



**Equip.  
Empower.  
Expect More.**

Unit 1 Student Notes Page 1  
[Table of Contents Link](#)

# **AP Biology**

## **Unit 1**

### **Student Notes**

# Unit 1 Student Notes

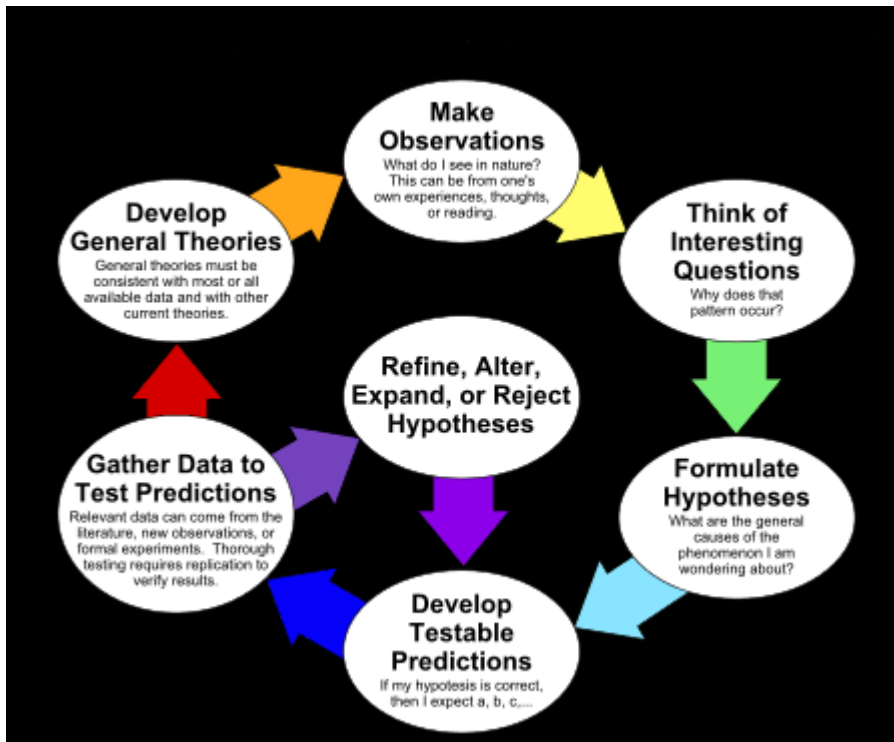
## Table of Contents

- A. [Scientific Method/Experimental Design—Pages 3-5](#)
- B. [Graphing—Pages 5-6](#)
- C. [Free Response Writing Tips—Pages 7-9](#)
- D. [Data Analysis/Statistics—Pages 10-25](#)
- E. [Graphs With Error Bars—Pages 12-16](#)
- F. [Hypothesis Testing—Pages 16-25](#)
- G. [Chi Square Analysis—Pages 16-19](#)
- H. [t-tests—Pages 20-24](#)
- I. [Chemistry Basics—Pages 25-26](#)
- J. [Biochemistry of Water—Pages 27-29](#)
- K. [Biochemistry of Carbon—Pages 30-31](#)
- L. [Carbohydrates—Page 32](#)
- M. [Lipids—Pages 32-35](#)
- N. [Proteins—Pages 35-39](#)
- O. [Nitrogen Cycle—Pages 39-40](#)
- P. [Nucleic Acids—Pages 41-43](#)
- Q. [Phosphorus Cycle—Pages 43-44](#)

## Unit 1 Student Notes Content Outline: The Scientific Method

### The Scientific Method

This is a *series of steps* followed to *solve problems*. The steps are not always the same for each question you are researching and may not be followed in a linear pattern. The scientific method might better be illustrated by the diagram included below:



#### Step 1: State your Problem/Question

1. Develop a question or problem that *can be solved through experimentation*.
2. Make sure it is something that interests you.

#### Step 2: Make Observations/Do Research

1. Make observations – the act of *seeing* an object or an event and *noting* the physical characteristics or points in the event. Observation is an extension of our *senses*; when we observe, we *record what is seen, smelled, tasted, heard, and touched*.
  - a. **Qualitative observations** - These *describe* an object's characteristics, properties, or attributes. (For example, in the state, "The apple is red," red is a qualitative observation of the apple's appearance.)

- b. **Quantitative observations** – These involve a *quantity or an amount*. (In the statement, “The apple weighs 125 grams,” 125 grams is a quantitative observation of the apple’s appearance.) Quantitative data refers to numerical or measured data.
  - c. **Inferences** – *conclusions based on observations*. Inferences go beyond what we can directly sense. (Example: You make an inference when you use clues from a story to figure out something the author *doesn’t* tell you.)
  - d. **Predictions**- using observations, inferences, and/or trends in data to predict what will happen in the *future*.  
(Example: If, on a sunny day, you observe a massive line of dark clouds quickly advancing, what prediction can you make?)
2. Do research – In this step, we are talking about doing literature research, not lab-based research. Scientists should read about the research that has already been done on the topic by searching the Internet and scientific journals. Good quality research helps in developing an excellent hypothesis.

### Step 3: Formulate a Hypothesis

1. A **hypothesis** is a *prediction or possible answer* to the problem or question.
2. It is a relationship between the **Independent variable** and **Dependent variables**.
  - a. **Independent Variable (manipulated variable)** – the factor that is *intentionally varied/tested* by the experimenter.
  - b. **Dependent Variable (responding variable)** – the factor that *may change* as a result of changes made in the independent variable (the outcome).
  - c. **Example:** Let’s say that a scientist wanted to know if the use of miracle-gro affected the height of tomato plants. The independent variable in the experiment would be the amount of miracle-gro applied to the plants. The dependent variable would be the height of the plants.
  - d. The hypothesis needs to be written as an “If...then” statement.  
The “If” part of the statement *should describe what is done to the independent variable*.  
The “then” part of your statement *is the prediction of what will happen to the dependent variable*.  
**Example:** If miracle-gro is applied to tomato plants, then they will grow taller.

### Step 4: Experiment

- A. The scientist must develop and follow a *procedure* that anyone can follow.
  1. Use *precise* directions.
  2. Include a *detailed materials list*.
  3. The outcome must be **quantifiable** (measurable).
  4. The experiment must have a **control group**.
    - a. The **control group** may be a “no treatment” or an “experimenter selected” group to use as a *standard of comparison* for the independent variable.  
**Example:** In the miracle-gro experiment described above, the control group would consist of plants that are not exposed to any miracle-gro.
    - b. The control group is *exposed to all of the same factors* as the **experimental group(s)** except for the independent variable being tested.  
**Experimental group** – group or groups that have the *independent variable applied/manipulated*.  
**Example:** In the miracle-gro experiment, the experimental group would consist of a group of plants that are treated with miracle-gro. We might treat different subgroups with different amounts of miracle-gro to test the effect of concentration.  
**Constants** – all the *factors* that the experimenter attempts to keep the *same/control* in all of the groups in the experiment.

**Example:** In the miracle-gro experiment, we would want to ensure that all of the plants are of the same species, growing in the same type of soil, exposed to the same amount of light, given the same amount of water, and grown at the same temperature.

### Step 5: Collect Data

- A. You must write down results (measurements, observations, temperatures, times, etc.) *as you perform your experiment.*
  1. **Qualitative Data** - observations (using senses) written in note form.
  2. **Quantitative Data**- numerical measurements and calculations.
    - a. SI Units must be included on all measurements.
2. Must be kept *orderly in a table or chart.*
3. Modify the procedure if needed.

### Step 6: Analyze Data

- A. *Confirm* the results by *retesting*, if possible.
- B. **Trials** – the *number of times* you repeat the experiment.
  1. The more trials you can do, the *more reliable* the results.
- C. Convert results to a graph that is appropriate for the experiment.
- D. Use both descriptive and inferential statistics to help make a conclusion.

### Step 7: Conclusion

- A. The *written results* of the experiment.
- B. Include a *statement if the hypothesis was supported or refuted.*
- C. Make *recommendations for further study and possible improvements* to the procedure.

### Step 8: Communicate Results

- A. Be prepared to present the project to an audience. *Scientists share information* through media, journal articles, and lectures.

## Graphing

Graphs and charts communicate information visually. They can show patterns, help scientists identify correlations, and get the point of the experiment across quickly.

The **independent variable** is plotted on the **x-axis**

The **dependent variable** is plotted on the **y-axis**.

The mnemonic **DRY MIX**, for “dependent, responding, y-axis” and “manipulated, independent, x-axis,” can help you remember this pattern.

Label both axes (independent variable on the X-axis and dependent variable on the Y-axis)

Include units on both axes. Enclose the unit in parentheses.

Provide a descriptive title. Use the pattern, “The Effect of the *independent variable* on the *dependent variable*”. For example if you were graphing the miracle-gro concentration against plant height. The title of the graph might be “The Effect of Miracle-Gro Concentration on Plant Height”.

If the instruction is to plot rather than graph the data points, no line needs to be drawn.

If a line is drawn, do not extend the line beyond the last point plotted (unless asked to make a prediction) or connect the line from the origin (unless there is a time zero reading.)

If multiple lines are drawn on the same graph, label each line clearly.

Use a line of best fit when appropriate.

### **Which Graph Type To Use?**

**A) Line Graph:** are used for looking at the relationship between two continuous types of data. Typically, both the independent and dependent variables are numerical.

**B) Bar Graphs:** are used for making comparisons between discrete cases or to look for trends, such as over space or time. The independent variable is usually a category and the dependent variable is usually an average, percentage, or frequency.

**C) Scatter Plot:** Scatter plots are used for examining relationships between two types of data. These are very similar to line graphs, just without the line.

Line graphs provide an excellent way to map independent and dependent variables that are both quantitative. When both variables are quantitative, the line segment that connects two points on the graph expresses a slope, which can be interpreted visually relative to the slope of other lines or expressed as a precise mathematical formula. Scatter plots are similar to line graphs in that they start with mapping quantitative data points. The difference is that with a scatter plot, the decision is made that the individual points should not be connected directly together with a line but, instead express a trend. This trend can be seen directly through the distribution of points or with the addition of a regression line or line of best fit.

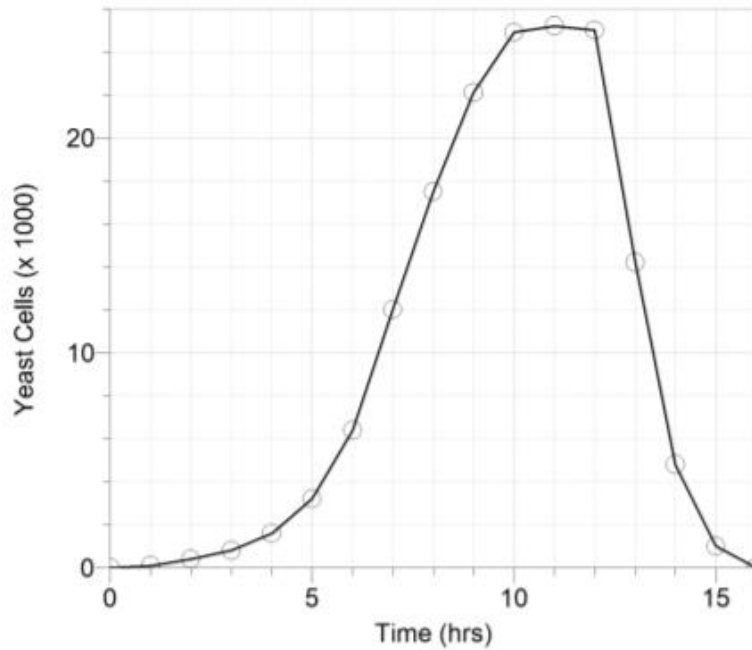
### **Determining the rate from a graph.**

The AP Biology exam often asks students to find the rate of a process or reaction between two points on a graph.

In order to find the rate, calculate the slope of the best fit line that connects the two points. Use  $m = \frac{y_2 - y_1}{x_2 - x_1}$

Be sure to include a unit with your answer. The unit for the slope/rate should be the y axis unit divided by the x axis unit.

Example:



You are asked to calculate the yeast population growth rate between  $t=5$  hours and  $t=10$  hours.

First, determine your coordinates at those times. The coordinates should be (5, 3.5) and (10, 25).

Next, calculate the slope:  $m=(25-3.5)/(10-5)=4.3$  yeast cells/hour

### Free Response Writing Tips

The following tips were adapted from:

[www4.bluevalleyk12.org/.../apbiology/.../ap\\_biology\\_free\\_response\\_writing\\_tips.pdf](http://www4.bluevalleyk12.org/.../apbiology/.../ap_biology_free_response_writing_tips.pdf)

The AP Biology Exam consists of two sections: multiple choice and free response. Both sections include questions that assess your understanding of the big ideas, enduring understandings, and essential knowledge, and application of these through the science practices.

Question Type	Number of Questions	Timing
<b>Section I</b>		
Part A: Multiple Choice	63	90 minutes
Part B: Grid-In	6	
<b>Section II</b>		
Long Free Response	2	80 minutes + 10-minute reading period
Short Free Response	6	

**General Tips:**

- **You must write all answers in complete sentences!** There is room on the test for you to create an outline to guide your answer, but outlines are not graded. That being said, perfect essay writing is not expected. There are no deductions for grammar or spelling mishaps (provided the spelling is close enough to determine the word you are trying to write).
- Diagrams are helpful. However, if you draw a diagram, be sure to refer to it in your essay. You will not earn points for diagrams that stand by themselves. **You must explain all diagrams and drawings.**
- Points are not deducted from your essay score if you give an incorrect statement. You just do not receive points for incorrect statements. However, you must be careful not to contradict yourself. **If you state something correctly but then later state the opposite, you will not earn the point.**

**Tips for AP Lab Free Response Questions:****Design an Experiment Free Response Questions**

The AP Biology exam will often ask you to design an experiment to address a certain topic/question. When asked to design an experiment, always include the following elements in your answer:

Form a concise hypothesis which is testable. State It Clearly! Use the if...then format. For example: If tomato plants are exposed to increased amounts of Miracle-Gro, then they will grow taller.

Describe the control group that you will use for comparison with the experiment.

Identify the independent variable.

Identify the dependent variable.

Identify at least 3-4 variables that you will use as constants.

Describe the basic procedure that you will use. Describe the measurements that you will take, the # or subjects that you will use, and how the subjects were assigned to either the control or experimental groups. This should usually be random.

Stress the importance of a large sample sizes.

Mention that you will conduct multiple trials.

Describe the statistical tests that you will use to interpret the data (Chi Square, rate determination...).

Graph: Choose an appropriate graph. Use the guidelines included above.

Make a prediction about the expected outcome and a rationale for your prediction.

**DOS and DON'TS on Exam Day****DO THIS on Exam Day:**



- **DO** use the ten minute reading time advantageously. Carefully read all of the free response questions and map out/outline your answers. These maps will NOT be graded, but you can use them to write your responses.
  - Read the prompt thoroughly, then read the prompt again, then read the prompt, then read the prompt again, then read the prompt, then...
  - Jot down the big ideas. Make sure you clearly understand what you are being asked to do.
  - Use this time to create a mindmap or bullet points of the main terms you want to elaborate on.
  - Outline your answer to organize your thoughts.
  - Remain focused and on task.
  - Answer the prompt and only the prompt.
  - Feel free to write on the exam/question booklet.
- **DO underline the important terms** in the question such as “OR” and “CHOOSE 2” and the power verbs such as “DESCRIBE,” “IDENTIFY,” “LABEL,” “CONSTRUCT,” “DESIGN,” or “EXPLAIN.” The verbs usually indicate where points can be earned.  
**DO** use the **80 minutes** to write thorough responses to all eight questions. Use all of your time. Don’t give up.
- **DO** stay focused on what the prompt is requiring you to do. Pay particular attention to words like:
  - **Discuss:** give reasoning pro and con; analyze carefully
  - **Analyze:** summarize in detail with a selected focus
  - **Explain:** clarify and interpret; give reasons for differences, analyze causes
  - **Compare/contrast:** emphasize similarities **and** differences
  - **Relate:** show how ideas or concepts are connected to each other**DO** use the outline, mindmap or bullet points that you developed during the 10 minute reading time.
- **DO** write as legibly as possible, using **black** ink. The papers are shuffled quite a bit when they are scored, and answers written in pencil may be smeared. If the person scoring your essays cannot read what you have written, then you will not earn any points. Do use a pencil to create all graphs. When the graph is complete and correct, outline over the response in pen.
- **DO** answer in the format of the question so that you do not slow the reader down.
  - Use the format of the free response to write your answer so that the reader has an easy time finding your responses to each section of each essay.
  - Organize the free response answers using the format of the question—write „1a“ then respond to 1a; write „1b“ then respond to 1b, etc...
  - It is best not to skip around when responding to sub-questions in one question.
- **DO** apply the language of science, show depth, elaboration, and give examples.
  - Pull, tie, link and loop together your ideas—show how ideas connect.
  - Use a scientific term and then explain what it means.
  - Write for clarity, accuracy, thoroughness, and breadth (not just factual regurgitation).
- **DO** use graphs or diagrams when it will enhance your essay response. However, unless the prompt specifically asks for drawings/graphs, every thought you hope to convey must also be put in writing.
- **DO** clearly mark your answer sheet with the free response question you are answering. Write freely on the response sheet—use several sheets as needed. Usually, the longer the answer to the question the more points you will earn! Write! Write! Write!  
**DO** answer ALL subunits of a question thoroughly—to ensure you will gain maximum points for your response.
- **DO** label all graphs correctly.
  - Include a graph title.
  - Include a key identifying lines and data points.
  - Label axes (including units).
- **DO** use the time at the end to re-read responses—underlining key concepts, checking for clarity, accuracy and thoroughness.

**DON'T DO THIS on Exam Day:**

- **DON'T** leave any free response questions blank.
  - Even if the question seems odd or you draw a temporary blank, find the “main idea” being addressed and elaborate on it.

[Table of Contents Link](#)

- Remember that all students in the nation will be in the same boat with a difficult or unclear question.
- **DON'T** obsess over correct grammar. There are no deductions for grammatical imperfections.
- **DON'T** write introductory or closing paragraphs. No points are earned for thesis statements or topic sentences.
- **DON'T** ramble. Get to the point. Do not waste time describing your feelings about how glad you are that the AP College Board asked you about photosynthesis. If anything, this will annoy the reader.  
**DON'T** write only in outline format. Your answers must be written in complete sentences.
- **DON'T** over-answer the sub-questions of a free response question.
  - Remember that for any given question requiring sub-question responses, each response is allotted a maximum number of points. Writing more than is necessary will not earn you more points.

### Unit 1 Student Notes

#### Content: Data Analysis/Statistics

#### Adapted from the AP Biology Quantitative Skills Guide, Using Biointeractive Resources to Teach Mathematics in Statistics in Biology

As you start making observations and collecting data for a lab investigation, you will probably notice patterns. These patterns may or may not be real, or valid. **Quantitative data analysis** is one of the first steps toward determining whether an observed pattern has validity.

Data analysis also helps distinguish among multiple working hypotheses. Every AP Biology laboratory activity will require data collection and analysis. This analysis will help you to discover meaningful patterns relevant to your investigation. In AP Biology, we will use both descriptive and inferential statistics to analyze our lab data.

**Descriptive statistics** is used to estimate important parameters of the sample data set. Examples include sample standard deviation, which describes the variability in the data; measurements of central tendencies such as mean, median, and mode; standard error of the sample mean, which helps you determine your confidence in the sample mean and how well the sample mean represents the true population mean. The same parameters (mean, standard deviation, etc.) can also describe the entire or true population that you are studying, but collecting the data to compute these statistics is most often not possible. That's where inferential statistics comes in.

**Inferential statistics** includes tools and methods (statistical tests) that rely on probability theory and an understanding of distributions to determine precise estimates of the true population parameters from the sample data. This is a key part of data analysis and allows you to support and draw conclusions from your data about the true population.

Most of the data collected during AP Biology experiments will be **parametric data**. Parametric data follows an approximate normal distribution/bell curve distribution.

For a normal distribution, the appropriate descriptive statistics for the data set include the mean (average) ( $\bar{x}$ ), sample size ( $n$ ), standard deviation ( $S$ ), and standard error ( $SE_x$ ). Each is important.

The mean ( $\bar{x}$ ) of the sample is the average (the sum of the numbers in the sample divided by the total number in the sample). The AP Biology formula sheet lists the following formula for finding the mean:

## Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Don't over-complicate the mean calculation. The mean is simply the average of your sample.

The mean summarizes the entire sample and might provide an estimate of the entire population's true mean.

The sample size ( $n$ ) refers to how many members of the population are included in the study. Sample size is important when students try to estimate how confident they can be that the sample set they are trying to analyze represents the entire population.

Both the **standard deviation** measure and the **standard error** measure define boundaries of probabilities.

The sample standard deviation ( $S$ ) is a tool for measuring the spread (variance) in the sample population, which in turn provides an estimate of the variation in the entire sample set. A large sample standard deviation indicates that the data have a lot of variability. A small sample standard deviation indicates that the data are clustered close to the sample mean.

The AP Biology formula sheet lists the following formula for calculating standard deviation:

## Standard Deviation\*

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Most likely, you will not be asked to calculate the standard deviation on an AP exam, but you will be expected to be able to interpret its meaning and be able to use it to analyze your data and construct appropriate graphs.

In a **normal distribution**, a little more than two-thirds of the data points will fall between +1 standard deviation and -1 standard deviation from the sample mean. More than 95% of the data falls between  $\pm 2$  standard deviations from the sample mean.

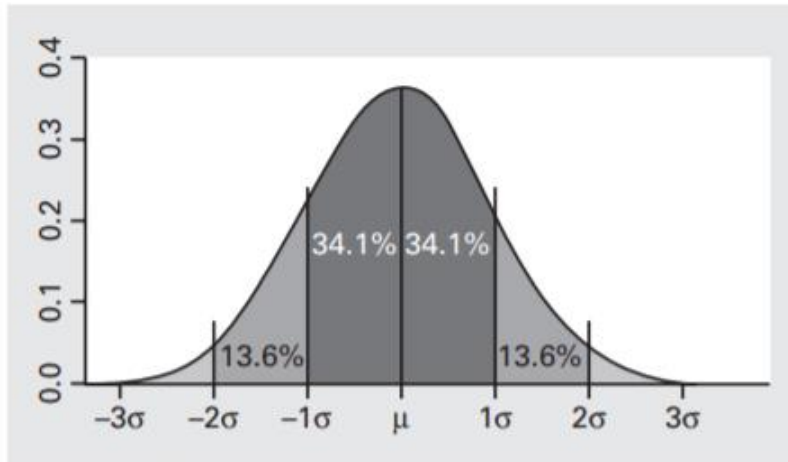


Figure 6. Standard Deviations and a Normal Distribution

**Sample standard error ( $SE_x$ )** is a statistic that allows students to make an inference about how well the **sample mean** matches up to the **true population mean**. The standard error of the mean utilizes the standard deviation of the sample and the sample size to estimate how closely the sample data approximates the data that would be collected if the entire population were measured. If one were to take a large number of samples (at least 30) from a population, the means for each sample would form an approximately normal distribution—a distribution of sample means. Normally, you would not do hundreds of individual investigations on a population. This distribution of sample means, then, is a theoretical construct that helps us define our boundaries of confidence in our sample. This distribution also has parameters, such as a standard deviation. Standard error is the equivalent of the standard deviation of the sampling distribution of the means and is calculated from the following formula:

### Standard Error of the Mean\*

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

You will probably not be asked to calculate the standard error of the mean on an AP Biology exam, but you will be expected to use the standard error to analyze your data and construct appropriate graphs.

An interval within  $\pm 1 SE_x$  of the sample mean describes the range of values about which an investigator can have approximately 67% confidence that the range includes the true population mean. Even better, a sample interval within  $\pm 2 SE_x$  of the sample mean defines a range of values with approximately a 95% certainty that the true population mean falls within the interval. This interval is often referred to as a **95% confidence interval**. You will be asked to graph 95% confidence intervals on a regular basis in this course.

The 95% confidence interval technique is an inference; it is a statistic that allows investigators to gauge just how good their estimate of the true population mean actually is. With this understanding, the investigator can establish ahead of time a reasonable sample size for this population and the degree of confidence needed. Note: The larger the sample size, the smaller the standard error and the more confident the researcher can be about the reliability of the data.

#### Creating and Interpreting Graphs with Error Bars

The **Standard Error of the Mean** and the **95% Confidence Interval** are used to construct error bars for graphs showing the mean values of data sets. In AP Biology, the error bars usually show the range 2 standard errors above and 2 below the mean value.

To create a graph with error bars, graph the means of each data set using a bar chart. Starting at the mean of the first bar, move directly up 2 times the  $SE_x$ . At that point, draw a horizontal line centered over the bar. Go back to the mean,

and move directly down 2 times the  $SE_x$ . Again, draw a horizontal line that is in line with the line you drew above the bar. Connect the two horizontal lines with a vertical line through their centers.

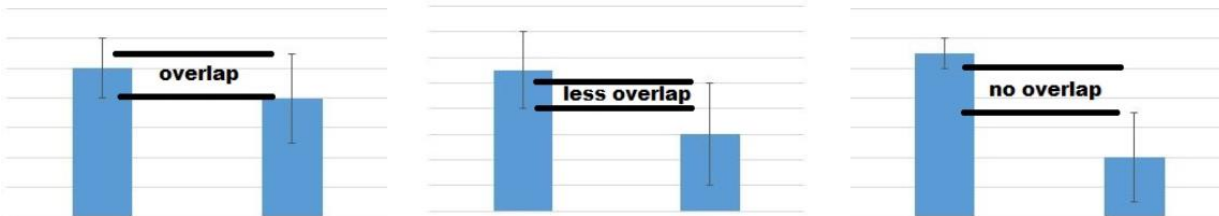
Repeat the process for the other bars included in the chart.

The vertical space between the two horizontal lines represents a 95% confidence interval. This means that you can be 95% sure that the mean for the entire population falls within this interval.

As a researcher, you may want to use your collected data to compare samples/groups and to determine if the groups are "significantly different".

A "**significant difference**" means that the results that are seen in the data are most likely not due to chance or sampling error. In any experiment or observation that involves sampling from a population, there is always the possibility that an observed effect could have occurred due to sampling error alone. But if a result is "significant," the investigator may conclude that the observed effect actually reflects the characteristics of the population rather than just sampling error or chance.

The standard error bars on a graph can be used to get a sense for whether or not a difference in two groups/samples is significant. Look for overlap between the error bars:



When errors bars overlap quite a bit, it's likely that the difference between the two groups is not statistically significant and the differences are probably due to chance or sampling error. You must actually perform a statistical test to draw a valid conclusion, **but in AP Biology you can simply say that if the error bars overlap that the two groups/samples are not statistically different.**

**On the other hand, if there is no overlap between the error bars, the differences between the two groups is likely to be statistically significant.**

#### Example: Descriptive Statistics and Error Bars

The data included below is from research done by Peter and Rosemary Grant on Daphne Major in the Galápagos Islands. The data shows the change in beak depth of a population of finches following a drought year (1977). Think of the band numbers like names for the individual birds.

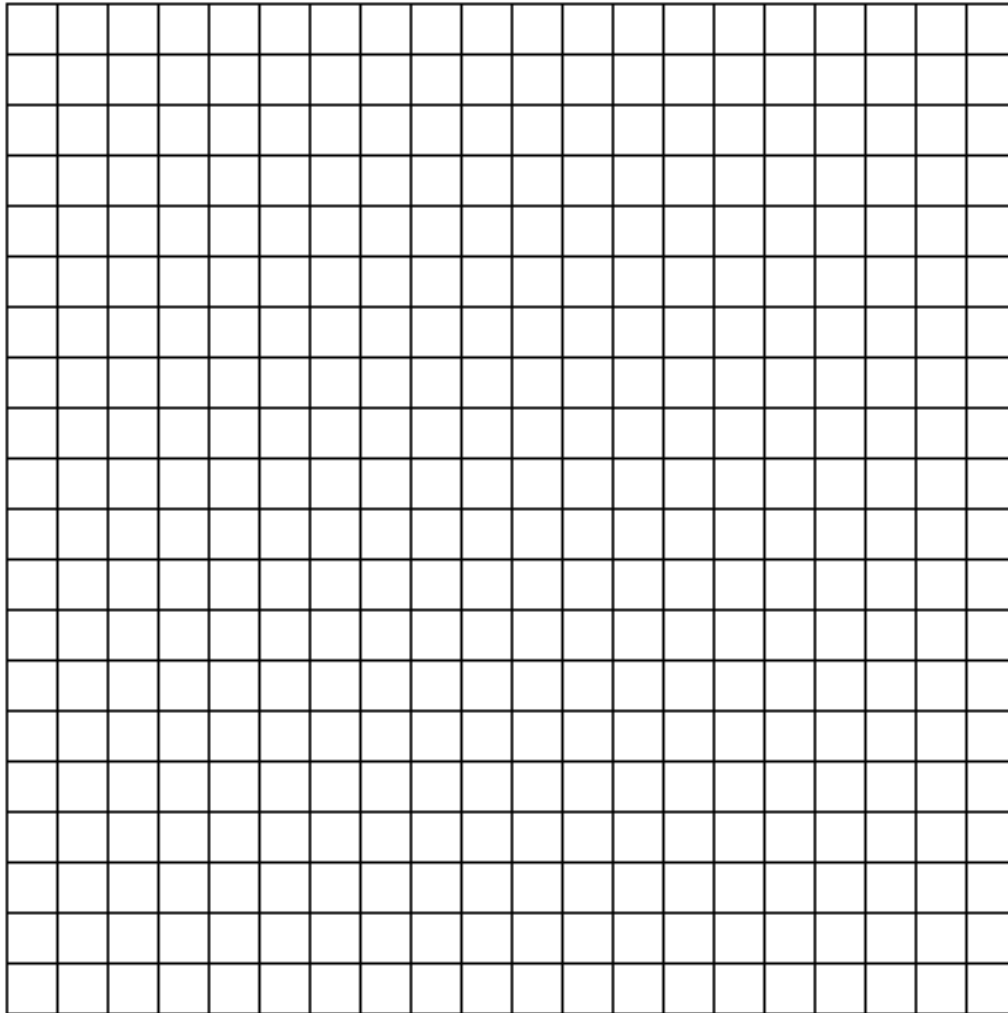
[Table of Contents Link](#)

Died in Drought		Survived Drought	
Band	Beak Depth (mm)	Band	Beak Depth (mm)
283	11.2	1019	11.21
278	10.6	1919	11.2
294	10.5	2244	11.01
609	10.5	8191	10.86
674	10.5	1659	10.78
422	10.3	1861	10.7
428	10.2	1599	10.7
561	10.2	2249	10.68
605	10.2	1426	10.61
461	9.8	2206	10.5

Use the data to fill in the chart included below:

	Died in Drought	Survived Drought
<i>mean</i>		
<b>S</b>		
<b>SEM</b>		
<b>95% CL</b>		

Graph the data as a bar chart of the means showing error bars that create a 95% Confidence Interval.

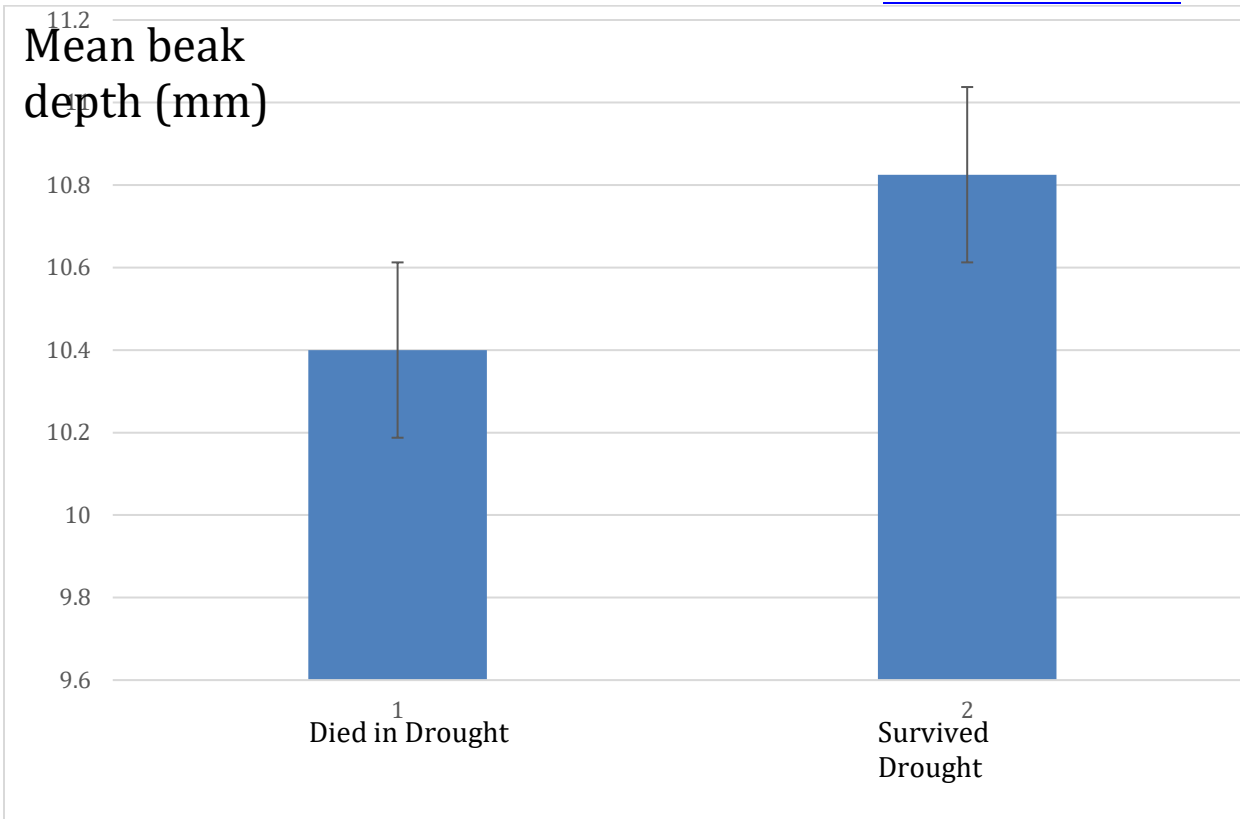


Was the mean beak depth before the drought different from the mean beak depth after the drought?

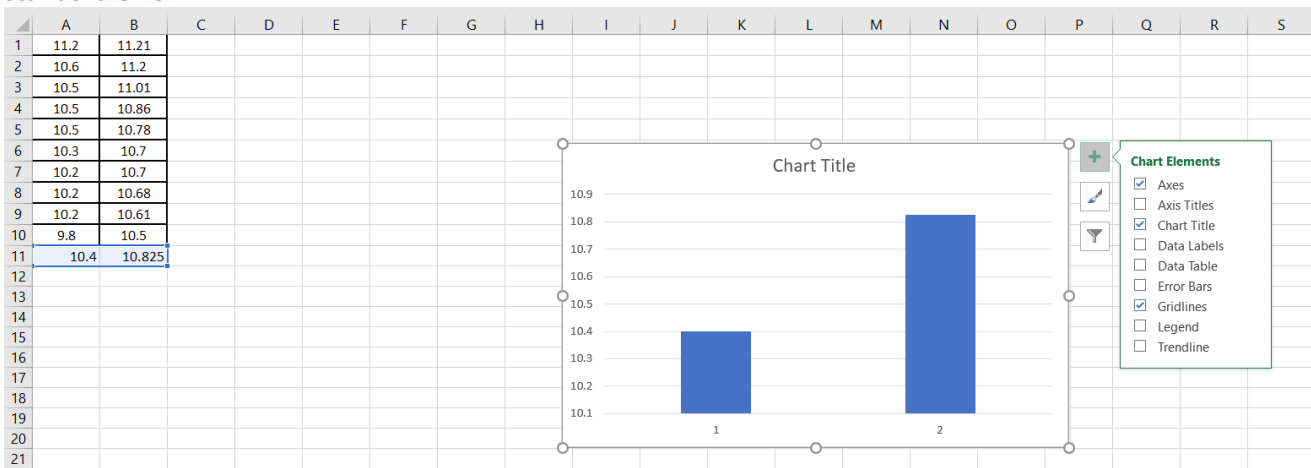
Answer to Example:

	Died in Drought	Survived Drought
mean	10.4	10.825
S	0.365148	0.242773
SEM	0.11547	0.076772
95% CL	10.169--10.631	10.671--10.979

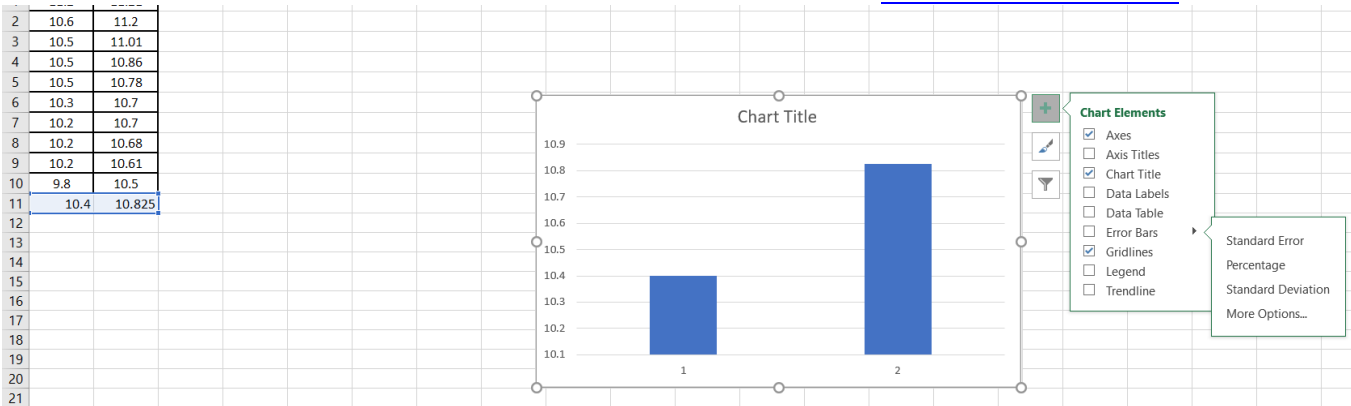
The mean beak depth after the drought was different in a statistically significant amount from the beak depth before the drought. This can be determined by looking at the error bars for the two conditions. Since the error bars don't overlap, the differences are likely to be statistically significant.



Note: To add error bars to an excel graph, click on the “chart elements” button. From the drop down menu, choose standard error.







## Hypothesis Testing

A hypothesis is a statement explaining that a causal relationship exists between an underlying factor (variable) and an observable phenomenon. Often, after making an observation, you might propose some sort of tentative explanation for the phenomenon; this could be called your working hypothesis. Because absolute proof is not possible, **statistical hypothesis testing** focuses on trying to reject a **null hypothesis**.

A **null hypothesis ( $H_0$ )** is a statement explaining that the underlying factor or variable is independent of the observed phenomenon—there is no causal relationship.

Stated another way, a null hypothesis ( $H_0$ ) is usually a statement asserting that there is no difference or no association between variables. The null hypothesis is a tool that makes it possible to use certain statistical tests to figure out if another hypothesis of interest is likely to be accurate or not. For example, if you were testing the idea that sugar makes kids hyperactive, your null hypothesis might be that there is no difference in the amount of time that kids previously given a sugary drink and kids previously given a sugar-substitute drink are able to sit still. After making your observations, you would then perform a statistical test to determine whether or not there is a significant difference between the two groups of kids in terms of time spent sitting still.

The **alternative hypothesis ( $H_A$ )** to the null hypothesis might be that there is a difference between the two groups of kids in terms of time spent sitting still. Usually (but not always), an investigator is trying to find an alternative to the null hypothesis—evidence that supports the alternative hypothesis by rejecting the null (based on statistical tests). If the null hypothesis (that there is no difference between the two groups of kids in terms of time spent sitting still) can be rejected, then that is support for this alternative hypothesis.

It is important to realize that hypothesis testing does not allow proof, or even acceptance, of the alternative to the null hypothesis. Typically, the decision comes down to whether there is enough evidence to reject the null hypothesis. **If evidence to reject the null hypothesis is sufficient, what can be said is that the investigator rejects the null hypothesis—not that the investigation has proven the alternative hypothesis.** This is a crucial concept for students to understand. In data analysis, investigators determine the size and confidence they have in various population parameters that were measured, counted, or calculated during the course of the investigation. Hypothesis testing asks the question, Is there something to these measurements? or Is the effect real?

### Types of Statistical Tests

There are a wide range of statistical tests which can be used for hypothesis testing. For AP Biology, we will focus on only two of these methods: **chi square analysis and the t-test.**

### Chi Square Analysis

The Chi-square test is a statistical method that **makes a comparison between the data collected in an experiment and the data an investigator expected to find.** The Chi square test is a way to evaluate the variability that is always present in the real world to get an idea if the difference between the real and expected results is due to random chance or if some other factor is involved.

Chi square analysis can be used when you are comparing two or more categories of data. In AP Biology, the categories will usually be the observed and expected data. **The actual data under these categories will typically be in the form of either counts or percentages. Chi square analysis should not be used to compare averages.** For example, chi square analysis can be used to test how well the results of genetic crosses fit predicted outcomes based on Mendel's laws of inheritance or to see how well measured gene frequencies in a population match up to Hardy-Weinberg predictions. When the Chi-square test is applied in these kinds of analyses, the goal is to determine whether or not the variation in the results from the expected values is due to chance. In these analyses, students are trying to confirm a theoretical expectation about their data, and they hope to quantify the contribution due to chance events. Here researchers hope to fail to reject the null hypothesis, i.e., that there is no evidence of a significant difference between the expected and observed results. This approach might also be used in the M& M lab in which students compare the percentages of M&Ms of each color in a bag to the theoretical percentages produced at the factory.

In other investigations, however, students may ask a question that requires a different application of the Chi-square test. For example, in a pill bug environmental choice experiment, students may wish to know if pill bugs actually choose one environment over another, or whether they just randomly move about. With this type of investigation, students are trying to discover and verify that an actual pattern exists as opposed to the random variation that often characterizes natural systems. Here students hope to reject the null hypothesis, indicating that their observed results are significantly different from the ones they expected.

The Chi Square statistic can be calculated using the table formula included on the next page. My advice for calculating chi square is to setup a table like the one included at the bottom of the next page.

**Standard Error of the Mean\***

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

**Chi-Square**

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

**Chi-Square Table**

p value	Degrees of Freedom							
	1	2	3	4	5	6	7	8
0.05	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51
0.01	6.64	9.21	11.34	13.28	15.09	16.81	18.48	20.09

Phenotype	O	E	(O-E)	(O-E) <sup>2</sup>	(O-E) <sup>2</sup> /E

Let's walk you through the process of calculating chi square by working through an example problem.

- A student wanted to know if pillbugs have a preference for wet or dry environments. The student setup a choice chamber with a wet and a dry side. He placed 10 pillbugs on each side of the chamber and after 2 hours found 14 pillbugs on the wet side and 6 on the dry side.
- The student's null hypothesis was that pillbugs had no preference for either wet or dry
- His phenotypes or groups for this test are "wet" and "dry"
- His expected values on each side are "10".
- His observed values are "14" on the wet side and "6" on the dry side.
- Fill in the columns in the chart by performing the mathematical calculations shown at the top of each row. See the chart below as an example.
- The chi square statistic is calculated by find the sum of the last column in the table. Remember that chi square is equal to:

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

- 
- In this case, chi square is equal to 3.2.

[Table of Contents Link](#)

- There are two different ways to interpret the meaning of the chi square statistic. One way is to compare it to a **critical value**. Use the chi square table, included above to find this critical value. First, determine your degrees of freedom. The degrees of freedom are equal to your number of phenotypes/categories minus one. In this case, our degrees of freedom are equal to 1.
- We are always going to use the 0.05 significance level in Biology. The significance level, also denoted as alpha or  $\alpha$ , is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference. This is essentially an error range.
- Use the 0.05 significance level and the 1 degree of freedom to find a critical value of “3.84” in the chi square table.
- Compare the critical value “3.84” to the chi square statistic (3.2) that we calculated. **If the chi square statistic is greater than the critical value, we will reject the null hypothesis.** In this experiment, the chi square statistic is less than the critical value. **This means that we must fail to reject the null hypothesis. This essentially means that the differences between our observed and expected values are small and likely due to chance.**
- The other way to interpret the chi square statistic is to use the **p-value approach**. Move along row 1 (1 degree of freedom) in the chi square distribution table included below until you can find the Chi-square value of 3.2. It is somewhere between the 0.10 column and the 0.05 column. This means that the p-value for this data is between 0.10 and 0.05. Remember, the probability of whether the results of an investigation differ from the null results by chance alone is called the p-value. A p-value of 0.05 means that there is a 5% chance that the difference between the observed and the expected data is a random difference and a 95% chance that the difference is real and repeatable—in other words, a significant difference. Therefore, **if an investigator’s p-value is greater than 0.05, he or she would fail to reject the null hypothesis—that the difference between the observed results and the expected results is due to random chance and is not significant.** Our data fall into this category. This means that our data doesn’t indicate that pillbugs prefer a wet environment over a dry environment. On the other hand, **if the p-value had been less than 0.05, we would have rejected the null hypothesis.** This would have indicated that the differences between our observed and expected data were significant and likely due to something other than chance. **This same p-value approach is used with the t-test and ANOVA testing.**

Table 9. Chi-square Distribution Table

Degrees of Freedom	Probability (p) Value									
	ACCEPT NULL HYPOTHESIS								REJECT	
	0.99	0.95	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
1	0.001	0.004	0.06	0.15	0.46	1.07	1.64	2.71	<b>3.84</b>	6.64
2	0.02	0.10	0.45	0.71	1.30	2.41	3.22	4.60	<b>5.99</b>	9.21
3	0.12	0.35	1.00	1.42	2.37	3.67	4.64	6.25	<b>7.82</b>	11.34
4	0.30	0.71	1.65	2.20	3.36	4.88	5.99	7.78	<b>9.49</b>	13.28
5	0.55	1.14	2.34	3.00	4.35	6.06	7.29	9.24	<b>11.07</b>	15.09
6	0.87	1.64	3.07	3.38	5.35	7.23	8.56	10.65	<b>12.59</b>	16.81
7	1.24	2.17	3.84	4.67	6.35	8.38	9.80	12.02	<b>14.07</b>	18.48

- See the [chi square Powerpoint](#) and [chi square video](#) for more practice with chi square analysis.

### t-Test

A t-test is commonly used to determine whether the mean of a population significantly differs from the mean of another population. This is useful if you need to compare the means of control and experimental groups. In most cases in AP Biology, you are going to assume that the data is parametric (follows a normal distribution) and that the two samples are independent of each other.

An excellent place to use the t-test in AP Biology would be in comparing the mean number of trichomes in the different fast plant generations (from the Artificial Selection Lab). We will use the t-test to analyze some simulated artificial selection lab data. Our null hypothesis for this test is that, "The mean number of trichomes in the generation 2 sample is the same as the mean of the generation 1 sample."

Trichome numbers Generation 1	Trichome Number Generation 2	
20	18	
8	56	
24	37	
12	22	
4	65	
31	33	
16	82	
16.42857	44.71429	mean
9.378293	23.64822	Standard deviation (S)
87.9524	559.2383	Variance (S <sup>2</sup> )

Note: The sample size (n) for both samples is equal to 7.

Thus, the complete equation for the t-test is

$$t_{\text{obs}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Calculation steps:

1. Calculate the mean of each sample population and subtract one from the other. Take the absolute value of this difference.

2. Calculate the standard error, SE. To compute it, calculate the variance (standard deviation squared) ( $S^2$ ) of each sample, and divide it by the number of measured values in that sample ( $n$ , the sample size). Add these two values and then take the square root.
3. Divide the difference between the means by the standard error to get a value for the t-statistic. This calculation yields “2.9417” for our data.
4. Compare the calculated value to the appropriate critical t-value in the table included below. Table 8 shows the critical values for different degrees of freedom at a significance value of 0.05. The degrees of freedom are calculated by adding the number of data points in the two groups combined, minus 2. In our situation that would be  $(7 + 7 - 2)$ . We should use 12 degrees of freedom for our trichome data. Note that you do not have to have the same number of data points in each group. **If the calculated t-value is greater than the appropriate critical t-value, this indicates that you should reject the null hypothesis and that you have enough evidence to support the hypothesis that the means of the two samples are significantly different at the probability value listed (in this case, 0.05). If the calculated t is smaller than the critical value, then you cannot reject the null hypothesis that there is no significant difference.**

**Table 8. Critical t-Values for a Significance Level  $\alpha = 0.05$**

Degrees of Freedom (df)	$t_{crit} (\alpha = 0.05)$
1	12.71
2	4.30
3	3.18
4	2.78
5	2.57
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
11	2.20
12	2.18
13	2.16
14	2.14
15	2.13
16	2.12
17	2.11
18	2.10
19	2.09
20	2.09
21	2.08
22	2.07
23	2.07
24	2.06
25	2.06
26	2.06
27	2.05
28	2.05
29	2.04
30	2.04
40	2.02
60	2.00
120	1.98
Infinity	1.96

In our example the t-test statistic of “2.9417” is greater than the critical value of “2.18”. This means that we should reject the null hypothesis. There is evidence that the mean number of trichomes in the generation 2 sample is different than the mean number of trichomes in the generation 1 sample.



[Table of Contents Link](#)

Just as there is with chi square analysis, there is another way to interpret the t-test data using p-values. Move along row 12 (12 degrees of freedom) in the t distribution table included below until you can find the t value of .2.9417. It is somewhere between the 0.02 column and the 0.01 column. This means that the p-value for this data is between 0.02 and 0.01. Remember, the probability of whether the results of an investigation differ from the null results by chance alone is called the p-value. A p-value of 0.05 means that there is a 5% chance that the difference between the observed and the expected data is a random difference and a 95% chance that the difference is real and repeatable—in other words, a significant difference. Therefore, if an investigator’s p-value is greater than 0.05, he or she would fail to reject the null hypothesis—that the difference between the observed results and the expected results is due to random chance and is not significant. **In our case, the p-value is less than 0.05. This means that we should reject the null hypothesis and that the differences between the means of the two generations is statistically different. There is less than a 5% chance that the differences between the means is due to chance.**

df	PROPORTION IN TWO TAILS COMBINED					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Table III of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. London: Longman Group Ltd., 1964 (previously published by Oliver and Boyd Ltd., Edinburgh). Adapted and reprinted with permission of the Addison Wesley Longman Publishing Co.

Visit the following links for more information about the t-test.

[AP Biology Statistics Teacher Guide](#)

[AP Quantitative Skills Guide](#)

[Bozeman Science t-test video](#)

[Performing a t-test in Google Sheets](#)

The t-test calculations are very laborious. In most cases, you won’t do them by hand. Below are directions for calculating a t-test with Excel, a TI calculator, and Google Sheets.

**Performing a t-test with Excel**

Excel calculates a T-test in a slightly different way. Rather than giving you the t value and comparing it to a table of critical values, Excel simply tells you the probability that the means are different simply due to chance, the “P value.” Follow these steps to calculate a P value using a t-test with Excel:

1. Create two columns, side by side, for the data of interest. Each sample’s data should be in separate columns
2. Click on another blank cell where you wish the P value to appear.
3. Then click “fx” on the Excel Formulas toolbar.
4. In the box, search for the "T test" function and choose “T.TEST” from the list. Hit OK. You will need to set the t-test parameters:

Period	Cats	Dogs
1	12	6
2	7	14
3	2	11
4	9	9
5	5	14
6	11	8
MEAN	8	10
SD	4	3

Formulas tab selected on ribbon. Insert Function dialog box open with T.TEST selected. P value cell contains "P value= =".

- For “Array1” highlight the data from one sample; for “Array2”, highlight the data in the second sample.
- Enter “2” in the box for “Tails.”
- Lastly, you will have to select the “Type” of t-test. For our purposes, we will mostly use type “2.” Although, if you are measuring the same sample at two points in time (for example before and after treatment) then you would have a type “1.”

5. After answering these questions click “OK” and the P value will appear. The P value will fall between zero and one.



Number of Student Pets in Each Class Period of IB Biology			
	Period	Cats	Dogs
	1	12	6
	2	7	14
	3	2	11
	4	9	9
	5	5	14
	6	11	8
	MEAN	8	10
	SD	4	3
P value=	=:D8,2,2)		

Function Arguments

T.TEST

Array1 C3:C8 = {12;7;2;9;5;11}

Array2 D3:D8 = {6;14;11;9;14;8}

Tails 2 = 2

Type 2 = 2

= 0.220095424

Returns the probability associated with a Student's t-Test.

Type is the kind of t-test: paired = 1, two-sample equal variance (homoscedastic) = 2, two-sample unequal variance = 3.

Formula result = 0.220095424

[Help on this function](#) OK Cancel

What does my P value mean? Excel gives the chance that the differences between the two samples are due to random chance alone. If Excel calculates a P value of 0.22, it means that there is a 22% likelihood that the difference in the means of your two data sets is due to random chance. If the calculated p-value is **.05 or less the differences between the two groups is significant and you should reject the null hypothesis**. If the P value is greater than 0.05, we can accept the null hypothesis and conclude that there is no significant difference between the two groups.

### A. Performing a T-test with the TI-83/84

1. Hit the STAT button on the calculator
2. Select option 4 to clear any past lists of data.
3. Select option 1 to EDIT your lists.
4. Enter your data for each group as List 1 and List 2
5. Hit STAT button and use the arrow key to move over to the TESTS option
6. Scroll down to option 4, the 2-sample T test and hit ENTER
7. Scroll to the bottom of the screen and hit ENTER over the CALCULATE option
8. Your results are given. Compare the calculate t-statistic to the critical value from the table 8. If the t statistic is greater than the critical value reject the null hypothesis. If the t statistic is less than the critical value, fail to reject the null hypothesis.

```

EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:Interval...
    
```

```

2-SampTTest
List1:L1
List2:L2
Freq1:1
Freq2:1
μ1:≠μ2 <μ2
Pooled:No
Calculate Draw
    
```

```

2-SampTTest
μ1≠μ2
t=3.799893167
P=6.5585123E-4
df=18
x̄1=35.5
x̄2=25.6
    
```

**Performing a t-test with Google Sheets**

1. Enter the data from your two samples in two separate columns.
2. Insert the formula: =TTEST(A1:A4, B1:B4, 2, 2). Replace A1:A4 with your data from the first sample and B1:B4 with your data from the second sample.
3. As with Excel, this calculation will give you the p-value, not the t-test statistic. If the p-value is less than 0.05, reject the null hypothesis. If the p-value is greater than 0.05, accept the null hypothesis.

**AP Biology****Biochemistry – Chemistry Basics****Atom**

The smallest *unit of an element* that maintains the chemical properties of the element.

**Subatomic Particles**

**Proton** – Subatomic particles that carry a positive charge. They are located in the nucleus of an atom.

The number of protons *never* changes in an element. **The number of protons determines the identity of the element and defines the atomic number of the element.**

**Neutron** - These particles carry NO charge (are neutral). They are also located in the nucleus of an atom and are similar in size to the proton. The number of neutrons *can* change. (Atoms of the same element with different numbers of neutrons are called **Isotopes**.)

**Electrons** - These subatomic particles carry a negative charge. They are located in the “Electron cloud”. The electrons are attracted to the positive protons in the nucleus, but can move within the electron cloud. The number of electrons associated with an atom *can* change. (Atoms with different numbers of electrons than the *normal amount* for that element are called **Ions**.)

**Molecule**

Two or more *atoms* that are covalently bonded together.

**E levels or e- shells** – Where the electrons are located within an atom or molecule.

*Adding energy* to the electrons makes them move farther out, away from the nucleus; *losing energy* causes them to move inward, toward the nucleus.

**Valence Shell**- Where the *outer most electrons* are located on an atom.

**Valence e-** - Refers to the outer most electrons. (**These are the most important for chemical bonds and the chemical properties of an element or molecule.**) Most elements need 8 valence electrons (an octet) in order to be chemically stable. Atoms react with other elements in order to obtain a total of 8 valence electrons and to become chemically stable.

Chemical Bonds (These occur between atoms.)

**Covalent Bonds**

A type of intramolecular bond.

Results from the *sharing of valence electrons* between atoms.

Atoms held together by covalent bonds are called molecules.

**Polar molecules** carry a slight *electrical charge* at opposite poles (poles refers to the “ends” of the molecule) and **non-polar** molecules *do not have an electrical charge*.

### Electronegativity

Refers to an atom's *desire to acquire* electrons.

Hydrogen is the least electronegative atom.

Oxygen and Nitrogen are the most biologically important molecules with a high electronegativity.

Molecules which contain oxygen and nitrogen are likely to be polar.

### Ionic Bonds

Ionic bonds form between metal and non-metal atoms.

They form when the metal atoms loses electrons and the non-metal atom gains electrons.

Both atoms do this in order to have 8 valence electrons.

**Compounds held together by ionic bonds are called salts.**

Ionic Bonds are very strong when the compound is dry.

Ionic bonds are easily broken in water. This is why many salts dissolve easily in water to form ions.

**Cations** – Ions which possess a *positive charge* because they have more protons than electrons. The metal atoms in a salt typically become cations.

**Anions** – Ions which possess a *negative charge* because they have more electrons than protons. The non-metal atoms in a salt typically become anions.

### Hydrogen Bonds

*Hydrogen Bonds are fairly weak (compared to covalent and ionic bonds)* intermolecular attractions that occur between polar molecules. They are often depicted as dots in chemical diagrams.

Hydrogen bonding is very important in water due to its polar nature.

### Van der Waals Interactions or London Dispersion Forces

These are *temporary intermolecular attractions*. (Usually a fraction of a second.)

These *interactions* are “created” when electrons clump on one side of an atom making that side temporarily “negative” and the other side “positive”. This allows other charged particles to attach *momentarily*. The electrons eventually unclump and the van der waals interactions disappear.

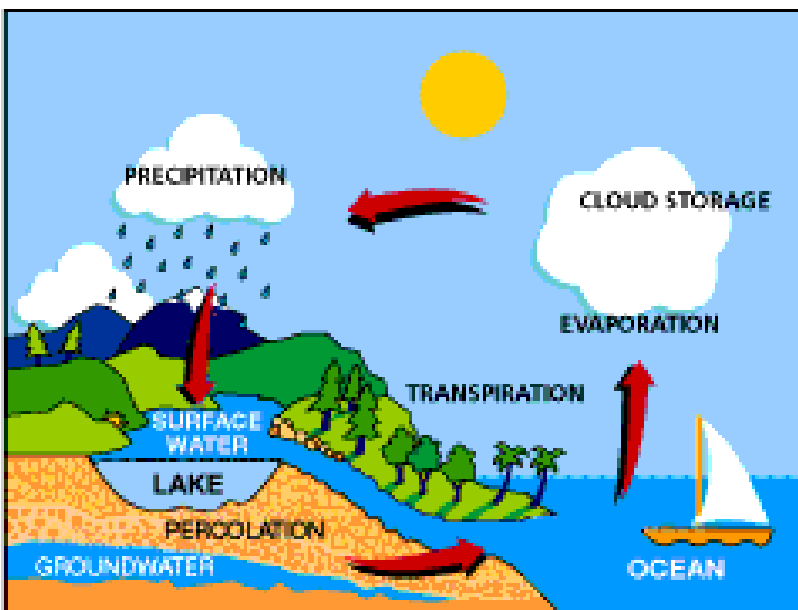
## AP Biology Biochemistry of Water

Water *supports life* on Earth.

Water makes up over 70% of the bodies of most organisms.

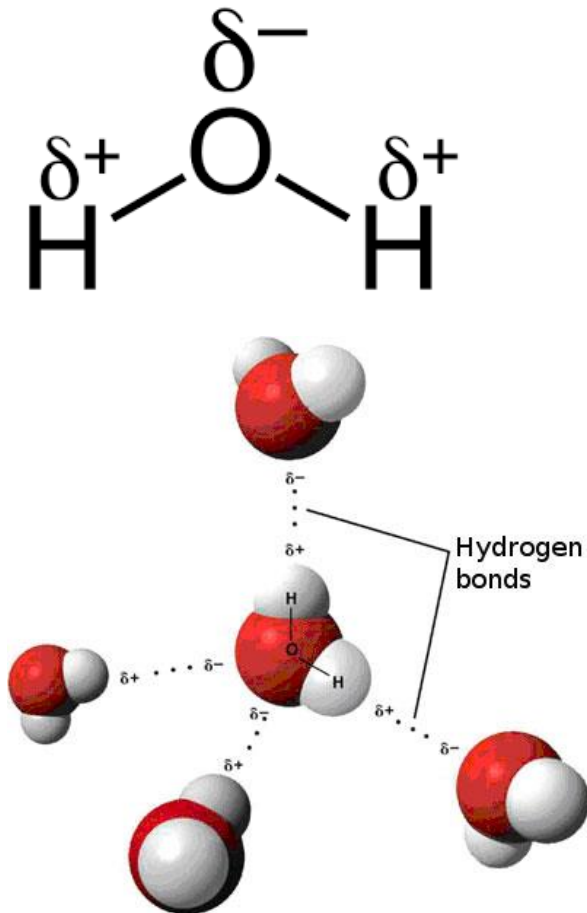
**Biogeochemical Cycles**--These refer to the cycling of matter.

**Water cycle** – Water vapor is *generated* by the sun causing the evaporation of water from oceans, lakes, rivers, trees, etc.... This water vapor rises and condenses to form rain or snow (referred to as precipitation) and is returned to the land or ocean. Eventually, the water that lands back on the land, makes its way to plants or rivers and streams that lead back to the oceans. Plants take in the water and use it for photosynthesis but also can lose it in the form of transpiration to the air.



Water is a **Polar** Molecule.

Because of the high electronegativity of oxygen and the low electronegativity of hydrogen, the oxygen end of the water molecule has a slight negative charge, while the hydrogen end has a slight positive charge. The water molecule's shape is said to be "bent". This shape means that one side of the molecule "the hydrogen side" is positive, while the other side "the oxygen side" is negative. The polar water molecules form hydrogen bonds with each other. These bonds affect many of the biologically important properties of water.



### Important Properties of Water for Biology

#### Water has a high specific heat.

Specific heat refers to the amount of heat needed to change the temperature of 1 gram of a substance by 1 degree. Water's specific heat is very high due to the hydrogen bonds that hold the water molecules together.

Due to its high specific heat, water is excellent for helping to maintain a constant internal body temperature. Large bodies of water have fairly constant temperatures. Water also helps to moderate the air temperatures of land masses located near large bodies of water.

#### Water is an excellent solvent.

Water is often referred to as the "Universal Solvent". This means that it is great at dissolving other materials (solutes).

Due to its polarity, water is best at dissolving salts and polar molecules.

This is important in Biology, because it allows vital nutrients and gases to be dissolved and transported through the body and through the ground.

**Water has a high heat of vaporization.**

Heat of vaporization refers to the amount of heat needed to convert a liquid to a gas. Due to the many hydrogen bonds which hold water molecules together, the heat of vaporization is very high for water. When the water does eventually evaporate it carries heat away with it.

This is important in Biology, because it allows processes like sweating and transpiration to cool off organisms through the process of evaporative cooling.

**Water is cohesive and adhesive.**

Due to its polarity, water molecules stick to each other (cohesion) and to other polar molecules (adhesion). This property allows transpiration, the movement of water through the xylem of plants, to move water from the ground, through the plant, and eventually out into the air.

**Water expands as it freezes.**

Like most other compounds, the volume of water contracts as it cools down, but unlike most other materials, the volume of water begins to expand after it cools to temperatures below 4 degrees Celsius. This causes ice to be less dense than liquid water. This means that ice floats on liquid water.

This happens because the **HYDROGEN** bonds force the chains of water molecules further apart as the molecules cool down and slow down. This increased volume causes a decrease in density. This property is important to life, because as bodies of water freeze, the ice floats and insulates the liquid water underneath. This allows aquatic life to survive long, cold winters.

## AP Biology

### Biochemistry: Carbon Properties

#### Organic Chemistry

*Branch of science* dealing with the element carbon and its many properties.

Most of the compounds found in living things, other than water, are organic.

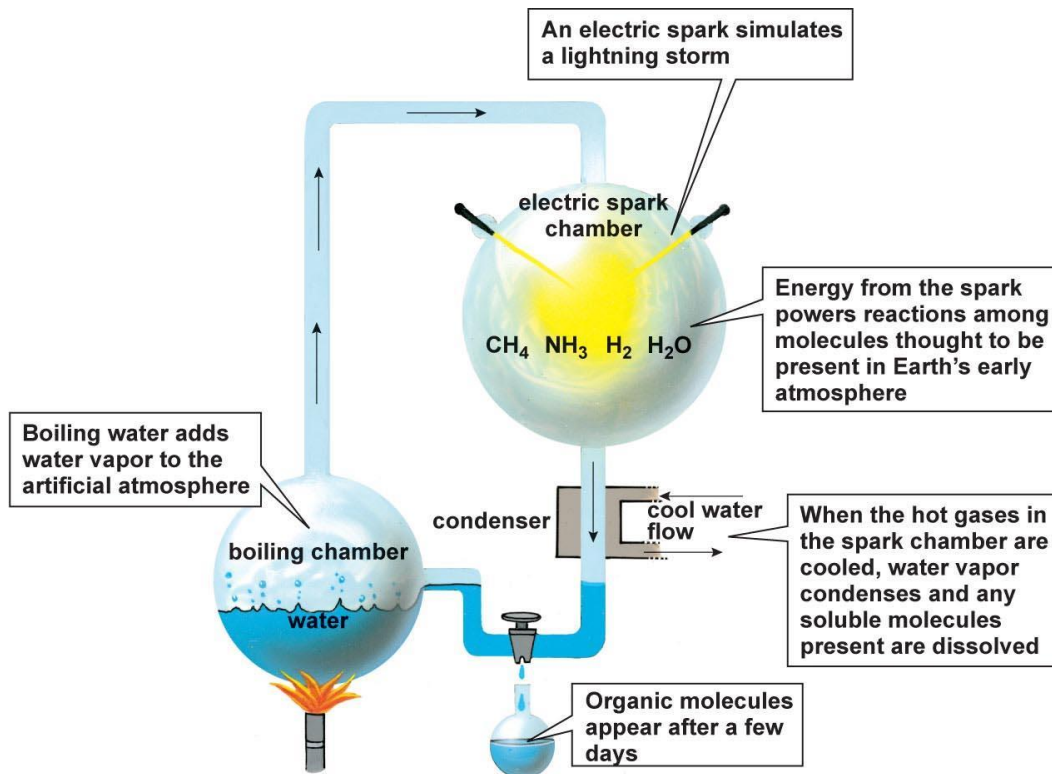
About 30% of an organism's *dry weight* (**Biomass**) is composed of Carbon found in the body's organic molecules.

Carbon helps to make the 4 major groups of organic macromolecules: *Carbohydrates, Lipids, Proteins, and Nucleic Acids.*

The *original source* for Carbon in *all life forms* is Carbon Dioxide. (CO<sub>2</sub> - Photosynthesis)

#### Miller/Urey Experiment (Took place in 1953.)

Miller/Urey took *inorganic substances* that were thought to have been present in Earth's early atmosphere (H<sub>2</sub>O vapor, H<sub>2</sub>, NH<sub>3</sub>, CH<sub>4</sub>) and created *organic* amino acids and oils. (CO<sub>2</sub> and CH<sub>4</sub> are *not* considered organic compounds, even though they contain Carbon.) Miller wanted to show that organic molecules, which are necessary for life, could be created by non-living things.



## Important Properties of Carbon for Biology

### Carbon has 4 valence electrons

Since carbon has 4 valence electrons, it can form four covalent bonds. This allows carbon to bond with a variety of molecules and form molecules with an almost infinite variety of shapes and functions.

### Carbon atoms are small

Because carbon is a small atom, its valence electrons are near the nucleus. This means that when carbon forms covalent bonds, the bonds are strong. This makes carbon an excellent building material for life.

### Carbon is abundant on Earth

Carbon isn't the most abundant element on Earth, but there is a lot of carbon found on the planet. This makes it a good choice for the building block of life.

## Organic Macromolecules

### Polymers

These are large molecules that are formed by combining/bonding *individual units* called **monomers**.

The monomers are linked together by *covalent bonds*. Remember that covalent bonds are strong.

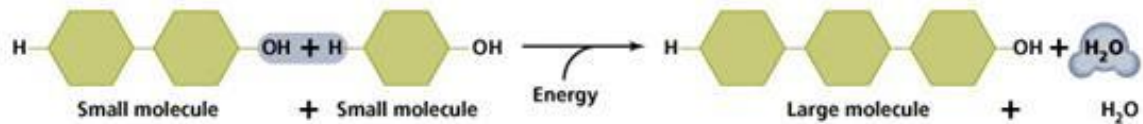
Macromolecules are formed by **Dehydration Synthesis (also called Condensation) Reactions**.

During these reactions, monomers are covalently bonded together.

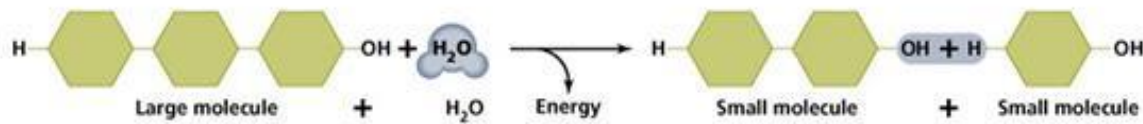


The reactions release water as a byproduct.

Macromolecules are broken apart into individual monomers by **Hydrolysis reactions**. (“lysis” means “to split.”). In these reactions, water is used as a reactant.



(a) Dehydration synthesis



(b) Hydrolysis

## Groups of Organic Macromolecules

### Carbohydrates

Carbohydrates are *sugars*. In living things, carbohydrates serve as sources of quick energy and as structural/building materials. Most carbohydrate names end in the letters “-ose”.

**Monosaccharides** - Are the monomers or “building blocks” of carbohydrates. (“sacch” means “sugar”.) Glucose, fructose, galactose, ribose, and deoxyribose are common monosaccharides. The general molecular formula for a monosaccharide is  $(C_NH_{2N}O_N)$ .

**Polysaccharides** - Are the polymers. These compounds are formed by the bonding together of several monosaccharides. Biologically important polysaccharides include:

**Starch** – Used as an energy storage molecule in plants.

**Glycogen** – Used as an energy storage molecule in Animals.

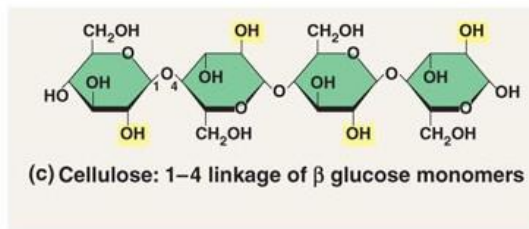
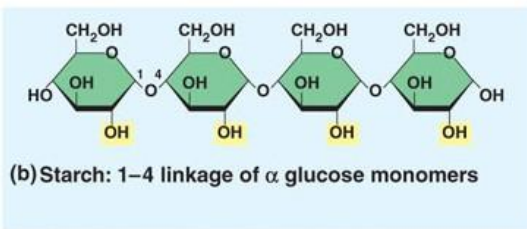
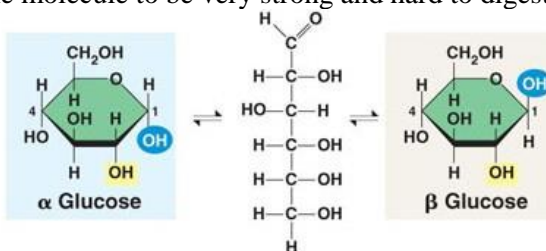
**Cellulose** – Used as a structural component of plant cell walls.

Cellulose is the most abundant organic compound on Earth. Due to the hydrogen bonds that occur in cellulose, it is very hard for most organisms to digest.

**Chitin** – Used as a component of the exoskeleton of some animals and also fungal cell walls.

Although starch, cellulose, and glycogen are all polymers of glucose, the molecules have different shapes. Because of the orientation of the glucose monomers, many hydrogen bonds form in cellulose and cause the molecule to be very strong and hard to digest.

(a)  $\alpha$  and  $\beta$  glucose ring structures



Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

## Lipids

These macromolecules are *fats, oils, waxes, and steroids*.

Most lipids are *hydrophobic* molecules. They contain little oxygen and are mostly nonpolar in nature.

Two Main parts of a lipid:

**Fatty Acid**—long chain of carbon and hydrogen atoms.

3 Carbon **Glycerol molecule** (alcohol) to *hold the whole molecule together*.

Lipids use a *covalent bond* called an **Ester Linkage** to hold the fatty acids and glycerol together.

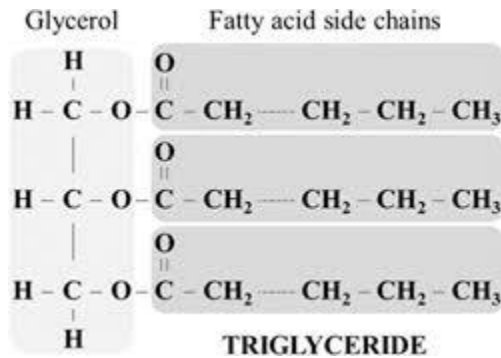
Major Types of lipids:

**Triglycerols or Triglycerides** (These are your basic fat or oil.)

**Saturated fats**--These fatty acids are saturated with Hydrogen atoms. The molecule has *no open bonds* to put any more Hydrogen on. There are no carbon—carbon double or triple bonds in these chains. Saturated fats are *solids at room temp* and they usually are *associated with animals*. Saturated fats contribute to coronary artery disease.

**Unsaturated fats**--These fatty acid chains contain some carbon-carbon *double or triple bonds* that “could be broken” to add more Hydrogen to the fatty acid. Saturated fats tend to be *liquids at room temperature* and they usually are *associated with plants* (vegetable oil, sunflower oil, or peanut oil). Unsaturated fats are typically viewed as a healthy part of the diet.

**Polyunsaturated fats**--These fats have *many double or triple bonds* in their fatty acid chains.  
**Hydrogenated or Trans fats**. These are *oils turned solid fats* by **adding Hydrogen** and by *breaking the double or triple bonds* in the fatty acids. This was done in the past to change the texture, stability, and shelf life of fats. Trans fats were common in processed foods. They have been shown to increase LDL cholesterol levels and to contribute to coronary artery disease.

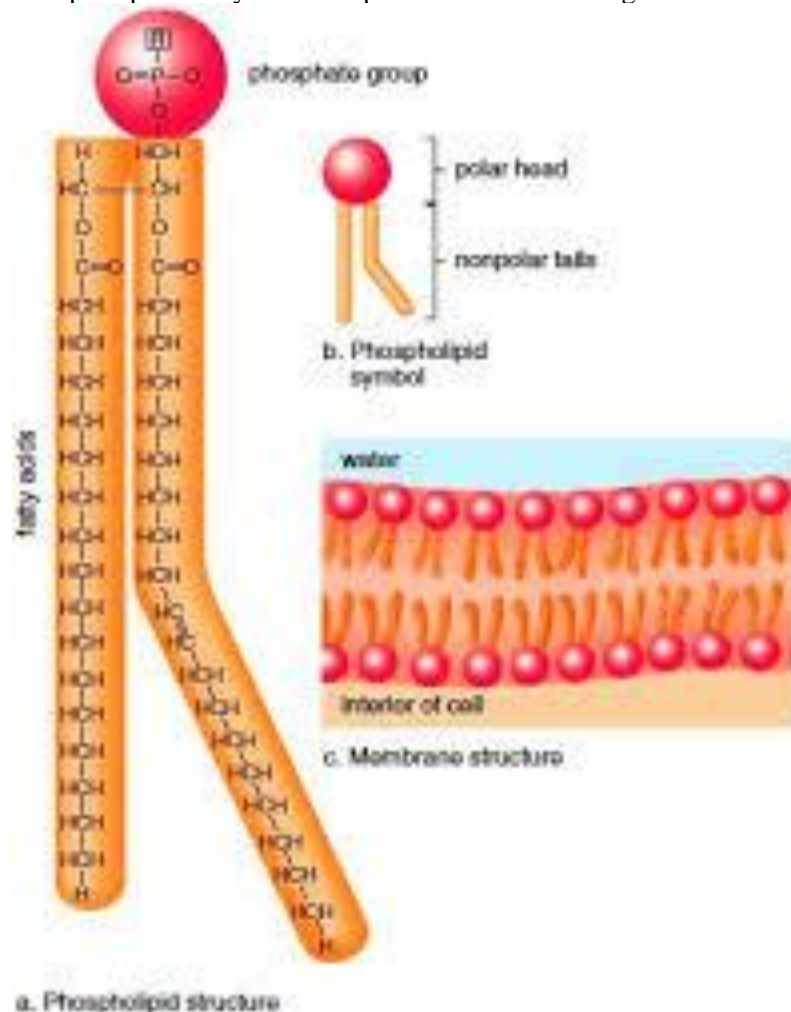


### Phospholipids

These molecules *replace* a single fatty acid chain from a triglyceride with a single Phosphate ion. The phosphate portion of the molecule is **Hydrophilic** because of the charge on the phosphate ion.

The remaining fatty acids chains are hydrophobic. They are mostly composed of long chains of carbon atoms that are also bonded to hydrogens. They are neutral and are not attracted to water.

Phospholipids are said to be **amphipathic** since they have both polar and nonpolar sides. Phospholipid Bi-layers are important for the building of *cell and organelle membranes*.



### Steroids

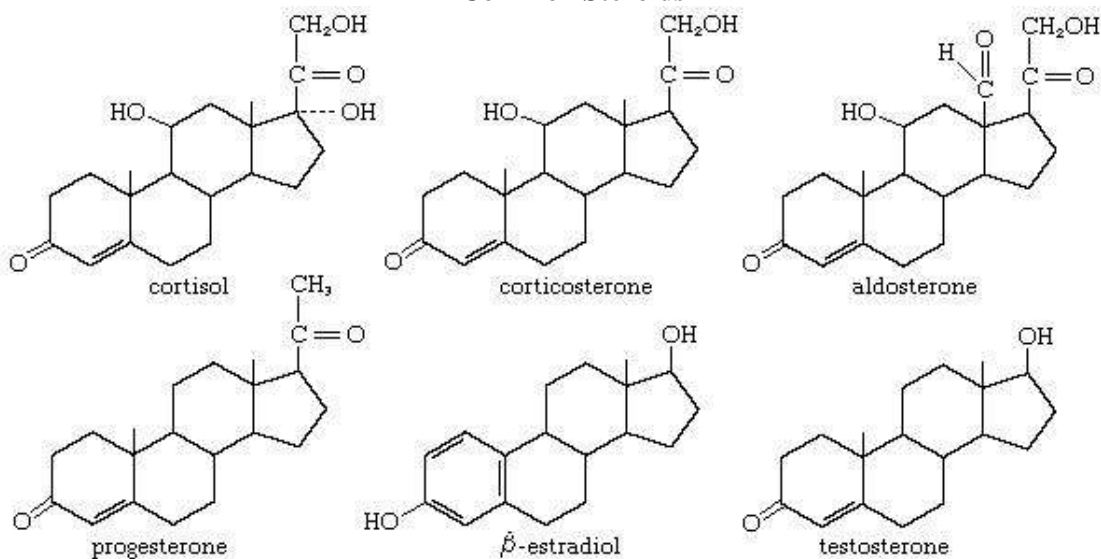
[Table of Contents Link](#)

A steroid is a lipid composed of 4 carbon rings. Common steroids found in the body include testosterone, estrogen, progesterone, and cholesterol.

What makes them different from each other are the *attached functional groups*. These functional groups help determine the *function* of the steroid. Most of the steroids are built from cholesterol. Steroids function in the body as cell signals/hormones. Due to their lipid/nonpolar nature they can penetrate the cell membrane and bind to intracellular receptors. They often act to regulate the expression of certain genes.

Cholesterol is an important component of the cell membrane. It helps with *cell membrane flexibility*.

### Common Steroids



### Proteins

Proteins make up greater than 50% of an organism's dry weight or *biomass*.

The names of most proteins end in "lin" while the names of most enzymes (which are composed of proteins) usually end in "ase".

Proteins are large macromolecules which are composed of *monomers* "building blocks" called **Amino Acids**. There are 20 different Amino Acids used by living things to make proteins. Proteins and enzymes usually have hundreds to thousands of Amino acids in their structure.

**Amino Acids** have 4 different parts to them:

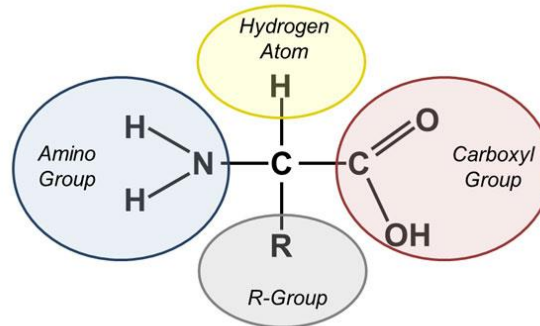
**Carboxyl end** (COOH) – This part *acts as an acid* because it can release a hydrogen ion.

**Amine end** (NH<sub>2</sub>) – This end *can act as a base* by accepting a Hydrogen ion.

**Alpha (α) Carbon** – This is the *central carbon atom* to which all of the other functional groups are attached.

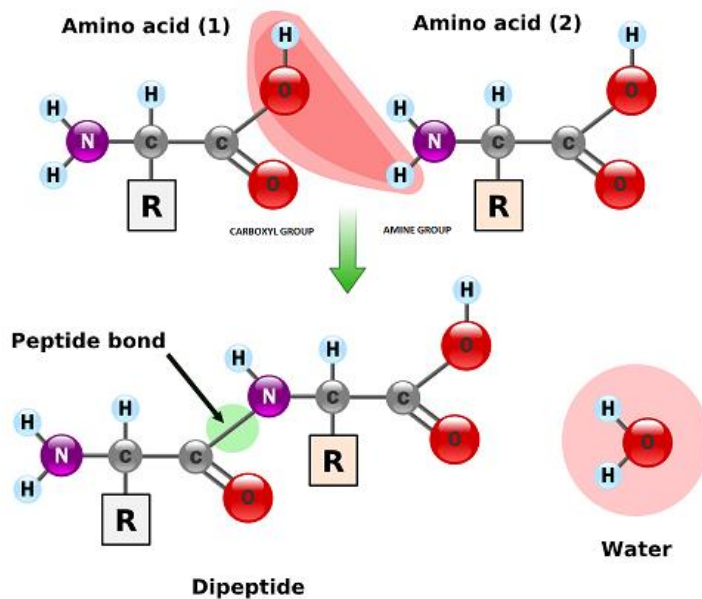
**R group** - This *part* is the only part that is different in the 20 different amino acids. The R group gives different amino acids their distinctive properties.

### Structure of an Amino Acid



Individual Amino Acids (monomers) are bonded together by a *covalent* bond called a **peptide bond**. The peptide bond is a covalent bond and is strong and hard to break.

In order to create a peptide bond, the *amino end* of one amino acid is positioned to combine with the Carboxyl end of the second amino acid. The amino acids are joined during *dehydration synthesis reactions* (also known as “*condensation reactions*”).



Two amino acids bonded together are referred to as a dipeptide. When more than two are bonded, the structure is referred to as a polypeptide chain.

A protein is typically made from several polypeptide chains that are wrapped together into a large 3-D unit.

The function of a protein is largely determined by its shape/structure.

The information included below discusses the four levels of protein structure.

### Levels of Protein Structure

**Primary Structure** (Represented by the symbol - 1' .)

This refers to the **sequence of bonded Amino Acids (which amino acids are present and what order are they in)**. **The primary structure of each protein is coded for by DNA.**

**A mutation in the DNA can change the primary structure of a protein. This change can vastly change the overall shape of the protein and cause it to lose its function.**

**Secondary Structure** (2' )

**This refers to the shape of specific sections of each polypeptide chain.**

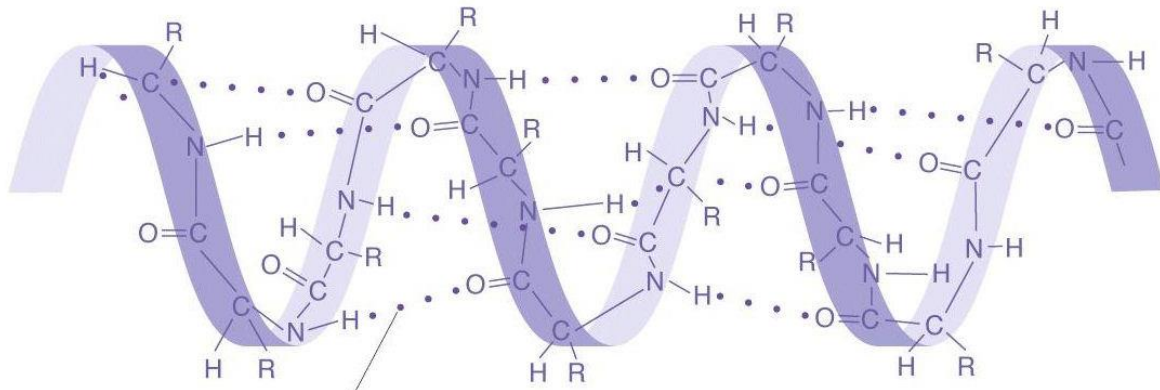
**HYDROGEN** bonding between polar amino acids allows for overlapping and coiling to occur in the “folding” of the protein into its 3-D shape. **All proteins must be “folded” in order to work. The polarity/nonpolarity of the “R” groups of the protein’s amino acids determine how the protein will fold.**

**For proteins that are built by ribosomes located on the rough ER, this folding takes place within the Rough Endoplasmic Reticulum. These proteins are usually exported from the cell through the process of exocytosis (secretion).**

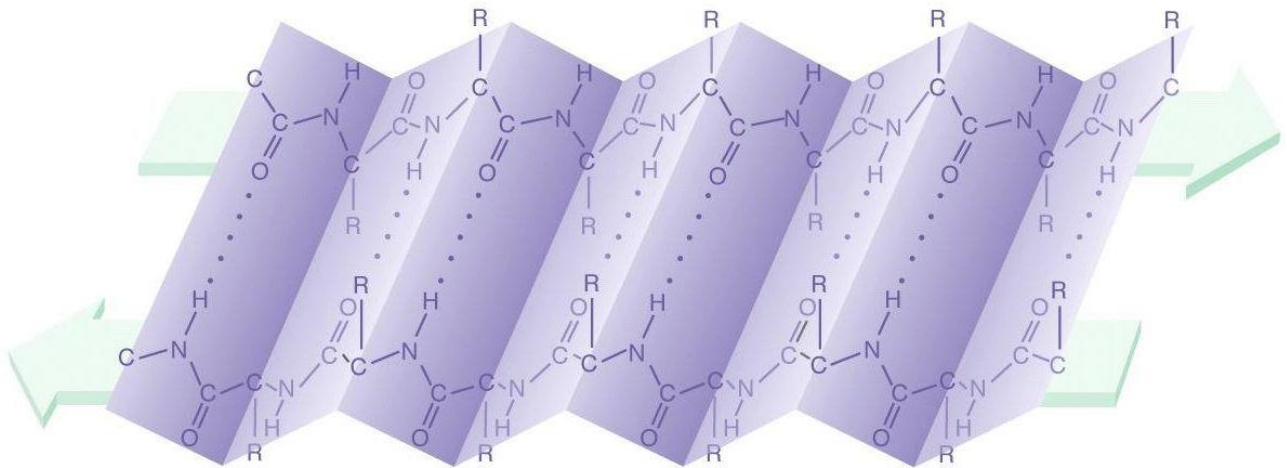
**Other proteins are constructed by ribosomes which float freely in the cytosol. These proteins typically remain within the cell in which they were made. The folding of these proteins usually takes place within a chaperonin. Chaperonins are protective structures that allows proteins to fold *without water present*.**

Two common types of secondary structure are the alpha helix and the beta pleated sheet.





Hydrogen bonds between amino acids at different locations in polypeptide chain  $\alpha$  helix



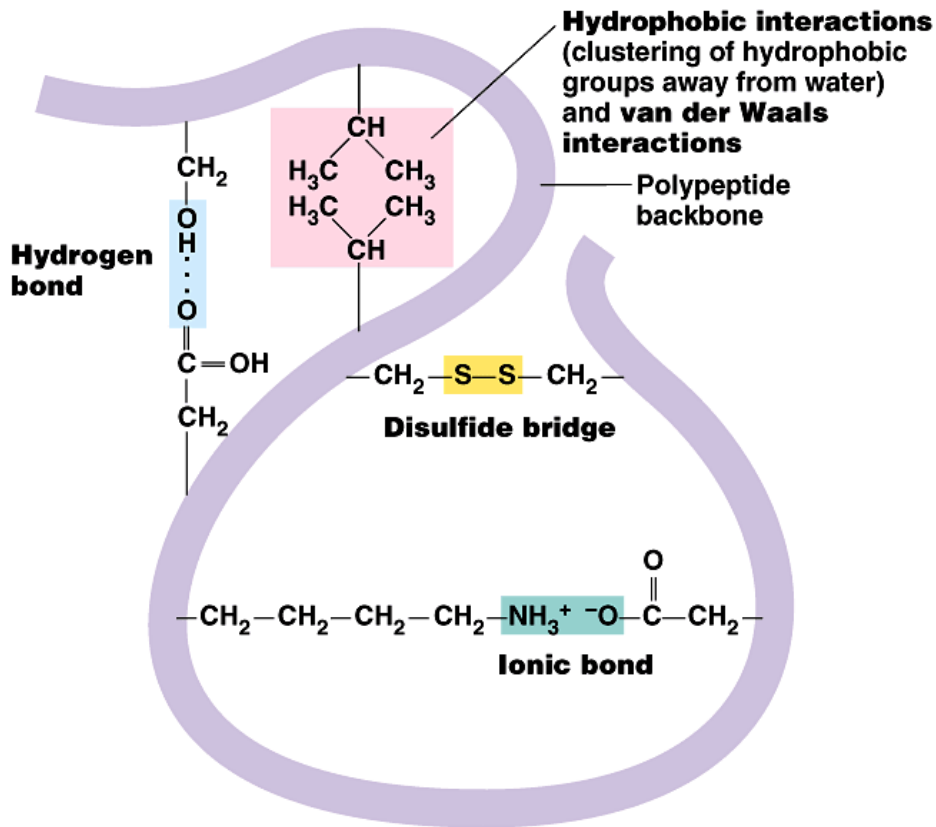
Pleated sheet

### Tertiary Structure (3') ("Tert" means "third")

This refers to the overall shape of each individual polypeptide chain.

Di-sulfide bridges help *stabilize* the protein's folded structure. These bonds form when cysteines (a type of amino acid) are near each other in a polypeptide. These cysteines form a disulfide bond or bridge which is much stronger than a hydrogen bond.

The tertiary structure is also partially determined by ionic interactions between certain "R" groups and by hydrophobic interactions between certain R groups and water.



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

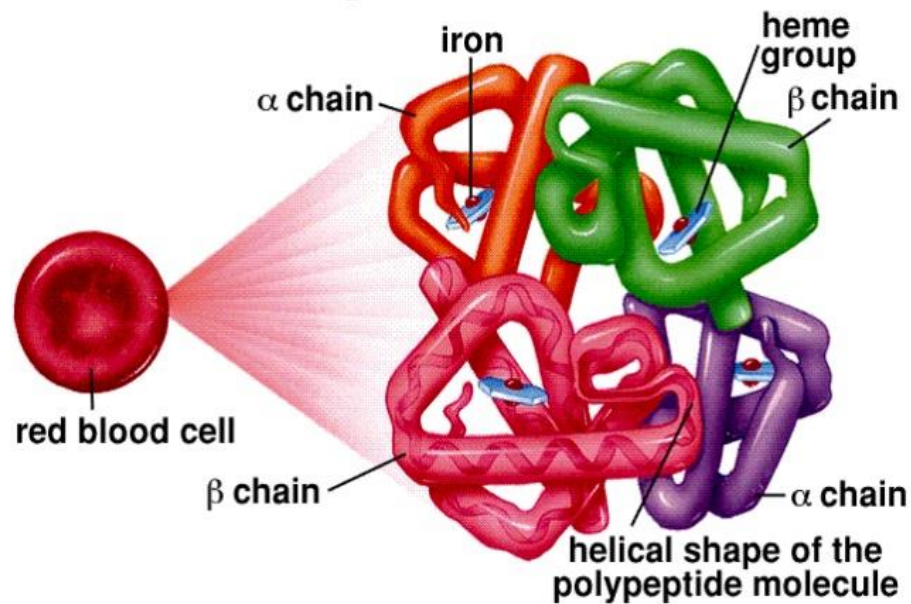
### Quaternary Structure (4')

This is when two or more polypeptides are woven together to form the overall structure of a protein.

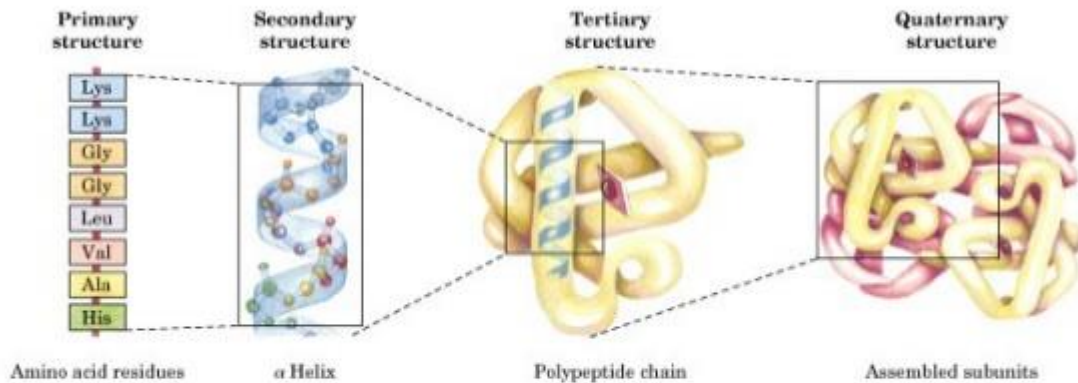
**Example: Hemoglobin** -has four polypeptides woven together to make it.



## Hemoglobin Molecule



## 4 levels of protein structure



- Primary – sequence of amino acids
- Secondary – interactions between adjacent amino acids
- Tertiary – 3D folding of the polypeptide
- Quaternary – arrangements of multiple polypeptides

### Denaturation

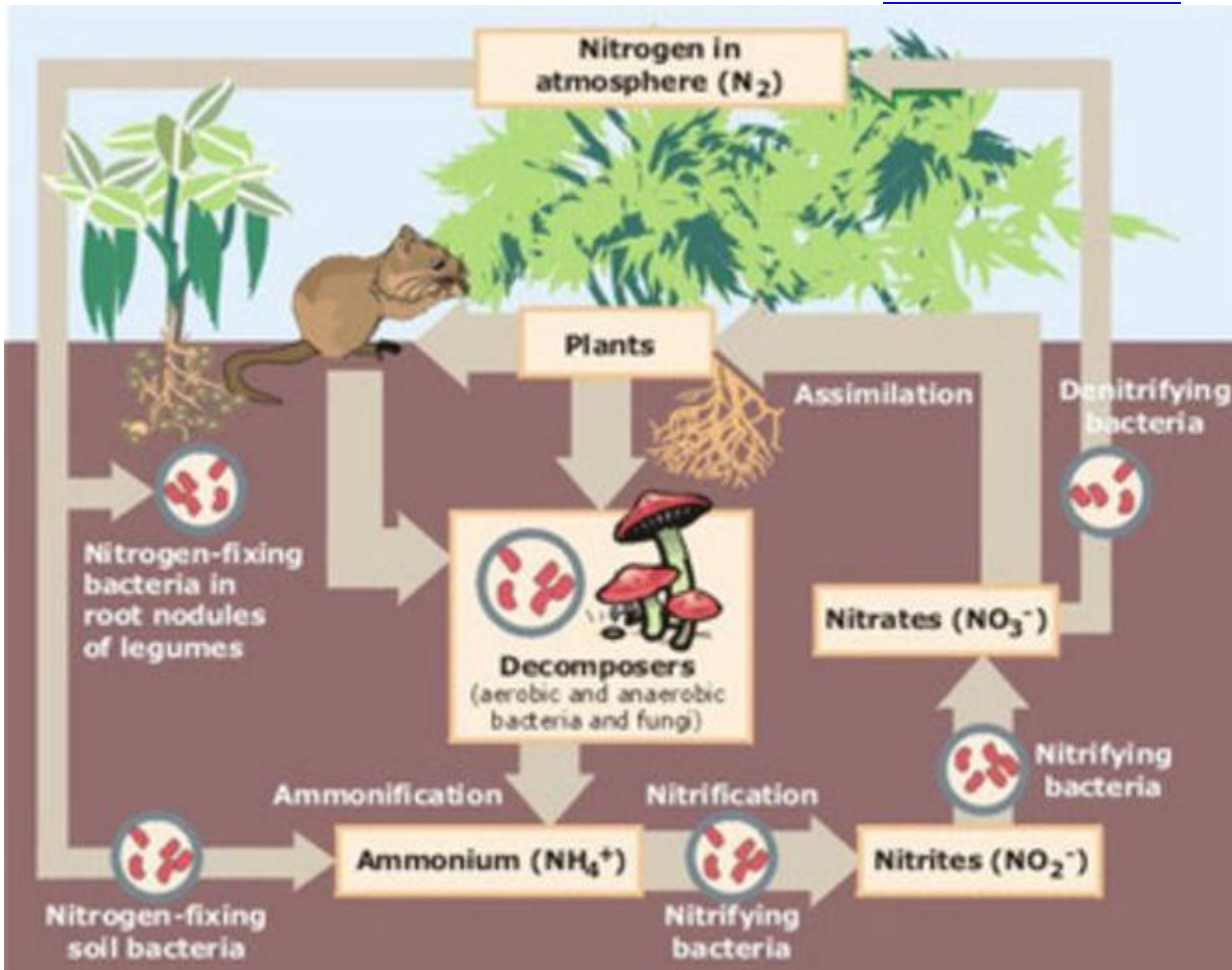
The “unraveling” or “unfolding” of a protein or enzyme causing it not to function. Denaturation can be caused by *pH changes, salt concentration changes, and increases in temperature.*

*Denaturation usually disrupts the secondary and tertiary levels of protein structure by affecting either the hydrogen bonds or disulfide bonds which stabilize the structure.*

### Nitrogen Cycle

**Although there is plenty of nitrogen found in the air, the amount of usable nitrogen available for life on Earth is limited. Since the nitrogen atoms in nitrogen gas are held together by triple covalent bonds, atmospheric nitrogen is difficult for life forms to break apart and use. The nitrogen cycle is the process via which nitrogen moves from the atmosphere into living things and ultimately back to the atmosphere. Cycles like the nitrogen cycle, the carbon cycle, the water cycle, and the phosphorus cycle are collectively known as the biogeochemical cycles. These processes continuously cycle elements throughout the Earth.**

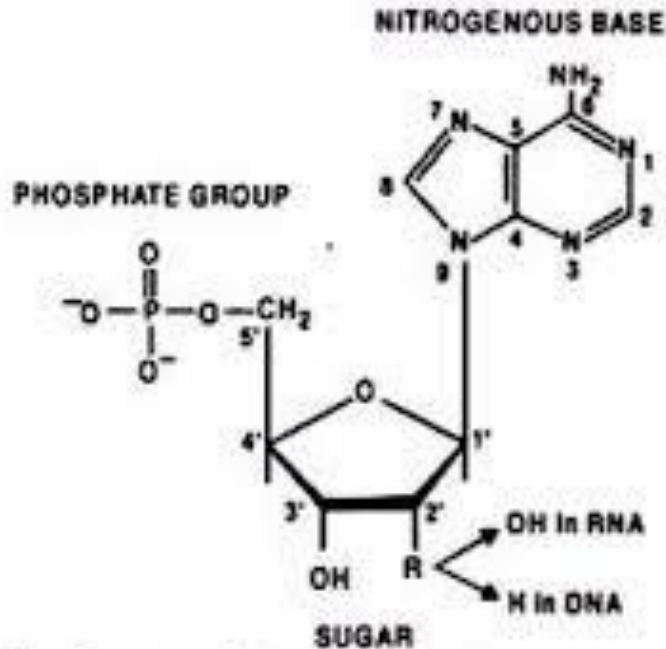
Nitrogen is an essential element for the construction of proteins and amino acids. Nitrogen is also an important component of DNA, RNA, and ATP. Some nitrogen is removed from the air by rainwater. Remember, water is the universal solvent, so the gas is dissolved in the rain. The Nitrogen in the water can be consumed by **Nitrogen Fixing bacteria**, in the soil, that convert it into Ammonium ions ( $\text{NH}_4^+$ ). This process is referred to as **Nitrogen Fixation**. The Ammonium ions can then be absorbed by plants to help make proteins and DNA or RNA. Some Ammonium in the soil is also consumed by **Nitrifying Bacteria**, and converted to Nitrite ( $\text{NO}_2^-$ ) first and then ultimately into Nitrate ( $\text{NO}_3^-$ ). This process is called **Nitrification**. The Nitrates are also absorbed by the plants, just as was the Ammonium ions. Other bacteria, called denitrifying bacteria, in the soil can also absorb the nitrates and convert them back into Oxygen gas ( $\text{O}_2$ ) and Nitrogen gas ( $\text{N}_2$ ). These molecules are both returned to the air to be used again. This process is called **Denitrification**. As plants are eaten by animals, the Nitrogen travels through the food chain. When all life forms die, the bodies decompose and create Ammonia ( $\text{NH}_3$ ), which is why they stink. The Ammonia is converted by bacteria into Ammonium to be used again by plants and bacteria. This conversion is called **Ammonification**. Some Nitrogen is also released by animals in their urine. It too undergoes Ammonification.



### Nucleic Acids

Nucleic Acids function to store genetic information and/or to store and transfer energy. Common nucleic acids found living organisms include: DNA, RNA, ATP, cAMP, NADH, and NADPH.

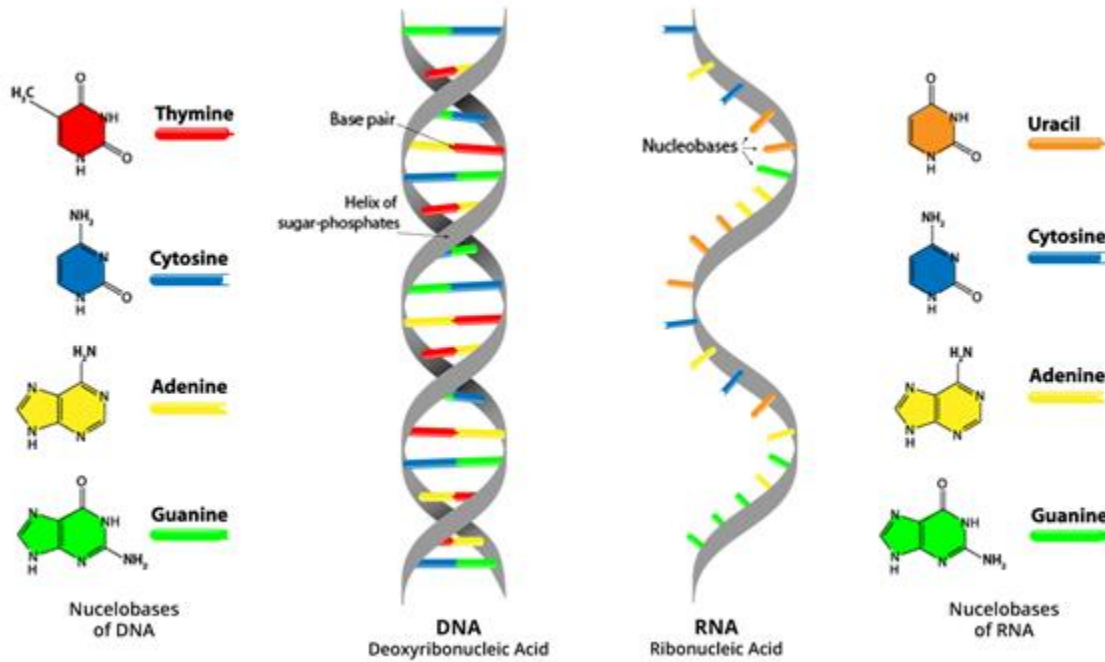
*The monomers of nucleic acids* are called **nucleotides**. A nucleotide consists of a 5 carbon (pentose) sugar bonded to a phosphate group and a nitrogenous base.



**Fig. 6.4** Structure of a nucleotide. The nucleotide subunits of DNA and RNA are made up of three elements : a five-carbon sugar, an organic nitrogenous base (adenine is shown here), and a phosphate group.

**DNA and mRNA are both polymers.** DNA and RNA are the primary sources of genes and hereditary information.

1. DNA has *Deoxyribose* as its 5 Carbon sugar. DNA is double stranded. In eukaryotic cells, DNA is always stored inside a nuclear membrane or envelope. **DNA's function is to code for proteins.** The sequence of the nitrogenous bases in the DNA determines the order of the amino acids in each of the body's proteins.
2. RNA has *Ribose* as its 5 Carbon sugar. RNA is single stranded. There are several types of RNA. Messenger RNA (mRNA) is made from the DNA template during the process of transcription. mRNA's job is to transmit the protein building directions from the DNA in the nucleus to the ribosomes in the cytoplasm. Transfer RNA's (tRNA) job is to deliver and place the appropriate amino acids into the proteins that are built by the ribosomes. Ribosomal RNA (rRNA) is one of the main building components of the cell's ribosomes.



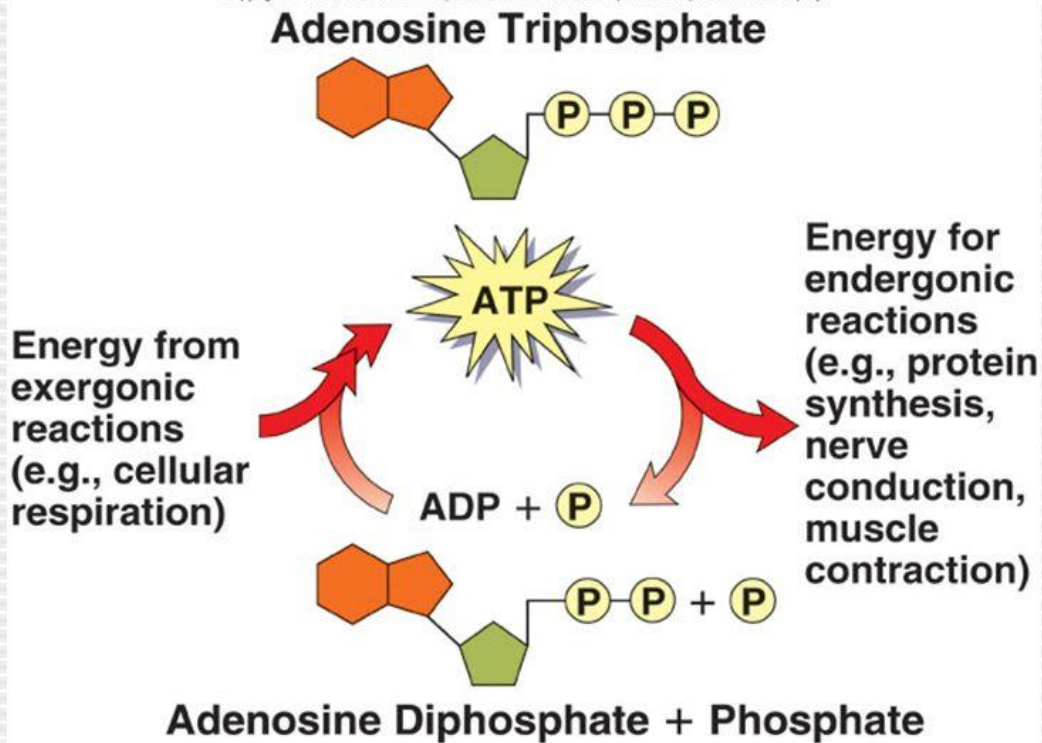
Scientists can now sequence the nucleotide/nitrogenous bases found in genes of an organism and compare this sequence to the sequence of the same gene found in another organism. The more similar the two sequences are, the more related the two organisms are.

**ATP (Adenosine Triphosphate)** is another important nucleic acid. An ATP molecule is composed of a single nucleotide which consists of the sugar (ribose) bonded to a nitrogenous base (always adenine), and three phosphate groups. ATP's role in the body is to store and transfer energy. ATP is made during the process of cellular respiration. It functions to power almost every activity that occurs in the cell.



# The ATP cycle

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



## Phosphorus Cycle

The phosphorus cycle is another important biogeochemical cycle. Phosphorus is an important component of DNA, RNA, ATP, and bone. Most of the Earth's phosphorus is found in rock. As the rock weathers, some of the phosphorus is released into the soil. Some dissolves into the water as the rains pass through the soil. This phosphorus makes its way into bodies of water and is available for producers (phytoplankton) to use to make organic compounds such as phospholipids, DNA, RNA, ATP, etc... Plants can also retrieve the phosphorus directly from the soil and use it to make organic compounds. When organisms die, decomposers break down the bodies and return the phosphorus to the soil so that it can be reused.

