametes will be genotype AB and ab, and in the cross on the right, the recombinant gametes will be Ab and aB. (Original-Deyholos-CC:AN)

Let us now consider a complete experiment in which our objective is to measure recombination frequency (Figure **25.5.8**). We need at least two alleles for each of two genes, and we must know which combinations of alleles were present in the parental gametes. The simplest way to do this is to start with pure-breeding lines that have contrasting alleles at two loci. For example, we could

cross short-tailed mice, brown mice (*aaBB*) with long-tailed, white mice (*AAbb*). Based on the genotypes of the parents, we know that the parental gametes will be *aB* or *Ab* (but not *ab* or *AB*), and all of the progeny will be dihybrids, *AaBb*. We do not know at this point whether the two loci are on different pairs of homologous chromosomes, or whether they are on the same chromosome, and if so, how close together they are.

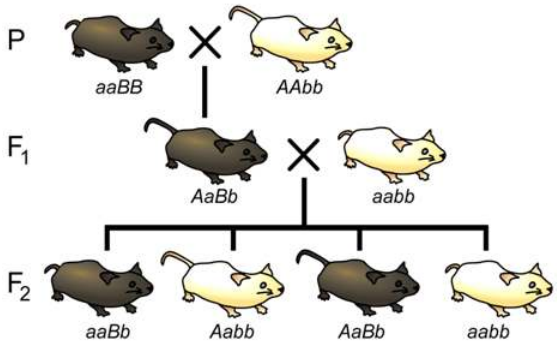


Figure **25.5.8**: An experiment to measure recombination frequency between two loci. The loci affect coat color (B/b) and tail length (A/a). (Wikipedia-Modified Deyholos-CC:AN)

The recombination events that may be detected will occur during meiosis in the dihybrid individual. If the loci are completely or partially linked, then prior to meiosis, alleles *aB* will be located on one chromosome, and alleles *Ab* will be on the other chromosome (based on our knowledge of the genotypes of the gametes that produced the dihybrid). Thus, recombinant gametes produced by the dihybrid will have the genotypes *ab* or *AB*, and non-recombinant (i.e. parental) gametes will have the genotypes *aB* or *Ab*.

How do we determine the genotype of the gametes produced by the dihybrid individual? The most practical method is to use a testcross (Figure **25.5.8**), in other words to mate *AaBb* to an individual that has only recessive alleles at both loci (*aabb*). This will give a different phenotype in the F_2 generation for each of the four possible combinations of alleles in the gametes of the dihybrid. We can then infer unambiguously the genotype of the gametes produced by the dihybrid individual, and therefore calculate the recombination frequency between these two loci. For example, if only two phenotypic classes were observed in the F_2 (i.e. short tails and brown fur (*aaBb*), and white fur with long tails (*Aabb*) we would know that the only gametes produced following meiosis of the dihybrid individual were of the parental type: *aB* and *Ab*, and the recombination frequency would therefore be 0%. Alternatively, we may observe multiple classes of phenotypes in the F_2 in ratios such as shown in Table **25.5.1**:

Table **25.5.1**: An example of quantitative data that may be observed in a genetic mapping experiment involving two loci. The data correspond to the F_2 generation in the cross shown in Figure **25.5.8**.

tail phenotype	fur phenotype	number of progeny	gamete from dihybrid	genotype of F_2 from test cross	(P)arental or (R)ecombinant
short	brown	48	*aB*	*aaBb*	P
long	white	42	*Ab*	*Aabb*	P
short	white	13	*ab*	*aabb*	R
long	brown	17	*AB*	*AaBb*	R

Given the data in Table **25.5.1**, the calculation of recombination frequency is straightforward:

$$\text{recombination frequency} = \frac{\text{number of recombinant gametes}}{\text{total number of gametes scored}} \tag{25.5.1}$$

$$\text{R.F.} = \frac{13+17}{48+42+13+17} \tag{25.5.2}$$

$$= 25\% \tag{25.5.3}$$

This page titled 25.5: Inferring Recombination From Genetic Data is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.5: Inferring Recombination From Genetic Data** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

25.6: Genetic Mapping

Because the frequency of recombination between two loci (up to 50%) is roughly proportional to the chromosomal distance between them, we can use recombination frequencies to produce genetic maps of all the loci along a chromosome and ultimately in the whole genome. The units of genetic distance are called **map units** (mu) or **centiMorgans** (cM), in honor of **Thomas Hunt Morgan** by his student, **Alfred Sturtevant**, who developed the concept. Geneticists routinely convert recombination frequencies into cM: the recombination frequency in percent is approximately the same as the map distance in cM. For example, if two loci have a recombination frequency of 25% they are said to be ~25cM apart on a chromosome (Figure **25.6.9**). Note: this approximation works well for small distances (RF<30%) but progressively fails at longer distances because the RF reaches a maximum at 50%. Some chromosomes are >100 cM long but loci at the tips only have an RF of 50%. The method for mapping of these long chromosomes is shown below.

Note that the map distance of two loci alone does not tell us anything about the orientation of these loci relative to other features, such as centromeres or telomeres, on the chromosome.

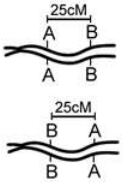


Figure **25.6.9**: Two genetic maps consistent with a recombination frequency of 25% between A and B. Note the location of the centromere. (Original-Deyholos-CC:AN)

Map distances are always calculated for one pair of loci at a time. However, by combining the results of multiple pairwise calculations, a **genetic map** of many loci on a chromosome can be produced (Figure **25.6.10**). A genetic map shows the map distance, in cM, that separates any two loci, and the position of these loci relative to all other mapped loci. The genetic map distance is roughly proportional to the physical distance, i.e. the amount of DNA between two loci. For example, in *Arabidopsis*, 1.0 cM corresponds to approximately 150,000bp and contains approximately 50 genes. The exact number of DNA bases in a cM depends on the organism, and on the particular position in the chromosome; some parts of chromosomes ("crossover hot spots") have higher rates of recombination than others, while other regions have reduced crossing over and often correspond to large regions of heterochromatin.

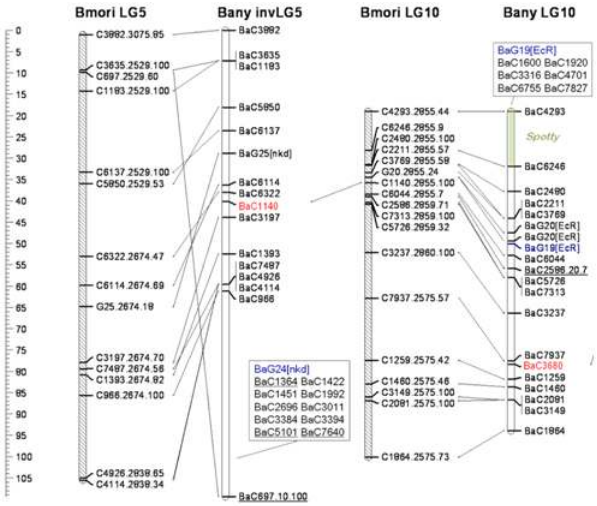


Figure **25.6.10**: Genetic maps for regions of two chromosomes from two species of the moth, Bombyx. The scale at left shows distance in cM, and the position of various loci is indicated on each chromosome. Diagonal lines connecting loci on different chromosomes show the position of corresponding loci in different species. This is referred to as regions of **conserved synteny**. *(NCBI-NIH-PD)*

When a novel gene or locus is identified by mutation or polymorphism, its approximate position on a chromosome can be determined by crossing it with previously mapped genes, and then calculating the recombination frequency. If the novel gene and

the previously mapped genes show complete or partial linkage, the recombination frequency will indicate the approximate position of the novel gene within the genetic map. This information is useful in isolating (i.e. cloning) the specific fragment of DNA that encodes the novel gene, through a process called **map-based cloning**.

Genetic maps are also useful to track genes/alleles in breeding crops and animals, in studying evolutionary relationships between species, and in determining the causes and individual susceptibility of some human diseases.

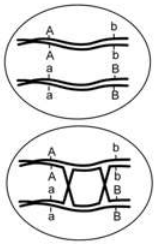


Figure **25.6.11**: A double crossover between two loci will produce gametes with parental genotypes. (Original-Deyholos-CC:AN)

Genetic maps are useful for showing the order of loci along a chromosome, but the distances are only an approximation. The correlation between recombination frequency and actual chromosomal distance is more accurate for short distances (low RF values) than long distances. Observed recombination frequencies between two relatively distant markers tend to underestimate the actual number of crossovers that occurred. This is because as the distance between loci increases, so does the possibility of having a second (or more) crossovers occur between the loci. This is a problem for geneticists, because with respect to the loci being studied, these **double-crossovers** produce gametes with the same genotypes as if no recombination events had occurred (Figure **25.6.11**) – they have parental genotypes. Thus a double crossover will appear to be a parental type and not be counted as a recombinant, despite having two (or more) crossovers. Geneticists will sometimes use specific mathematical formulae to adjust large recombination frequencies to account for the possibility of multiple crossovers and thus get a better estimate of the actual distance between two loci.

This page titled 25.6: Genetic Mapping is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.6: Genetic Mapping** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

CHAPTER OVERVIEW

26: Genetic Maps

26.1: Mapping With Three-Point Crosses

Thubmail: A three point cross for loci affecting tail length, fur color, and whisker length. (Original-Modified Deyholos-CC:AN)

26: Genetic Maps is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

26.1: Mapping With Three-Point Crosses

A particularly efficient method of mapping three genes at once is the **three-point cross**, which allows the order and distance between three potentially linked genes to be determined in a single cross experiment (Figure **26.1.12**). This is particularly useful when mapping a new mutation with an unknown location to two previously mapped loci. The basic strategy is the same as for the dihybrid mapping experiment; pure breeding lines with contrasting genotypes are crossed to produce an individual heterozygous at three loci (a trihybrid), which is then testcrossed to determine the recombination frequency between each pair of genes.

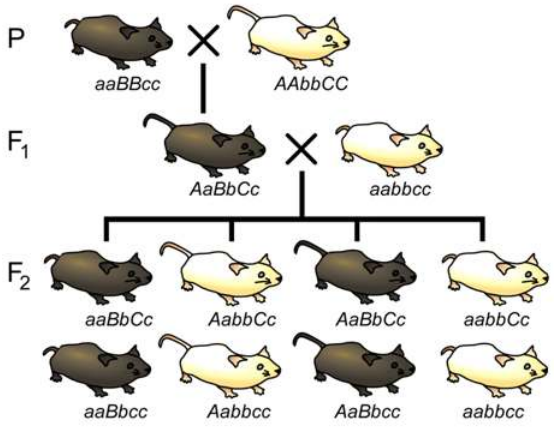


Figure **26.1.12**: A three point cross for loci affecting tail length, fur color, and whisker length. (Original-Modified Deyholos-CC:AN)

One useful feature of the three-point cross is that the order of the loci relative to each other can usually be determined by a simple visual inspection of the F_2 segregation data. If the genes are linked, there will often be two phenotypic classes that are much more infrequent than any of the others. In these cases, the rare phenotypic classes are usually those that arose from two crossover events, in which the locus in the middle is flanked by a crossover on either side of it. Thus, among the two rarest recombinant phenotypic classes, the one allele that differs from the other two alleles relative to the parental genotypes likely represents the locus that is in the middle of the other two loci. For example, based on the phenotypes of the pure-breeding parents in Figure **26.1.12**, the parental genotypes are *aBC* and *AbC* (remember the order of the loci is unknown, and it is not necessarily the alphabetical order in which we wrote the genotypes). Because we can deduce from the outcome of the testcross (Table **26.1.2**) that the rarest genotypes were *abC* and *ABc*, we can conclude that locus *A* that is most likely located between the other two loci, since it would require a recombination event between both *A* and *B* and between *A* and *C* in order to generate these gametes. Thus, the order of loci is *BAC* (which is equivalent to *CAB*).

Table **26.1.2**: An example of data that might be obtained from the F2 generation of the three-point cross shown in Figure **26.1.12**. The rarest phenotypic classes correspond to double recombinant gametes ABc and abC. Each phenotypic class and the gamete from the trihybrid that produced it can also be classified as parental (P) or recombinant (R) with respect to each pair of loci (A,B), (A,C), (B,C) analyzed in the experiment.

tail phenotype	fur phenotype	whisker phenotype	number of progeny	gamete from trihybrid	genotype of F_2 from test cross	loci A, B	loci A, C	loci B, C
short	brown	long	5	*aBC*	*aaBbCc*	P	R	R
long	white	long	38	*AbC*	*AabbCc*	P	P	P
short	white	long	1	*abC*	*aabbCc*	R	R	P
long	brown	long	16	*ABC*	*AaBbCc*	R	P	R
short	brown	short	42	*aBc*	*aaBbcc*	P	P	P
long	white	short	5	*Abc*	*Aabbcc*	P	R	R
short	white	short	12	*abc*	*aabbcc*	R	P	R
long	brown	short	1	*ABc*	*AaBbcc*	R	R	P

Recombination frequencies may be calculated for each pair of loci in the three-point cross as we did before for one pair of loci in our dihybrid (Figure 7. 8).

$$\text{loci A,B R.F.} = \frac{1+16+12+1}{120} = 25\% \tag{26.1.1}$$

$$\text{loci A,C R.F.} = \frac{1+5+1+5}{120} = 10\% \tag{26.1.2}$$

$$\text{loci B,C R.F.} = \frac{5+16+12+5}{120} = 32\% \tag{26.1.3}$$

$$\text{(not corrected for double crossovers)} \tag{26.1.4}$$

(26.1.5)

However, note that in the three point cross, the sum of the distances between A-B and A-C (35%) is less than the distance calculated for B-C (32%)(Figure **26.1.13**). this is because of **double crossovers** between B and C, which were undetected when we considered only pairwise data for B and C. We can easily account for some of these double crossovers, and include them in calculating the map distance between B and C, as follows. We already deduced that the map order must be *BAC* (or *CAB*), based on the genotypes of the two rarest phenotypic classes in Table **26.1.2**. However, these double recombinants, *ABc* and *abC*, were not included in our calculations of recombination frequency between loci *B* and *C*. If we included these double recombinant classes (multiplied by 2, since they each represent two recombination events), the calculation of recombination frequency between B and C is as follows, and the result is now more consistent with the sum of map distances between A-B and A-C.

$$\text{loci B,C R.F.} = \frac{5+16+12+5+2(1)+2(1)}{120} = 35\% \tag{26.1.6}$$

$$\text{(corrected for double recombinants)} \tag{26.1.7}$$

(26.1.8)

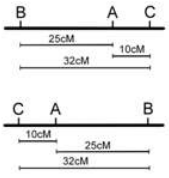


Figure **26.1.13**: Equivalent maps based on the data in Table **26.1.2** (without correction for double crossovers). (Original-Deyholos-CC:AN)

Thus, the three point cross was useful for:

1. determining the order of three loci relative to each other,
2. calculating map distances between the loci, and
3. detecting some of the double crossover events that would otherwise lead to an underestimation of map distance.

However, it is possible that other, double crossovers events remain undetected, for example double crossovers between loci A,B or between loci A,C. Geneticists have developed a variety of mathematical procedures to try to correct for things like double crossovers during large-scale mapping experiments.

As more and more genes are mapped a better genetic map can be constructed. Then, when a new gene is discovered, it can be mapped relative to other genes of known location to determine its location. All that is needed to map a gene is two alleles, a wild type allele (e.g. A) and a mutant allele (e.g. 'a').

This page titled 26.1: Mapping With Three-Point Crosses is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by Todd Nickle and Isabelle Barrette-Ng via source content that was edited to the style and standards of the LibreTexts platform.

- **7.7: Mapping With Three-Point Crosses** by Todd Nickle and Isabelle Barrette-Ng is licensed CC BY-SA 3.0. Original source: http://opengenetics.net/open_genetics.html.

Index

A

ABO blood group
7.1: Relationships Between Genes, Genotypes and Phenotypes
absolute linkage
25.3: Linkage Reduces Recombination Frequency
Adding probabilities
4.2: Probability
allele
3.3: Mendel's First Law
Autosomal Dominant (AD)
6.2: Inferring the Mode of Inheritance
Autosomal recessive (AR)
6.2: Inferring the Mode of Inheritance

B

blending inheritance
3.1: Genes are the Basic Units of Inheritance

C

cAMP binding protein (CAP)
16.1: The lac Operon
Chargaff's Rules
10.2: The Structure of DNA
chimera
5.2: Sex-Linkage- An Exception to Mendel's First Law
cloning
3.5: Crossing Techniques Used in Classical Genetics
codominance
7.1: Relationships Between Genes, Genotypes and Phenotypes
complete linkage
25.3: Linkage Reduces Recombination Frequency
constitutive mutants
16.2: The Use of Mutants to Study the lac Operon
coupling configuration
25.5: Inferring Recombination From Genetic Data
cytokinesis
2.1: Mitosis

D

dideoxy nucleotides
23.2: DNA Sequencing
dihybrids
3.6: Dihybrid Crosses
DNA ligation
11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
DNA microarray
24.2: Genomic Approaches- The DNA Microarray
DNA mutations
21.2: DNA Mutations
DNA sequencing
23.2: DNA Sequencing
dominance
3.4: Biochemical Basis of Dominance
dominant
7.1: Relationships Between Genes, Genotypes and Phenotypes

E

episome
16.2: The Use of Mutants to Study the lac Operon
ethidium bromide
12.2: DNA Analysis - Gel Electrophoresis

G

gel electrophoresis
11.4: PCR and Gel Electrophoresis
12.2: DNA Analysis - Gel Electrophoresis
gene therapy
18.5: Transgenic organisms
genetic pathway
9.1: Genetic Pathways
gynandromorph
5.2: Sex-Linkage- An Exception to Mendel's First Law

H

haploinsufficient
3.4: Biochemical Basis of Dominance
haplosufficient
3.4: Biochemical Basis of Dominance
hemizygous
3.3: Mendel's First Law
heterozygous
3.3: Mendel's First Law

I

incomplete dominance
7.1: Relationships Between Genes, Genotypes and Phenotypes
Isolating Genomic DNA
11.2: Isolating Genomic DNA

K

karyogram
5.1: Karyotypes Describe Chromosome Number and Structure
karyotype
5.1: Karyotypes Describe Chromosome Number and Structure

L

lac operon
16.1: The lac Operon
lacI
16.1: The lac Operon
Law of Equal Segregation
3.3: Mendel's First Law
Law of Independent Assortment
3.6: Dihybrid Crosses
linkage
25.1: Linkage
linked genes
3.6: Dihybrid Crosses
locus
7.1: Relationships Between Genes, Genotypes and Phenotypes

M

meiocytes
2.2: Meiosis
Mendel's First Law
3.3: Mendel's First Law
Mendel's Second Law
3.6: Dihybrid Crosses
Mitochondrial Inheritance
6.2: Inferring the Mode of Inheritance
mitosis
2.1: Mitosis
monohybrid crosses
3.5: Crossing Techniques Used in Classical Genetics
3.6: Dihybrid Crosses
mosaic
5.2: Sex-Linkage- An Exception to Mendel's First Law
Multiplying probabilities
4.2: Probability

N

northern blot
13.1: DNA Analysis- Blotting and Hybridization
Nucleases
11.2: Isolating Genomic DNA

O

OGG
1.1: The Cell Cycle and Changes in DNA Content
2.1: Mitosis
2.2: Meiosis
3.1: Genes are the Basic Units of Inheritance
3.2: Model Organisms Facilitate Genetic Advances
3.3: Mendel's First Law
3.4: Biochemical Basis of Dominance
3.5: Crossing Techniques Used in Classical Genetics
4.1: Phenotypic Ratios May Not Be As Expected
5.1: Karyotypes Describe Chromosome Number and Structure
5.2: Sex-Linkage- An Exception to Mendel's First Law
7.1: Relationships Between Genes, Genotypes and Phenotypes
10.1: DNA is the Genetic Material
10.2: The Structure of DNA
20.1: The Function of Genes
oncogene
22.5: Oncogenes

OOG
1.1: The Cell Cycle and Changes in DNA Content
2.1: Mitosis
2.2: Meiosis
3.1: Genes are the Basic Units of Inheritance
3.2: Model Organisms Facilitate Genetic Advances
3.3: Mendel's First Law
3.4: Biochemical Basis of Dominance
3.5: Crossing Techniques Used in Classical Genetics
4.1: Phenotypic Ratios May Not Be As Expected
5.1: Karyotypes Describe Chromosome Number and Structure
5.2: Sex-Linkage- An Exception to Mendel's First Law
7.1: Relationships Between Genes, Genotypes and Phenotypes
10.1: DNA is the Genetic Material
10.2: The Structure of DNA
20.1: The Function of Genes

P
PCR
11.4: PCR and Gel Electrophoresis
Pedigree charts
6.1: Pedigree Analysis
plasmids
12.1: Cloning DNA - Plasmid Vectors
polymerase chain reaction (PCR)
11.3: Isolating or Detecting a Specific Sequence by PCR
11.11: The Polymerase Chain Reaction (PCR)
primer
11.3: Isolating or Detecting a Specific Sequence by PCR
product rule
6.4: Calculating Probabilities
Punnett square
3.5: Crossing Techniques Used in Classical Genetics

R
recessive
7.1: Relationships Between Genes, Genotypes and Phenotypes
reciprocal crosses
5.2: Sex-Linkage- An Exception to Mendel's First Law
recombinant DNA
12.3: Recombinant DNA Technology
repulsion configuration
25.5: Inferring Recombination From Genetic Data
restriction endonucleases,
11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
restriction enzymes
11.7: Cutting and Pasting DNA- Restriction Digests and DNA Ligation
12.3: Recombinant DNA Technology

S
sampling effects
4.1: Phenotypic Ratios May Not Be As Expected
Southern blot
13.1: DNA Analysis- Blotting and Hybridization
STARLINK AFFAIR
11.3: Isolating or Detecting a Specific Sequence by PCR
sum rule
6.4: Calculating Probabilities

T
thermal cycler
11.4: PCR and Gel Electrophoresis
transcription
17.2: Gene Expression- Transcription
transfection
18.5: Transgenic organisms
transformation
12.1: Cloning DNA - Plasmid Vectors
18.5: Transgenic organisms
transgene
18.5: Transgenic organisms
transgenesis
18.5: Transgenic organisms
Transgenic organism
18.5: Transgenic organisms
transgenic plant
18.5: Transgenic organisms
translation
20.5: Gene Expression- Translation

W
Western blot
13.1: DNA Analysis- Blotting and Hybridization

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Loading...