

Proteins, Amino Acids, and The Bioinformatics of Insulin (Duration: 2 class periods)

Standards

National Content Standard C: Life Science

- Most cell functions involve chemical reactions;
- Smaller precursor molecules (subunits) can be used to assemble larger molecules with biological activity including proteins;
- Students should understand the chemical basis of life.

California State Standards for Biology

- 4.c. *Students know* how mutations in the DNA sequence of a gene may or may not affect the expression of the gene or the sequence of amino acids in an encoded protein.
- 4.e. *Students know* proteins can differ from one another in the number and sequence of amino acids.
- 4.f. *Students know* why proteins having different amino acid sequences typically have different shapes and chemical properties.

Review and Revisit

- 1.d. *Students know* the central dogma of molecular biology outlines the flow of information from transcription of ribonucleic acid (RNA) in the nucleus to translation of proteins on ribosomes in the cytoplasm.

Objectives

Students will:

- Understand that the sequence of amino acids of a protein represent a biological information unit;
- Understand that this information creates the unique structure of a protein;
- Understand that the unique structure of the protein determines its activity (chemical properties) in the organism;
- Understand that organisms have proteins in common;
- Understand that those shared proteins are slightly modified over time through mutations and the differences are referred to as variations or protein polymorphisms (poly = many, Morph = structure);
- Understand that changes (mutations) in the DNA can result in changes in the protein and those changes can produce disease in the organism.

Prerequisite Knowledge: Transcription and Translation

Background Information for the Teacher

Students will quickly review Transcription and Translation in this lesson and then be introduced to protein structure and function, and the relationship between proteins and disease. They will learn how to use computer-based Bioinformatics tools to analyze protein sequences in different organisms.

The 20 different amino acids are the subunits that make up proteins. Each amino acid is covalently bonded to the next in the linear sequence with a peptide bond. The sequence of amino acids coded by Messenger RNA and formed during Translation is the primary structure of a protein.

The primary protein structure, the amino acid sequence, is a biological information unit that determines the structure and function of a protein. The properties of each amino acid in the sequence influence protein folding into the secondary structure - alpha helix and beta sheet – and tertiary structure folding. Some proteins are functional as a tertiary structure. Other proteins require a quaternary structure: two or more tertiary structures (subunits) bonded together. The quaternary structure of some proteins is composed of two or more of the same subunit. Each functional protein with its unique amino acid sequence has a unique final tertiary or quaternary structure and a unique function dependent on that structure.

Bioinformatics is: “Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.” “Bioinformatics applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and useful,” (National Institute of Health, 2000).

Biological organisms have many proteins in common. In this lesson, students will begin to understand that those shared proteins are slightly modified over time through mutations so that each organism has a slightly different version of the same protein with the same function.

Students will look closely at the amino acid sequence of Insulin – the information unit. They will use bioinformatics tools to begin to understand that changes in the protein can produce disease in the organism.

Whole Group Discussion

Focusing Advanced Organizer: What makes them different?

http://www.hhmi.org/biointeractive/obesity/maramouse_clip.mov

Questions:

What gives you energy?

Do you know of any diseases that cause people to have low energy?

What is diabetes? **Diagram with text:** Normal Regulation of Blood Glucose:
<http://www.endocrineweb.com/insulin.html>

What kinds of diseases do you know about that you do not catch from someone or something else?

What causes those diseases?

Which of those diseases are inherited?

What is it that you inherit?

Your DNA is the code that you inherited but what is it that the genes on your DNA code for?

DNA codes for RNA.

RNA codes for Proteins.

Transcription/Translation Review: Protein Synthesis

http://www.accessexcellence.org/RC/VL/GG/protein_synthesis.html

http://biop.ox.ac.uk/www/mol_of_life/ProtSynth.html

<http://learn.genetics.utah.edu/units/basics/transcribe/>

What would you have to know to find a cure for inherited diseases?

Protein Web-Based Introductory Activities

Many diseases involve proteins. Proteins are large biological chemicals (macromolecules) with complex shapes that are built up from small molecules called **amino acids**. When a chemical, such as a protein, is composed of smaller repeating units it is called a **polymer**. The small repeating units for protein polymers are amino acids.

Polymer Activity: <http://folding.stanford.edu/education/paperclip.html>

Amino Acids: http://www.biology.arizona.edu/biochemistry/problem_sets/aa/aa.html

Amino Acid Quiz:

http://www.biology.arizona.edu/biochemistry/problem_sets/aa/Problems/Quiz.html

Each protein has a unique linear sequence of amino acids, also called a **polypeptide** that is coded for by a certain RNA strand during Translation. Each amino acid is held to the next in the sequence by a **peptide bond**. That amino acid sequence contains information that guides the protein to fold up into a unique shape. Many complete protein structures require more than one polypeptide. The unique complete structure of each protein is required for the specific task that the protein must perform in a living organism.

Protein Structure:

http://www.biology.arizona.edu/biochemistry/problem_sets/large_molecules/05t.html

Primary Structure:

<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/PrimaryStructure.html>

Secondary Structure:

<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/S/SecondaryStructure.html>

Tertiary Structure:

<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/TertiaryStructure.html>

Quaternary Structure:

<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/Q/QuaternaryStructure.html>

A **gene** is a length of DNA that codes for a specific RNA strand that codes for one polypeptide. Changes, **mutations**, in a gene may result in changes to the polypeptide that the gene codes for. Those changes may result in a protein that does not fold correctly and so will not be able to do its job exactly right.

What happens when a protein does not fold correctly?

<http://folding.stanford.edu/education/diseases.html>

Mutations: Definition: <http://www.mrothery.co.uk/module2/keyfactsmutation.htm>

Point Mutations: http://www.pnl.gov/BERC/glossary/gd_PointMutation.html

Activity: Bioinformatics of Insulin **Developed and Presented by D. Jeff Milton** **at the San Diego Super Computer Center for TeacherTECH**

A hormone is a protein that acts as a messenger bringing important information to the cells throughout your body. Insulin is a type of hormone that is created in the pancreas and released in the blood after meals. It spreads throughout the body and tells liver, muscles and fat cells about the state of the sugar content in the blood. This then tells the body to take sugar, or glucose, out of the blood and store it as glycogen (the stored form of glucose) in the muscles and liver. By understanding the 3-D structure of this protein researchers can get an idea of how it works and find proteins with related functions.

Open a web browser and go the Protein Data Bank web site: <http://www.pdb.org>

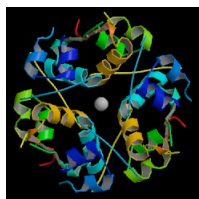
In the PDB keyword search window:



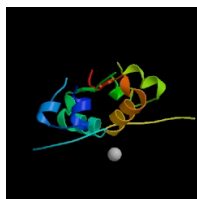
In the text window enter the PDB ID "1MSO" and click the search button.
(1 = the number one and the letters are not case specific.)

The browser will load what is called the "Structure Summary" page. This provides a general description of the human insulin structure and the experiment that was used to determine the atomic coordinates.

You can look at the structure of the protein by looking at the image on the right side of the structure summary page. There are two images: the asymmetric unit and the biological molecule. The asymmetric unit is what the experiment was able to resolve, and the biological molecule is what the experimenter assumes is the actual structure in the living organism.



Biological Molecule



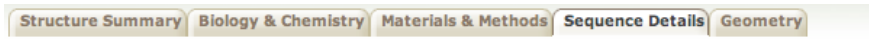
Asymmetric Unit

1. Use protein workshop to visualize this protein (the asymmetric unit) and highlight each chain visible with a different color.

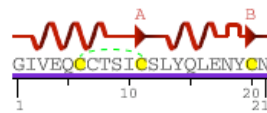
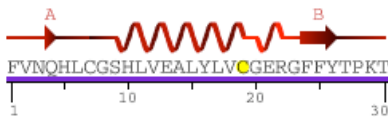
2. What secondary structures are present?

How many beta sheets are present in the asymmetric unit? _____

Next click on the Sequence Details option in the tab panel:



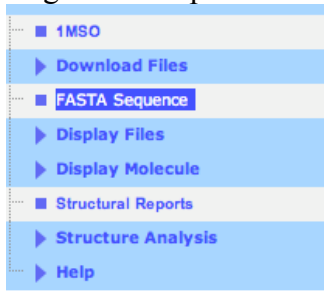
Scroll down until you see the sequence of amino acids in the protein. The two chains look something like the following:



3. How many amino acids make up each unique chain of the structure?

How many total amino acids make up the entire protein? (Hint: how many of each unique chain is in the biological molecule).

On the left side of the browser page click on the link labeled “FASTA Sequence”. FASTA is a type of file format that biologists use to look at the sequence of amino acids. It always has the same format: a comment line preceded by the ‘>’ character followed by single letter representations of amino acids



Copy the first four lines of the amino acid sequence:

```
>1MSO:A|PDBID|CHAIN|SEQUENCE  
GIVEQCCTSI CSLYQLENYCN  
>1MSO:B|PDBID|CHAIN|SEQUENCE  
FVNQHLCGSHLVEALYLVCGERGFFYTPKT  
>1MSO:C|PDBID|CHAIN|SEQUENCE  
GIVEQCCTSI CSLYQLENYCN  
>1MSO:D|PDBID|CHAIN|SEQUENCE  
FVNQHLCGSHLVEALYLVCGERGFFYTPKT
```

Paste them into a text editor such as Microsoft word and save these for later.

In the top search bar enter the PDB ID “2a3g” and click the “Search” button.

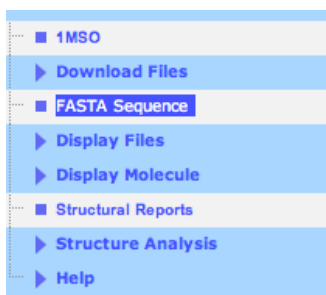


This is a crystal structure of an insulin molecule extracted from a cow. As you can see this structure is very similar to the human insulin.

4. Compare this 3D structure with that of the human insulin (1MSO) using Protein Workshop.

Are there any noticeable differences?

Take a look at the sequence of this structure by clicking on the FASTA file link on the left side of the browser window.



Copy the first four lines of this sequence file, as you did with the human insulin sequence. Paste these lines into your text editor below the previous sequences. Your file should look like the following:

```
>1MSO:C|PDBID|CHAIN|SEQUENCE
GIVEQCCTSICSLYQLENYCN
>1MSO:D|PDBID|CHAIN|SEQUENCE
FVNQHLCGSHLVEALYLVCGERGFFYTPKT

|>2A3G:A|PDBID|CHAIN|SEQUENCE
GIVEQCCASVCSLYQLENYCN
>2A3G:B|PDBID|CHAIN|SEQUENCE
FVNQHLCGSHLVEALYLVCGERGFFYTPKA
```

Take the first chain of each structure and align them on successive lines. Your file should look like the following:

GIVEQCCTS ICSLYQLENYCN
GIVEQCCASVCSLYQLENYCN

FVNQHLCGSHLVEALYLVCGERGFFYTPKT
FVNQHLCGSHLVEALYLVCGERGFFYTPKA

Make sure the human insulin (1MSO) is on the top and the cow insulin (2A3G) is on the bottom of each chain pair.

5. How many amino acids are different in both chains?

Diabetes is a disease where the production or use of insulin in the bloodstream is impaired. As a result the body cannot metabolize sugar and leads to very serious medical conditions. Because of the structural similarity of this protein between species humans can use insulin from cows and pigs as a replacement. However, some diabetics have allergic reactions to these molecules.

6. Use the PDB keyword search to find a pig (porcine) insulin structure. Compare the sequence of this structure to humans (PDB ID 1MSO).

Based on this information and that of the cow insulin (Bovine) which molecule will be more likely to provoke an immune response? Explain your answer.

Homework

Completely answer each of the following questions.

Include **ALL** of the terms listed below.

Be sure to define the terms when you use them.

You may use any of the websites in the lesson or those below to review for answers but put your responses in your own words.

1. What is meant by: The sequence of amino acids of a protein represent a biological information unit?
2. What information in the protein primary structure creates the unique structure of a functional protein?
3. What is meant by: The unique structure of the protein determines its activity in the organism?
4. Pigs, and cows and humans all have the protein insulin.

Why do humans with diabetes often have an immune response to the insulin from the other two?

What causes diabetes?

Vocabulary to define and use when answering the questions above:

Amino Acids: Structure, Side Chain/R-Group, Peptide Bond

Protein Primary Structure

Protein Secondary Structure: Alpha Helix, Beta Sheet

Protein Tertiary Structure

Protein Quaternary Structure

Mutation: variation, protein polymorphisms

Diabetes: Pancreas, hormone, insulin, glucose, glycogen,

Websites for Review: (Refer to the websites in the lesson also)

Translation/Transcription Review: Click on the Blue Box and follow directions

<http://learn.genetics.utah.edu/units/basics/transcribe/>

What Proteins Do: <http://learn.genetics.utah.edu/units/basics/tour/>

Follow the direction and then select: What is a Protein

Amino Acids and Protein Primary Structure Review:

<http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/Polypeptides.html>

Insulin/Diabetes Background:

www.pdb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb14_1.html

Think About It

1. Read the following two article excerpts.
2. Answer the questions below each article.
3. Use the National Institute of Health on-line talking glossary when needed:

<http://www.genome.gov/page.cfm?pageID=10002096>

Science News Online

Week of Sept. 30, 2000; Vol. 158, No. 14

Gene Tied to Heightened Diabetes Risk By Nathan Seppa

People with certain common variations of a newly identified gene called CAPN10 face a sharply increased risk of getting adult-onset, or type II, diabetes, research now suggests. If further studies confirm that these variants contribute to the disease, the finding could have landmark implications for diabetes diagnosis and prevention, scientists say.

The gene, which sits on chromosome 2, encodes an enzyme called calpain-10. Calpains are proteases - proteins that cleave (cut up) other proteins. Some proteases have well-defined roles, but scientists admit they know little about calpains.

The gene for calpain-10 can come with dozens of variations, or polymorphisms. In the October Nature Genetics, scientists report that having a particular form of CAPN10 tripled the diabetes risk in a group of Mexican-Americans. The most troublesome version of the gene contains three specific variations, or single nucleotide polymorphisms (SNPs), within its DNA sequence.

The three variations don't appear to change calpain-10 itself, but they may cause its supply to fluctuate, says study coauthor Nancy J. Cox, a geneticist at the University of Chicago. This, in turn, seems to influence susceptibility to type II diabetes, she says.

Questions:

What does an increased risk of getting the disease mean?

What other factors besides genetics might contribute to getting or avoiding type II diabetes for people who carry these variations?

Science News Online Week of Oct. 23, 2004; Vol. 166, No. 17

Single gene turns flu deadly

By Christen Brownlee

The 1918 Spanish-influenza outbreak remains the worst pandemic (world-wide epidemic) in recorded history, killing more than 20 million people worldwide. With an eye toward preventing similar health disasters, researchers have long speculated about why the 1918 outbreak was so deadly (SN: 9/28/02, p. 196: <http://www.sciencenews.org/articles/20020928/fob3.asp>). Now, findings reported in the Oct. 7 Nature suggest that differences in a single viral gene made the 1918 flu strain particularly virulent (infectious).

Using a method called reverse genetics, Yoshihiro Kawaoka of the University of Wisconsin-Madison and the University of Tokyo and his colleagues engineered a relatively mild flu virus to carry two genes from the 1918 strain. In 1990, scientists had sequenced these genes from preserved lung tissue obtained from victims of the 1918 pandemic. The genes code for the proteins hemagglutinin and neuraminidase, which help flu viruses enter and infect cells.

Mice infected with the engineered virus quickly sickened and died. By inoculating mice with engineered viruses that carried just one or the other of the two genes, Kawaoka's team found that the hemagglutinin gene alone was enough to increase the virus' pathogenicity (ability to cause disease).

Within days, mice inoculated with this variety came down with severe and deadly infections. Mice infected with a virus carrying only the neuraminidase gene showed just mild flu symptoms.

Although all flu viruses carry a variant of the hemagglutinin gene, Kawaoka says that the 1918 version "does something different, but we don't know how it does it."

Questions:

When the article refers to "differences in a single viral gene, what were those differences called in the lesson?"

What viruses have you heard about in the news that scientists think could cause another pandemic?

Why are those viruses not currently causing pandemics?
