Insulin mRNA to Protein Kit©

A 3DMD Paper BioInformatics and Mini-Toober Folding Activity©

Teacher Key

www.3dmoleculardesigns.com
Why Is Insulin Important?

Insulin is a protein (peptide hormone) that plays a major role in glucose homeostasis – the regulation of your blood sugar levels. After you eat a meal insulin is normally released into your blood and triggers your liver, muscle, and fat cells to take up glucose from your bloodstream. Once inside these cells, the glucose can be used to fuel the production of ATP (adenosine triphosphate). ATP is frequently called the universal molecular currency because it transfers energy in our cells. See the animation on the CD included in the kit for more information on the role insulin plays in regulating blood sugar and the uptake of insulin.
Insulin Paper BioInformatics Activity

In this activity, you will explore the steps involved in the synthesis of the protein, starting with insulin mRNA. Specifically, you will consider how this mRNA is translated by the ribosome into a precursor form of insulin, and then how the precursor is processed to create the final, functional protein. As the final step in this activity, you will create a physical model of insulin by folding two Mini-Toobers (foam-covered wires) into the precise 3-D shape of the A-Chain and the B-Chain of this protein.

Becoming Familiar with the Data

A gene encoded within the DNA of a chromosome is transcribed into mRNA in the nucleus of a cell. The mRNA is then transported into the cytoplasm, where a ribosome reads the code and builds a protein (translation). This activity focuses on how the insulin mRNA is translated into the insulin protein.

1. Unroll your Insulin mRNA Map and look at the green-colored sequence of letters at the top of the map.
   a. What different letters appear in this sequence?
      A, U, G, C
      __________________________________________________________
   b. What do these letters represent?
      The four nucleotides of RNA
      __________________________________________________________
      __________________________________________________________
      __________________________________________________________
The Standard Genetic Code

When RNA polymerase initially transcribed the insulin gene into messenger RNA, two introns – totaling 966 additional nucleotides – were included in the precursor form of the insulin mRNA. These intron sequences were removed from the mRNA in a splicing reaction as the mRNA was being transported out of the nucleus of the cell. You might want to discuss why almost all eukaryotic genes contain introns.

<table>
<thead>
<tr>
<th>First Letter</th>
<th>Second Letter</th>
<th>Third Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>U</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>G</td>
</tr>
</tbody>
</table>

Translation Reading Frames

2. Look at the three blue sequences at the bottom of the Insulin mRNA map.

a. What different letters appear in these blue sequences? How many different letters appear in these sequences?

Twenty (20) different letters are shown: A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y
b. What do these letters represent?

The 20 amino acids

_______________________________________________________________________
_______________________________________________________________________

c. What is the relationship between the green letters at the top of the strip to the blue letters at the bottom?

The blue letters represent the amino acids that are encoded in the triplets of green nucleotides of the mRNA.

_______________________________________________________________________
_______________________________________________________________________
d. Why are there three blue sequences?

These three sequences represent the three possible reading frames in which the mRNA could be decoded.

_______________________________________________________________________
_______________________________________________________________________
e. What do you think the asterisks (*) represent in the blue sequences?

Any one of the three translation stop codons (UAA, UGA, UAG)

_______________________________________________________________________

Identifying the A-Chain and the B-Chain

The insulin protein actually consists of two separate chains, known as the A-Chain and the B-Chain. The amino acid sequences of the two chains are shown below:

A-Chain
G I V E O C C T S I C S L Y Q L E N Y C N

B-Chain
F V N Q H L C G S H L V E A L Y L V C G E R G F F Y T P K T
Identifying the A-Chain and the B-Chain (continued)

3. Locate, highlight and label the A-Chain and the B-Chain amino acid sequences on your Insulin mRNA Map.

   a. What do you notice about the location of the A-Chain and B-Chain amino acid sequences within the BioInformatics Map?

   The B-Chain comes first, followed by the A-Chain.

   Thirty-five (35) amino acids separate the B-Chain and the A-Chain sequences.

Note: The subunit composition of insulin (two chains) was known before the sequence of the gene was determined. Unfortunately, when the gene was sequenced, it was discovered that the B-Chain was encoded before the A-Chain – which has been confusing biology students ever since!

Translating mRNA Into Protein

To translate an mRNA into protein, the ribosome recognizes an AUG codon – and begins decoding the mRNA as it moves from left to right (5’ to 3’) down the mRNA sequence. As a result, all proteins begin with the amino acid methionine (Met, M) at their N-terminal end.

In humans and other eukaryotes the ribosome begins synthesizing proteins at the first AUG codon from the 5’ end of the mRNA.
4. Highlight the protein that is synthesized by a ribosome. The ribosome binds to the first AUG located downstream (to the right) of the 5' end of the mRNA to begin synthesis. See the Insulin mRNA - Teacher Map.

   a. Where does the protein stop?
      
      **At the STOP Codon (UAG) consisting of mRNA base numbers 390-392**

   b. How many amino acids are in the insulin protein?
      
      **110 amino acids**
Insulin is synthesized in beta islet cells of the pancreas. Following a meal, it is secreted from these cells into the bloodstream. Proteins that are destined to be released from the cell travel through the endoplasmic reticulum and Golgi apparatus of pancreatic cells to the cell surface where they can be secreted.
Preproinsulin - the Precursor Form of Insulin

Precursor Insulin

The precursor (inactive) form of insulin is known as *preproinsulin*. The first 24 amino acids of preproinsulin make up the Endoplasmic Reticulum (ER) Signal Sequence. As the protein is being synthesized, this signal sequence begins to emerge from the ribosome. Other proteins in the cell recognize this peptide and dock the ribosome onto the ER. As the rest of the protein is synthesized, it is directed through this membrane, into the lumen of the ER. From there, the preproinsulin is further processed (cleaved into four pieces) as it moves through the ER to the Golgi, and to the cell surface.

5. Locate, highlight and label the ER Signal Sequence on your Insulin Bioinformatics strip. See the Insulin mRNA - Teacher Map.

Signal Peptide

MaLwMrLLPlLLaLLaLWGPDPAAA

a. Referring to the Standard Genetic Code table, categorize the chemical properties of each of the 24 amino acids that make up the ER Signal Peptide (hydrophobic, hydrophilic, positive charge, or negative charge). What is notable about the chemical properties of the amino acids that make up the ER Signal Peptide?

Twenty (20) of the 24 amino acids are hydrophobic.
Soon after the ribosome that is synthesizing preproinsulin is docked onto the ER, a protease in the ER cuts the precursor protein between amino acids 24 and 25 (Alanine, Ala, A and Phenylalanine, Phe, F). The 24 amino acid signal peptide is rapidly degraded, while the remaining 86 amino acid proinsulin begins its journey toward the Golgi and cell surface.

Proinsulin consists of the B-Chain (30 amino acids) and the A-Chain (21 amino acids), separated by the 35 amino-acid C-Peptide.

As proinsulin spontaneously folds into its final 3-D shape in the ER, another protease cuts the protein at two sites: between amino acids 54 and 55 (Threonine, Thr, T and Arginine, Arg, R) and between amino acids 89 and 90 (Agrinine, Arg, R and Glycine, Gly, G). As the C-Peptide is released from the folded B-Chain and A-Chain complex, it floats away and is degraded.
6. Locate, highlight, and label the C-Peptide on your Insulin BioInformatics Map.
See the Insulin mRNA - Teacher Map.

   a. Since the C-Peptide is cut out of proinsulin to create the final mature insulin (Chain B and Chain A) what role do you think the C-Peptide might play in the biosynthesis of the mature insulin protein?

   One possible suggestion is that the C-Peptide is required for the initial folding of the B-Chain and A-Chain into its final 3-dimensional shape. After this folding has occurred, a protease encountered in the Golgi cuts the C-Peptide free from the A-Chain and B-Chain.

As with many secreted proteins that must make their living in the harsh environment outside the cell, insulin is stabilized by two covalent disulfide bonds that join the B-Chain to the A-Chain. Each chain contributes one cysteine amino acid (Cys, C) to each disulfide bond.

   Cys7 of the B-Chain forms a disulfide bond with Cys7 of the A-Chain.
   Cys19 of the B-Chain forms a disulfide bond with Cys20 of the A-Chain.
   A third disulfide bond forms between Cys6 and Cys11, both from the A-Chain.

7. Circle each Cys on your Insulin mRNA to Protein map that participates in disulfide bond formation, and connect (with a line) the pairs that interact to form each disulfide bond. See the Insulin mRNA - Teacher Map.
Like all proteins, insulin folds into a specific 3-D shape, following basic principles of chemistry. It is this 3-D shape that allows it to bind to the insulin receptor protein on the surface of liver, muscle, and fat cells to trigger the uptake of glucose from the bloodstream. In this final activity, you will shape two Mini-Toobers into the 3-D shape of the insulin protein.

1. Gather all of the parts you need:
   - Insulin Mini-Toober Folding Map
   - Orange and Purple Mini-Toobers
   - Bag with parts for Mini-Toobers
   - Sidechains
   - Support Posts
   - Metal Clips
   - White Dots
   - Plastic Markers
   - Endcaps

2. Unroll your Insulin Mini-Toober Folding Map and identify the N-terminus (blue) and the C-terminus (red) of each protein chain by putting one red and one blue end cap onto the ends of each Mini-Toober.

3. Using the Map, locate the cysteine amino acids on each protein chain. Write the number of each of the six cysteines on the white dots and add these numbered dots to six metal clips.

3. Carefully align each Mini-Toober with the corresponding chain on the Insulin Mini-Toober Folding Map matching the end caps to the images of the end caps on the Map. Add the appropriately numbered metal clips to the Mini-Toober. The metal clips represent the alpha-carbon of each cysteine amino acid.
4. Indicate where the α-helices are on each protein chain by placing the red plastic markers at the beginning and the end of each α-helix. Indicate where the β-sheets are on each protein chain – by placing the yellow plastic markers on the Mini-Toober at the beginning and the end of each β-sheet shown on the Map.

5. Fold the Mini-Toobers to create:
   a. The α-helices (right-handed) and the β-sheet strands (extended zig-zag) in each protein chain, and then
   b. The overall 3-D shape of each protein chain. Use the Jmol visualization tool and/or the images at the end of the Map to fold your insulin

6. Add the Cysteine sidechains to each of the six Cysteine alpha-carbon atoms.

7. Assemble the two chains into the final insulin model by positioning the chains as shown in the photo above, using the images on the map and/or the Jmol visualization tool located on the CD.

**Note:** The three pairs of Cysteine amino acids that form covalent disulfide bonds should be close to each other in the final model. Use the three plastic support posts to stabilize the protein at these three disulfide bonds.
• The insulin gene is located on the short arm of chromosome 11 in humans.

• The insulin gene is transcribed into an insulin mRNA molecule in the nucleus of the beta islet cells of the pancreas.

• The insulin mRNA is then transported to the cytoplasm of the cell where a ribosome recognizes the first AUG near the 5’-end of the mRNA and begins translating the protein, starting with methionine.

• The ribosome synthesizes a precursor form of insulin, known as preproinsulin.

• Preproinsulin is processed to become mature, functional insulin as it proceeds through the endoplasmic reticulum and Golgi apparatus, moving toward the cell membrane where it can be secreted from the cell.

• When there are high levels of sugar in the blood, insulin is released from the beta cells. It binds to receptors on the surface of liver, muscle, and fat cells. Binding of insulin results in a series of reactions within the cell (a signal cascade), leading to the fusion of vesicles containing glucose transporter proteins (GLUTS) with the membrane. The GLUTS transport glucose into the cells, where it is stored.
Introduction

The current well-documented and related epidemics of obesity and diabetes in the US provide a compelling reason to teach insulin and glucose homeostasis in high school and college biology curricula. Many teachers have found that insulin can serve as a unifying theme across many areas of a biology curriculum including:

- **Molecular Biology** – the flow of genetic information from DNA $\rightarrow$ RNA $\rightarrow$ Protein
- **Biochemistry** – the structure of the insulin protein
- **Cell Biology** – the role of the endoplasmic reticulum and Golgi in insulin secretion
- **Biotechnology and Recombinant DNA** – expression of human insulin in bacteria
- **Diabetes** – the molecular basis of insulin resistance in Type II diabetes
- **Nutrition** – the universal role for glucose in energy metabolism

The Insulin mRNA to Protein Kit© addresses the first two topics – molecular biology and biochemistry.

All of these insulin-related topics are highlighted in the Insulin Poster which is also available from 3D Molecular Designs (3DMD) at www.3dmoleculardesigns.com.
In reference to pages 1-4:

**Bioinformatics.** The new field of bioinformatics focuses on the analysis of nucleotide and amino acid sequences – with the goal to understand what their coded, digital information means. When, to what degree, and under what circumstances is this information expressed? While real bioinformatics involves the use of computer algorithms to automatically search and analyze long stretches of sequences, we believe that your students should first experience the manual analysis of nucleotide and amino acid sequence data, before progressing to the use of the computer tools. The manual analysis will provide them with a better foundation for understanding and an appreciation for the computer analysis.

**Introns and Exons.** Your students will probably not know about introns and exons unless you have exposed them to this idea in advance. You could introduce your students to this concept using 3DMD’s introduction to bioinformatics: Map of the β-Globin Gene. This activity allows your students to discover that the gene for the β-globin protein is not continuous. Rather, the coding region of the gene (exon) is interrupted by two intervening sequences (introns).

Introns are copied into the precursor mRNA by the RNA polymerase that transcribes the β-globin gene. These introns are then spliced out of the precursor mRNA as it is being transported to the cytoplasm where the mature mRNA will be translated into protein by the ribosomes. **In this 3DMD Insulin BioInformatics activity, we begin with the mature (spliced) mRNA.**

**The 3’ poly(A) End.** Note that the last 25 nucleotides at the 3’-end of the mRNA sequence are A. This represents the poly(A) sequence that is added to all eukaryotic (but not prokaryotic) mRNAs in a post-transcriptional processing event. This poly A is never translated into poly-leucine because a translation stop codon terminates translation upstream from the poly(A) sequence.
In reference to pages 4-6:

**DNA/Protein Relationship.** One “Big Idea” of biology is the relationship between DNA and protein. In teaching this big idea, we suggest that you focus first on proteins. Once students understand that proteins are long linear polymers of amino acids – joined together in a unique sequence – they are then ready to consider: How does the cell remember the correct order in which to join the amino acids together to make this protein? The answer, of course, is that the cell encodes the correct sequence of amino acids in the sequence of nucleotides that make up the gene for that protein. The genetic code is a triplet code that allows us to infer the amino acid sequence of a protein from the nucleotide sequence of its gene.

**Three Different Reading Frames.** The three different amino acid sequences shown (in blue) at the bottom of the strip represent the three different possible amino acid sequences that could be decoded from the mRNA sequence. Be careful to not teach a misconception: the three blue sequences do not represent proteins that are actually made by the ribosome. Rather they represent three different proteins that could be made – depending on the reading frame used by the ribosome. The ribosome binds to the initiation codon in such a way that only the correct reading frame (in this case, the first one) is used during decoding – to make the precursor form of insulin.

**Three Stop Codons mean STOP!** Stop codons are represented as astericks (*) in the sequence.
**Teacher Notes (continued)**

**In reference to pages 5-6:**

**Initiation of Protein Synthesis.** In the first step of translating an mRNA into protein, the small ribosome subunit binds to the mRNA at an AUG initiation codon. This is a critical event that establishes (i) the beginning of the protein and (ii) the correct reading frame in which the mRNA sequence will be decoded. You might want to challenge your students to think more deeply about the importance of this initiation event by posing some of the following questions:

- What prevents the ribosome from assembling on other AUG codons located elsewhere in the gene?

  In eukaryotes, the ribosome starts at the 5’end of the mRNA and scans down the RNA looking for the first AUG codon.

- How do you think a ribosome is able to recognize the AUG codon and bind?

  The small ribosome subunit, composed of RNA and protein, binds to the mRNA in much the same way as an enzyme binds to its substrate. The structure and chemical composition of the AUG binding pocket in the ribosome and the AUG codon of the mRNA are complementary to each other. The complex is stabilized by hydrogen bonds, electrostatic interactions, and hydrophobic interactions.

- What additional steps (after binding of the small ribosome subunit to the initiation AUG codon) occur to complete the initiation complex?

  A tRNA charged with the amino acid methionine must bind to the initiation complex, such that the anti-codon of the tRNA is base-paired with the AUG codon of the mRNA.
**In reference to pages 6-8:**

**The Hydrophobic Signal Peptide.** It is very unusual to encounter such a clustering of hydrophobic amino acids in a protein. You should challenge your students to propose a reason that the signal peptide is so hydrophobic. As this hydrophobic peptide begins to emerge from the ribosome, the ribosome is docked onto the membrane of the endoplasmic reticulum and this very hydrophobic signal peptide is buried in the hydrophobic membrane. As the rest of the insulin precursor protein is made, it is extruded into the lumen of the ER. A protease then cuts the precursor insulin protein between amino acids 24 and 25. The remaining soluble proinsulin is released from the membrane and begins to move through the ER, toward the Golgi.

**An Illustration of this Secretory Pathway.** David Goodsell has created a stunning painting of a section of an antibody-producing B cell that illustrates this process: an antibody gene is transcribed into precursor mRNA in the nucleus, the mRNA is spliced in the nucleus and then transported through a nuclear pore to the cytoplasm where it is bound to ribosomes that are docked onto the ER. The protein is shown as it travels through the ER to the Golgi and then is pinched off in vesicles that are actively transported to the cell surface where the vesicles fuse with the cell membrane. A poster of the Panorama of a Human Cell is available from 3D Molecular Designs. It can also be seen on four consecutive pages in David Goodsell’s book, *The Machinery of Life*, Second Edition (ISBN: 978-0-387-84924-9), which is available at [http://www.springer.com](http://www.springer.com).
Protein Folding and Assembly into Multi-Subunit Complexes. A linear chain of amino acids (a protein) will spontaneously fold up into a compact 3-D shape following basic principles of chemistry. These principles dictate that hydrophobic amino acids will cluster together in the center of a globular protein where they are protected from the polar water molecules that surround the protein. Polar or charged amino acids are oriented on the outside of the protein – where they can easily interact with water. The two chains of insulin are only 30 and 21 amino acids long. By synthesizing these two chains as part of a longer 86 amino acid precursor form of insulin (B-Chain plus C-Peptide plus A-Chain) these basic principles of chemistry are brought to bear on more amino acids – and the final structures of the B-Chain and A-Chain of insulin are more easily attained.

As an added bonus, once the C-Peptide is clipped and released from proinsulin, the B-Chain and A-Chain already exist as their functional complex. Imagine how much harder it would be if the B-Chain and the A-Chain were made on different ribosomes and then had to diffuse around the cell looking for each other before they could interact to form the functional insulin protein.

Maybe the most interesting question to ask is simply – Why does the cell bother to remove the C-Peptide at all? Why not leave it as part of a single-chain insulin?

Stabilizing Proteins. Disulfide bonds play an important role in stabilizing the structure of some proteins. However, be careful that you don’t teach a misconception – that all proteins require disulfide bonds for stability. Many proteins do just fine without disulfide bonds. It appears that secreted proteins are more apt to be stabilized by disulfide bonds than proteins that remain inside the cell. Students might be interested to know that there are enzymes in the cell known as disulfide isomerases. These enzymes function by reducing and oxidizing (making and breaking) disulfide bonds. They are often necessary for proteins to generate the correct pattern of disulfide bonds among multiple cysteine amino acids.
In reference to pages 11-12:

This exercise in exploring the complexity of 3-D space will be easy for some students – and very difficult for others. In evaluating the physical models that are created, we suggest that you only focus on a couple of key points:

1. **Are the alpha helices right-handed?** (As you walk “up” the spiral staircase represented by the alpha helix, your right hand should be on the outside rail.)

2. **Do the pairs of cysteine amino acids that form disulfide bonds end up next to each other**— so that students could imagine the formation of the covalent disulfide bonds? If this is true for the model you are evaluating, the students who created this model have **discovered** the complexity of 3D space— but they have not yet **mastered** it.

The important point to emphasize to students is that basic principles of chemistry act on all of the amino acid sidechains and backbone atoms to drive the shape of each chain into its final 3-D form. This final shape is one in which the overall free energy of the system is minimized. In this way, protein folding obeys basic laws of thermodynamics.
Related 3DMD products. 3D Molecular Designs offers several additional instructional kits that can be used to enhance and extend the ideas presented in the Insulin mRNA to Protein Kit™. These can be found at www.3dmoleculardesigns.com.

Water Kit™. Let your students explore the polar nature of water to lay the foundation for understanding the molecular interactions of proteins in the aqueous environment of the cell.

Amino Acid Starter Kit™. Let your students explore the basic principles of chemistry that drive the folding of proteins into complex shapes.

The Map of β-Globin Gene™. Show your students that human genes are split into exons and introns.

DNA Discovery Kit™. Let your students discover the structure of DNA – just like James Watson and Francis Crick.

David Goodsell’s Panorama of the Human Cell. Show your students the complex environment of the human cell that the insulin protein must navigate -- as only David Goodsell’s art can illustrate.
MSOE Center for BioMolecular Modeling Developed Insulin mRNA to Protein Kit®

All of the activities, graphics, and related materials for the Insulin mRNA to Protein Kit® originated with and were extensively field-tested by 3D Molecular Design’s sister organization, the MSOE Center for BioMolecular Modeling (CBM). More than 75 teachers, who participated in the CBM’s summer course, Modeling the Molecular World, provided input and feedback over a three year period during development and field-testing. Director of the CBM, Tim Herman, Ph.D.* led this effort and was assisted by Program Directors Margaret Franzen and Shannon Colton, and Lead Designer Mark Hoelzer.

All of us at 3D Molecular Designs thank the CBM staff and all of the teachers who were involved with the development of the Insulin mRNA to Protein Kit®

CBM Website: http://cbm.msoe.edu
CBM Lending Library: http://cbm.msoe.edu/teachRes/library
Glucose Homeostasis Animation: http://cbm.msoe.edu/teachRes/animations/insulin.html
Jmol: http://cbm.msoe.edu/teachRes/jmol/insulin.html
3D Molecular Designs: www.3dmoleculardesigns.com

*Tim Herman founded 3D Molecular Designs, LLC as a way to broadly disseminate the best activities developed by the CBM. In addition, the Milwaukee School of Engineering Library, the Center for BioMolecular Modeling and 3D Molecular Designs provide the MSOE Model Lending Library for teachers whose budgets don’t permit them to purchase kits and models or who want to test kits before purchasing.