Aipotu II: Biochemistry

Introduction:

The Biological Phenomenon Under Study

In this lab, you will continue to explore the biological mechanisms behind the expression of flower color in a hypothetical plant. These flowers can be white, red, orange, yellow, green, blue, purple, or black.

Scenario:

You are the chief biologist for a breeder of fine flowers. Your company sells seeds that customers plant in their gardens. Since most of your customers expect that the flowers will grow each year from seeds produced the previous year, you try to produce true-breeding plants whenever you can.

You've found a new species of flower with an attractive shape. You've collected four plants from the wild: two green, one red, and one white. Your customers would really like to have purple flowers from this plant. You set out to create a true-breeding purple flower.

In Part I, you found the alleles involved in color production. You went on to describe the colors produced by different allele combinations. You now need to expand your understanding to *how* these colors are produced and how they interact at a *Biochemical level.*

In this lab session, you will use a version of the Protein Investigator that computes the color of the proteins you fold. You should know that no real protein works this simply; however, the *way you figure out* how structure leads to color is the same as a biochemist would use in the lab.

Hypothesis Testing

In the three Aipotu labs, you will use a process much like that used by practicing scientists as they conduct research. Although this process almost never follows a formula, it often proceeds as follows:

1. **Observe Patterns**. Observe the natural world and look for patterns, exceptional events, etc. For example, you might observe that red proteins tend to have long thin shapes.
2. **Develop hypotheses**. From the observations, you define testable hypotheses – statements or questions that can be addressed experimentally. Continuing the example, you might reasonably hypothesize that long thin proteins will be red.
3. **Test hypotheses**. You then set up experiments or observations that will collect data that bear on your hypothesis. In the example, you might type in a sequence of amino acids that would be expected to fold into a long thin shape, fold the protein, and observe its color. If your hypothesis is correct, it will be red. If you get another result, your hypothesis is incorrect.
4. **Revise hypotheses as necessary**. If your results do not match your prediction, you need to revise your hypothesis and go to Step (3) again until they do match.

You have already been doing this informally in the VGL and Aipotu II labs. Today, you will do it much more formally.

Important Note: It is always important to keep in mind, the 'scientist's mantra': always be asking yourself "How could I be being fooled by this?" To continue the example from the previous page, consider the following:

Suppose that the long thin protein were red, you might congratulate yourself that you had found the connection between shape and color. However, what if the real mechanism is that proteins containing arginine are red and your long thin protein just happened to be made with arginine. The red color would be fooling you into thinking you had it right.

How do you avoid this trap? Even the best scientists sometimes fall into traps like this. The answer is to *always be thinking of alternative explanations for your results*. In the case above, one long thin red protein does not mean that "long & thin = red". You have to collect more data: proteins that aren't long and thin; long and thin proteins with different amino acids; etc.

Today, more than in Aipotu I, it is the *process of science* rather than the *answer* that is most important. You will use a blog to collaborate as a class to solve this scientific problem.

Tasks:
Work together as a class to:
- Determine the differences in amino acid sequence between the proteins produced by the alleles you found in Part I.
- Determine how the amino acid sequence of a pigment protein determines its color.
- Explain, in terms of the proteins present, the interactions between the alleles you found in part I.
    - Why is the color phenotype of some pigment proteins dominant while others are recessive?
    - How do the pigment proteins combine to produce the overall color of the plant?
- Construct a purple protein to demonstrate your understanding of this process.

As in real science, these tasks are too big to be solved by one group alone. If you think of real research as solving an enormous jigsaw puzzle, each researcher works on only one little corner of the puzzle. Scientists publish papers and present findings at conferences in order to connect the corners of the puzzle that each is working on.

In this lab, you will use a shared **Google Document** to share hypotheses, data, and conclusions among your lab-mates so that you can solve this problem in the required time. Individual contributions can be small, or even negative ("I know that it *can't* be…"), but you will be able to accomplish the tasks above if you work together. In research, no one 'owns' data – the point is to figure out how the world works, not who got the result.

Using the tool:

Once you start Aipotu, you can switch to the tool for this section by clicking the "Biochemistry" tab near the top of the window. You will see something like this:

This is a reference for the names and properties of the amino acids.
- The single-letter code is shown below the three-letter code.
- White circles are hydrophilic; gray are intermediate; and black are hydrophobic.
- A red (-) indicates a negatively-charged side chain
- A blue (+) indicates a positively-charged side chain
- A green (*) indicates a side chain that can make a hydrogen bond.

This part of the program uses the one-letter code for the 20 amino acids:

Amino Acid	3-letter code	1-letter code	Mnemonic
Alanine	Ala	A	Alanine
Arginine	Arg	R	aRginine
Asparagine	Asn	N	asparagiNe
Aspartic acid	Asp	D	asparDic acid
Cystine	Cys	C	Cystine
Glutamine	Gln	Q	Q-tamine
Glutamic Acid	Glu	E	glu-tE-amic acid
Glycine	Gly	G	Glycine
Histidine	His	H	Histidine
Isoleucine	Ile	I	Isoleucine
Leucine	Leu	L	Leucine
Lysine	Lys	K	lysinK
Methionine	Met	M	Methionine
Phenylalanine	Phe	F	Fenylalanine
Proline	Pro	P	Proline
Serine	Ser	S	Serine
Threonine	Thr	T	Threonine
Tryptophan	Trp	W	tWptophan
Tyrosine	Tyr	Y	tYrosine
Valine	Val	V	Valine

- Click in the **Amino Acid Sequence Box** at the top of the **Upper Folding Window**. Type a short sequence of letters and you will see a short amino acid sequence appear in the window. This tool converts the single-letter code to the three-letter code automatically.
- Click the "FOLD" button and a two-dimensional version of your amino acid sequence will appear in the **Folded Protein** window.

There are several important things to note about this folding process:

This is a highly-simplified model of protein folding. It is not intended to predict the correct structures of any proteins; it is designed to illustrate the major principles involved in that process. The important features of proteins that this software retains are as follows:
- Amino acids have side-chains of varying hydrophobicity, charge, and hydrogen bonding capacity.
- The amino acids are connected in an un-branched chain that can bend.
- Hydrophobic amino acids will tend to avoid the water that surrounds the protein; hydrophilic amino acids will bind to the water.
- Amino acids that can form hydrogen bonds will tend to form hydrogen bonds if they can.
- Positively-charged amino acids will tend to form ionic bonds with negatively-charged amino acids if they can.
- Like-charged amino acids will repel each other if they can.
- Ionic interactions are stronger than hydrogen bonds, which are stronger than hydrophobic interactions.

Even though this software provides some important insights into protein folding, you should always keep in mind that this is an approximation. The most important "gotcha's" to be aware of are:
- This program folds proteins in 2-dimensions only.
- This program treats all amino acids as equal-sized circles.
- This program models an environment where disulfide bonds do not form.
- This program folds the protein based on the interactions between the side chains only.
- This program does not model secondary or quaternary structure.
- This program assumes that all side chains with hydrogen bonding capability can bond with each other.

These simplifications are necessary for two reasons. The first is technical: it turns out to be extremely difficult to predict the full 3-d folded structure of a protein given only its amino acid sequence. As of the writing of this lab manual, it takes a super-computer several days to predict the fully-folded shape of even a small protein like lysozyme. Even then, the predictions don't always match known structures. Given the computers we have, it might take years….

The second reason is educational. Proteins are complex 3-dimensional molecules; thus, it can be hard to find your way around when inside one. Likewise, it would be very difficult to visually compare two protein molecules to observe the effects of changes to their amino acid sequence. It would be easy to miss the forest (the forces that control protein structure) for the trees (the tiny details of the structures).

For these reasons, we will use this simplification. It retains the properties of amino acids that are important for this lab while being simple and fast.

There are three kinds of experiments you can perform with this tool. The following sections use examples to show you how to do each; you will need to devise your own experiments to carry out the tasks from the previous page.

I) Examine the Pigment Proteins Present in an Organism from the **Greenhouse**. This simulates extracting the pigment protein(s) produced by the two alleles of the pigment protein gene that an organism possesses, displaying their two-dimensional structures, and displaying their colors.

> 1) Double-click on the **Green-2** organism in the **Greenhouse**. You should see this:
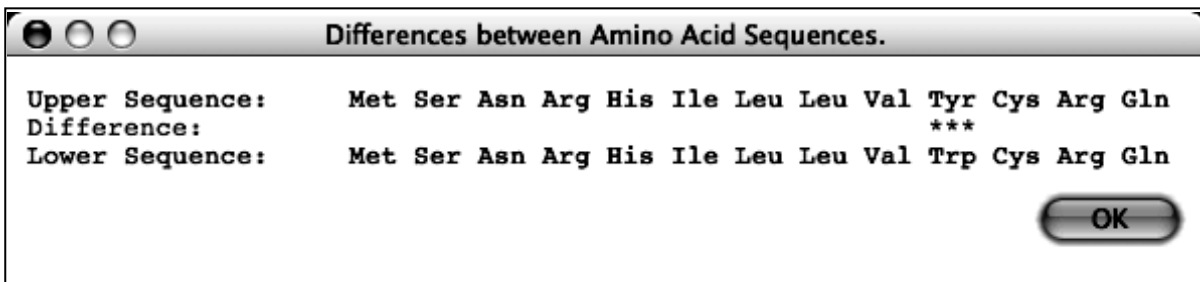

The **Green** organism contains two alleles of the pigment protein gene. Each of these alleles produces a different protein. One of these proteins is shown in the **Upper Folding Window**; it is a blue-colored protein as shown by the blue square next to the "Color:" label. The other protein is shown in the **Lower Folding Window**; this is yellow-colored protein. The combined color of the two proteins is green as shown by the **Combined Color** in between the two **Folding Window**s.

II. Examining Pigment Proteins From the Mutant Organism(s) You Made in the Aipotu I Lab. You can go back to your saved organism(s) in the greenhouse.

III) Compare the amino acid sequences of two pigment proteins. This aligns the two amino acid sequences so that the highest number of matching amino acids is obtained and then finds the remaining differences.

1) Double-click on the **Green** organism in the **Greenhouse**. You should see that the **Upper Folding Window** shows a blue protein and the **Lower Folding Window** shows a yellow protein.

2) You can compare the amino acid sequence of these two proteins by clicking on the "Compare" menu and choosing "Upper vs. Lower". A window will appear showing the differences between the two sequences. This is shown below:

```
● ○ ○          Differences between Amino Acid Sequences.

Upper Sequence:     Met Ser Asn Arg His Ile Leu Leu Val Tyr Cys Arg Gln
Difference:                                                 ***
Lower Sequence:     Met Ser Asn Arg His Ile Leu Leu Val Trp Cys Arg Gln

                                                         ( OK )
```

This shows that the only difference is that, in the upper (blue) protein, amino acid 10 is tyrosine, while in the lower (yellow) protein, amino acid 10 is tryptophan.

⇒You can also copy the sequence of a particular protein to the clipboard using the options in the **Edit** menu. You can then **Compare** a sequence to the one in the clipboard.

IV) Edit a Protein Sequence or Create a New Protein Sequence and Determine its Two-Dimensional Structure and Color. You can edit the sequence in either of the **Amino Acid Sequence** boxes and click the "Fold" button to predict the two-dimensional structure and color of the protein. The tool will also give the color that results from the combination of the colors in the Upper and Lower windows.

For example, click anywhere in the "Tyr" corresponding to amino acid 10 in the **Upper Amino Acid Sequence** box. Click the "delete" key and that amino acid will disappear. Type an "L" (the one letter code for leucine) and the amino acid sequence should be:

Met Ser Asn Arg His Ile Leu Leu Val Val Cys Arg Gln

Click the "FOLD" button in the **Upper Folding Window** (or click the return key). You will see that the color of the new protein is white as shown by the "Color:" in the **Upper Folding Window**. You should also notice that:
- the "Combined Color" at the center of the window is now yellow.
- there is now an entry in the **History List** with your new protein. The background of **History List** entry is white to show the color of this protein.

You can also click the "Load Sample Protein" button. This will load a sample amino acid sequence that folds to a white-colored protein with a shape that is similar to many colored proteins.

If you double-click an entry in the **History List**, you will you get a pop-up menu with a list of useful options:
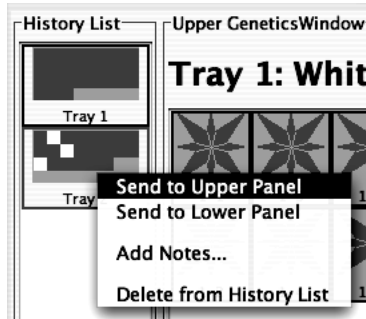
- **Send to Upper Panel**: Sends this **Tray** to the **Upper Panel** so you can cross those organisms.
- **Send to Lower Panel**: Sends this **Tray** to the **Lower Panel** so you can cross those organisms.
- **Add Notes...**: Allows you to add notes to the **Tray** in the **History List**. These notes will appear if you leave the cursor over the **Tray** for a few seconds.
- **Delete from History List**: deletes the **Tray** from the **History List**; this is cannot be undone.

⇒ You can also *take a Snapshot of either Work Panel* In order to make entries in the data blog, you will need to take snapshots of particular proteins. A typical snapshot looks something like this; it shows the protein's shape, amino acid sequence, and color:
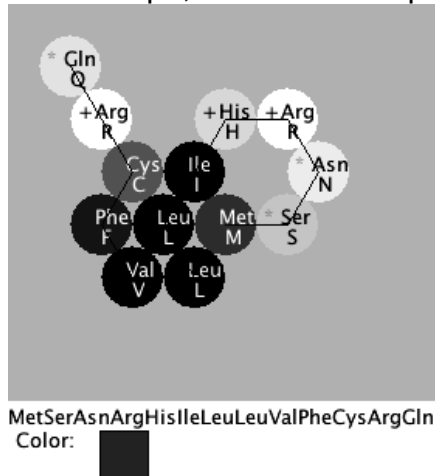
You can take a snapshot in either of two ways:
- *Save a snapshot as a picture*. In the **File** menu, choose either **Save Image of Upper Panel…** or **Save Image of Lower Panel…** . You will then be asked to give the file a name and it will be (typically) saved to the desktop as *name*.png; you can then import that file into a document.
- *Take a snapshot to the clipboard*. In the **Edit** menu, choose either **Copy Image of Upper Panel to Clipboard** or **Copy Image of Lower Panel to Clipboard**. You can then paste the resulting image into another program, like Microsoft Word or Google Documents.

Using the Google Document to collaborate during class
The Google Document will allow all the members of the class to share hypotheses, data, and conclusions in a scaled-down scientific community.

As you work, you should post to the document. You should not post all your data; just the interesting bits. It is probably better to err on the side of including something rather than not; you never know if it will turn out to be interesting. You should also consult the document frequently to see if there is data there that will help you figure out what's going on.

Using the document may be tricky at first, but it will soon become easier.

I) Logging In – type in the URL given in class.

II) Posting – you should do this when you have an interesting result to share with the class. Again, it is better to err on the side of posting than not posting.

A post should contain the following elements:

- **Title**. It should contain your group name and a short phrase indicating the subject of your post. For example, "Brian, Tina, & Ling: This long protein is red!"

- **An Image of the protein**.
    - *Copy and paste an image to the document*. Using Aipotu's "**Copy Image of … Panel…**" from the **Edit** menu.
    - *Next, upload it to the document*.

- **A Hypothesis**. Using the format buttons, select bold type and type "Hypothesis:". Go to the next line and, using regular type, write the hypothesis you were exploring. For example, "Long thin proteins will be red".

- **The Experiment**. "Experiment:" should be bold and on it's own line. Follow with a brief description of the experiment; for example, "We designed a long thin protein".

- **The Result**. "Result:" should be bold and on it's own line. Follow with a brief statement describing the result; for example, "The protein was colorless".

- **The Conclusion**. "Conclusion:" should be bold and on it's own line. Follow with a brief conclusion; in this case, "The hypothesis is incorrect; long thin proteins are not necessarily red."
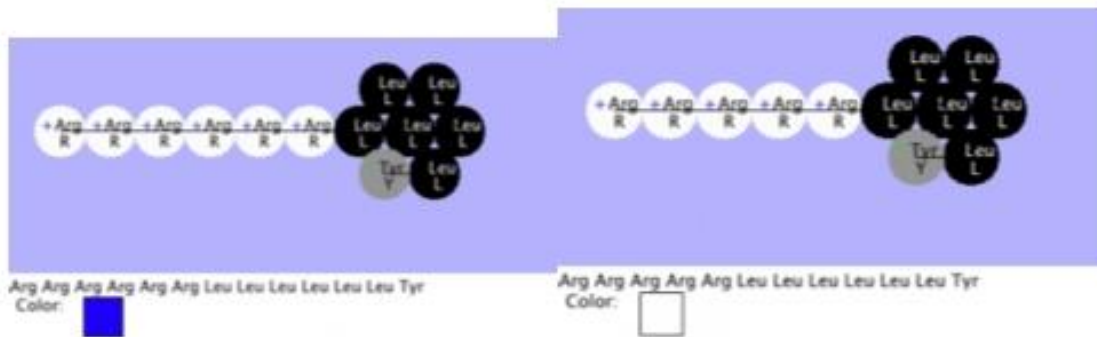
A sample post can be found on the next page:

A Sample post:

Matt,Paul,Quyen,Mike

October 28th, 2008 11:53 am

Hypothesis: At least 6 hydrophilic amino acids are required for any color to show.

Experiment: We constructed a protein consisting of a hydrophilic group of 6 amino acids and 7 hydrophobic amino acids. We added and subtracted hydrophilic amino acids. Under 6, no color showed. Over 6, color returned.

Arg Arg Arg Arg Arg Arg Leu Leu Leu Leu Leu Leu Tyr
Color:

Arg Arg Arg Arg Arg Leu Leu Leu Leu Leu Leu Tyr
Color:

Result: When there were 6 total hydrophilic amino acids, color was shown. When there were under 6 amino acids, we saw no color.

Conclusion. Our hypothesis is   supported

IMPORTANT NOTES:

This software is under development. Please treat it gently and be patient. Please report any bugs. **You should save your Greenhouse regularly, especially if you save a large number of organisms. You should also post pictures to the blog as soon as you can.**

Specific Tasks for this section

Work as a class, using the Google Document to answer:
  a) What are the differences in the amino acid sequences of the proteins produced by the alleles you define in Part I? Hint: use the **Compare** menu to find the difference(s) between the amino acid sequences.
  b) What features of the amino acid sequence make a protein pigmented?
  c) What features of the amino acid sequence make a protein a particular color?
  d) How do the colors combine to produce an overall color? How does this explain the genotype-phenotype rules you found in part (I)?
  e) Which proteins are found in each of the four starting organisms?
  f) Using this knowledge, construct a purple protein.

Hints:

A) It may be useful, before formulating any hypotheses, to look for *patterns* in the data. Which features do colored proteins have in common that uncolored proteins lack?
   - Try comparing the amino acid sequences of proteins with different colors.
   - Here are some additional interesting sequences to try:
     - `FFFFFFFRRRRRR`
     - `RRRFFFFFFFFRRR`
     - `KKKKKKLLLLLLF`
     - `KKKKKKLLLLLLL`
     - `SLQLNITMEVDFW`
     - `EEEWWWWWWWEEE`

B) Scientists, including yourselves, often find it useful to use mutation to study phenomena like this. Go to **Genetics** and make some mutants. Save any ones with interesting colors to the **Greenhouse**. Switch back to **Biochemistry** and look at the proteins they have.

Procedure:

1. Compare the proteins found in the starting strains to one another to answer questions (a) and (b) on the following pages.

2. Compare the proteins found in the starting strains to the "sample protein" that you get by clicking the **Load Sample Protein** button on one of the **Folding Windows** and then choosing from the **Compare** menu. Compare sequence and shape.

3. Based on these data, as a class, make several *specific* hypotheses that can be tested.

4. Each group should work on one or more of their hypotheses and post them to the document.

5. You may also discuss information with other groups but be sure to put any findings in the document for all to see and refer to later.

6. You will then be able to complete parts (d) through (f).

Put your data in the tables below:

(a) Which proteins are found in each of the four starting organisms?

Green-1

Green-2

Red

White

(b) **allele**   **color**   **amino acid sequence** (highlight differences)

(c) What features of a protein make it colored?

(d) What features of the amino acid sequence make a protein a particular color?

(e) How do the colors combine to produce an overall color? How does this explain the genotype-phenotype rules you found in part (I)?

(f) Show me that you have made a purple protein. For full credit, you need to explain *why* it is purple.

Lab report
- Must be typed; handwritten reports will not be accepted.
- Due during the week listed in the syllabus at the start of the lab session you are currently in. This is a firm deadline.
- Although you will perform these experiments as a group, each member of the group must turn in an individual lab report and this report must be in your own words.

Your lab report must include:

A description of one place where one hypothesis and some data interacted to lead to a firm conclusion. You should **not** describe all the experiments you did; just one place where you were able to draw a firm conclusion. The conclusion can be positive ("My hypothesis was supported because…") or negative ("My hypothesis was not supported because…"). This need not be an experiment you did, but it must have happened during your lab session.

1) <u>Hypothesis</u> – the hypothesis being tested. You must pick a clear and specific hypothesis that can be clearly and decisively tested by the experiment(s) you describe below. Your hypothesis need not be correct.

2) <u>Experiments</u> – a description of the experiment(s) you tried that addressed the hypothesis. Do not include all your experiments; only those relevant to the hypothesis from part (1). These must clearly and decisively test your hypothesis.

3) <u>Results</u> – pictures of the results of the experiments. You can get these from your section's blog page. Just drag the picture from the web page to the word processor. You may need to note the color(s) in text since most printers print in black and white.

4) <u>Conclusions</u> – do the data support the hypothesis or not. Explain your reasoning.

You will be graded on the quality of your argument: how clearly you described your hypothesis, experiment(s), and result(s) as well as how clearly and completely your conclusions are explained.