3-D Molecular Designs
. . . where molecules become real ™

Amino Acid Starter Kit

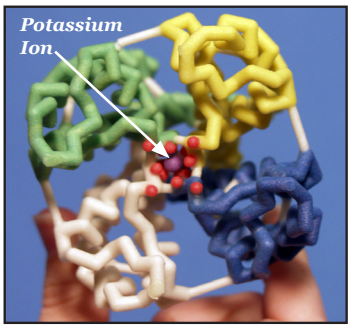
Key for Student Handout 1

Amino Acids - Building Blocks of Proteins

Introduction

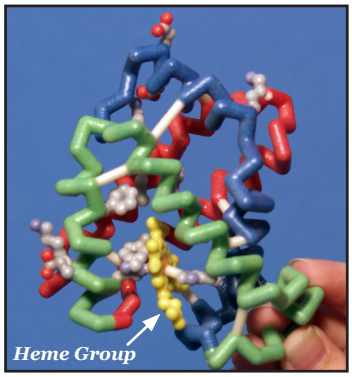
Proteins are more than an important part of your diet. Proteins are complex molecular machines that are involved in nearly all of your cellular functions. Each protein has a specific shape **(structure)** that enables it to carry out its specific job **(function)**.

A **core idea** in the life sciences is that ***there is a fundamental relationship between a biological structure and the function it must perform.*** At the macro level, Darwin recognized that the structure of a finch's beak was related to the food it ate. This fundamental structure-function relationship is also true at all levels below the macro level, including proteins and other structures at the molecular level. *For two examples of proteins and their functions, see the photos and cutlines at the right.*



The potassium channel (above) spans cell membranes and regulates the passage of potassium ions in and out of cells. It folds into a "pore" for the potassium ion to pass through.

The β-globin protein (below) transports oxygen in blood. It accomplishes this with the heme group (yellow structure in photo) in which an iron atom binds to O2. Other proteins perform other functions.



In this activity, you will explore the structure of proteins and the chemical interactions that drive each protein to fold into its specific structure, as noted below.

- Each protein is made of a specific sequence of **amino acids**. There are 20 amino acids found in proteins.
- Each amino acid consists of two parts — a **backbone** and a **sidechain**. The backbone is the same in all 20 amino acids and the sidechain is different in each one.
- Each sidechain consists of a unique combination of atoms which determines its 3D shape and its chemical properties.
- Based on the atoms in each amino acid sidechain, it could be ***hydrophobic, hydrophilic, acidic (negatively charged), or basic (positively charged)***.
- When different amino acids join together to make a protein, the unique properties of each amino acid determine how the protein folds into its final 3D shape. The shape of the protein makes it possible to perform a specific function in our cells.

1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

3-D Molecular Designs
. . . where molecules become real ™

Amino Acid Starter Kit

Chemical Properties Circle & Amino Acid Chart

Hydrophobic and Hydrophilic Properties

What do you think hydrophobic means? Separate the word 'hydrophobic' into its two parts — hydro and phobic. Hydro means water and phobia means fear or dislike, so hydrophobic sidechains don't like water. Hydrophobic sidechains are also referred to as non-polar sidechains.

Now can you guess what hydrophilic means? Philic means likes or attracted to, so hydrophilic sidechains like water. Hydrophilic sidechains are also referred to as polar sidechains.

Acidic (Negatively Charged) and Basic (Positively Charged) Properties

Can you think of acids you have around your house? Lemon and fruit juices, vinegar and phosphoric acid (in dark sodas) are common household acids. Acids taste sour and are typically liquids.

Can you think of bases you have around your house? Tums®, baking soda, drain cleaner and soap are common bases. Bases taste bitter and can be a liquid or solid.

What happens when you mix lemon juice or vinegar with baking soda? They neutralize each other, in a bubbling chemical reaction.

Preparation

The activities described in this handout primarily focus on amino acid sidechains. They will help you understand how the unique properties of each sidechain contribute to the structure and function of a protein.

First look at the components in your Amino Acid Starter Kit. Make sure your 1-group set has:

1. 1 **Chemical Properties Circle**
2. 1 Laminated **Amino Acid Sidechain List**
3. 4' **Mini-Toober**
4. 1 Set of **Red and Blue Endcaps**
5. 22 Clear **Bumpers**
6. 22 **Amino Acid Sidechains**
   - -1 each of the 20 Amino Acids
   - -1 additional cysteine and
   - -1 additional histidine
7. 22 Plastic **Clips**
   - -8 yellow
   - -8 white
   - -2 blue
   - -2 red
   - -2 green
8. 6 **Hydrogen Bond Connectors**

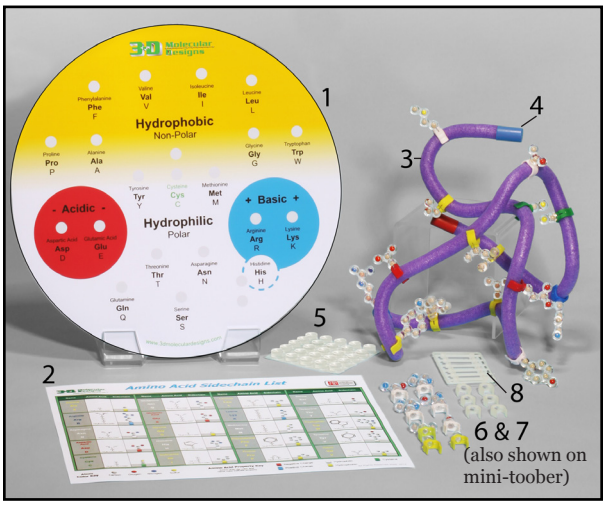


Photo shows a 1-Group Amino Acid Starter Kit.

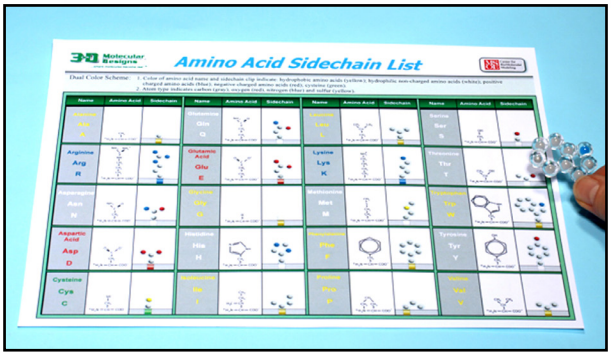
1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

Chemical Properties Circle (continued)

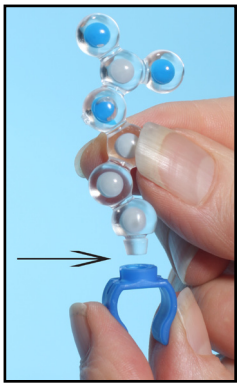
The colored areas on the Chemical Properties Circle, the color coding on the Amino Acid Sidechain List, the key below and the colored clips show the chemical properties of sidechains.

KEY

Hydrophobic Sidechains are **Yellow**
Hydrophilic Sidechains are **White**
Acidic Sidechains are **Red**
Basic Sidechains are **Blue**
Cysteine Sidechains are **Green**



Amino Acid Sidechain List.



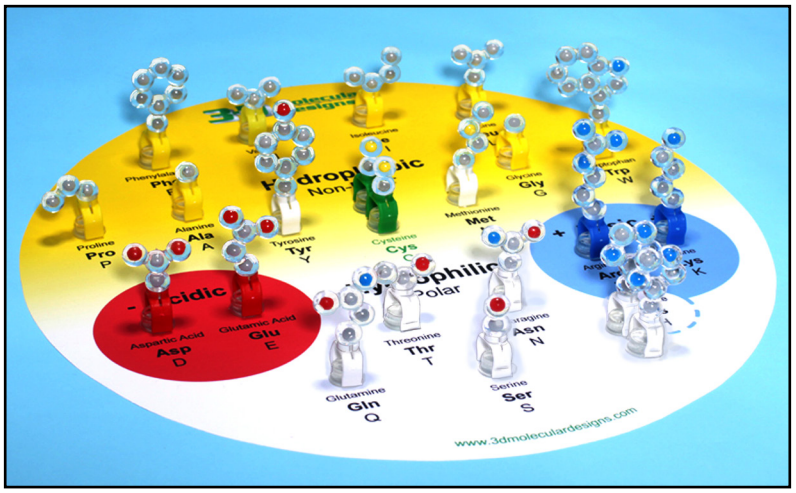
Insert sidechain into clip.

Directions

Select any sidechain and a colored clip that corresponds to the property of the sidechain. Insert the sidechain into the clip.

Place each amino acid sidechain attached to its clip on the bumper near its name and abbreviations. You will need to consult the Amino Acid Sidechain List in your kit to find the name of each sidechain, so you can position it correctly on the circle.

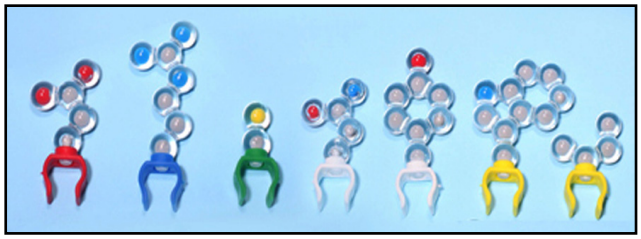
Place clip (with sidechain attached) onto the bumper.



Chemical Properties Circle with sidechains and clips.

Chemical Properties Circle (continued)

After each sidechain has been correctly positioned on the circle, look at the colored sheres in each sidechain. Scientists established a CPK coloring scheme (see chart below) to make it easier to identify specific atoms in models of molecular structures.



KEY

Carbon is **Gray**
Oxygen is **Red**
Nitrogen is **Blue**
Hydrogen is **White**
Sulfur is **Yellow**

Did you notice similarities of patterns in each group of sidechains? Describe Your Observations.

• Hydrophobic sidechains primarily contain _____carbon_____ atoms.

• Acidic sidechains contain two _____oxygen_____ atoms. This is called a carboxylic acid functional group.

• Basic sidechains contain ____one or two nitrogen____ atoms. This is called an amino functional group.

• Hydrophilic sidechains have various combinations of
oxygen, nitrogen and sulfur and carbon atoms.

• An exception to the above observation is:
Tryptophan - a hydrophobic amino acid that contains a nitrogen atom.

• **Optional Activity** - Amino Acids Jmol (see AASK Lessons on website)

Folding a 15-Amino Acid Protein

Once you have explored the chemical properties and atomic composition of each sidechain, think about how proteins spontaneously fold into their 3D shapes.

Predict what causes proteins to fold into their 3D shapes.

• Which sidechains might position themselves on the interior of a protein, where they are shielded from water?

The hydrophobic amino acids - tryptophan, leucine, isoleucine, valine, proline, alanine and glycine.

• From your experience with static electricity, which sidechains might be attracted to each other?

The basic amino acids (+ charge) and the acidic amino acids (- charge).

• Would the final shape of a protein be a high energy state or a low energy state for all of the atoms in the structure?

A low energy state.

Why?

A low energy state is more stable than a high energy state.

1. Unwind the 4-foot mini-toober (foam-covered wire) that is in your kit. Place a blue end cap on one end and the red end cap on the other end. The blue end cap represents the N-terminus (the beginning) of the protein and the red end cap represents the C-terminus (the end) of the protein (see photo on next page).

2. Choose 15 sidechains from the chemical properties circle as indicated in the chart below.

   Mix the Sidechains together and place them (in any order you choose) on your mini-toober.

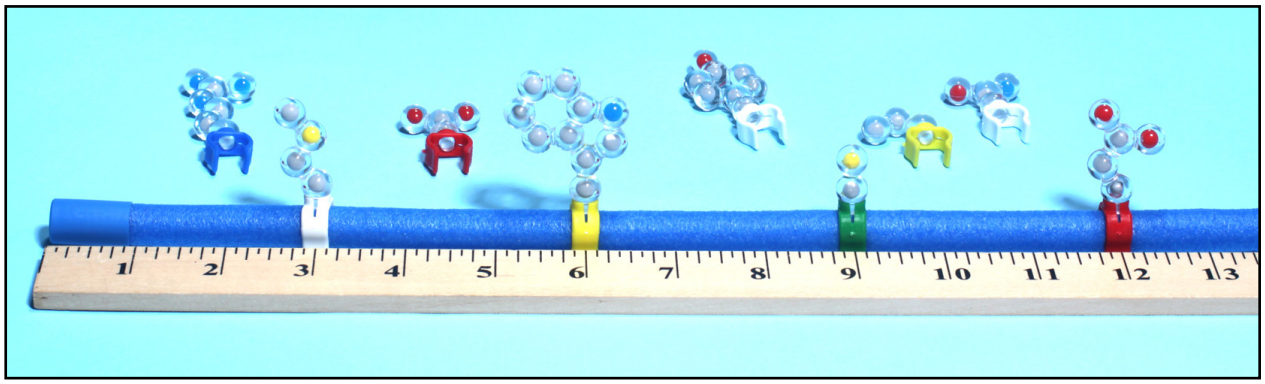
KEY

6 **Hydrophobic** sidechains
2 **Acidic** sidechains
2 **Basic** sidechains
2 **Cysteine** sidechains
3 **Hydrophilic** sidechains

Folding a 15-Amino Acid Protein (continued)

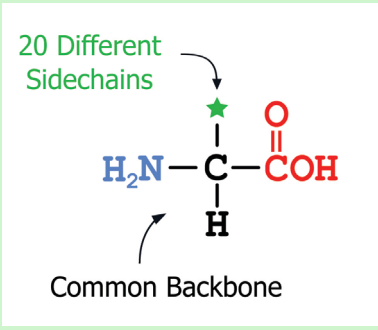
3. You may want to use a ruler to place your sidechains on you mini-toober.

Beginning at the N-terminus of your mini-toober, measure about three inches from the end of your mini-toober and slide the first colored clip with its sidechain onto the mini-toober. (See photo.) Place the rest of the clips three inches apart on your mini-toober until all are attached to the mini-toober.

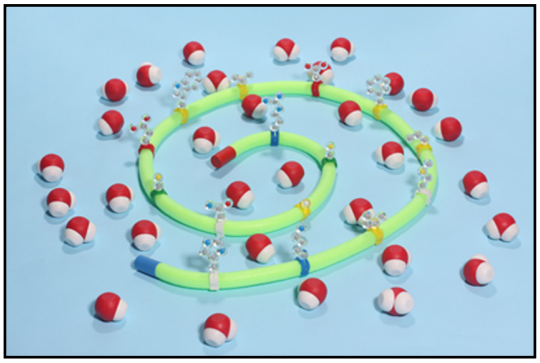


• This drawing represents the backbone section of an amino acid. What do you think the clips represent?

The alpha-carbon, because this is the atom that the sidechains are bonded to.



The *sequence* of amino acid sidechains that you determined when placing them on the mini-toober is called the **primary structure** of your protein. As a general rule the final shape of a protein is determined by its **primary structure**. Remember that protein folding happens in the watery environment of the cell.



Folding a 15-Amino Acid Protein (continued)

4. Now you can begin to fold your 15-amino acid protein according to the chemical properties of its sidechains. Remember all of these chemical properties affect the protein at the same time.

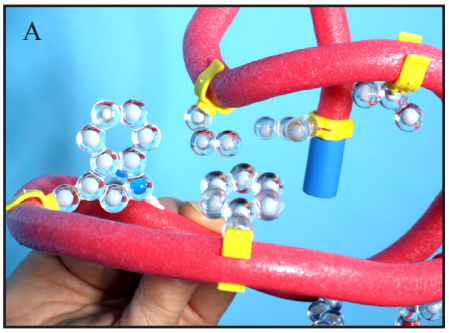


Photo A — Hydrophobic Sidechains
Start by folding your protein so that all of the hydrophobic (non-polar) sidechains are buried on the inside of your protein, where they will be hidden from polar water molecules.

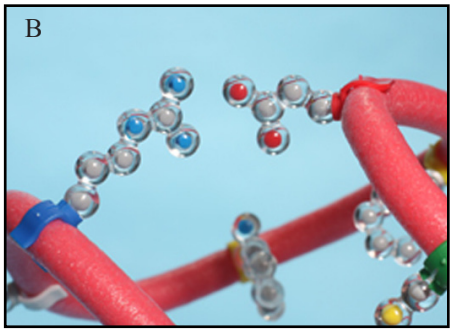


Photo B — Acidic & Basic Sidechains
Fold your protein so the acidic and basic **(charged)** sidechains are on the outside surface of the protein. Place one negative (acidic) sidechain with one positive (basic) sidechain so that they come within one inch of each other and neutralize each other. This positive-negative pairing helps stabilize your protein.

Note: As you continue to fold your protein and apply each new property listed below, you will probably find that some of the sidechains you previously positioned are no longer in place. For example, when you paired a negatively charged sidechain with a positively charged one, some of the hydrophobic sidechains probably moved to the outer surface of your protein. Continue to fold until the hydrophobic ones are buried on the inside again. Find a shape in which all the properties apply ***simultaneously***.

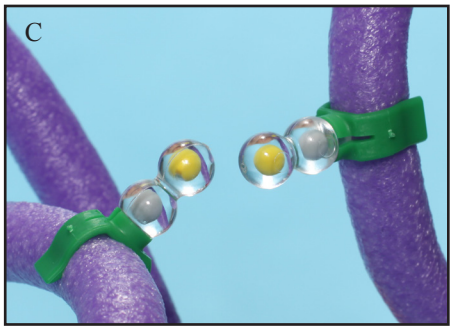


Photo C — Cysteine Sidechains
Fold your protein so that the two cysteine sidechains are positioned opposite each other on the inside of the protein where they can form a covalent-disulfide bond that helps stabilize your protein.

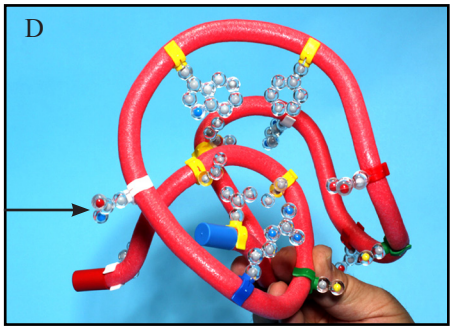


Photo D — Hydrophilic Sidechains
Continue to fold you protein making sure that your hydrophilic (polar) sidechains are also on the outside surface of your protein where they can hydrogen bond with water.

The final shape of your protein when it is folded is called the **tertiary structure.**

15-Amino Acid Protein Questions

• What happened as you continued to fold your protein and applied each new chemical property to your protein?

It became more compact and more complicated.

• Were you able to fold your protein, so that all of the chemical properties were in effect at the same time?

Yes. (Note to teachers: some students may answer "No".)

• If not, do you have any ideas why you weren't able to fold your protein in a way that allowed all of the chemical properties to be in effect simultaneously?

Some sequences simply do not allow for a single shape that simultaneously satisfies all the principles of chemistry that drive protein folding.

• Did your protein look like the proteins other students folded? No
Explain.

Because everyone had a different sequence of amino acids.

• How many different proteins, 15 amino acid long, could you make given an unlimited number of each of the 20 amino acids?

$20^{15} = 3.28 \times 10^{19}$

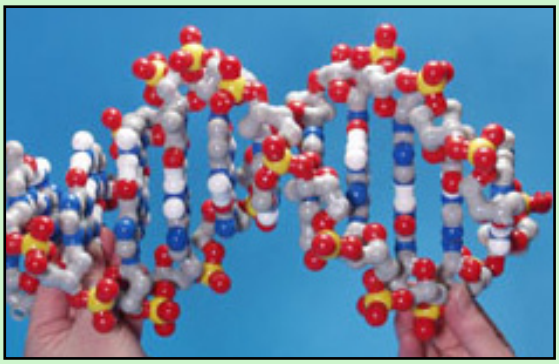
• Most real proteins are actually in the range of 300 amino acids long. How many different possible proteins, 300 amino acids in length, could exist?

$20^{300} = 2 \times 10^{390}$

15-Amino Acid Protein Questions (continued)

• Research how many different proteins are found in the human body. Hint: how many different genes are there in the human genome*?

25,000 or 2.5 x 10^4



• Assuming that all human proteins are 300 amino acids long, what fraction of the total number of possible different proteins is found in the human body?

1 x 10^{-386} = miniscule!

• Why do you think there are fewer actual proteins than possible ones?

Because only a relatively small number of amino acid sequences can adopt a stable shape that simultaneously satisfies all of the principles of chemistry.

* Completed in 2003, the Human Genome Project (HGP) was a 13-year project coordinated by the U.S. Department of Energy and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others.

Project goals were to:

- Identify all of the approximately 20,000-25,000 genes in human DNA,
- Determine the sequences of the 3 billion chemical base pairs that make up human DNA,
- Store this information in databases,
- Improve tools for data analysis,
- Transfer related technologies to the private sector, and
- Address the ethical, legal, and social issues (ELSI) that may arise from the project.**

*** U.S. Department of Energy Genome Programs website http://ornl.gov/sci/techresources/Human_Genome/home.shtml*

• **Optional Discussion:** Genes can code for multiple proteins through the process of alternative splicing.

15-Amino Acid Protein Questions (continued)

Record the sequence of amino acids in your protein, starting with the N-terminus (blue end cap). Use the single letter abbreviation for each amino acid (Methionine = M).

This is the **primary structure** of your protein.

In the space below, sketch the **tertiary structure** of your protein.

Discussion

Proteins perform critical functions in all our cells. Without proteins, life wouldn't exist. With your group or class, can you think of some of some specific proteins and describe what function they perform? *Proteins are involved in your metabolism, cell structure, immune system, DNA expression, protein folding, transport, movement, communication and storing energy.*

- **Optional Jmol Activity**
  - Basic Priciples of Chemistry that Drive Protein Folding Part 1 Jmol
  - Basic Priciples of Chemistry that Drive Protein Folding Part 2 Jmol

  (See AASK Lessons on website.)

The next student handout provides folding activities and information that will help you understand the **secondary structure** of proteins.

3-D Molecular Designs

. . . where molecules become real ™

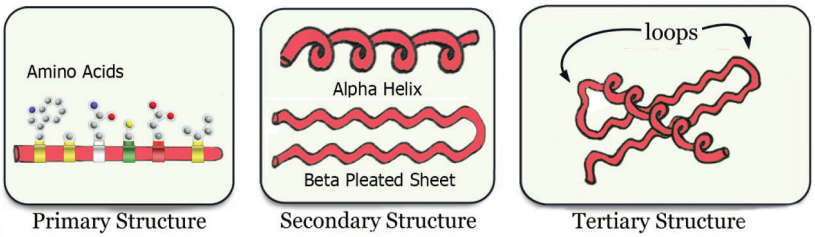
Amino Acid Starter Kit

Key for Student Handout 2

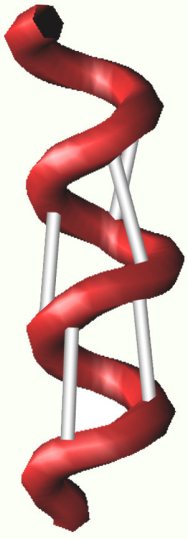
Secondary Structure

In the previous protein folding activity, you created a hypothetical 15-amino acid protein and learned that basic principles of chemistry determine how each protein spontaneously folds into its characteristic 3-dimensional shape. You learned that the sequence of amino acids in a protein (from N-terminus to C-terminus) is called its **primary structure**. The final folded, 3D shape of your protein is called its **tertiary structure**.

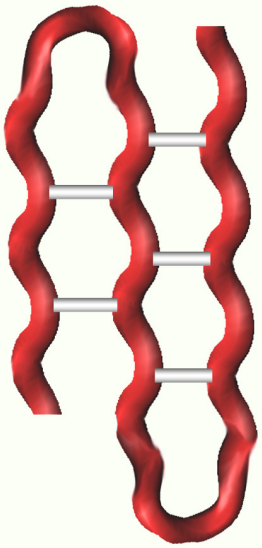
In this second protein-folding activity, you will learn about the **secondary structure** of proteins. This secondary structure consists of alpha helices and/or beta sheets. Proteins commonly contain a combination of alpha helices and beta sheets. Proteins can be described as a series of alpha helices and beta sheets, joined by **loops** of less regular protein structure.



Primary Structure | Secondary Structure | Tertiary Structure



An **alpha helix** is a compact right-handed helix, with 3.6 amino acids per turn of the helix. The amino acid sidechains are bonded to the alpha carbon of each amino acid and radiate outward from the helix. The alpha helix is stabilized by hydrogen bonds – weak bonds between the amino nitrogen of one amino acid (x), and the carbonyl oxygen of another amino acid (x+4) located four sidechains further along the chain.



A **beta sheet** is an extended, zig-zag structure in which individual strands are positioned parallel or anti-parallel to each other to form flat sheets in proteins. Since the amino acid sidechains are bonded to the alpha carbons of each amino acid, they are alternately orientated above and below the plane of the sheet. The beta sheet is stabilized by hydrogen bonds between the amino nitrogen of one amino acid and the carbonyl oxygen of another amino acid in an adjacent beta strand.

1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

Folding a Toober Model of the Zinc Finger

In this activity, you will fold a model of the first of three zinc fingers of the Zif268 protein. Zinc finger proteins regulate the transcription of DNA into mRNA – by binding to DNA and attracting RNA polymerase. A zinc finger protein contains two cysteine amino acids and two histidine amino acids which simultaneously bind to a single zinc atom. These four amino acids are contained within a 30 amino acid sequence that folds into a two-stranded beta sheet and short alpha helix. Many zinc finger proteins (like zif268) are composed of three consecutive fingers with similar features (motifs) which bind to a nine base pair sequence of double-stranded DNA.

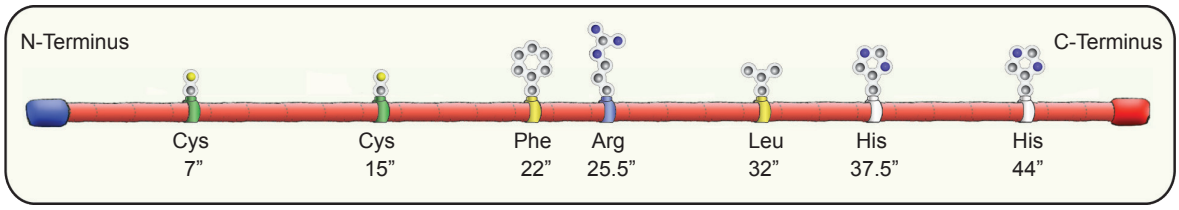
The primary structure of this zinc finger is below.

N-Terminus P Y A (C) P V E S (C) D R R (F) S (R) S D E (L) T R (H) I R I (H) T G C-Terminus

The sidechains of the seven circled amino acids in the above sequence will be included in the model you fold.

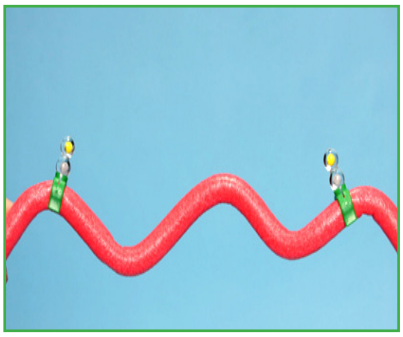
1. ***Primary Structure***

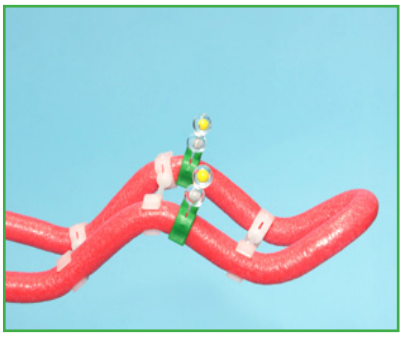
Map the positions of the seven amino acids on your mini-toober. Since the toober is 48 inches long and the zinc finger is 28 amino acids long, each amino acid occupies 1.7 inches of toober. Using a ruler, measure the distances shown below and add the appropriate sidechains to the mini-toober at each position.



2. ***Secondary Structure***

Fold the toober into its secondary structure. The first 13 amino acids (the first 22 inches from the N-terminus) should be folded into a 2-stranded beta sheet. This can be made by creating a zig-zag structure that is bent in the middle as shown in the photos below. Add the plastic hydrogen bonds connectors to your model as shown in the far right photo below.





Folding a Toober Model of the Zinc Finger (continued)

The last 15 amino acids of the zinc finger exist as a compact, right-handed alpha helix. This can be made by wrapping the mini-toober around your finger or an empty paper towel tube to create three full turns as shown in the photos below. Loosen the loops and add the hydrogen bond connectors as shown in the far right photo.

Your mini-toober should look similar to the one shown below.

1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

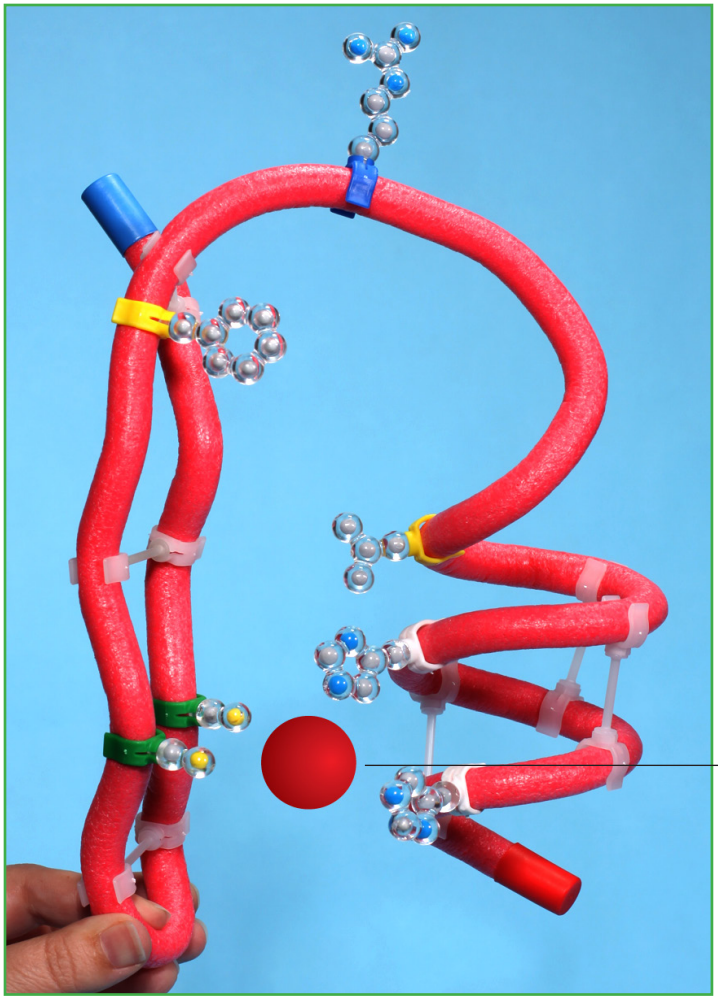
Folding a Toober Model of the Zinc Finger (continued)

3. *Tertiary Structure*

Fold the beta sheet and alpha helix into the final tertiary structure of the zinc finger.

In its final tertiary structure, the seven sidechains will be positioned such that:

- The two cysteine and two histidine sidechains will be oriented to simultaneously bind to a single zinc atom (not included) in the center of the structure (see photo).
- The two hydrophobic amino acid sidechains phenylalanine and leucine will be orientated toward the inside of the structure.
- The positively-charged arginine sidechain will be exposed at the top of the alpha helix, where it is available to bind to the negatively-charged phosphate backbone of DNA.



The zinc ion (not included with the kit) binds simultaneously to the two histidines and two cysteines.

As a folding guide, you can either use the photo shown below or the interactive Jmol image of a zinc finger on website (www.3dmoleculardesigns.com/resources.php).

Note: As you fold your mini-toober, you may need to rotate the sidechains around the mini-toober to make them adopt to the desired final shape.

Folding a Toober Model of the Zinc Finger (Questions)

1. Both alpha helices and beta sheets are stabilized by hydrogen bonds.
   - Which atoms share the hydrogen in these weak bonds?

   The nitrogen of an amino group and the oxygen of a carbonyl group.

   - Are these backbone atoms or sidechain atoms?

   Backbone atoms.

2. Describe the secondary structural elements that comprise a zinc finger:

   A 2-stranded beta sheet and a short alpha helix.

3. How is a zinc atom involved in the stabilization of the zinc finger motif?

   The zinc atom is simultaneously bound by the 2 cysteine and the 2 histidine sidechains.

4. Zinc fingers often bind to DNA. How might the arginine sidechain (positively-charged) shown on your model be involved in DNA binding?

   DNA has a negatively-charged phosphate backbone. Therefore, the positively-charged arginine of the zinc finger can bind to DNA via an electrostatic interaction.

- **Optional Activity** - Zinc Finger Jmol (see AASK Lessons on website)

Teaching Points on page 6.

Teaching Points

When proteins fold into their tertiary structures, there are often subdivisions within the protein, designated as domains, which are characterized by similar features or motifs. One such motif is the zinc finger in which a specific domain of the protein is arranged into a **finger-like** structure where two beta sheets and one alpha helix are positioned around a zinc ion. The zinc finger motif is commonly found in eukaryotic transcription factors, which are proteins that bind to specific sequences of DNA in order to regulate transcription.

One common class of zinc finger is the C_2H_2 class which is the one modeled in this collection. In this class of zinc fingers, the zinc ion is bound to two cysteine residues and two histidine residues.

3-D Molecular Designs

. . . where molecules become real ™

Amino Acid Starter Kit

Key for Student Handout 3

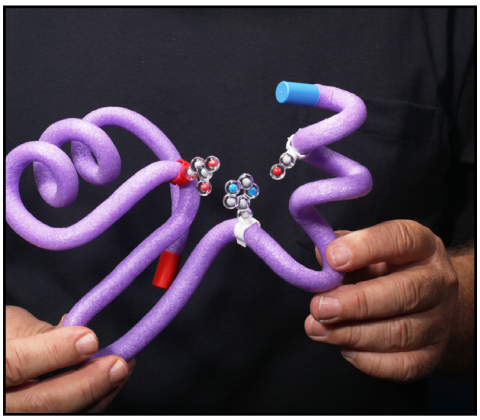
Understanding an Enzyme Active Site

In the first protein folding activity, you learned that a protein begins as a linear sequence (primary structure) of amino acids that spontaneously folds into a compact 3D shape (tertiary structure) following basic principles of chemistry.

In the second activity, you learned that the 3D shape of a protein consists of stretches of alpha helices and/or beta sheets (secondary structure) connected by short turns of less regular protein structure.

In the space below, draw and label examples of primary, secondary and tertiary structures.

Proteins perform many different functions in cells. Some proteins function as structural supports for the cell's architecture. Others transport small molecules — such as oxygen or neurotransmitters — between cells.



Enzyme Active Sites

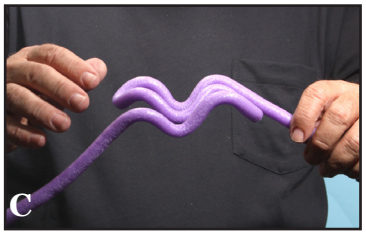
In this third activity, you will explore enzymes — a major class of proteins. Enzymes bind a specific small molecule — a substrate — and then catalyze a chemical reaction that changes the substrate in some way. The active site of an enzyme is the **region** of the protein that is able to bind a specific substrate (usually a small molecule) and then catalyze the reaction.

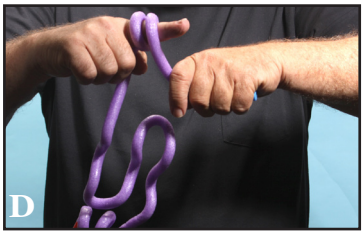
1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

Modeling an Active Site

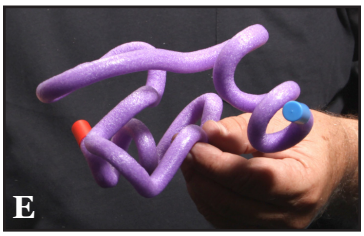
Imagine that your 4-foot mini-toober represents a protein consisting of 200 amino acids.

1. Begin folding your mini-toober into the shape of a protein by creating a three-stranded beta sheet and two short alpha helices. The beta sheet and alpha helices represent your protein's secondary structure. See photos A through D.



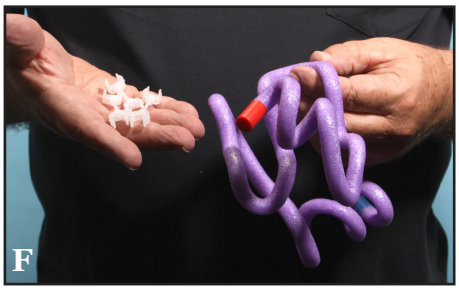


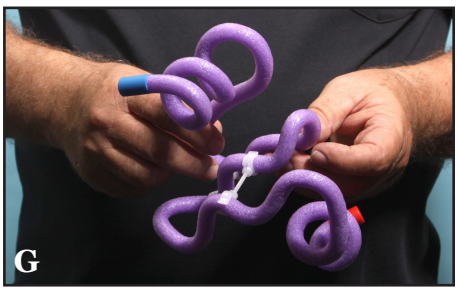
2. Fold the beta sheet and the alpha helices into a compact, globular shape. See photo E.



3. Use three connectors to stabilize the overall 3D shape of the folded protein. See photos F and G.

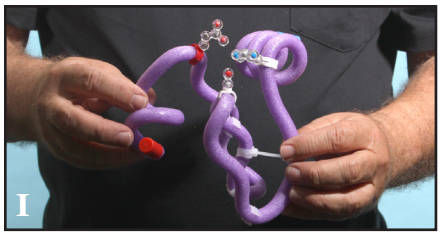
   These connectors stabilize your protein's structure in the same way that hydrogen bonds, which are present in alpha helices and beta sheets, stabilize the structure of a real protein. You now have a stable 3D structure – upon which you can precisely place three specific amino acid sidechains to create an enzyme active site.





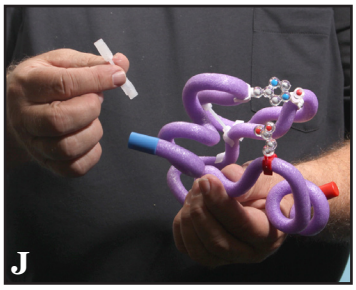
Modeling an Active Site (continued)

4. Create an active site in a shallow crevice on the surface of your protein by adding three amino acid sidechains – a serine, a histidine and a glutamic acid – to your mini-toober in such a way that all three sidechains are within 2 cm of each other. See photos H and I.

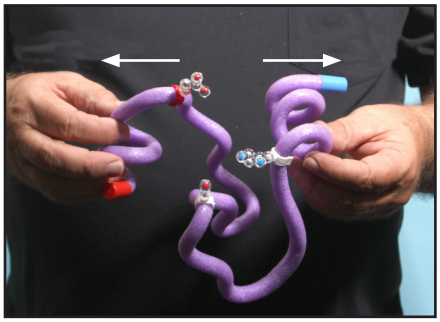


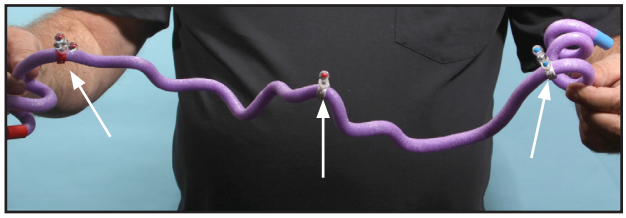
5. The three amino acid sidechains that make up your enzyme's active site interact with a substrate to catalyze a specific chemical reaction. This requires that the sidechains be precisely positioned in 3D space. Examine you protein, noting how its secondary and tertiary structure combines to provide a stable scaffolding, or framework, upon which the active site amino acids are precisely positioned relative to each other.

6. Now carefully remove the connectors that were stabilizing your folded protein. See photo J.



7. Holding your protein with one hand near the N-terminus end and the other near the C-terminus end, slowly move your hands away from each other – simulating the unfolding (denaturation) of your protein.

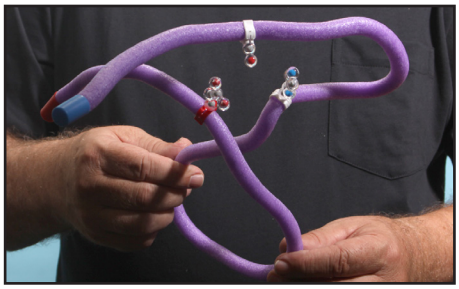


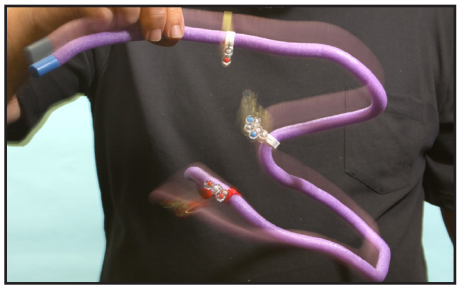


The 3 active site amino acids — that were close together in a folded enzyme — are now far apart in the linear sequence of the protein.

Modeling an Active Site (continued)

Notice that without the stabilizing effect of the hydrogen bonding in your protein's secondary structure, the normal thermal motion experienced by proteins would cause them to unfold (denature).





• Describe the kinds of interactions (bonds) that are present in your protein's secondary and tertiary structure that contribute to the stability of this scaffolding.

The protein's secondary structure (both alpha helices and beta sheets) are stabilized by **hydrogen bonds** — between the polar nitrogen and carbonyl oxygen atoms of the protein's backbone.

The protein's tertiary structure is stabilized by a variety of bonds and interactions between the amino acid sidechains that make up the protein. Bonds that stabilize the protein include: **hydrogen bonds** between polar sidechains and **electrostatic bonds** between oppositely-charged sidechains (acidic and basic sidechains). **Hydrophobic interaction** between hydrophobic sidechains — as they try to minimize their interaction with water — is another major force that stabilizes a protein's tertiary structure.

• Describe your observations of the distribution of the three active site amino acids in your enzyme?

The surprising thing about an enzyme active site is that the three amino acids — that were positioned very close together in the 3D shape of the protein — are actually very far apart in the linear sequence of the amino acids that make up the protein. The protein has to fold into its 3D shape for the sidechains that make the active site to come together, so they can perform their function.

• **Optional Jmol Activity** - Active Site Jmol (see AASK Lessons on website)

Teaching Points

Although most enzymes consist of 200 or more amino acids, the active site of an enzyme is made up of only 2 to 3 amino acids that are precisely positioned in 3D space. In this activity, your students will be asked to think about how all the other amino acids in the enzyme create a compact, stable scaffold upon which the 2-3 active site amino acids can be positioned. This activity will also demonstrate the role of protein secondary structure in achieving this stable scaffold. In addition your students may be surprised to discover that the three active site amino acids in this example are very far apart from each other in the linear sequence of amino acids that make up the protein.

Key Points

Enzyme active sites are composed of a small number (2-3) of amino acids that are precisely positioned in 3D space such that their sidechains create the chemistry needed to catalyze a reaction.

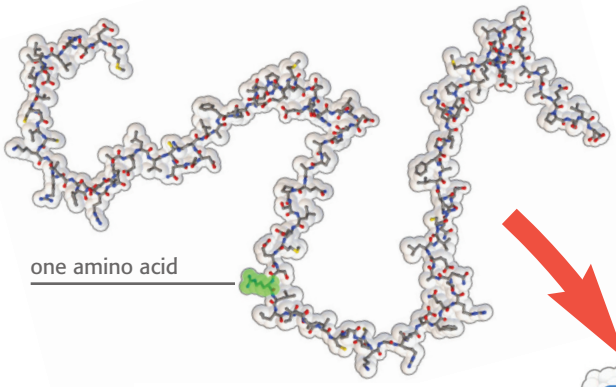
Protein secondary structure (alpha helices and beta sheets) provides that stable scaffolding upon which the critical active site amino acids can be precisely positioned in 3D space.

The 2-3 amino acids that come together in 3D space to create an enzyme active site are very far apart in the linear sequence of the amino acids that make up the protein.

What is a Protein?

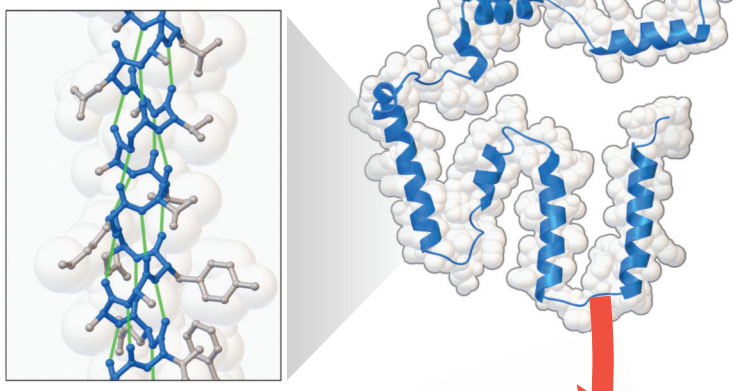
Proteins play countless roles throughout the biological world, from catalyzing chemical reactions to building the structures of all living things.

Despite this wide range of functions all proteins are made out of the same twenty amino acids, but combined in different ways. The way these twenty amino acids are arranged dictates the folding of the protein into its unique final shape. Since protein function is based on the ability to recognize and bind to specific molecules, having the correct shape is critical for proteins to do their jobs correctly.



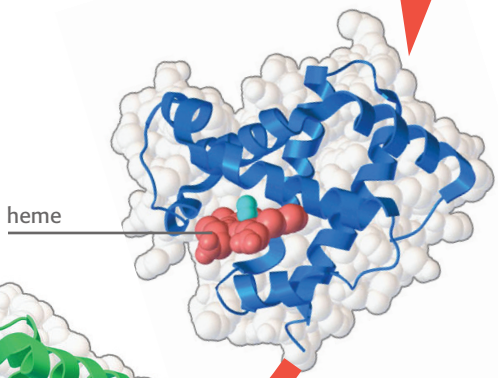
Primary Structure

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



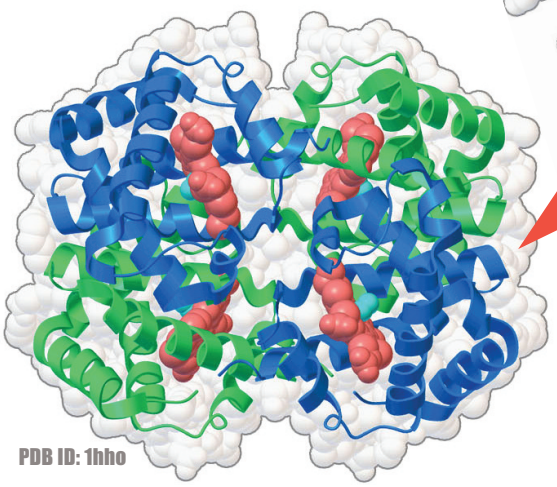
Secondary Structure

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



Tertiary Structure

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



Quaternary Structure

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

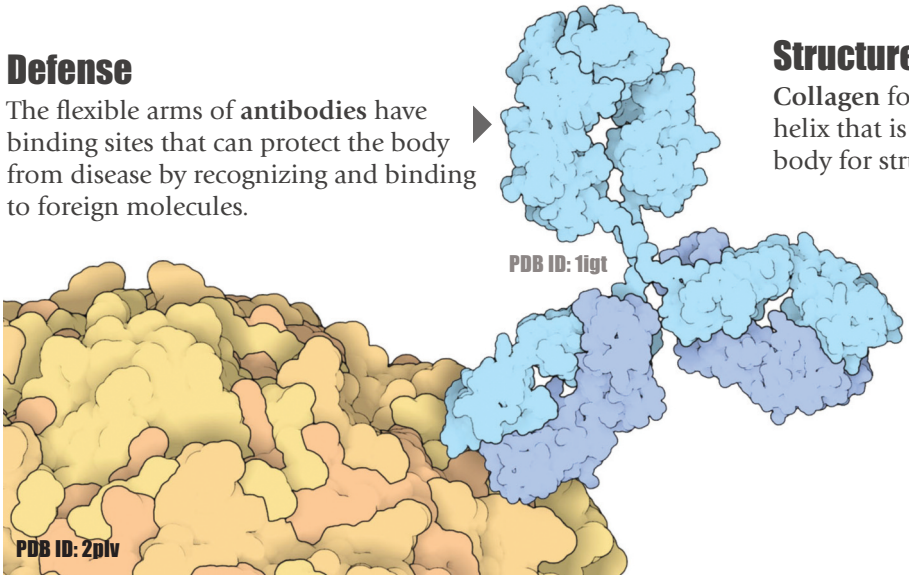
RCSB PDB
PROTEIN DATA BANK
www.rcsb.org • info@rcsb.org

Protein
Shape & Function

Specific amino acid sequences give proteins their distinct shapes and chemical characteristics. Protein shape is important because many proteins rely on the recognition of specific 3D molecular shapes to function correctly.

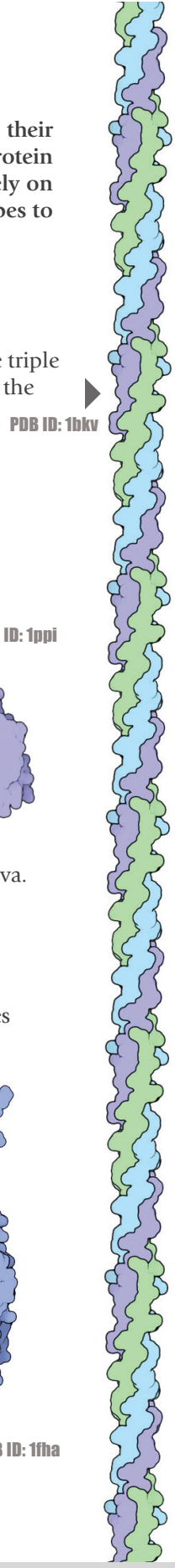
Defense

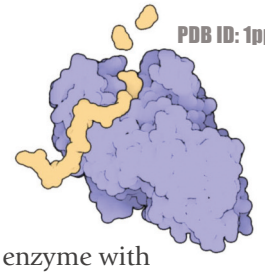
The flexible arms of **antibodies** have binding sites that can protect the body from disease by recognizing and binding to foreign molecules.



Structure

Collagen forms a strong and flexible triple helix that is widely used throughout the body for structural support.



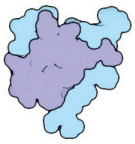


Enzymes

Alpha amylase is an enzyme with a specific catalytic site that begins the breakdown of carbohydrates in our saliva.

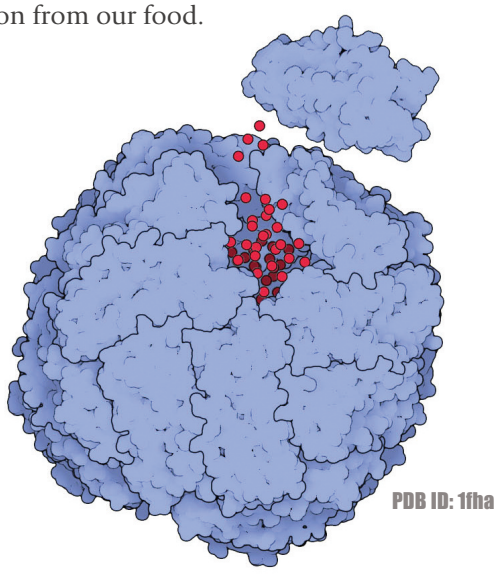
Communication

Insulin is a small, stable protein that can easily maintain its shape while traveling through the blood to regulate blood sugar levels.



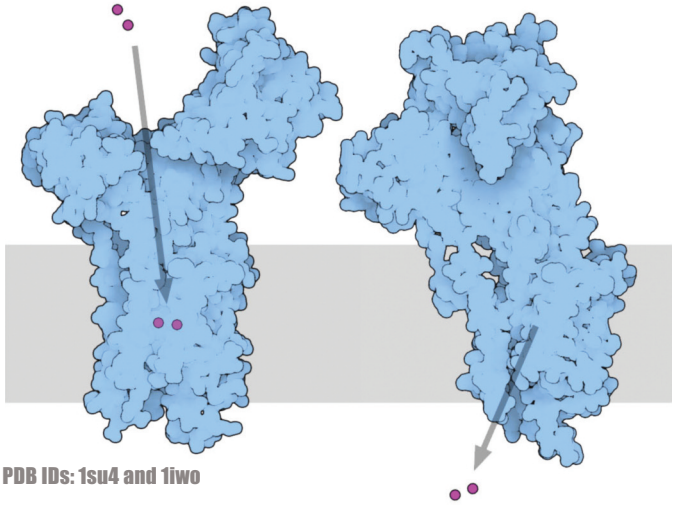
Storage

Ferritin forms a hollow shell that stores iron from our food.



Transport

The **calcium pump** moves ions across cell membranes allowing the synchronized contraction of muscle cells.



To learn more about these and other proteins please visit PDB-101 at **www.rcsb.org/pdb-101**

RCSB PDB-101

Text and illustrations by Jessica May and David S. Goodsell

3-D Molecular Designs

. . . where molecules become real ™

Teacher Notes

From Amino Acids to Proteins - in 4 Easy Steps

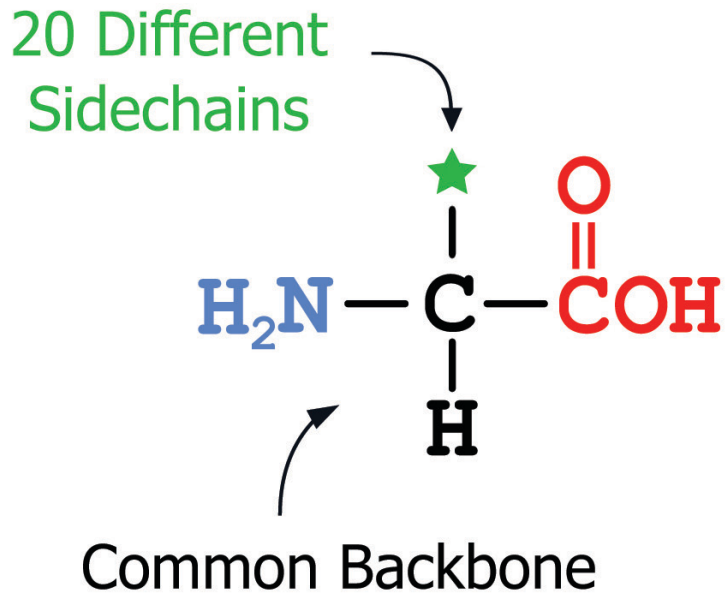
Although protein structure appears to be overwhelmingly complex, you can provide your students with a basic understanding of how proteins fold by focusing on the following four teaching points.

- The 20 amino acids are at the same time identical and different.
- In a single amino acid at neutral pH, the backbone amino group (NH_3^+) is positively charged, and the backbone carboxyl group (COO^-) is negatively charged.
- In a protein, the backbone amino group of the N-terminal amino acid is ***positively charged***, and the backbone carboxyl group of the C-terminal amino acid is ***negatively charged***. All other backbone charges have been ***neutralized*** by peptide bond formation.
- In a protein, the chemical properties of each sidechain are the major determinant of the final, folded 3D structure.

Four Easy Steps

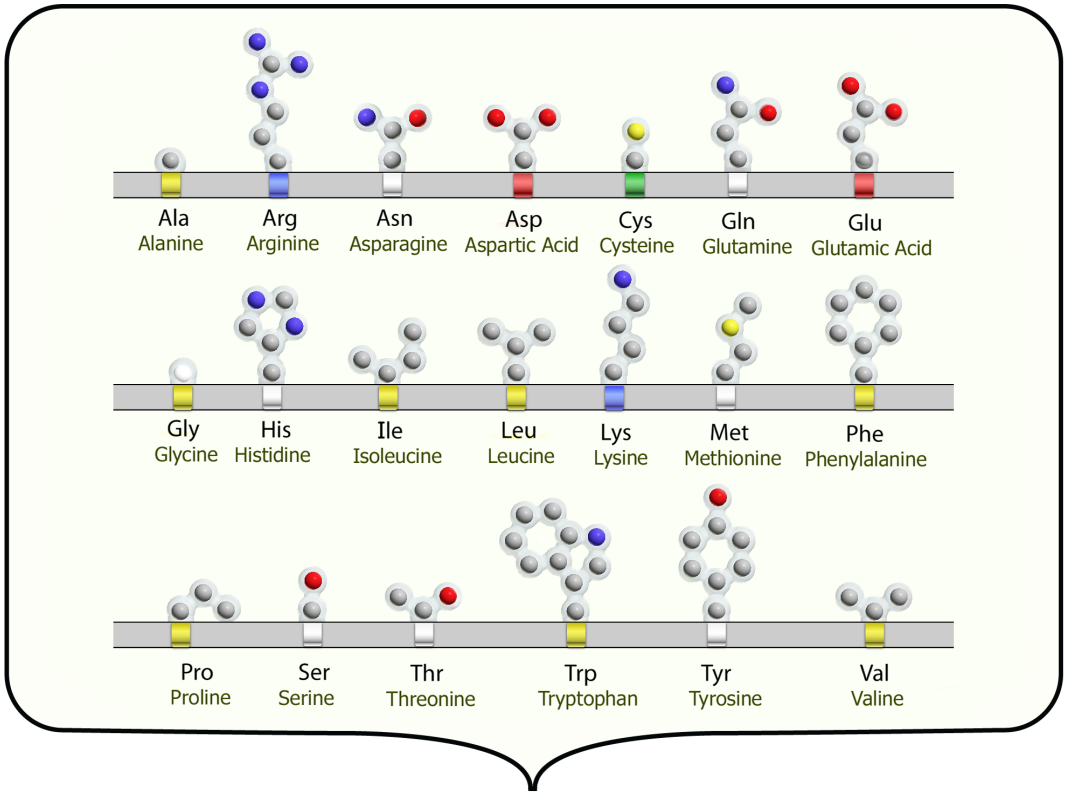
1. The 20 amino acids are at the same time identical and different. How can that be?

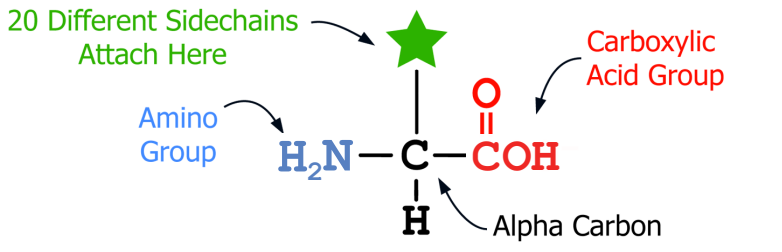
The 20 amino acids all share a common backbone and have different sidechains, each with different chemical properties.

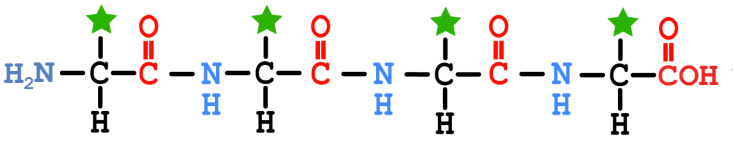


1050 North Market Street, Suite CC130A, Milwaukee, WI 53202
Phone 414-774-6562 Fax 414-774-3435
3dmoleculardesigns.com

From Amino Acids to Proteins (continued)



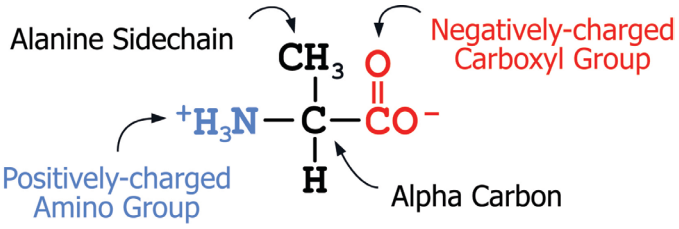


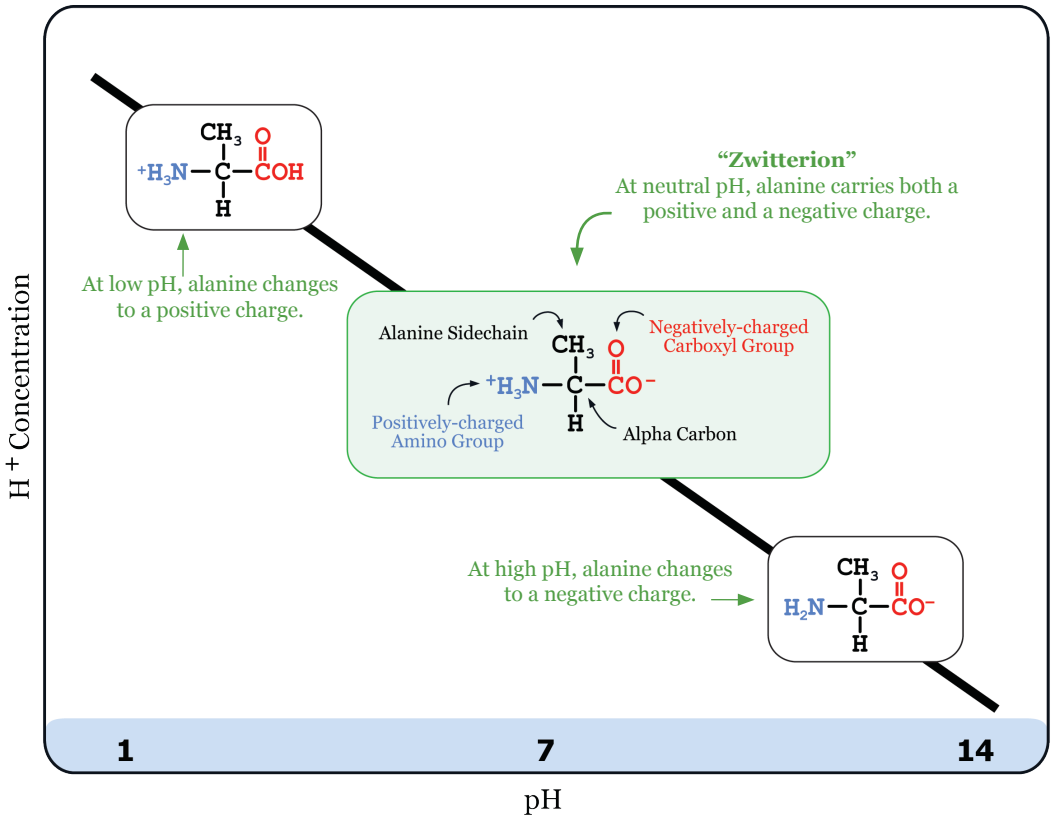


A polypeptide composed of 4 amino acids.

From Amino Acids to Proteins (continued)

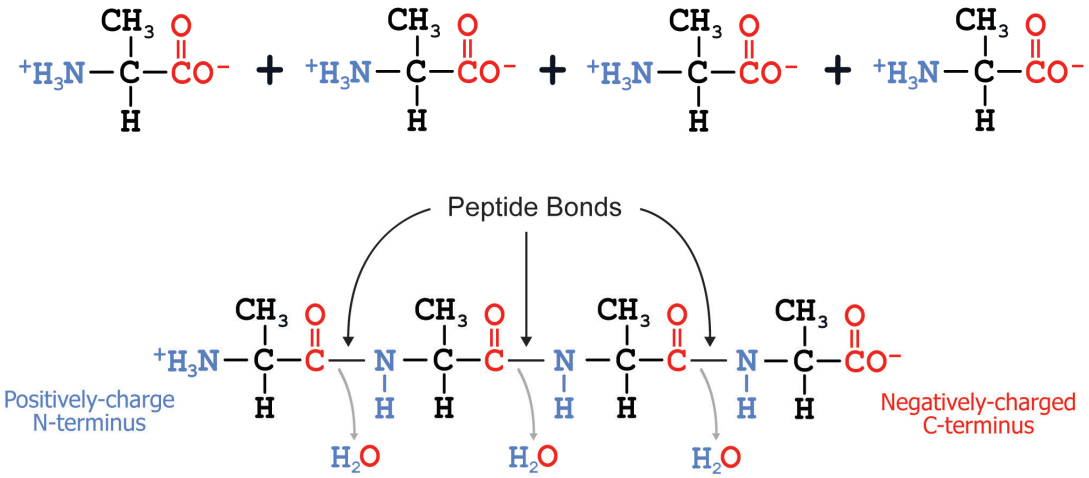
2. In a single amino acid at neutral pH, the backbone amino group (NH_3^+) is positively charged, and the backbone carboxyl group (COO^-) is negatively charged.



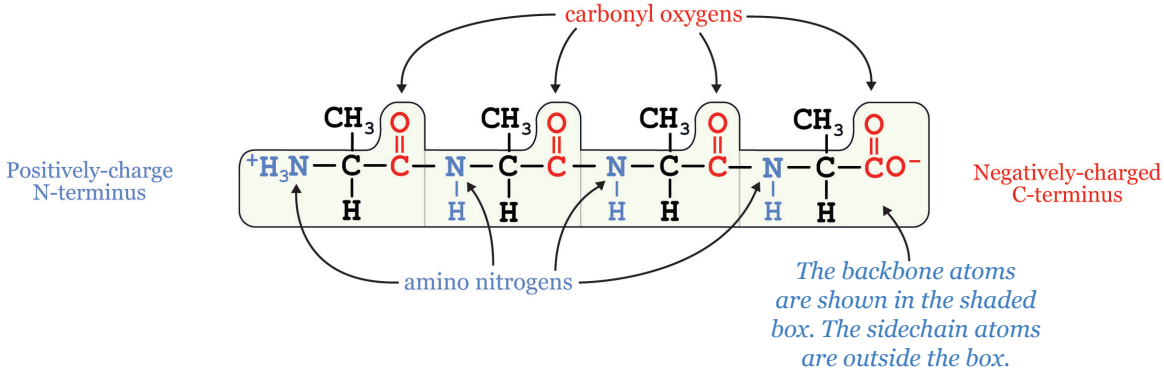


From Amino Acids to Proteins (continued)

3. In a protein, the backbone amino group of the N-terminal amino acid is ***positively-charged***, and the backbone carboxyl group of the C-terminal amino acid is ***negatively-charged***. All other backbone charges have been ***neutralized*** by peptide bond formation.



Note that the formation of each peptide bond results in the production of a water molecule. This is an example of a condensation reaction, also called dehydration synthesis.



From Amino Acids to Proteins (continued)

4. In a protein, the chemical properties of each sidechain are the major determinants of the final, folded 3D structure.

Basic Principles of Chemistry Drive Protein Folding

A. Hydrophobic amino acids are buried in the interior of a globular protein.

- Hydrophobic amino acids are composed primarily of carbon atoms, which cannot form hydrogen bonds with water. In order to form a hydrogen bond with water, a polar molecule, the amino acid sidechains must also be polar, or have an unequal distribution of electrons. Carbon atoms have a uniform distribution of electrons and create a non-polar sidechain. In a soluble, cytosolic protein, these amino acids can be found buried within the protein, where they will not interact with water.

B. Hydrophilic amino acids are usually exposed on the surface of globular proteins.

- Hydrophilic amino acids have oxygen and nitrogen atoms, which can form hydrogen bonds with water. These atoms have an unequal distribution of electrons, creating a polar molecule that can interact and form hydrogen bonds with water. These polar amino acids will be found on the surface of a soluble, cytosolic protein, where they can hydrogen bond with water.

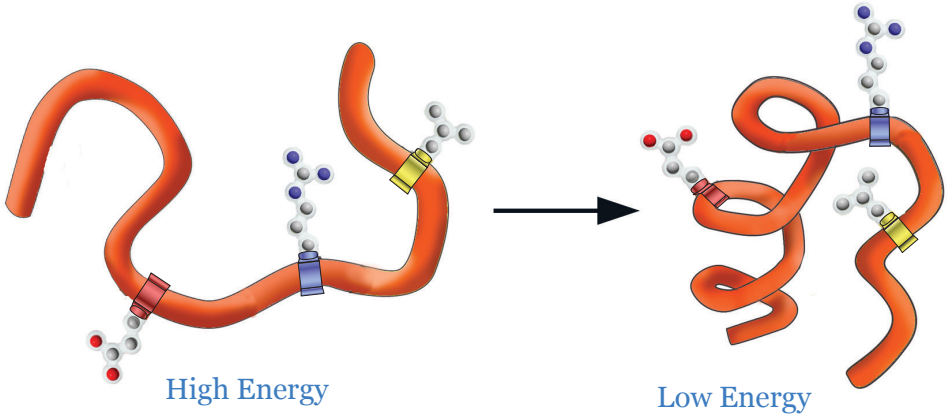
C. Acidic and basic amino acids can form salt bridges, or electrostatic interactions.

- Two of the polar amino acids (glutamic acid and aspartic acid) contain carboxylic acid functional groups and are therefore acidic (negatively charged).
- Two of the polar amino acids (lysine and arginine) contain amino functional groups and are therefore basic (positively charged).
- These two groups of amino acids (acidic and basic) are attracted to one another and can form electrostatic interactions.

D. Cysteine amino acids can form disulfide bonds.

- The cysteine sidechain contains a sulfur atom that can form a covalent disulfide bond with other cysteine sidechains. Disulfide bonds often stabilize the structure of secreted proteins.

When a protein is viewed as a system of interacting components, thermodynamic principles dictate the final shape should represent a low energy state for all of the atoms in the structure.

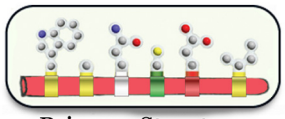


High Energy Low Energy

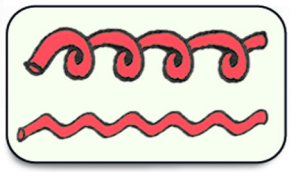
Protein Structure

The previous section focuses on the primary and the tertiary structures of proteins. However, it is useful to think about protein structure in a hierarchical manner, starting with the **primary structure**, and then proceeding to the **secondary**, **tertiary** and **quaternary structure**.

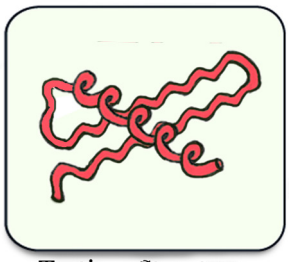
- **The primary structure of a protein** is simply the amino acid sequence of the protein. The final shape of a protein is encoded in its primary structure – the sequence of amino acids in a protein determines its final 3D structure.


Primary Structure

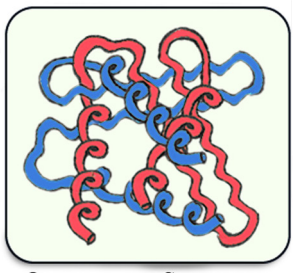
- **The secondary structure of a protein** refers to the alpha helices or beta sheets in the protein. These two common secondary structural elements are stabilized by hydrogen bonding between backbone atoms (the sidechains are not involved in protein secondary structure). A protein can be thought of as a collection of alpha helices and strands of beta sheet that are connected by loops.


Secondary Structure

- **The tertiary structure of a protein** refers to the overall 3D folded structure of a protein. This final folded structure represents a global low-energy state of all the atoms that make up the protein. The final tertiary structure of a protein is stabilized by a combination of many non-covalent interactions including hydrophobic forces, hydrogen bonds between polar atoms, ionic interactions between charged sidechains and Van der Waals forces. Covalant disulfide bonds can also provide stability in some proteins.


Tertiary Structure

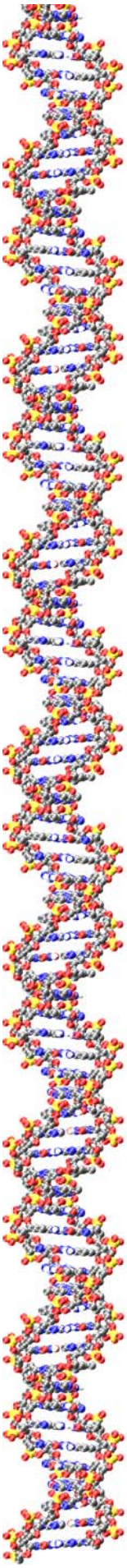
- **The quaternary structure of a protein** refers to protein complexes composed of more than one protein chain. Although some proteins exist as monomers (and therefore have no quaternary structure), many proteins interact to form multi-component protein complexes. Hemoglobin is a good example of a protein with quaternary structure. It is composed of two alpha chains and two beta chains.


Quaternary Structure

Summary

To construct a robust mental model of a protein, students will:

- **conclude** that the primary structure of a protein (its amino acid sequence) is a major determinant of its final 3D shape.
- **determine** that local regions of proteins first adopt a secondary structure (either alpha helices or beta sheets), which are stabilized by hydrogen bonding between backbone atoms.
- **establish** that the basic principles of chemistry act on the amino acid sidechains to determine the tertiary structure of the protein.
- **recognize** that many proteins assemble into quaternary structures, where they function as complex molecular machines.



3-D Molecular Designs

...where molecules become real™

The Discovery of DNA

On April 25, 1953, a one-page paper entitled, *A Structure for Deoxyribonucleic Acid,* appeared in the British journal, *Nature*. The authors of this paper were James Watson, a young American post-doctoral candidate who had recently received a Ph.D. from the University of Illinois, and Francis Crick, a physicist who was completing his doctoral dissertation at Cambridge University, England. The paper began; *"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D. N. A.). This structure has novel features which are of considerable biological interest."*

This initial description of the structure of DNA marked a major milestone in the development of molecular biology. In addition to reporting the correct structure of DNA, the paper also contained their classic understatement in scientific literature: *"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."* Their paper serves as an excellent example of what has become a recurring theme in the molecular biosciences — ***Forms Follows Function***. That is, the structure of a macromolecule often explains the macromolecule's function (how the macromolecule) works.

Watson and Crick's achievement is notable in several ways, including the fact that they determined the structure of DNA without performing a single experiment. They used the information from numerous other scientists who were investigating various properties of DNA. Modeling was the major approach Watson and Crick used. Using paper cut-outs of the shapes of the four nitrogenous bases (A,T, G and C), they were able to combine all of the different facts that had accumulated to that date into a plausible model for the structure of DNA.

> *...The structure has two helical chains coiled around the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining B-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite direction.*
>
> — Watson, J.D. and Crick, F.H.C., *Nature*, 171, 737-738 (1953)

Student Handout

The DNA Student Challenge

Your challenge today is to see if you can discover the correct structure of double-stranded DNA, just as Watson and Crick did over 50 years ago.

Your model should satisfy all of the pieces of experimental information that was known in 1953, as noted in the blue box below. Rather than using paper cut-outs to represent the DNA bases, you will use plastic models of the four deoxyribonucleotides whose 3D structures are based on known atomic coordinates of the B-form DNA. In these nucleotide models, magnets are used to represent both:

- the phosphodiester bonds that link the nucleotide units together into a long, linear polymer
- the hydrogen bonds that bond one base to another.

Information Available to Watson and Crick in 1953

DNA is a Polymer: Previous studies identified DNA as the genetic material of cells, and that DNA was a polymer consisting of three components:

- A nitrogenous base
- A pentose (5-carbon) sugar called deoxyribose
- A phosphate group.

Moreover, experiments suggested that the DNA molecule was unbelievably large, with molecular weights ranging from 25 x 10^6 to 3 x 10^9 daltons. (Since each nucleotide has a mass of 330 daltons, DNA molecules were believed to be composed of between 76,000 and 9,000,000 nucleotides.)

DNA is more dense than protein. At a density of 1.6 gm/cm3, DNA was known to be more dense than protein (1.3gm/cm3). This suggested that DNA was a densely packed structure.

Chargaff's Rules: In 1947, Erwin Chargaff demonstrated that while the four nucleotides were not present in equal amounts in the DNA from different organisms, the amount of adenine was the same as thymine, and the amount of guanine was the same as cytosine. This became known as *Chargaff's Rules:*

- The proportion of A always equals that of T, and the proportion of G always equals that of C. Thus, A = T and G = C.

X-ray Crystallography Data: In the laboratory of Maurice Wilkins, Rosalind Franklin used X-ray diffraction to analyze fibers of DNA. The pattern of spots on the X-ray diffraction pattern suggested that:

- Phosphate was on the outside, nitrogenous bases were on the inside.
- DNA was a double helix, made up of two strands.
- The two strands of DNA run in opposite directions (anti-parallel).
- There are 10 base pairs per turn of the double helix.

Background information for students

Each group of students should have physical models of the four nucleotides, separated into their component parts. These include:

- Phosphate group – which is negatively charged
- Deoxyribose group — which is a cyclic ring structure
- Four nitrogenous bases (A, G, C and T)

Each component of the nucleotides is color coded according to atom type, following the standard CPK coloring scheme:

Oxygen is RED **Nitrogen is BLUE**

Carbon is GRAY Hydrogen is WHITE

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (*Stockholm*), **2** (11) (1905).

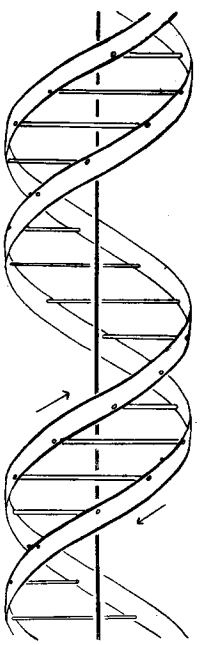
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline dooxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

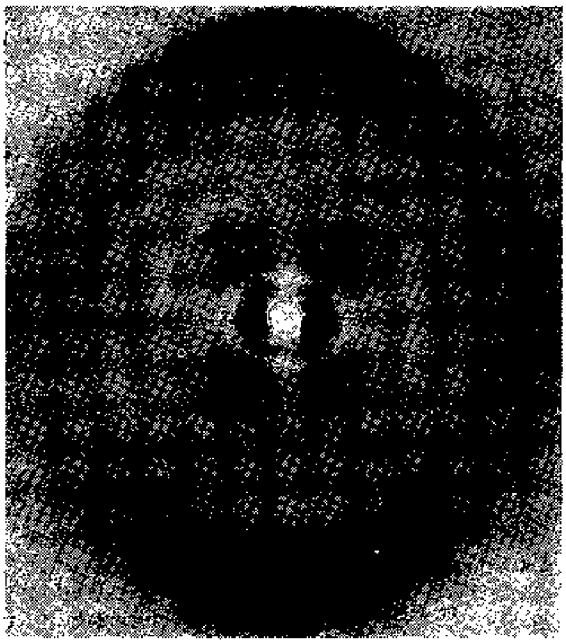


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_0^2) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[6] being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to *C* in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

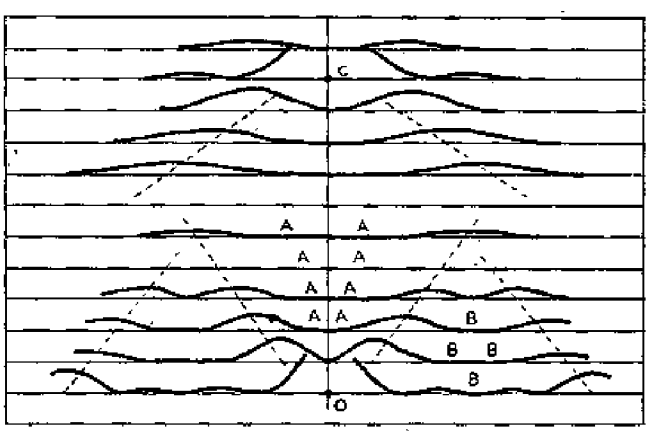


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About *C* on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

which has never been since surpassed. Dr. Schonland expressed disappointment that the membership in recent years has been but a little more than a thousand, for South Africa has expanded enormously since 1906 and with this expansion the need for, and potential value of, such a body as the Association. The general aims of the Association have not changed at all with the passing of years: "We exist," he said, "primarily to create and foster a scientific fraternity in South Africa, not to publish original work. We exist to provide a common meeting-ground for South African scientists and a forum for general discussion of the problems of this country from the scientific angle." He defended the use of Afrikaans by those who preferred it, for "we were intended by our founders to be parochial, and we should pride ourselves on being parochial. I would suggest that if we try to be anything else we will have mistaken our real aim".

Having thus firmly and, most people would agree, wisely placed the Association in its proper perspective, Dr. Schonland went on to make some concrete suggestions. The *South African Journal of Science* should have a series of semi-popular articles reviewing and surveying the new ideas of science and so bridge the gap between those who teach and do advanced research work and those who pay for it. This, he thought, is the proper function of the *Journal*, and it is but one aspect of the Association's duty, as representative of all sections of scientific opinion in South Africa, "to take a stronger, a more continuing and a more active interest in all scientific developments, national and university, in South Africa and to study carefully what is being done in other countries".

Besides his plea that the Association needs to form a standing committee to watch over scientific education in schools, Dr. Schonland suggested that the Association might consider taking a part in the formation of a body on the lines of the British Parliamentary and Scientific Committee and also help in the creation of better facilities for advanced research in South Africa. On this last-named point, he cited the instances of the National University in Canberra and the Institute for Advanced Studies in Dublin, but he made the interesting suggestion that a more acceptable solution might be the creation of a number of specialized institutes for advanced study, attached to and forming part of those universities which for one reason or another are best suited for them.

BASIS OF TECHNICAL EDUCATION

GENERAL education to-day should be planned so as to enable the ordinary citizen to adapt himself to the needs of technological society and to understand what is happening and what is required of him. This was the theme of an international conference convened by the United Nations Educational, Cultural and Scientific Organization at Unesco House in June 1950*.

Broadly, the Conference found that organized social foresight is essential to enable the educational system of a country to prepare children for the type of life and work they are likely to encounter, and that a substantial development of technical education is required at all levels: at present it is wholly inadequate for future needs, while the practical content of general education is also inadequate for the needs of future citizens of a technological society. The cultural content of technical education is also generally inadequate; technical education requires special consideration, and training for adaptability is an outstanding requirement in an age of ultra-rapid technological change. The education of women and girls also demands particular attention in view of their dual role as workers and home-makers, and improved administrative arrangements are essential if education is to fulfil its true function in such a society.

The report does not suggest that all these propositions apply equally to every country, though the Conference considered that, so far as its knowledge extended, they are generally valid for the world as a whole. The stress is laid on the need for adapting technology to man, not man to technology. The questions formulated in this report—and which merit attention in current discussions on the expansion of both technical and technological education in Great Britain—are raised in the belief that mastery of the machine by man is not an end in itself: it is a means to the development of man and of the whole society.

The distinction between technician and technologist is not always kept clear in this report, particularly in the chapter on the content of technical education. Nevertheless, the report directs attention to some fundamental issues which no sound policy for either type of education can disregard. In both fields it must be recognized that we are concerned not simply with the efficiency of production, but also with the fundamental attitude which the men and women of to-morrow will adopt in facing the problems of a technological society. Both, too, in seeking to foster flexibility, must recognize that flexibility is determined not only by education and training but also by social, economic and technical conditions; and the administrative measures required to ensure that education becomes more adapted to the needs of a changing technological society are themselves likely to be most effective when they are informal and varied rather than concentrated and uniform. The administrator, no less than the teacher and student, has need of frequent opportunities of contact with the industrial world, and requires experience of the difficulties and problems created by technological development in society; just as the teacher and student should keep abreast of developments in research and of practical applications in industry.

* Education in a Technological Society: a Preliminary International Survey of the Nature and Efficacy of Technical Education. (Tensions and Technology Series.) Pp. 76. (Paris: Unesco; London: H.M.S.O., 1952.) 200 francs; 4*s*.; 75 cents.

GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

By J. D. WATSON and F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

THE importance of deoxyribonucleic acid (DNA) within living cells is undisputed. It is found in all dividing cells, largely if not entirely in the nucleus, where it is an essential constituent of the chromosomes. Many lines of evidence indicate that it is the carrier of a part of (if not all) the genetic specificity of the chromosomes and thus of the gene itself.

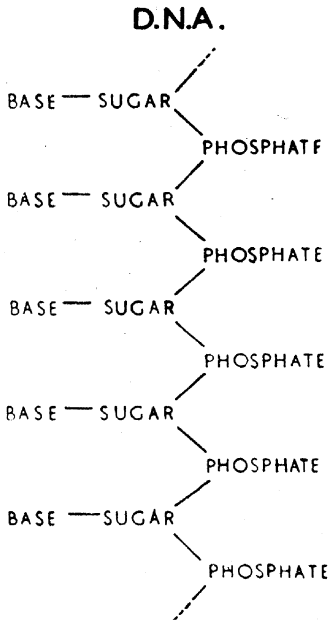


Fig. 1. Chemical formula of a single chain of deoxyribonucleic acid

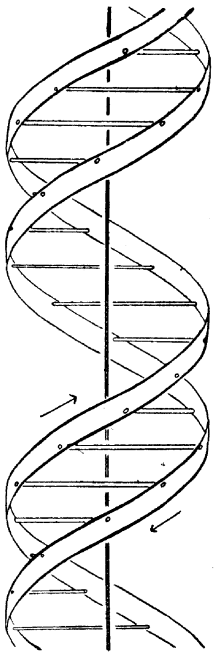


Fig. 2. This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

Until now, however, no evidence has been presented to show how it might carry out the essential operation required of a genetic material, that of exact self-duplication.

We have recently proposed a structure[1] for the salt of deoxyribonucleic acid which, if correct, immediately suggests a mechanism for its self-duplication. X-ray evidence obtained by the workers at King's College, London[2], and presented at the same time, gives qualitative support to our structure and is incompatible with all previously proposed structures[3]. Though the structure will not be completely proved until a more extensive comparison has been made with the X-ray data, we now feel sufficient confidence in its general correctness to discuss its genetical implications. In doing so we are assuming that fibres of the salt of deoxyribonucleic acid are not artefacts arising in the method of preparation, since it has been shown by Wilkins and his co-workers that similar X-ray patterns are obtained from both the isolated fibres and certain intact biological materials such as sperm head and bacteriophage particles[2,4].

The chemical formula of deoxyribonucleic acid is now well established. The molecule is a very long chain, the backbone of which consists of a regular alternation of sugar and phosphate groups, as shown in Fig. 1. To each sugar is attached a nitrogenous base, which can be of four different types. (We have considered 5-methyl cytosine to be equivalent to cytosine, since either can fit equally well into our structure.) Two of the possible bases—adenine and guanine—are purines, and the other two—thymine and cytosine—are pyrimidines. So far as is known, the sequence of bases along the chain is irregular. The monomer unit, consisting of phosphate, sugar and base, is known as a nucleotide.

The first feature of our structure which is of biological interest is that it consists not of one chain, but of two. These two chains are both coiled around a common fibre axis, as is shown diagrammatically in Fig. 2. It has often been assumed that since there was only one chain in the chemical formula there would only be one in the structural unit. However, the density, taken with the X-ray evidence[2], suggests very strongly that there are two.

The other biologically important feature is the manner in which the two chains are held together. This is done by hydrogen bonds between the bases, as shown schematically in Fig. 3. The bases are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other. The important point is that only certain pairs of bases will fit into the structure. One member of a pair must be a purine and the other a pyrimidine in order to bridge between the two chains. If a pair consisted of two purines, for example, there would not be room for it.

We believe that the bases will be present almost entirely in their most probable tautomeric forms. If this is true, the conditions for forming hydrogen bonds are more restrictive, and the only pairs of bases possible are :

adenine with thymine ;
guanine with cytosine.

The way in which these are joined together is shown in Figs. 4 and 5. A given pair can be either way round. Adenine, for example, can occur on either chain ; but when it does, its partner on the other chain must always be thymine.

This pairing is strongly supported by the recent analytical results[5], which show that for all sources of deoxyribonucleic acid examined the amount of adenine is close to the amount of thymine, and the amount of guanine close to the amount of cytosine, although the cross-ratio (the ratio of adenine to guanine) can vary from one source to another. Indeed, if the sequence of bases on one chain is irregular, it is difficult to explain these analytical results except by the sort of pairing we have suggested.

The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information. If the actual order of the

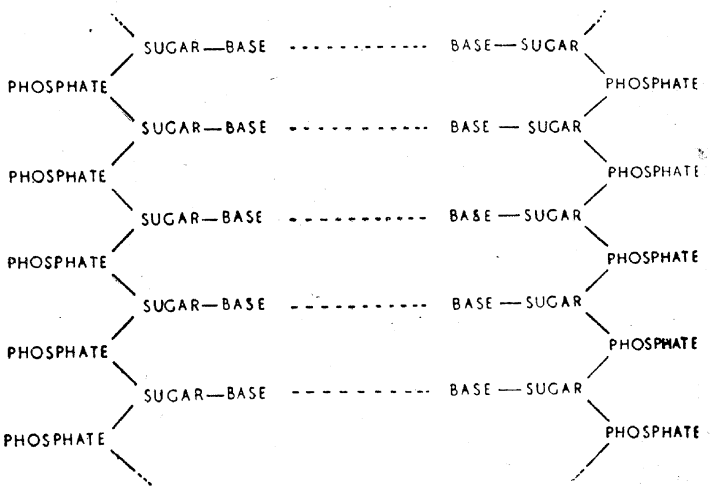


Fig. 3. Chemical formula of a pair of deoxyribonucleic acid chains. The hydrogen bonding is symbolized by dotted lines

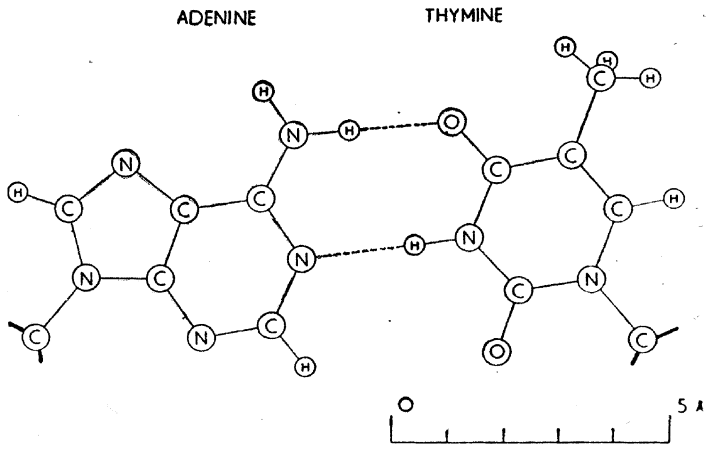


Fig. 4. Pairing of adenine and thymine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

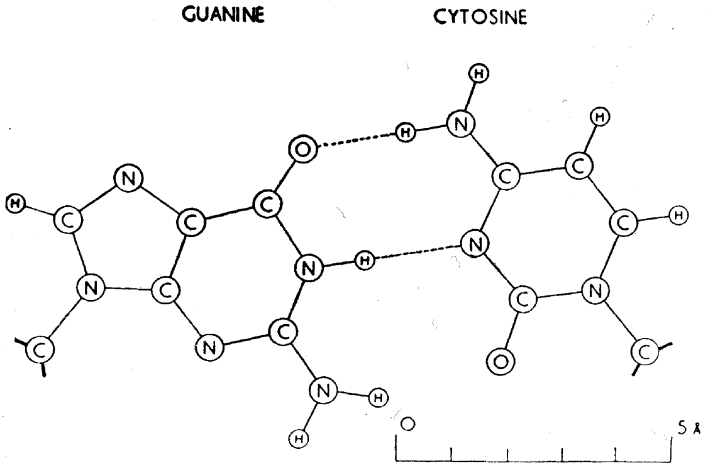


Fig. 5. Pairing of guanine and cytosine. Hydrogen bonds are shown dotted. One carbon atom of each sugar is shown

bases on one of the pair of chains were given, one could write down the exact order of the bases on the other one, because of the specific pairing. Thus one chain is, as it were, the complement of the other, and it is this feature which suggests how the deoxyribonucleic acid molecule might duplicate itself.

Previous discussions of self-duplication have usually involved the concept of a template, or mould. Either the template was supposed to copy itself directly or it was to produce a 'negative', which in its turn was to act as a template and produce the original 'positive' once again. In no case has it been explained in detail how it would do this in terms of atoms and molecules.

Now our model for deoxyribonucleic acid is, in effect, a *pair* of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken, and the two chains unwind and separate. Each chain then acts as a template for the formation on to itself of a new companion chain, so that eventually we shall have *two* pairs of chains, where we only had one before. Moreover, the sequence of the pairs of bases will have been duplicated exactly.

A study of our model suggests that this duplication could be done most simply if the single chain (or the relevant portion of it) takes up the helical configuration. We imagine that at this stage in the life of the cell, free nucleotides, strictly polynucleotide precursors, are available in quantity. From time to time the base of a free nucleotide will join up by hydrogen bonds to one of the bases on the chain already formed. We now postulate that the polymerization of these monomers to form a new chain is only possible if the resulting chain can form the proposed structure. This is plausible, because steric reasons would not allow nucleotides 'crystallized' on to the first chain to approach one another in such a way that they could be joined together into a new chain, unless they were those nucleotides which were necessary to form our structure. Whether a special enzyme is required to carry out the polymerization, or whether the single helical chain already formed acts effectively as an enzyme, remains to be seen.

Since the two chains in our model are intertwined, it is essential for them to untwist if they are to separate. As they make one complete turn around each other in 34 A., there will be about 150 turns per million molecular weight, so that whatever the precise structure of the chromosome a considerable amount of uncoiling would be necessary. It is well known from microscopic observation that much coiling and uncoiling occurs during mitosis, and though this is on a much larger scale it probably reflects similar processes on a molecular level. Although it is difficult at the moment to see how these processes occur without everything getting tangled, we do not feel that this objection will be insuperable.

Our structure, as described[1], is an open one. There is room between the pair of polynucleotide chains (see Fig. 2) for a polypeptide chain to wind around the same helical axis. It may be significant that the distance between adjacent phosphorus atoms, 7·1 A., is close to the repeat of a fully extended polypeptide chain. We think it probable that in the sperm head, and in artificial nucleoproteins, the polypeptide chain occupies this position. The relative weakness of the second layer-line in the published X-ray pictures[3a,4] is crudely compatible with such an idea. The function of the protein might well be to control the coiling and uncoiling, to assist in holding a single polynucleotide chain in a helical configuration, or some other non-specific function.

Our model suggests possible explanations for a number of other phenomena. For example, spontaneous mutation may be due to a base occasionally occurring in one of its less likely tautomeric forms. Again, the pairing between homologous chromosomes at meiosis may depend on pairing between specific bases. We shall discuss these ideas in detail elsewhere.

For the moment, the general scheme we have proposed for the reproduction of deoxyribonucleic acid must be regarded as speculative. Even if it is correct, it is clear from what we have said that much remains to be discovered before the picture of genetic duplication can be described in detail. What are the polynucleotide precursors ? What makes the pair of chains unwind and separate ? What is the precise role of the protein ? Is the chromosome one long pair of deoxyribonucleic acid chains, or does it consist of patches of the acid joined together by protein ?

Despite these uncertainties we feel that our proposed structure for deoxyribonucleic acid may help to solve one of the fundamental biological problems—the molecular basis of the template needed for genetic replication. The hypothesis we are suggesting is that the template is the pattern of bases formed by one chain of the deoxyribonucleic acid and that the gene contains a complementary pair of such templates.

One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis (U.S.A.).

[1] Watson, J. D., and Crick, F. H. C., *Nature*, **171**, 737 (1953).
[2] Wilkins, M. H. F., Stokes, A. R., and Wilson, H. R., *Nature*, **171**, 738 (1953). Franklin, R. E., and Gosling, R. G., *Nature*, **171**, 740 (1953).
[3] (*a*) Astbury, W. T., Symp. No. 1 Soc. Exp. Biol., 66 (1947). (*b*) Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952). (*c*) Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953). (*d*) Fraser, R. D. B. (in preparation).
[4] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).
[5] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952). Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).

GEOPHYSICAL AND METEOROLOGICAL CHANGES IN THE PERIOD JANUARY–APRIL 1949

IN a recent article[1] Lewis and McIntosh have considered the geophysical data for the period January–April 1949, which we presented in an earlier communication[2]. On the basis of certain probability criteria they appear to show that the apparent regular variations in ionospheric and meteorological phenomena which occurred in that period were not significant. We have studied their article and made a separate statistical analysis of the *unsmoothed* data, and conclude that in all respects our original suggestions seem to be valid.

In our original article we presented graphs showing five-day moving averages in four parameters: (*a*) ground pressure, p; (*b*) E-layer critical frequency, fE; (*c*) F-layer critical frequency, $fF2$; and (*d*) K-index of geomagnetic activity. The connexion between ionospheric and geomagnetic phenomena is well known. Thus, Appleton and Ingram[3] in 1935 established the correlation between geomagnetic activity and depressions in $fF2$. It is worthy of note that in the period under discussion the inverse correlation between K and $\Delta fF2$ is, as Lewis and McIntosh point out, considerably less striking than that between p and ΔfE (cf. Figs. 1 and 2 in our original article). It would seem, then, that if statistical analysis can be successfully applied to show that there is no significance between the variations in p and ΔfE, it is, *a fortiori*, evident that a similar analysis might, in the present instance, be used for discrediting the established relationship between K and $\Delta fF2$. Conversely, of course, the fact that a phenomenon appears to be statistically significant over a short period must likewise be treated with reserve. The need for the utmost care in the application and interpretation of statistical analyses to such a limited time series is thus clear.

From inspection of our graphs it seemed to us that, so far as p and ΔfE were concerned, the period was unusual in three respects: (i) there appeared to be four oscillations in ground pressure showing a progressive diminution of amplitude, with an average period of about 27 days; (ii) in like manner there appeared to be four marked oscillations of period about 27 days in ΔfE; (iii) oscillations (i) and (ii) appeared to be almost exactly out of phase. In addition, we noted that the period was characterized by an unusual 27-day recurrence of great sudden commencement (S.C.) magnetic storms.

In our original communication we merely directed attention to these matters, and suggested that there might be some connexion between them. We did not then suggest, nor do we now suggest, that from a period of length only four months any conclusions can be drawn regarding the general behaviour over a long period of any of the geophysical parameters considered. The severely limited number of observations available, together with the fact that there is considerable uncertainty about the correct statistical approach to time series analysis, seemed to us sufficient reason for not entering into an extended statistical analysis.

However, the contrary conclusions reached by Lewis and McIntosh (see below) have prompted us to re-examine the data. Briefly, their conclusions are: (i) the 27-day oscillation in ground pressure is of no significance, since the amplitude is no more than would be expected from mere chance considerations; (ii) the 27-day oscillation in ΔfE is probably significant; (iii) oscillations (i) and (ii) are exactly in anti-phase; (iv) there is no significant correlation coefficient between the p and ΔfE data; (v) our conclusions arise from smoothing of the data.

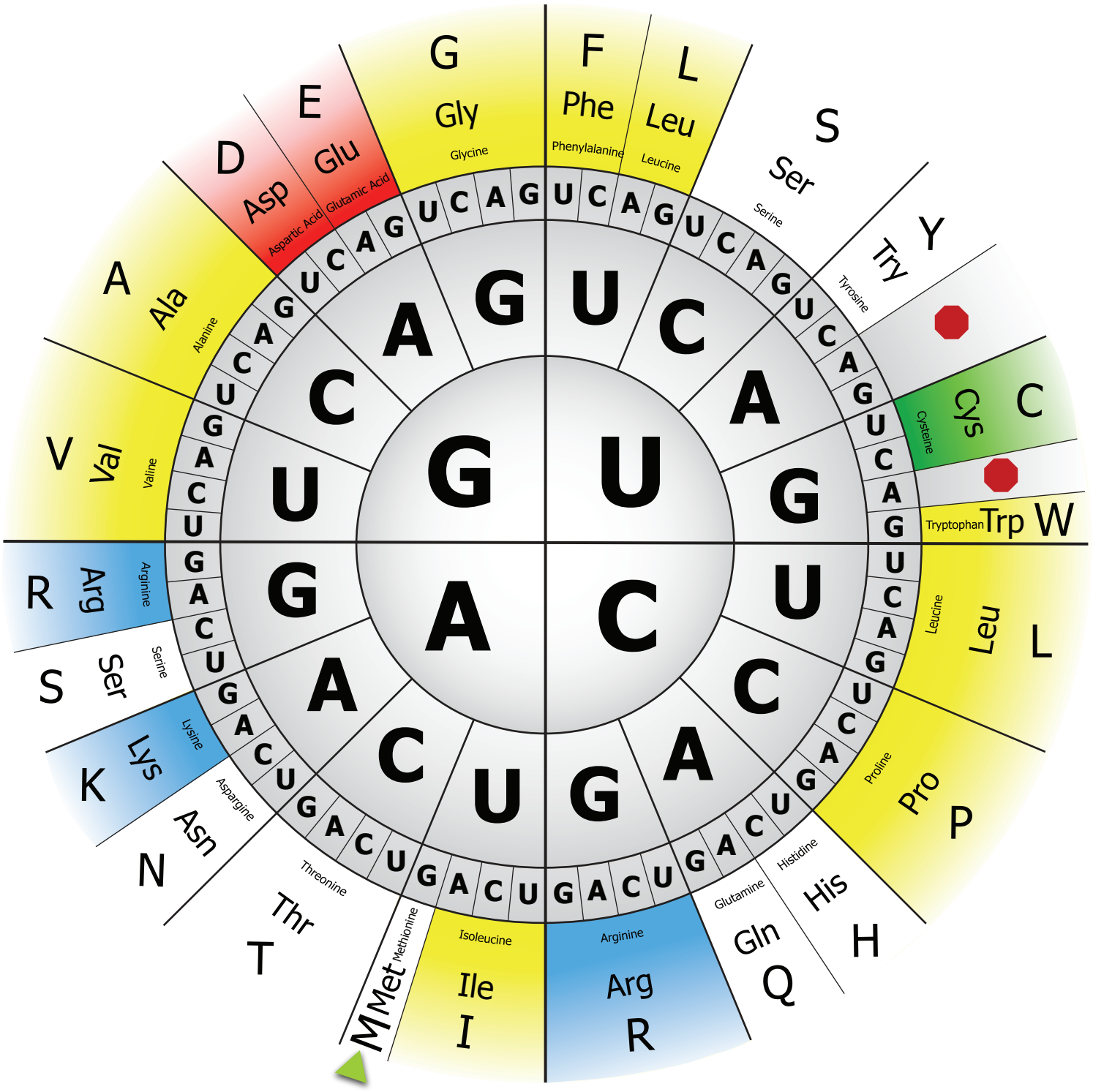
We shall now outline our own analysis. In various communications[4–6], Kendall has made it abundantly clear that most of the methods generally used for studying periodicities in time series (for example, periodograms, Fourier analysis, etc.) may yield very misleading results when applied to the kind of time series with which we are here concerned. He has also questioned the reliability of the usual significance tests for periodicities when applied in time series analysis. Kendall has shown that the most reliable approach is that of serial correlation coefficients as exhibited in the correlogram. He points out that although the correlogram may be insensitive, it does give a lower limit to the oscillatory effects, and that if it oscillates there is almost certainly some systematic oscillation in the primary series explored. Figs. 1 and 2 show the correlograms for Δp and ΔfE respectively for the period under consideration. In both of these the original *unsmoothed* data have been used.

It is important to note that there is a marked trend in the pressure data, and to eliminate this we have dealt with values of pressure departures, Δp (as with the fE data), rather than with the absolute magnitudes p. The oscillations in both correlograms are clear, with a maximum at 26–27 days in each case. These correlograms provide strong support for our original deductions (based, as they were, on simple inspection of graphs), and make it essential for us to repeat Lewis and McIntosh's calculations.

At the outset we must again stress that the pressure data exhibit a marked downward trend (approximately linear), and it is imperative initially to eliminate this before proceeding with any numerical analysis. It appears that Lewis and McIntosh have overlooked this point, and as a result have arrived at quite contrary conclusions. This will be clear from an examination of Table 1, in which we present the results of calculations made by us using (i) pressure, p, (ii) pressure departures, Δp, and (iii) fE departures, ΔfE. The nomenclature employed (c, φ, σ, etc.) is that used by Lewis and McIntosh.

Without going into details, it can be stated that there is little significant difference between the present results *using pressure*, p, and those given by Lewis and McIntosh. The slight differences in the values of amplitude c and first serial correlation coefficient r_1 are of no significance and can be ascribed to different ways of deducing the amplitude and phase

The Genetic Codon Chart©



Amino Acid Properties

- Translation Start Codon
- Hydrophilic / Polar
- Hydrophobic / Non-polar
- Cysteine
- Translation Stop Codon
- Negative Charge
- Positive Charge

3-D Molecular Designs ...where molecules become real™
3dmoleculardesigns.com

MSOE Center for BioMolecular Modeling
cbm.msoe.edu

 Version 3.0

3-D Molecular Designs ...where molecules become real™

The Standard Genetic Code

Center for BioMolecular Modeling

	U	C	A	G	
U	UUU → Phe	UCU → Ser	UAU → Tyr	UGU → Cys	U
	UUC → Phe	UCC → Ser	UAC → Tyr	UGC → Cys	C
	UUA → Leu	UCA → Ser	UAA → Stop	UGA → Stop	A
	UUG → Leu	UCG → Ser	UAG → Stop	UGG → Trp	G
C	CUU → Leu	CCU → Pro	CAU → His	CGU → Arg	U
	CUC → Leu	CCC → Pro	CAC → His	CGC → Arg	C
	CUA → Leu	CCA → Pro	CAA → Gln	CGA → Arg	A
	CUG → Leu	CCG → Pro	CAG → Gln	CGG → Arg	G
A	AUU → Ile	ACU → Thr	AAU → Asn	AGU → Ser	U
	AUC → Ile	ACC → Thr	AAC → Asn	AGC → Ser	C
	AUA → Ile	ACA → Thr	AAA → Lys	AGA → Arg	A
	AUG → Met	ACG → Thr	AAG → Lys	AGG → Arg	G
G	GUU → Val	GCU → Ala	GAU → Asp	GGU → Gly	U
	GUC → Val	GCC → Ala	GAC → Asp	GGC → Gly	C
	GUA → Val	GCA → Ala	GAA → Glu	GGA → Gly	A
	GUG → Val	GCG → Ala	GAG → Glu	GGG → Gly	G

translation start codon

translation stop codon

hydrophobic amino acids

hydrophilic non-charged amino acids

- charged amino acids

+ charged amino acids

cysteine

Pilot Study Version

DNA Starter Kit

1-Group Set

3-D Molecular Designs

...where molecules become real ™

DNA Replication and Transcription Activity

The purpose of this exercise is to explore the steps involved in DNA replication and transcription, simulating these activities using foam models of the nucleotides. At the end of this activity, you should be able to describe each process and identify the enzymes involved.

Open the bag labeled "DNA Starter Kit" and separate the parts. Each kit should contain:

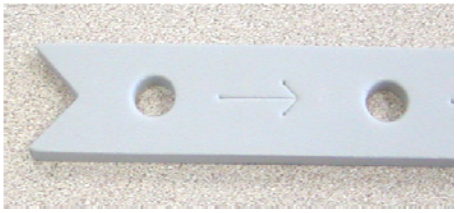
- 12 red adenine
- 12 blue cytosine
- 12 green guanine
- 12 yellow thymine
- 6 white uracil
- 4 long gray DNA backbone pieces – note the arrows on the backbone – point from 5' phosphate end to 3' OH end
- 24 gray sugar-phosphate backbone pieces (representing deoxyribose linked to a phosphate for DNA)
- 12 purple sugar phosphate backbone pieces (representing ribose linked to a phosphate for RNA)
- 56 orange phosphate pieces
- 48 orange phosphate linkers

If something is missing from your kit or you have too many pieces, notify your instructor.

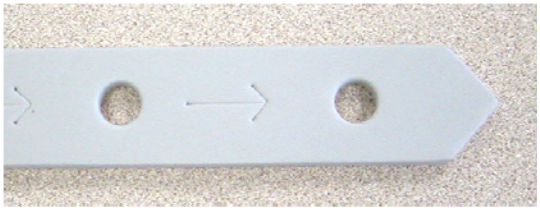
Construct Double-Stranded DNA

1. Look at the long gray backbone piece. Note the arrows on one side of these pieces:

5 ' phosphate end of the backbone



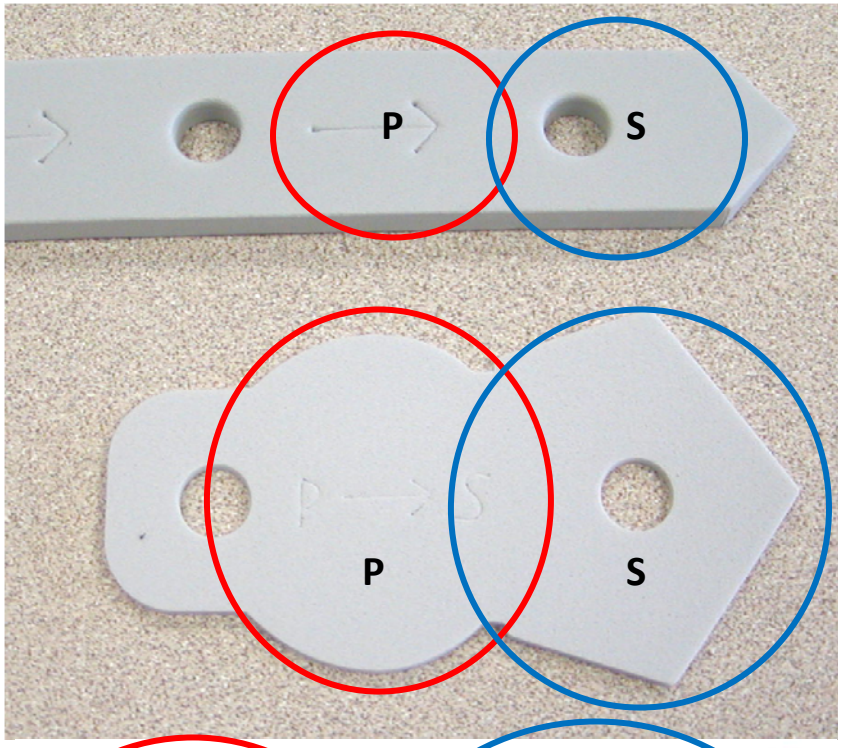
3' OH (sugar) end of the backbone



5' ⟶ 3'

This model is a schematic model – since the sugar and phosphate structures are not clearly evident. The holes in the backbone are the attachment sites for the nucleotides, so they represent the 1' carbon of the sugar molecules.

The arrows point from the 5' phosphate end of the backbone towards the 3' OH (sugar) end of the backbone. The smaller individual sugar-phosphate pieces also have an arrow that points from the 5' end to the 3' end of the piece:



In the smaller pieces, the shape of the sugar (pentagon) is more visible, and the phosphate is represented by the circle. You'll also see a stamp that shows P → S on this model as well.

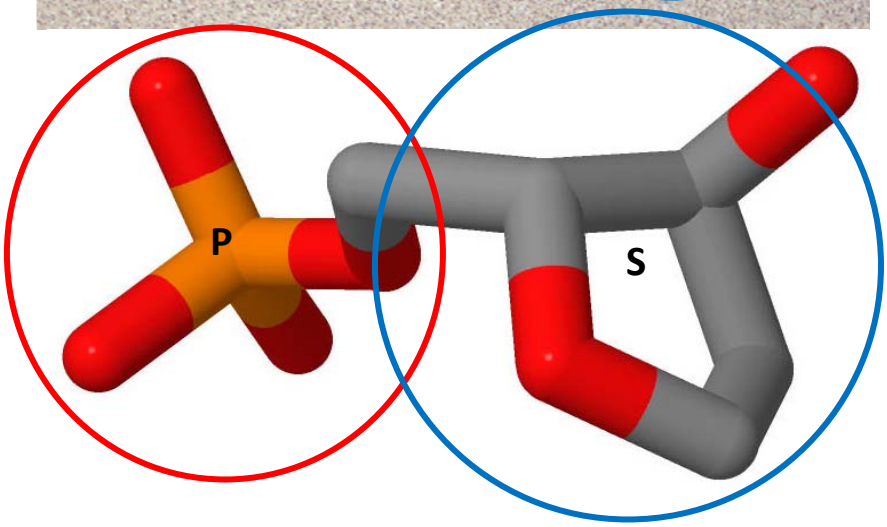
Below the images of the schematic phosphate-sugar backbone pieces is a more accurate representation of the backbone, based on the X-ray crystal structure, displayed in Jmol.

The Jmol image is color-coded, such that:

Gray is carbon
Red is oxygen
Orange is phosphorus

In each depiction, the phosphate group is circled in **red**, and the sugar group is circled in **blue**.

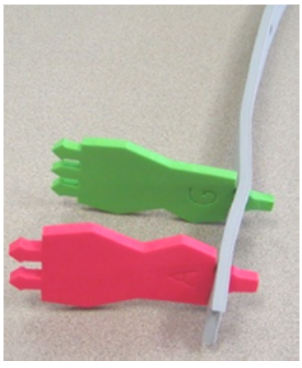
Each of these models, though different, represents the same structure.



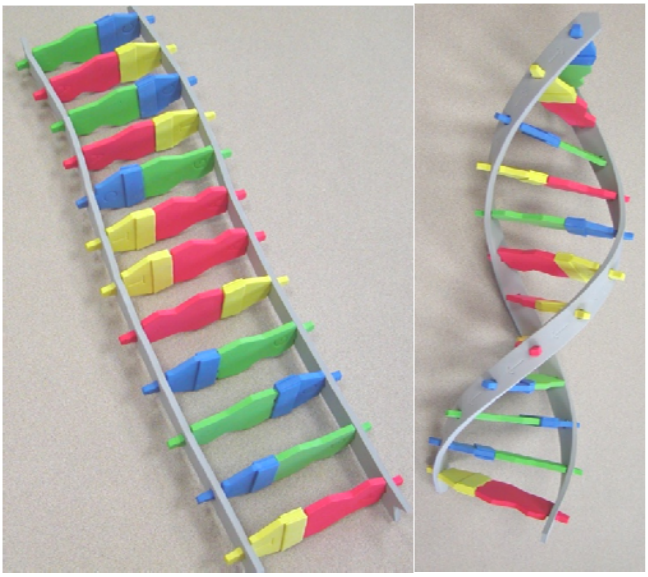
5' ⟶ **3'**

All nucleic acids (both DNA and RNA) are synthesized in the 5' to 3' direction – by polymerases. DNA polymerases make DNA, and RNA polymerases make RNA.

2. Using 3 each of adenine, cytosine, guanine and thymine and a long DNA backbone piece, construct a segment of single stranded DNA. Place the nucleotides in any order on the chain. (See figure at right.)



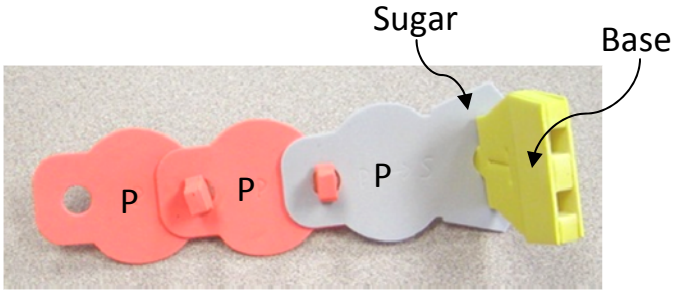
3. Next, build the second strand of DNA **complementary** to the first, using 3 of each of the nucleotides A, G, C and T, and a second long DNA backbone. Be sure the backbone is running **antiparallel** (in the opposite direction) to the first backbone. Add the diphosphate to the 5' end of this DNA strand, and join the base pairs.



4. Twist your DNA into a right handed helix by holding the two ends in your hands and twisting toward you with your right hand and away from you with your left hand. Verify that this is a right-handed helix. (Hint: Imagine walking up the DNA as a spiral staircase. Which hand is on the OUTSIDE railing?)

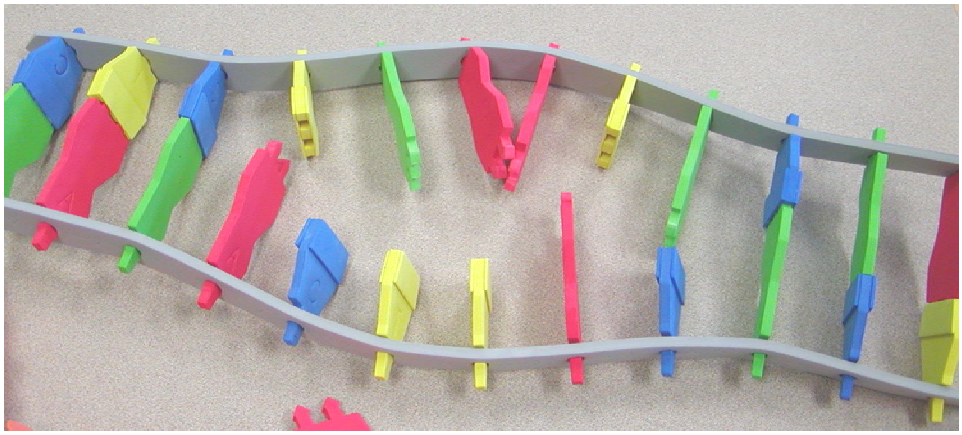
5. **Complete Part A of the DNA Replication and Transcription Worksheet before proceeding.**

DNA Replication



1. DNA is synthesized by DNA polymerase using nucleotide triphosphates as substrates. Build nucleotide triphosphates with the remaining A, G, C and T bases. Each of these will have a base (A, G, C or T) joined to a gray sugar (the pentagon side of the gray sugar phosphate). See illustration at right. The base is attached to the 1' carbon and the phosphate group is attached to the 5' carbon of the sugar molecule, but these individual atoms are not identifiable in this model of DNA. Two phosphate groups will be attached to the phosphate using the orange phosphate connectors.

2. Create replication bubble – pulling the middle 6 bases apart. The end bases will remain intact for the time being.



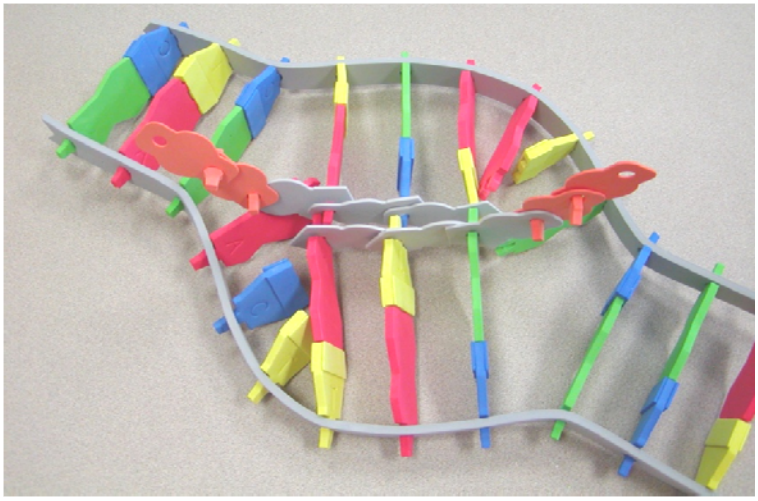
3. New DNA strands are synthesized in the 5’ to 3’ direction – so opposite directions on each strand.

4. Energy for each nucleotide addition comes from splitting off the two phosphates. Just as ATP is an energy carrier molecule, so are GTP, CTP and TTP.

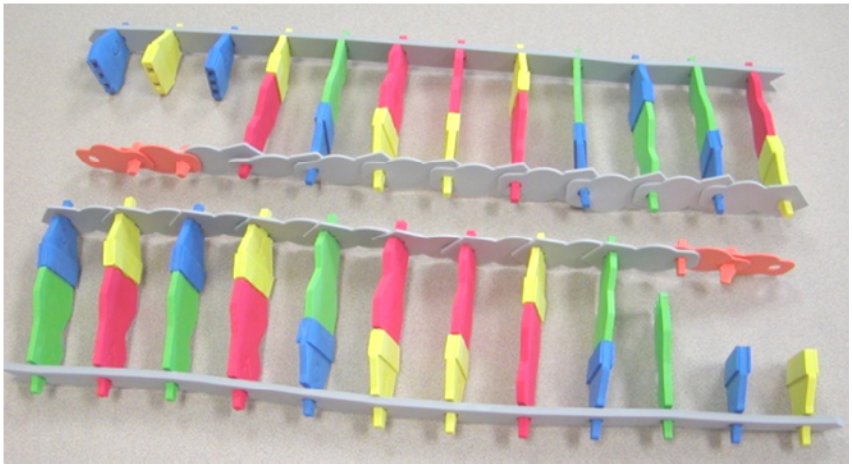
5. Initiation of DNA replication begins with a short RNA primer. Since you are working with such a short strand of DNA, we’re not going to simulate the RNA primer in this activity.

6. Add a nucleotide that is complementary to the nucleotide at the 3’ end of one of the template strands within the replication bubble. Be sure you orient the nucleotide so the arrows on the sugar-phosphate piece are running in the *opposite* direction from the arrows on the backbone of the template strand. Similarly, add a nucleotide complimentary to the nucleotide at the 3’ end of the *other* template strand. Make sure that the arrows on the backbone pieces on the two strands point in opposite directions. Note that replication on the two strands does not start in the same place.

7. Bring in the next nucleotide on each strand – being sure to ‘grow’ your new chain in the 5’ to 3’ direction. Remove 2 phosphate groups as you add each new nucleotide to the sugar of the preceding nucleotide. Note that the first nucleotide retains its phosphate groups. Recall that ATP is considered the “energy currency” of the cell, and that removing one or more phosphate groups from ATP releases energy to drive chemical reactions. Similarly, CTP, GTP and TTP also have stored energy in the phosphate bonds. Energy released from cleaving the phosphates from each of these molecules is used to form the phosphodiester bond in the DNA backbone.



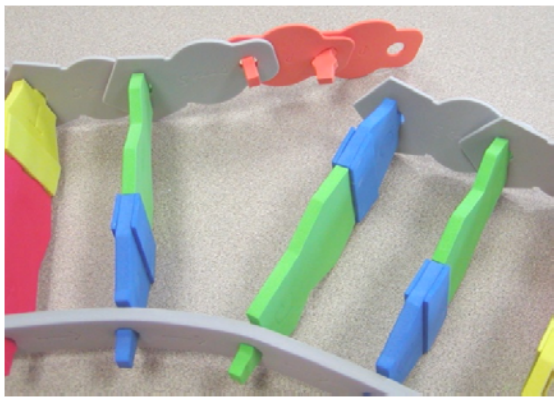
8. Build your two DNA strands, 5’ to 3’, opening the bubble and adding bases complimentary to the template strand.



9. When you are done, you should have two double stranded DNA sequences that lack two nucleotides on the 5’ end. Because the replication bubble proceeds in BOTH directions, and DNA synthesis can only occur from 5’ to 3’, for each replication fork, there is a leading strand (synthesized 5’ to 3’) and a lagging strand (running 3’ to 5’). In order to fill in the gaps on the lagging strand, synthesis is delayed, then build 5’ to 3’ in

segments of approximately 200 nucleotides. These short segments were discovered by Okazaki and coworkers, they are aptly named "Okazaki fragments".

10. Simulate the filling in of the Okazaki fragments by adding bases at the 5' end of your DNA chains. Be sure to add the bases 5' to 3'. Remember that DNA polymerase can add bases one at a time in the 5' to 3' direction. But DNA polymerase cannot join the backbone between two nucleotides that have already been added to the chain.



11. A separate enzyme, DNA ligase, joins the backbone between the Okazaki fragments. Simulate DNA ligase to create intact sugar-phosphate backbones on both of your DNA strands.

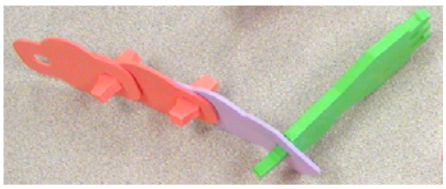
12. At this point you should have two nearly identical DNA double strands. What is the difference between the two strands? [Hint: Observe the solid vs. the pieced backbone pieces. Which strand is pieced on the two models?] The solid strand represents the ORIGINAL DNA, and the pieced strand represents the newly synthesized DNA. Back in the 1950's scientists didn't know exactly HOW DNA replication occurred – whether the old sequence remained intact and a completely new DNA was synthesized (termed "conservative replication") or whether each strand served as a template for a new strand, so that the daughter DNA contained one old strand and one new strand (termed "semi-conservative replication" since one half of the DNA is conserved in the new strand). Through a series of experiments using two isotopes of phosphorous, Meselson and Stahl were able to demonstrate that DNA replication is semi-conservative and that each strand serves as a template to build a new complementary strand of DNA. Your model demonstrated semi-conservative replication.

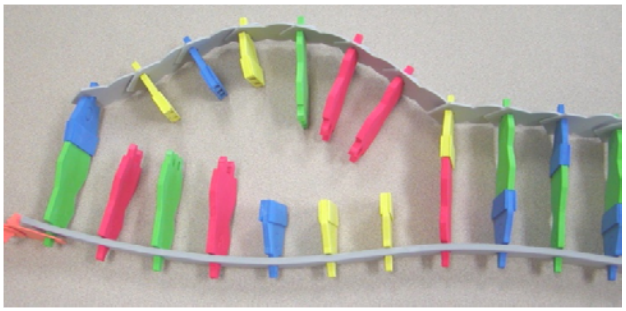
6. Complete Part B of the DNA Replication and Transcription Worksheet before proceeding.

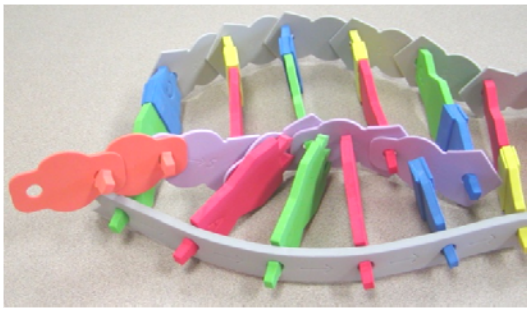
Transcription into mRNA

1. Human DNA contains sequences coding for genes, sometimes thousands of bases long, as well as regulatory sequences that are involved in turning genes on and off. Sometimes regulators function more like dimmer switches, fine tuning the expression of genes. Your segment of DNA is too short to model regulatory sequences. – Instead, you'll focus on the basics of transcription into RNA.

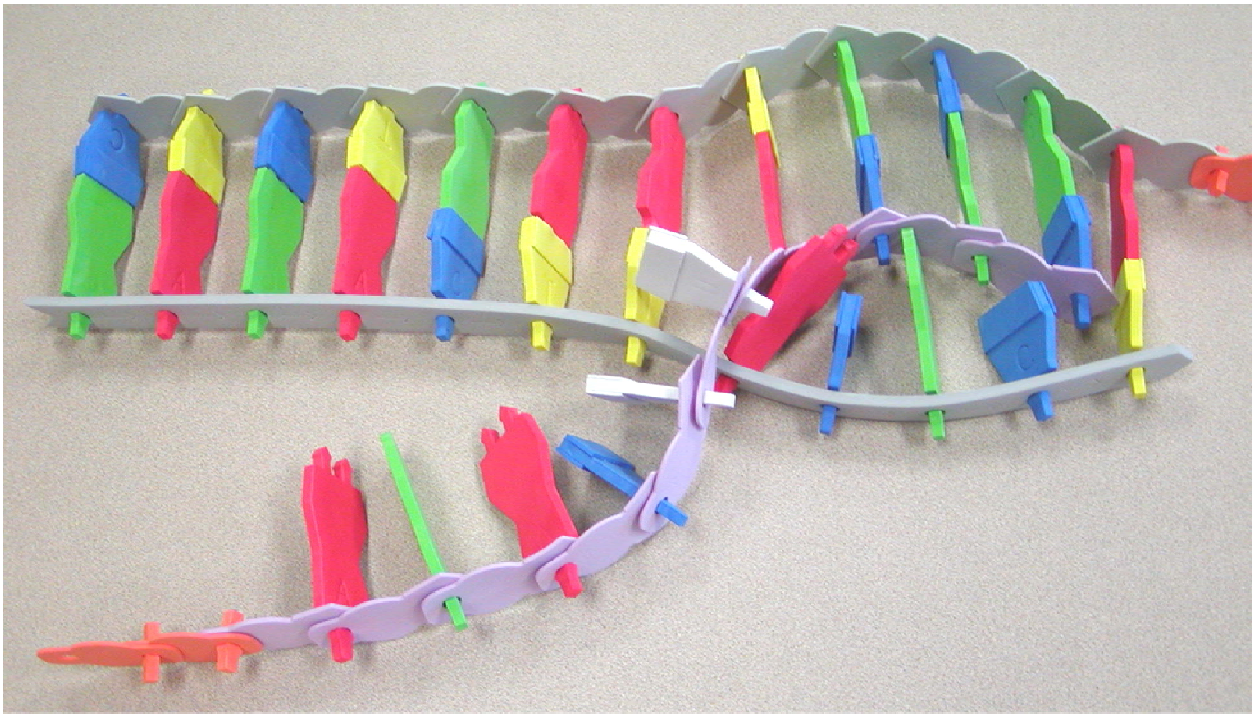
2. When DNA is transcribed into mRNA, only one strand of DNA is used. This is termed the template strand. Select one of your double-stranded DNA molecules to use. Take apart the other model of DNA. (You'll be using pieces of it for the next step.) Next, create ribonucleotide triphosphates by attaching the lavender ribose phosphate, then two additional orange phosphate groups, to each of three A (red), C (blue), G (green) and U (white) nucleotides.



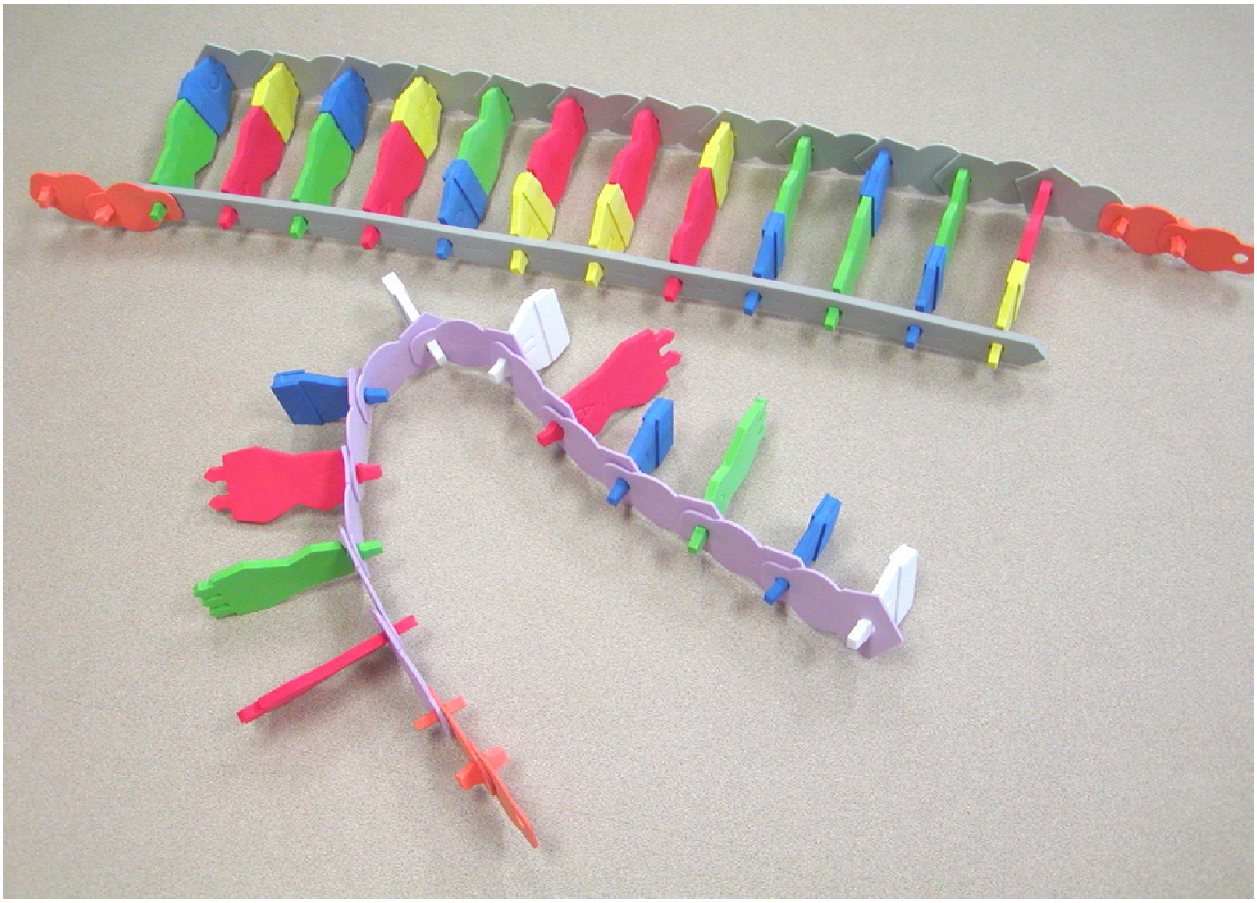




You will be using the newly synthesized strand of DNA (the one with the pieced backbone) as the template strand. The RNA polymerase binds to the DNA and unzips the DNA, forming a bubble similar to replication. In this simulation, you should form a bubble of 6 bases beginning two bases in from the 3' end of the template strand. Since RNA polymerase builds the RNA from 5' to 3', it must read the DNA from 3' to 5'. Using the template strand, build a complementary mRNA. As the transcription bubble moves along the DNA, each time you unzip one DNA base pair to add a ribonucleotide to the growing mRNA chain, you should remove one ribonucleotide from the start of the mRNA molecule and rezip the DNA. In this simulation, only 4 base pairs of DNA should be unzipped at once. (In reality, about 10 base pairs are unzipped and bound to mRNA at any given time.) You will have a double stranded DNA model with a few bases unzipped. One strand of the unzipped DNA will be bound to mRNA, which will have a single stranded "tail" that grows as RNA polymerase moves down the DNA.



3. Once you reach the end of the DNA, the mRNA will continue detaching from the DNA and the DNA will rezip. In the end, you should have a single-stranded mRNA molecule complementary to the DNA template strand, and a double-stranded DNA molecule.



4. **Complete Part C of the DNA Replication and Transcription Worksheet before proceeding.**

5. Disassemble your DNA and mRNA and place all the parts in the appropriate bags.

DNA Replication and Transcription Worksheet

Lending Library: DNA Starter Kit© (DS)

Part A: DNA Structure

1. Explain what is meant by each of the following terms as they relate to DNA structure:

   A. Complementary base pairing

   B. Antiparallel strands

   C. Sugar-phosphate backbone

2. Record the sequence of your DNA structure, and indicate 5’ phosphate and 3’ carbon ends of the strands. Draw lines between complementary base pairs:

Part B: DNA Replication

1. Explain what is meant by semi-conservative DNA replication.

2. Forming the phosphodiester linkage in the DNA backbone is an endergonic reaction. What is the source of the energy that drives this reaction?

3. In what direction is the newly synthesized DNA made?

4. Recall that the two DNA strands are antiparallel. In what direction is the DNA polymerase reading the *template* strand?

5. What is an Okazaki fragment?

6. What is the purpose of DNA ligase?

Part C: Transcription

1. What are two differences between DNA and RNA?

2. Complete the following table comparing and contrasting DNA replication with transcription:

DNA replication	Characteristic	Transcription
	How many strands of DNA are used as the template?	
	Direction of synthesis of the new strand?	
	Direction of movement of polymerase along the template strand?	
	What happens to the new strand as it is being made?	
	Enzyme(s) involved in the process	

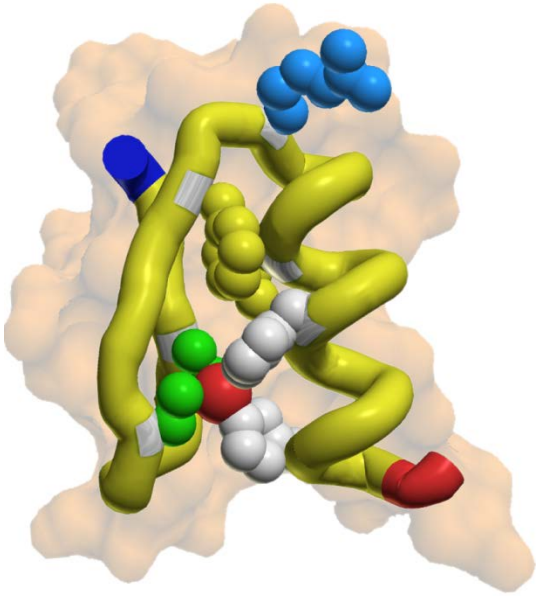
Center for BioMolecular Modeling

Zinc Finger Folding Activity

Based on amino acids 4-31 of 1zaa.pdb

Parts List

- Toober segment – 28 amino acids long (approx 72 cm)
- Sidechains
  - 2 Cys 2 His 1 Phe
  - 1 Arg 1 Leu
- 7 metal clips
- Zinc Atom
- Zinc Finger Folding Map
- 1 Blue end cap (for designating the amino terminus)
- 1 Red end cap (for designating the carboxy terminus)



Introduction

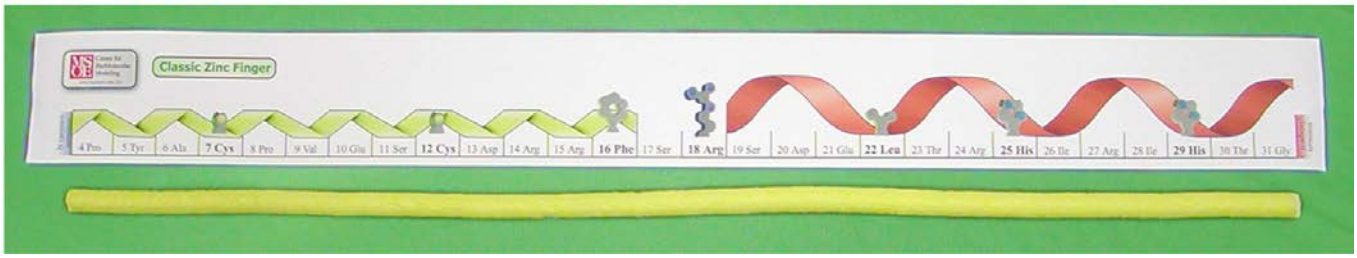
A C_2H_2 zinc finger is a 28 amino acid protein motif composed of a short alpha helix and a two-stranded beta sheet. The structure of the zinc finger is stabilized by a zinc atom that binds 2 cysteine and 2 histidine sidechains, and by hydrophobic amino acid sidechains that are buried on the inside of the folded motif. Zinc finger proteins function as regulators of gene expression. They bind to the negatively-charged backbone of DNA through a positively-charged arginine amino acid sidechain located at the beginning of the short alpha helix.

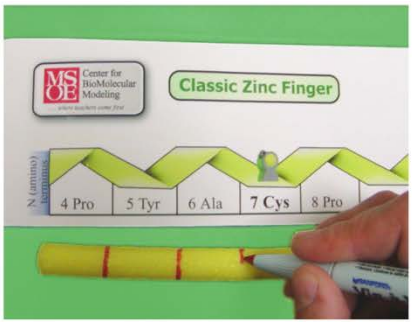
The construction of a physical model of the 3D structure of a zinc finger serves as a good example of how Toobers can be used to model protein structures.

This kit is based on 1ZAA.pdb and represents amino acids 4-31.

Getting Started: Laying out the primary sequence

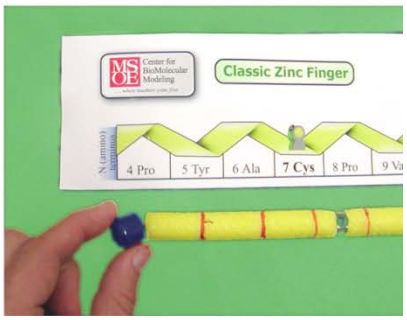
1. Lay out the protein folding map and position the toober below it.



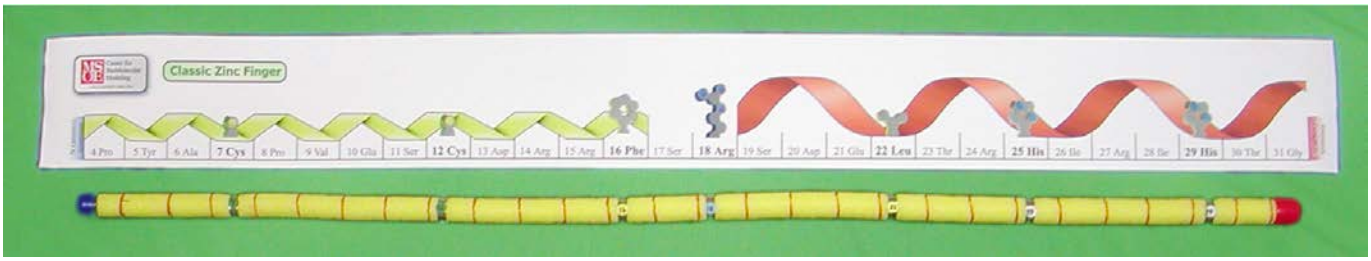


2. Align the toober with the protein folding map and draw a line for each amino acid.

3. Number and place the seven clips on the toober according to the protein folding map.

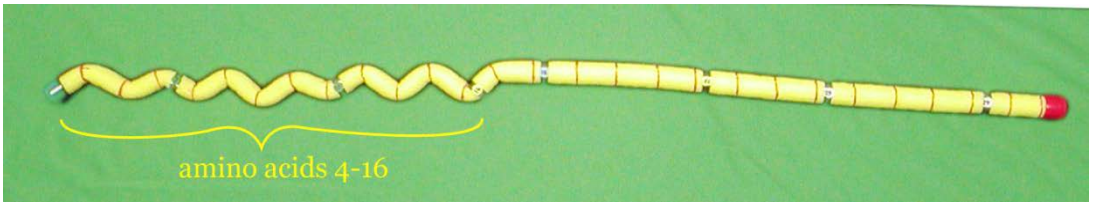


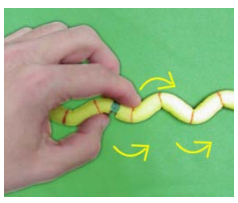
4. Add the blue and red end caps.



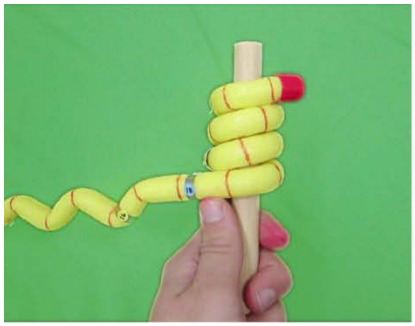
Folding the Protein

1. Bend the first 13 amino acids into a β-sheet (amino acids 4-16). You can create a β-sheet by bending the toober at every amino acid (approximately every 2 cm).

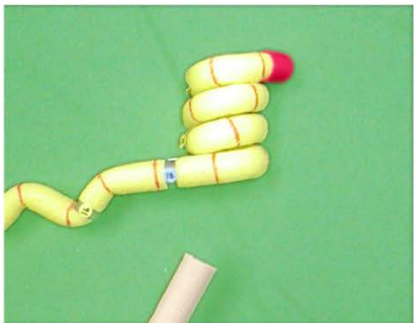




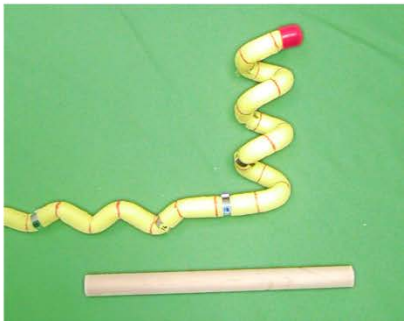
2. Next, create an α-helix with amino acids 19-31 (the last 13 amino acids of the protein). Do this by wrapping the toober segment around an alpha helix bending jig (wooden dowel or your finger) and then stretch it out so that so that there are approximately 3.6 amino acids per turn of the helix.



Wrap the toober around the jig.



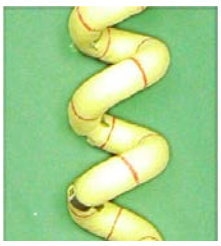
Slide the toober off the jig.

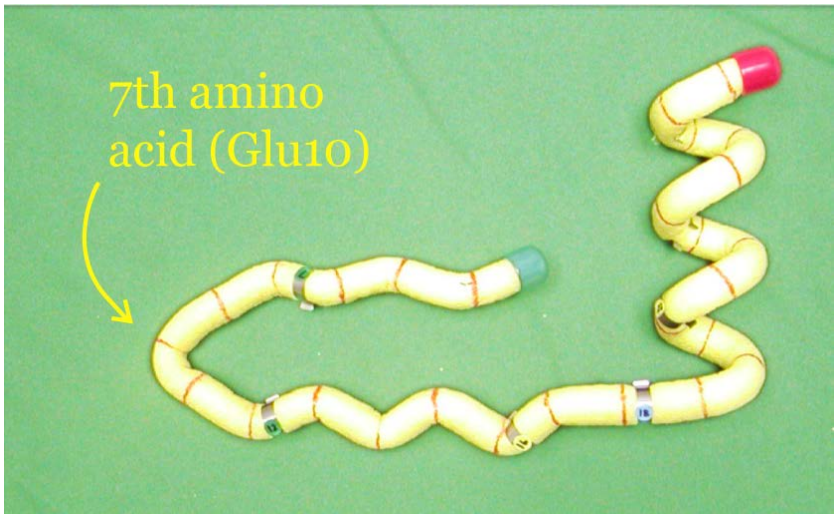


Stretch the toober.

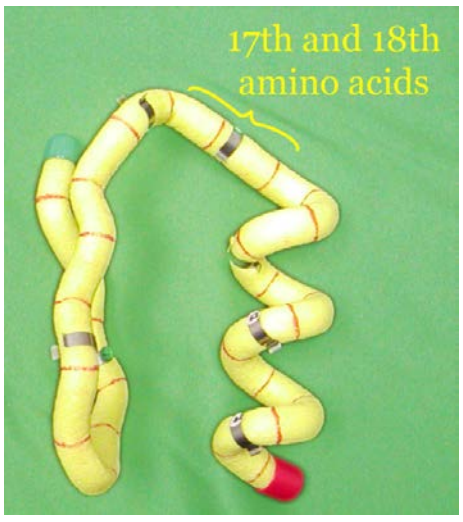
Is your helix right-handed or left-handed?

Alpha-helices are right-handed. Make sure that your model has a right-handed helix. To do this, imagine that your alpha helix is a spiral staircase. If you can climb that staircase with your right hand on the outside railing (the toober), then you have a right-handed helix.



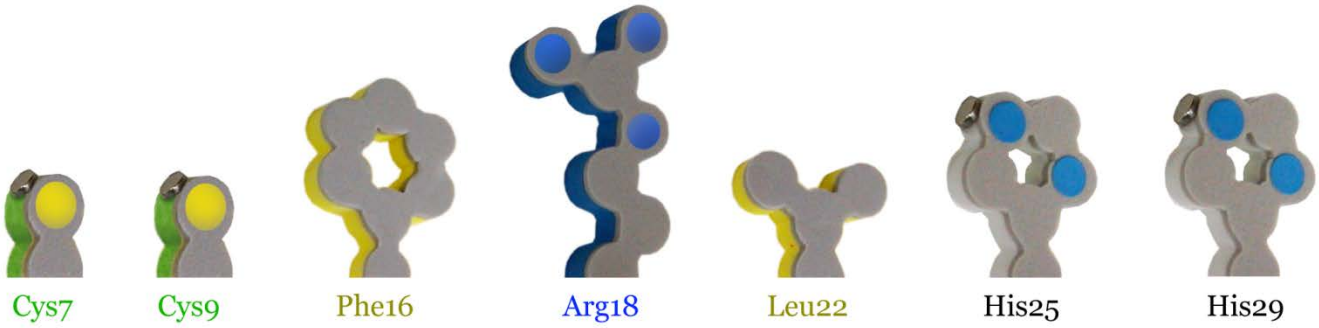


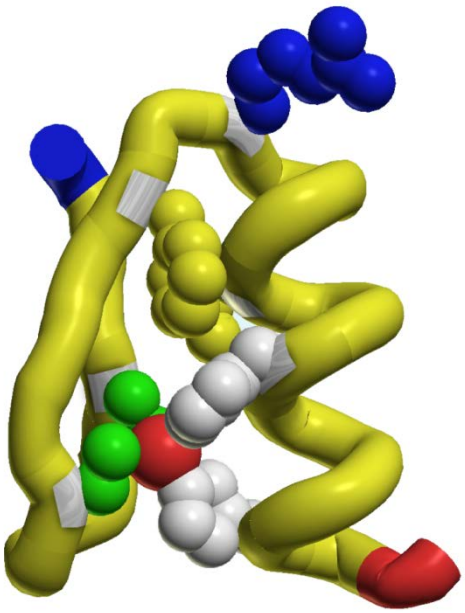
3. Now, bend the beta pleated strand in half at the 7th amino acid (Glu10).



4. Next bend the toober segment at the "turn" (between the 17th and 18th amino acids) so that it resembles the picture shown to the right.

5. Finally, using the zinc finger folding map included within this activity, decipher which amino acids have sidechains represented and attach the sidechains to the clips at their correct amino acid locations. There are two cysteines (positions 7 and 9), two histidines (positions 25 and 29), one phenylalanine (position 16), one arginine (position 18) and one leucine (position 22). Then position the zinc atom in the middle so that it is coordinated by the two histidines and two cysteines.





Milwaukee School of Engineering
Center for BioMolecular Modeling
http://cbm.msoe.edu

phone: 414-277-2806 fax: 414-277-7486
1025 North Broadway
Milwaukee, WI 53202

3-D Molecular Designs

...where molecules become real ™

Enzymes in Action Kit©

Creating Catalytic Connections with Models

Teacher Key

Objectives

You will use the model pieces in the kit to:

- **Simulate** enzymatic actions.
- **Explain** enzymatic specificity.
- **Investigate** two types of enzyme inhibitors used in regulating enzymatic activity.
- **Examine** how an enzyme may affect activation energy.

Introduction

Enzymes are specialized proteins that **catalyze** or speed up chemical reactions within cells. The substance upon which an enzyme acts is called a **substrate**. Substrates are small molecules.

Enzymes:

- Accomplish catalysis without being consumed in the reaction.
- Catalyzes a specific chemical reaction.

The Enzyme in Action Kit© allows you to explore how enzymatic reactions occur.

Catabolism

Model pieces needed

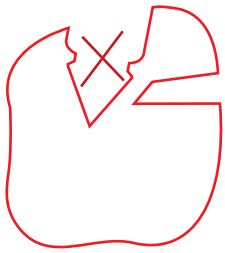
gray A foam piece without stickers

green B_1 and B_2 foam pieces

orange C_1 and C_2 foam pieces

1. The gray foam piece is a model of an **enzyme**. Place it with the **A** label facing up. Assemble the two green pieces (B_1 and B_2) into a single unit to model the **substrate** in this reaction.

2. Draw and label the **enzyme** and **substrate** before the enzymatic action.



enzyme

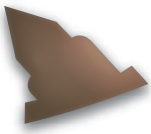
Substrate

1050 North Market Street • Suite CC130A • Milwaukee, WI 53202
Phone: (414) 774-6562 Fax: (414) 774-3435
3dmoleculardesigns.com

Enzyme Action Continues






3. In this first metabolic action, the enzyme will act on the substrate to break it apart. Experiment with the pieces to model how the enzyme and substrate might interact.

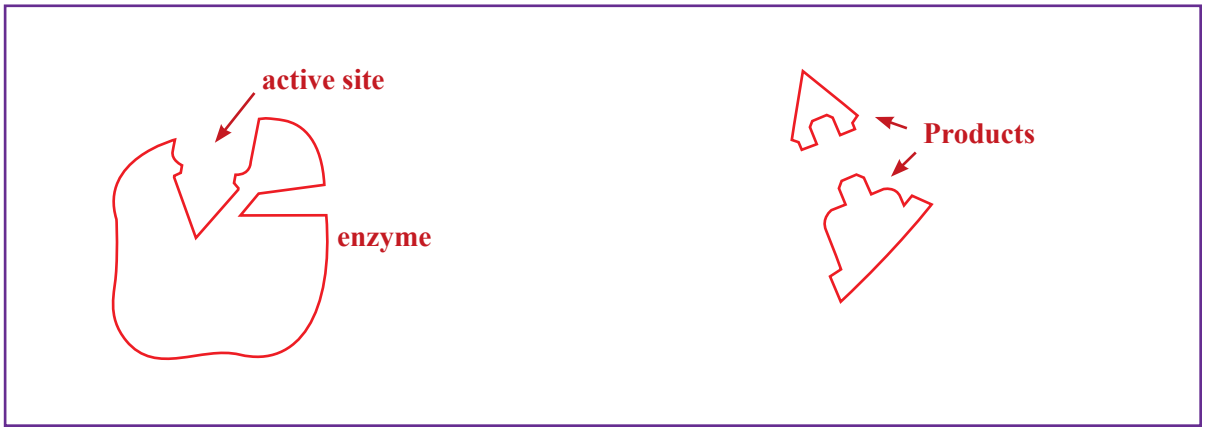
4. The substance an enzyme acts upon is referred to as the **substrate**.

5. Place an "**X**" on the drawing of the enzyme and substrate you created on page 1 to show where the substrate binds to the enzyme.

The part of the enzyme that binds the substrate to be acted on is referred to as the **active site**.

Once the substrate is locked into the enzyme, the two green substrate pieces may be easily pulled apart. This type of metabolic process is called **catabolism** (the breaking down of complex molecules into simpler molecules).

6. Draw and label the **enzyme**, **products** and **active site** after enzymatic action.



7. Although the substrate model changed in this reaction, what changes did you observe in the model of the enzyme during this reaction?

When the substrate was locked into place on the enzyme, a slight shape change took place in the enzyme to allow for the substrate to dock.

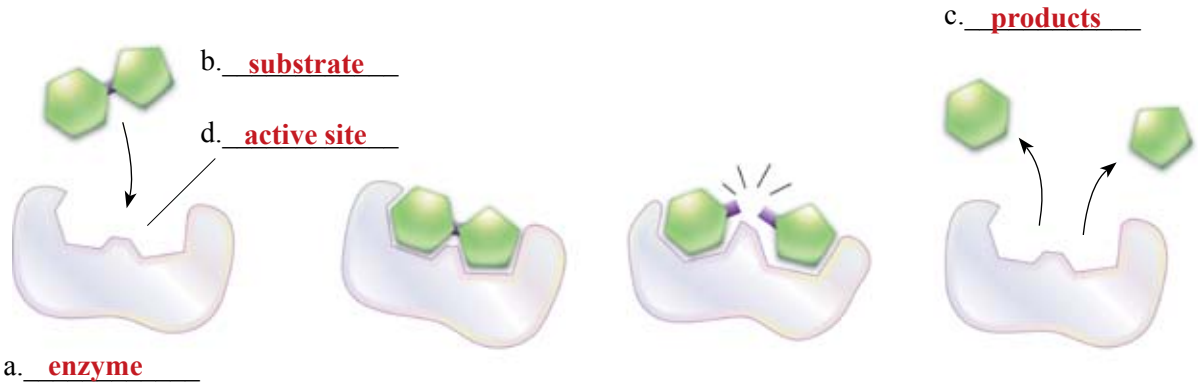
8. Why do you think it is an advantage for the enzyme to remain unchanged while catalyzing a chemical reaction?

Answers will vary but may include: The enzyme may be used over and over again reducing the amount of resources the cell would have to use to make more enzyme. The enzyme is not a direct reactant in the cellular chemical reaction and doesn't change in order to avoid interference with cellular reactions.

Enzyme Action Continues

Note: A real life example of catabolism occurs when the enzyme **sucrase** breaks down the substrate sucrose into glucose and fructose (monosaccharides).

9. Given what you now know about catabolism, identify the following components in the model illustrated below: enzyme, substrate, products and active site.



Induced Fit Model of Enzyme Action

In 1958 scientist Daniel Koshland, Jr., PhD., proposed the **induced fit model** to describe enzyme-substrate interaction. This model suggests that enzymes are flexible structures in which the binding of the substrate results in small changes to the shape of the active site, maximizing its interaction with the substrate.

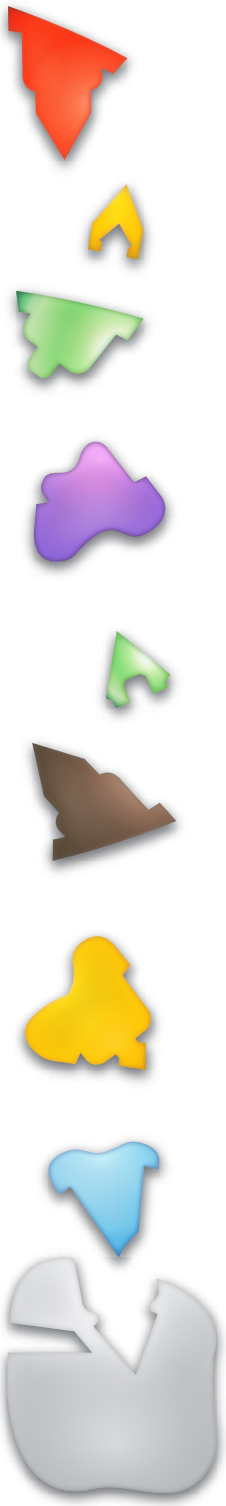
10. Describe how the foam catabolism model illustrates the induced fit model of enzyme-substrate interaction.

The enzyme shows flexibility as it reshapes slightly to allow for the substrate to dock into the active site.

11. Explain the difference between **catalysis** and **catabolism**.

Catabolism is defined as a metabolic process that breaks down complex molecules into simpler ones. Catalysis is the increase in the rate of a chemical reaction due to participation of a substance that can modify the rate of the reaction without being consumed in the process.

Enzyme Action Continues

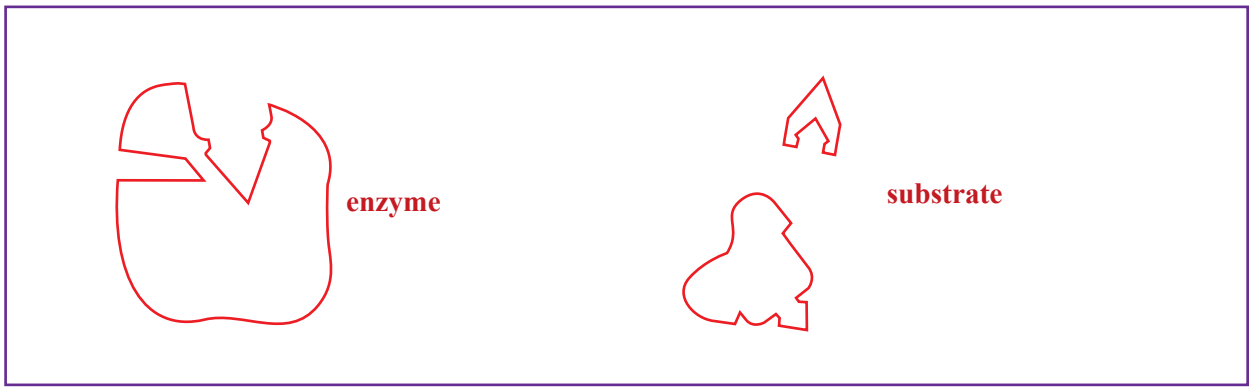


Anabolism

Enzymes may also bring substrates together to form a final product. This metabolic process is called **anabolism** (the building of complex molecules from simpler molecules).

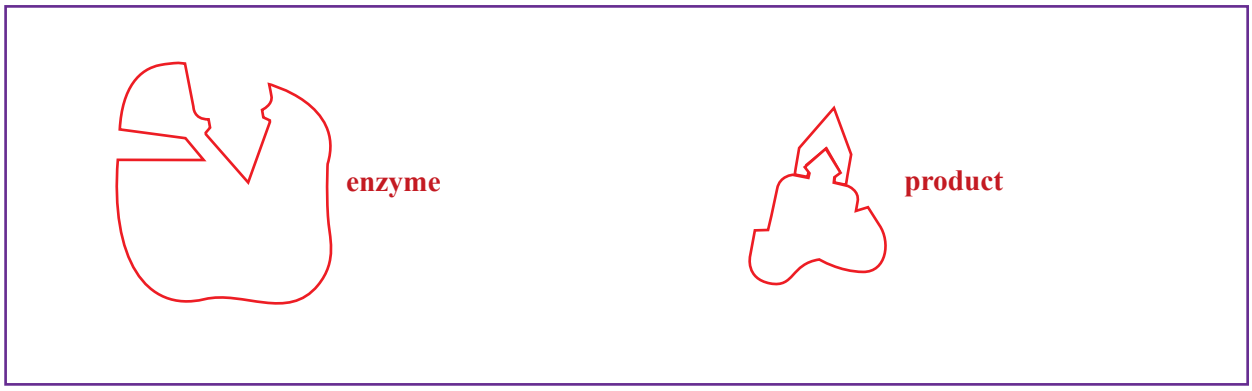
12. Use the gray foam piece and the orange foam pieces (C_1 and C_2) to simulate an anabolic process. The orange pieces should not be assembled prior to the anabolism action.

13. Sketch and label the **enzyme** and **substrate** prior to enzyme action in the space below.



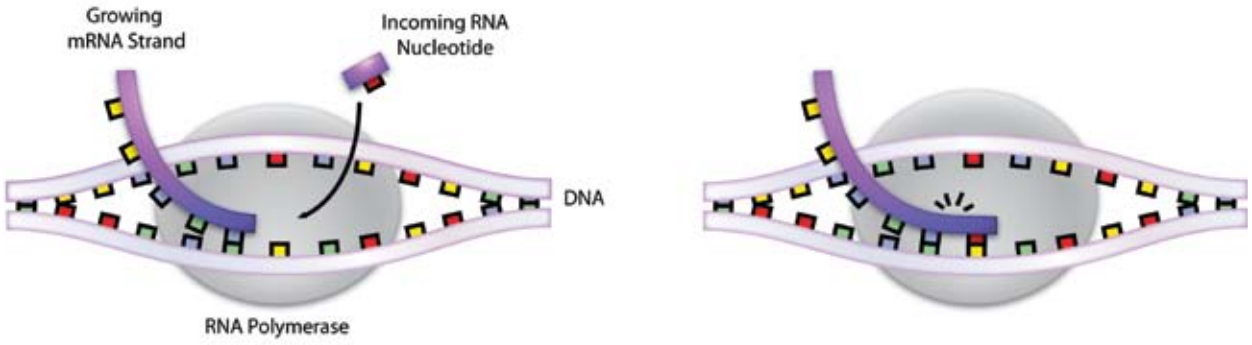
14. Place the small pointed orange piece (C_2) into the enzyme. Join the larger orange piece (C_1) to C_2. Note that the two pieces lock together to form a final product.

15. In the space below, sketch and label the **enzyme** and **products** after the enzyme has acted on the substrate.



Enzyme Action Continues

Note: A real life example of anabolism occurs when **RNA polymerase** links RNA nucleotides together by catalyzing the formation of a bond between the backbone sugar of one nucleotide to the backbone phosphate of another nucleotide during transcription.



16. Given what you now know about anabolism, identify the substrate in the above diagram.

The incoming RNA nucleotides.

17. Explain why the above process is an example of anabolism.

Anabolism is the process of bringing substrates together to form a product. The RNA polymerase brings the RNA nucleotides together to form an RNA molecule.

Lock and Key Model of Enzyme Action

In 1894 scientist Emil Fisher wrote, "To use a picture, I would like to say that enzyme and glucoside have to fit to each other like a lock and key in order to exert a chemical effect on each other." Fisher created a mental model of how an enzyme acts and referred to it as the ***Lock and Key Model of Enzyme Action***.

This model suggests that the enzyme and the substrate possess specific complementary geometric shapes that fit exactly into one another like a key into a lock.

18. Describe how the anabolic process you previously modeled illustrates the lock and key model of enzyme-substrate interaction.

The lock and key model suggests that the enzyme does not change shape to accommodate the substrate. In the anabolic activity, the enzyme model does not change shape when the substrate is bound.

Enzyme Action Continues

Note: Most enzymes catalyze either **catabolic** OR **anabolic** processes. There are a few enzymes that do both. ATP synthase and ATPase are the same protein but have different names because they function as enzymes in both catabolic and anabolic reactions.

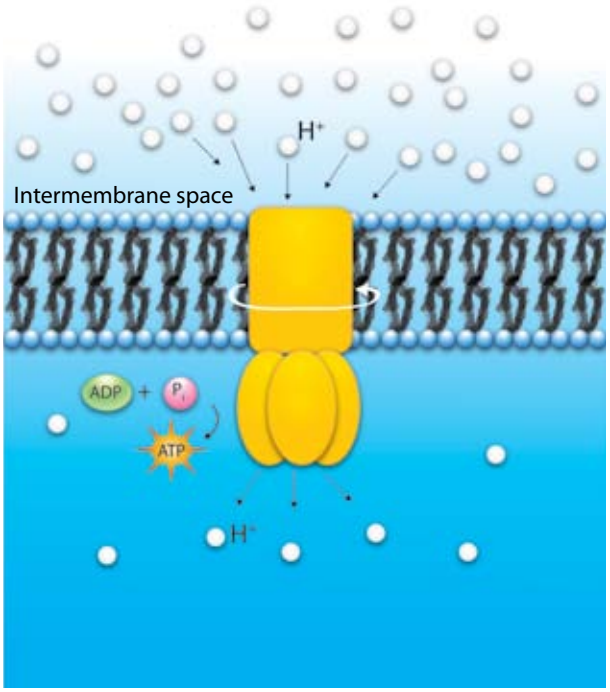


Diagram A — ATP Synthase

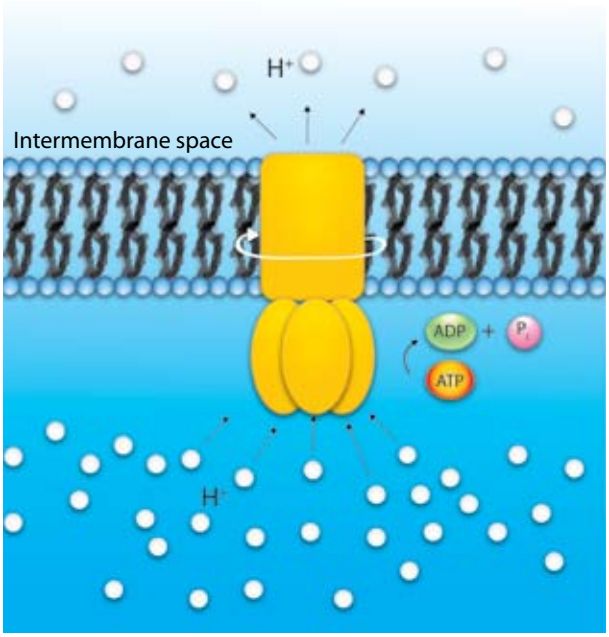


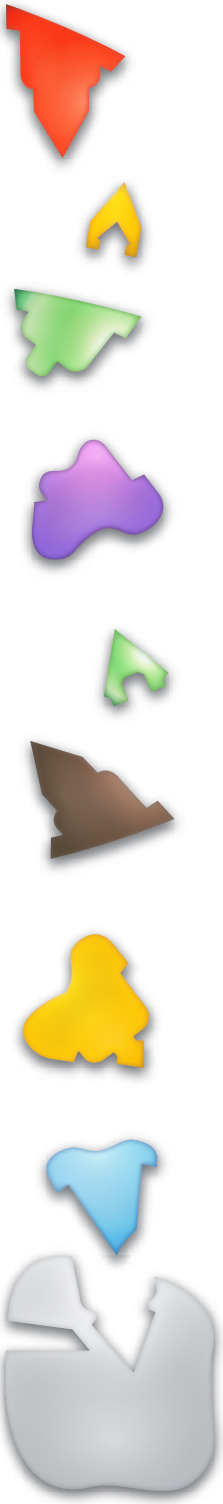
Diagram B — ATPase

19. Describe the action of the enzyme in diagram A. In your description, identify the substrate and enzyme.

Diagram A illustrates an anabolic process. The enzyme ATP synthase is putting together the substrates ADP and P_i to form ATP. Diagram B illustrates a catabolic process where the enzyme ATP synthase is helping to break apart the substrate ATP into the products ADP and P_i.

1050 North Market Street • Suite CC130A • Milwaukee, WI 53202
Phone: (414) 774-6562 Fax: (414) 774-3435
3dmoleculardesigns.com

Enzyme Specificity



The reaction catalyzed by an enzyme is very specific. Most enzymes are proteins with unique three-dimensional configurations based on their amino acid sequence. The specificity of an enzyme can be attributed to the compatibility between the shape of the enzyme's active site and the shape of the substrate.

Model pieces needed

gray foam piece with stickers

red D foam piece with stickers

tan E foam piece with stickers

1. Place the enzyme model with the sticker side facing up. Write your observation about the active site of the enzyme below.

Answers will vary but may include: The active site has a negative charge on the left and a positive charge on the right. There is a yellow region deep inside the active site.

2. What might these specialized areas in the enzyme represent?

The charged areas represent charged amino acids found in the active site. The yellow region represents the hydrophobic amino acids in the active site.

3. What do the red D and tan E foam pieces represent?

The red and tan pieces represent possible substrates for the enzyme to act upon.

4. How do the specialized areas of the red D piece interact with the specialized areas of the enzyme?

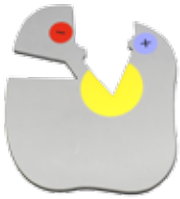
The positive charge of the enzyme matches with the negative charge of the red piece. The negative charge of the enzyme matches with the positive charge of the red piece. The yellow regions match together as well.

5. In order for enzymes to bind to the correct substrate, enzymes have specific active site configurations that allow for interaction with the substrate. Explain why the tan E substrate would not interact with the enzyme.

The tan substrate would not interact with the enzyme because the charged amino acids will tend to repel and the hydrophobic areas will not interact.

Enzyme Inhibition

Model pieces needed


gray foam piece with stickers


red D foam piece with stickers


purple (F) foam piece.


blue (G) foam piece.

Competitive Inhibition

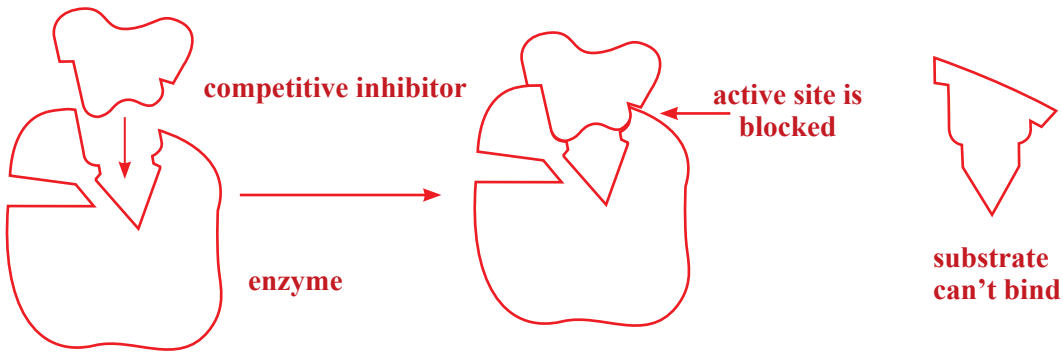
1. Place the gray A, red D, purple F and blue G foam pieces on your work surface. Which two pieces may fit into the active site?

the red and **the purple pieces**

2. Can the red D substrate bind to the active site if the purple F piece is bound to the enzyme? **No.**

A substance which binds in the active site and prohibits normal substrate interaction is called a **competitive inhibitor**.

3. Create a sketch using the foam models to illustrate **competitive inhibition**.



4. Predict what might happen in a cell if the concentration of competitive inhibitor exceeded that of the substrate.

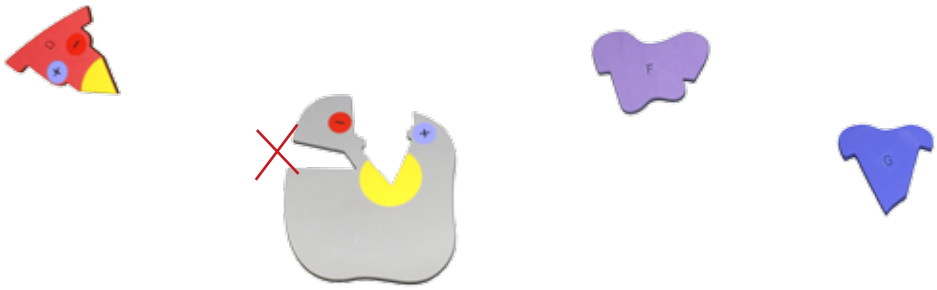
If the concentration of competitive inhibitor exceeded that of the substrate, the enzyme would not be able to bind the substrate and the reaction would slow down.

Enzyme Inhibition

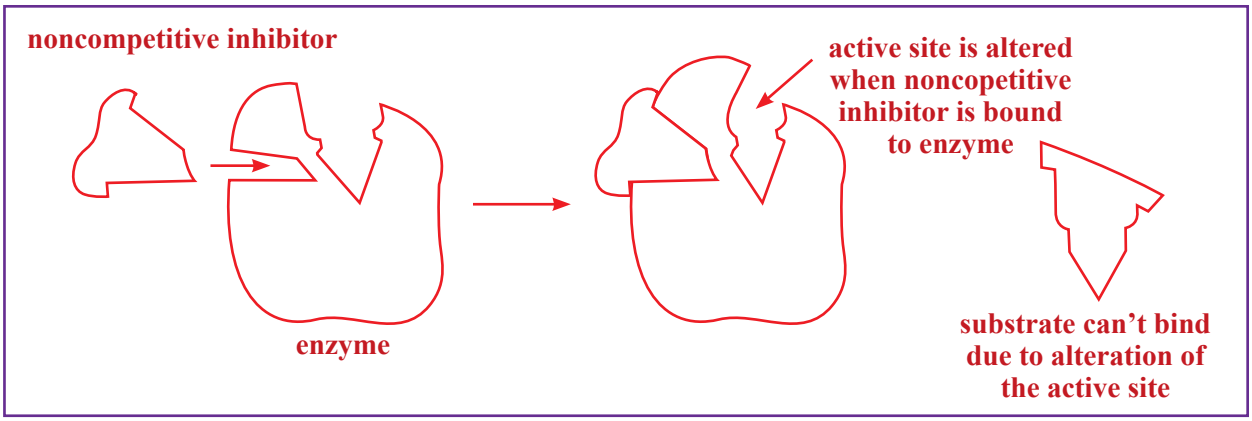
Noncompetitive Inhibition

A **noncompetitive inhibitor** impedes enzymatic action by binding to another part of the enzyme. This second site, known as the ***allosteric site***, is the place on an enzyme where a molecule that is not a substrate may bind, thus changing the shape of the enzyme and influencing its ability to be active.

5. In the diagram below, draw an "**X**" where the **blue G** piece may noncompetitively bind to the enzyme.



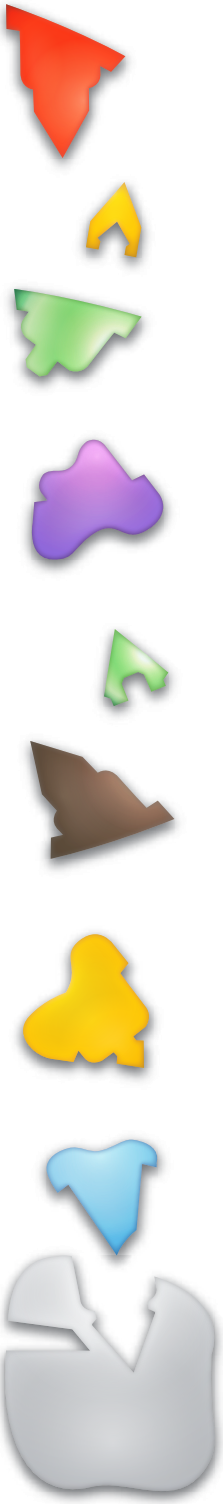
6. Sketch what happens to the shape of the enzyme when the blue piece is bound to the allosteric site.



7. How does this affect the binding of the substrate?

When the noncompetitive inhibitor is bound to the enzyme, the active site changes shape so that the substrate is unable to bind to the enzyme.

Activation Energy



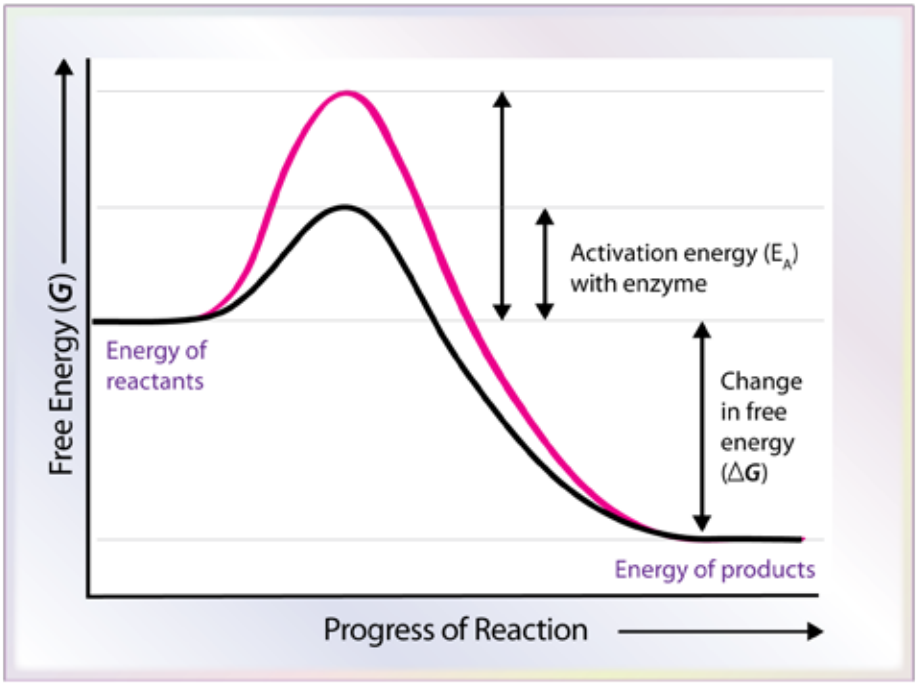
Activation energy may be defined as the minimum amount of energy required to get the reactants in a chemical reaction to the transition state, in which bonds are broken and new bonds are formed. The activation energy of a reaction is usually denoted by E_A. By now you know that enzymes are proteins that catalyze chemical reactions. Enzymes lower the activation energy needed to start a reaction.

You may use the foam pieces to simulate the activation energy needed in a reaction with and without an enzyme. Begin by connecting the green foam pieces. To illustrate the activation energy without the enzyme interaction, pull the apart the two green pieces with your hands.

Reconnect the green pieces. This time lock them into the active site on the gray A enzyme (without stickers) foam piece. With the help of the enzyme it takes less energy to pull the pieces apart. The activation energy has been lowered!

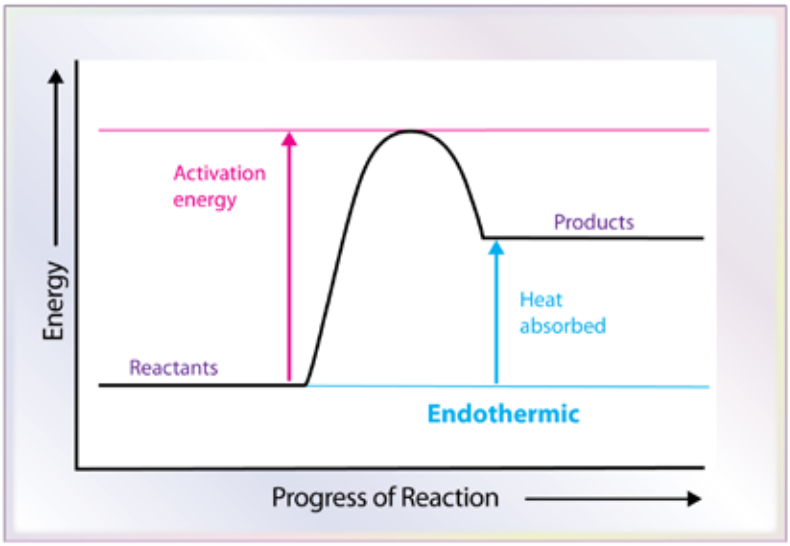
Notice in the graph below that the resulting products have less free energy than the reactants. In such a reaction, energy has been released and the reaction is said to be **exothermic**. A specific example of an exothermic reaction is the breakdown of hydrogen peroxide into hydrogen and oxygen. The enzyme used to facilitate this reaction is known as **catalase**.

$$2H_2O_2 \rightarrow 2H_2O + O_2$$



Activation Energy Continues

Conversely, in the graph below, the products have more free energy than the reactants. Reactions that absorb heat from the environment are known as **endothermic** reactions. A common example is a chemical ice pack which typically contains water and a packet of ammonium chloride. To activate the ice pack, the barrier separating the two substances must be physically broken so the two substances may react. Enzymes may also facilitate endothermic reactions.



1. Examine the graph below. Is the reaction depicted exothermic or endothermic? Explain your answer.

The reaction in the graph below is endothermic because the products have more energy than the reactants, indicating that energy has been taken in by the reaction.

The activation energy curve show below represents a non-enzyme catalyzed reaction.

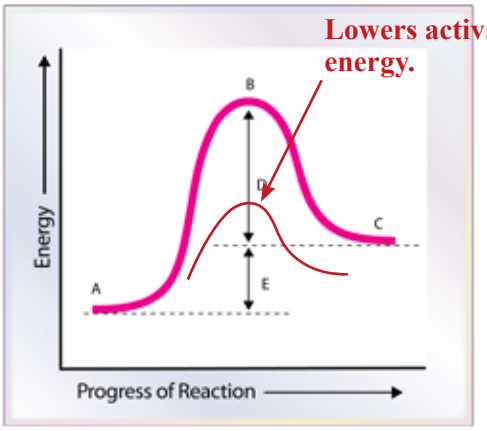
2. Draw a line on the graph indicating the activation energy in the presences of an enzyme.

3. Which letter depicts the activation energy without the enzyme present?

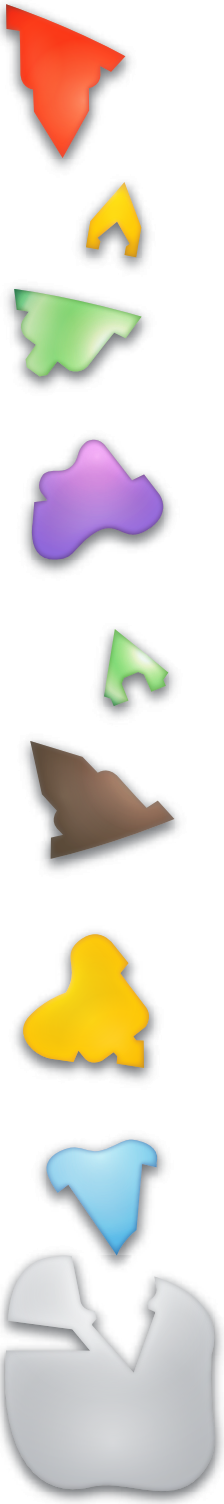
B

4. What does the letter 'E' represent?

Heat absorbed / Endothermic



Activation Energy Continues



Post Lab Questions

1. Predict what might happen if enzyme activity were not regulated within a cell's metabolic pathways.

Answers will vary but may include: If enzyme activity were not regulated within a cell's metabolic pathways, too much heat may be released in a reaction and the cell may sustain damage. Additionally, if the enzyme is not working to produce a product, those products may not be available for the next step in a cascade of reactions.

2. Contrast the action of a competitive inhibitor with that of a noncompetitive inhibitor.

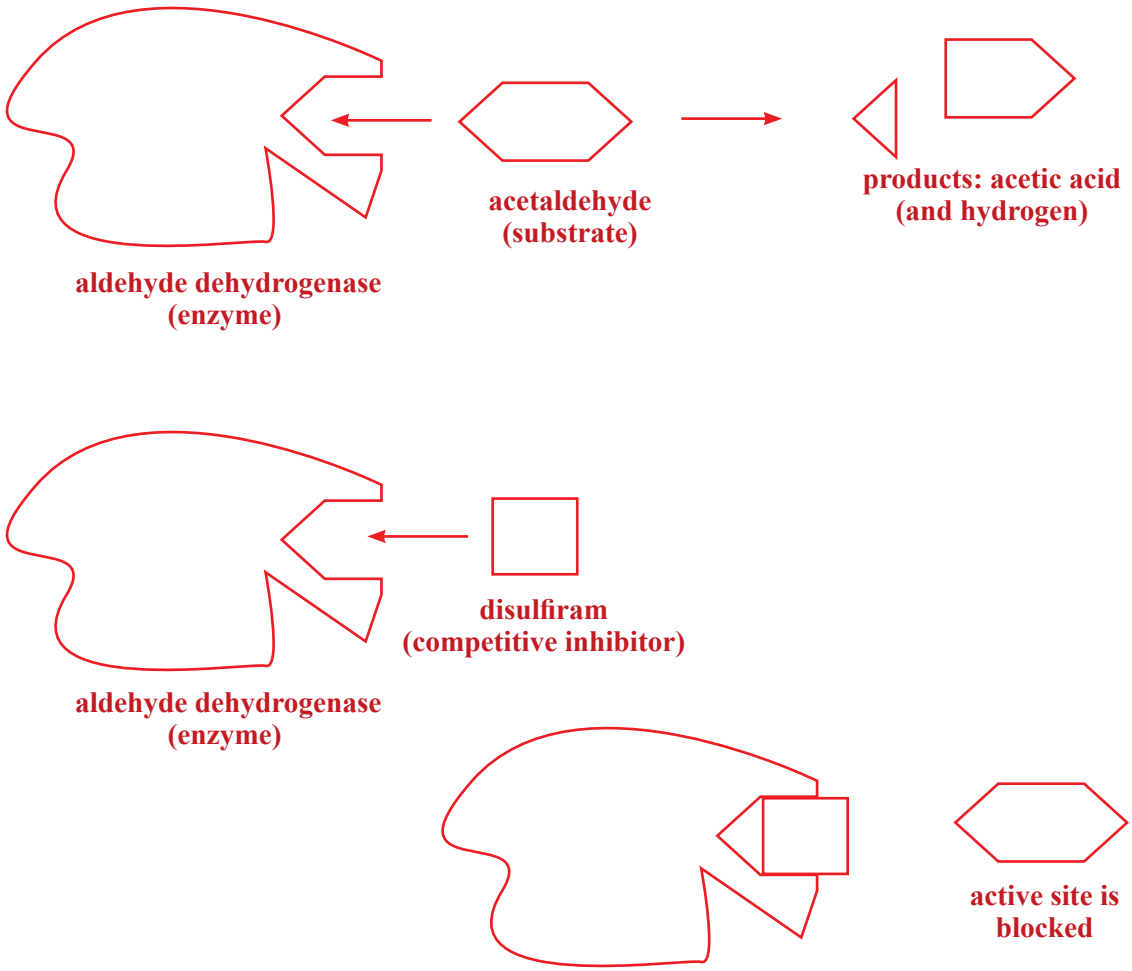
A competitive inhibitor binds in the active site preventing the substrate from interacting with the enzyme, while a noncompetitive inhibitor binds at another site on the enzyme which changes the active site preventing the substrate from interacting with the enzyme.

Activation Energy Continues

Ethanol is metabolized in the body into acetaldehyde. Normally, acetaldehyde does not accumulate in the body because aldehyde dehydrogenase rapidly oxidizes the acetaldehyde into acetic acid. The drug disulfiram inhibits aldehyde dehydrogenase, which causes an accumulation of acetaldehyde in the body with the subsequent unpleasant side effects of nausea and vomiting. Disulfiram is sometimes used to treat patients with a drinking habit.

3. Create a sketch to illustrate the action of the competitive inhibitor disulfiram. Be sure to label your diagram with the following terms, **aldehyde dehydrogenase**, **disulfiram**, **acetaldehyde**, **enzyme**, **competitive inhibitor**, **substrate** and **product**.

Example only: Student's answers will vary



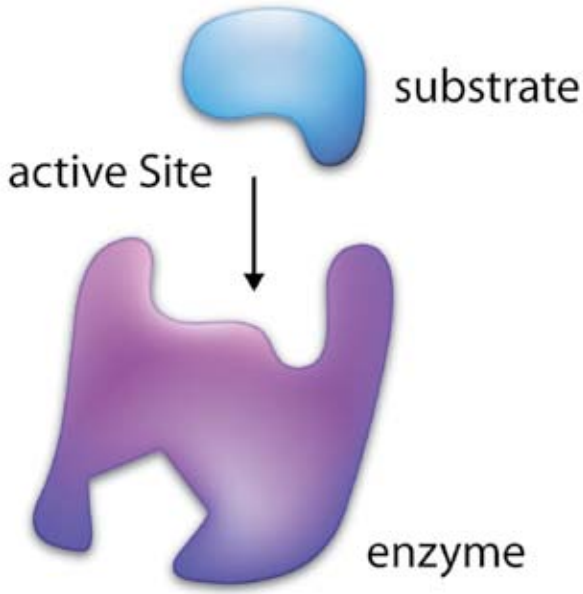
Activation Energy Continues

4. How is Koshland's theory of induced fit supported by noncompetitive inhibition of enzymes?

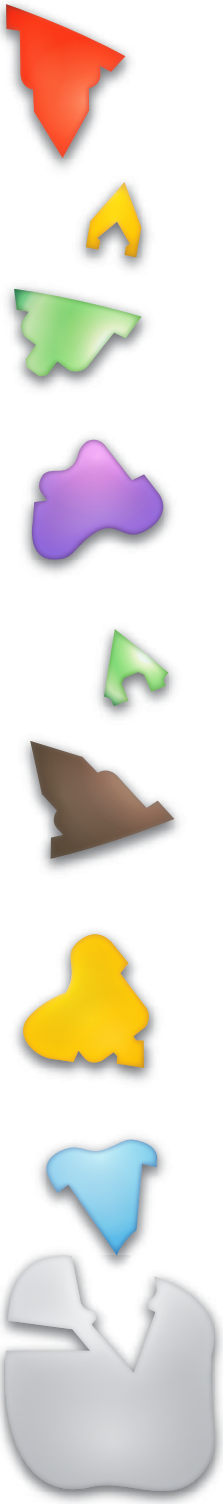
Koshland's theory of induced fit implies that the enzyme must change shape in order to interact with the substrate. Noncompetitive inhibition occurs when a substance binds to the enzyme at a site other than the active site while still causing a change in the active site.

5. Examine the model of the enzyme shown below. Design a competitive and noncompetitive inhibitor for this enzyme.

Answers will vary, but the competitive inhibitor should block the active site while the noncompetitive inhibitor blocks the allosteric site.



National Standards



Connections to: A Framework for K-12 Science Education

Practices, Crosscutting Concepts, and Core Ideas*

Dimension 1. Scientific and Engineering Practices

1. Asking Questions (for science) and Defining Problems (for engineering)
2. Developing and Using Models
6. Constructing Explanations (for science) and Designing Solutions (for engineering)

Dimension 2. Crosscutting Concepts

1. Patterns
2. Cause and Effect: Mechanism and Explanation
4. Systems and System Models
6. Structure and Function
7. Stability and Change

Dimension 3. Disciplinary Core Ideas

Physical Science

HS-PS1: Matter and its Interactions

HS-PS1-2: Construct and revise an explanation for the outcome of a simple chemical reaction based on the outermost electron states of atoms, trends in the periodic table, and knowledge of the patterns of chemical properties.

HS-PS1-4: Develop a model to illustrate that the release of absorption of energy from a chemical reaction system depends upon the changes in total bond energy.

HS-PS1-5: Apply scientific principles and evidence to provide an explanation about the effects of changing the temperature or concentration of the reacting particles on the rate at which a reaction occurs.

Life Science

LS 1: From Molecules to Organisms: Structures and Processes

HS-LS1-1: Construct an explanation based on evidence for how the structure of DNA determines the structure of proteins which carry out the essential functions of life through systems of specialized cells.

HS-LS1-2: Develop and use a model to illustrate the hierarchical organization of interacting systems that provide specific functions within multicellular organisms.

Engineering, Technology and Applications of Science

HS-ETS1: Engineering Design

HS-ETS1-4: Use a computer simulation to model the impact of proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem.

*The NSTA Reader's Guide to A Framework for K-12 Science Education, National Research Council (NRC), 2011. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, D.C.: National Academies Press.

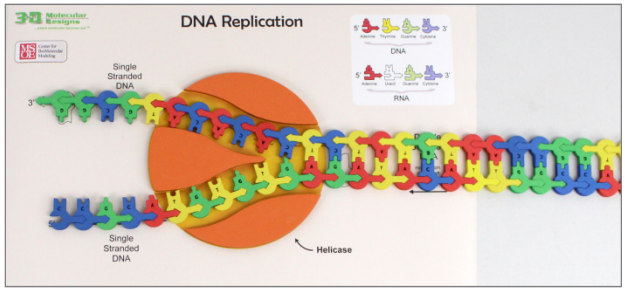
DNA Replication

The process by which a DNA molecule is copied

Template

Non-template

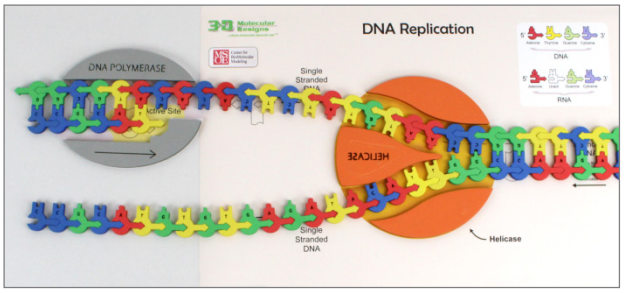
Diagram 1



Helicases unwind the DNA at the replication fork

Helicases are enzymes that untwist the double helix and separate the two strands.

Diagram 2



DNA polymerase catalyzes the synthesis of new DNA

Leading Strand: New DNA can elongate only in the 5′ → 3′ direction. The DNA strand that is made **continuously** is referred to as the leading strand. DNA polymerase moves toward the replication fork.

Diagram 3

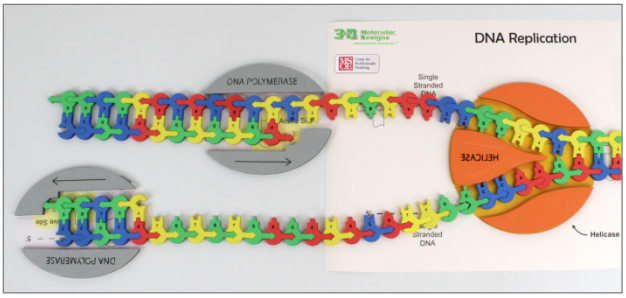
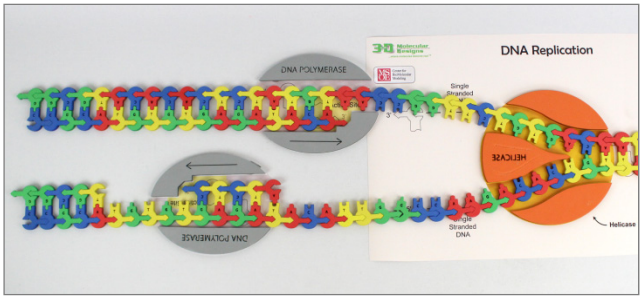


Diagram 4



Lagging Strand: DNA polymerase must move away from the fork instead of toward the fork as it did in the leading strand in order to maintain 5′ → 3′ DNA synthesis (diagram 3). The lagging strand is synthesized in a series of **discontinuous** fragments referred to as Okazaki fragments (diagram 4).

Note: Each of the two daughter molecules has one old strand from the parental molecule and one newly synthesized strand. This type of replication is referred to as **semiconservative replication**.