

Protein Investigator

Objectives

- To learn more about the interactions that govern protein structure.
- To test hypotheses regarding protein structure and function.
- To design proteins with specific shapes.

Introduction

In the lectures on Hemoglobin (and later in the Protein structure lab), you saw the 3-dimensional structure of proteins. This allowed you to see a protein in its full 3-dimensional form and to make hypotheses about the interactions between various side chains. However, you were not able to test your hypotheses because the program did not allow you to make changes to the protein sequence and observe their effects on the protein's structure and function.

In this lab, you will get an opportunity to interactively explore protein structure and function in a simplified system. 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 reconnected 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 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 in the Bio 111 labs, it might take years....

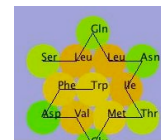
The second reason is educational. As you saw in the lysozyme lab, proteins are complex 3-dimensional molecules. It can be hard to find your way around when inside one. Likewise, it would be very difficult to visually compare two protein molecules to observe the effects of changes to their amino acid 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 in Bio 111 while being simple and fast.

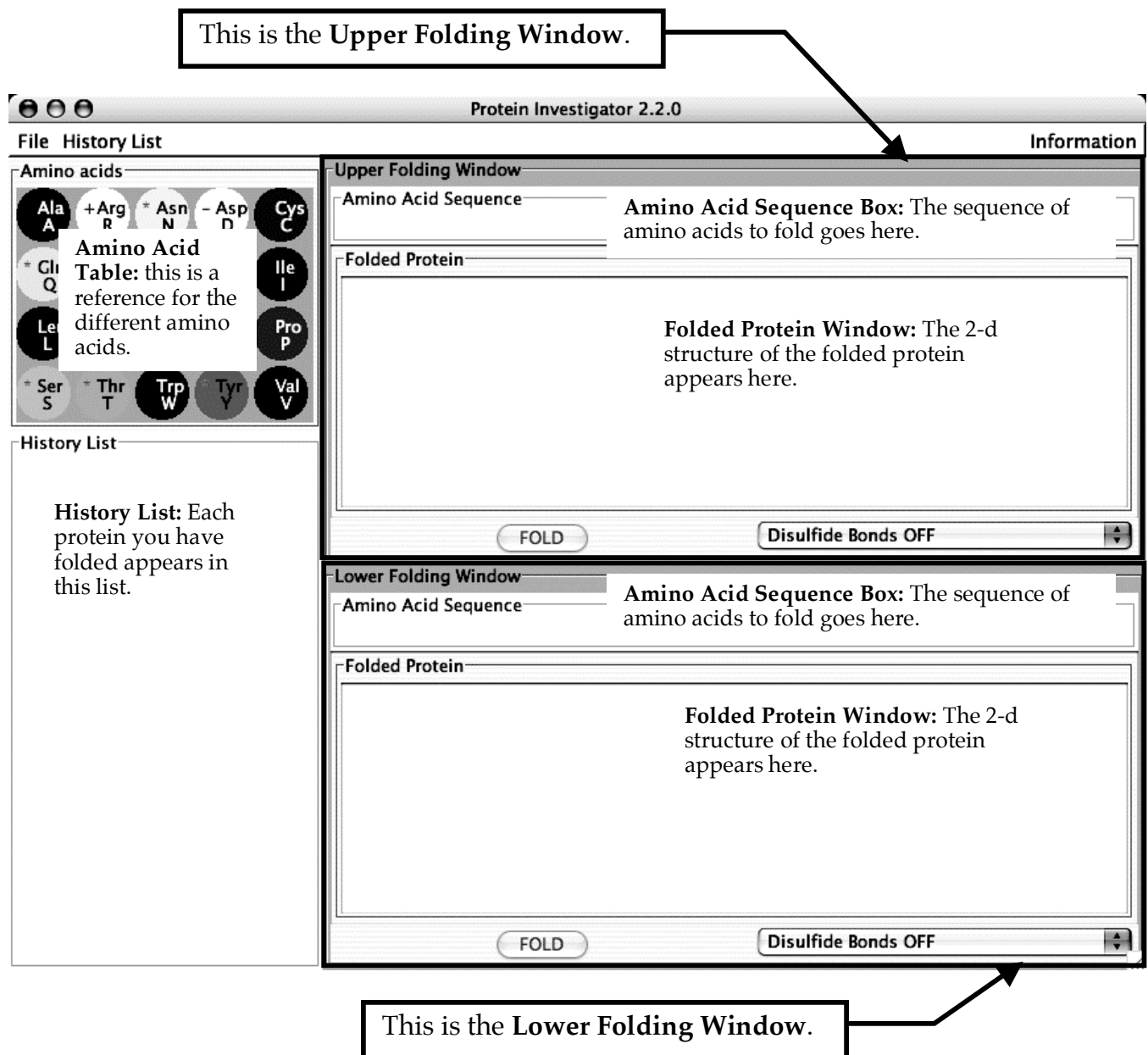
One other note: this software is under development. Please be patient and gentle with it and report any bugs and/or suggestions to your TA. You may also want to save your History List from time to time to protect against crashes (see page 6 for how to do this).

Procedure

1) Start the Protein Investigator. Click on the ProtInv icon in the Dock. It looks like this:



2) ProtInv should start. You should click in the lower right corner and drag the window so that it fills the screen. You will see this:



Using ProtInv

There are some important parts of the program that you should know about:

- **Menus:** there are three menus at the top of the ProtInv window:
 - **File:** this has two items:
 - **Print:** this prints the amino acid sequences and structures of the proteins in the upper and lower Folding Windows.
 - **Quit:** this quits the program.

- **Menus, continued:**
 - **History List:** this has five items related to the History List:
 - **Save:** this saves the current History List to a .histlist file. You will need to provide a name for this file.
 - **Save As...:** this saves the current History List to a .histlist file you name.
 - **Load:** this clears the History List and replaces it with the contents of a .histlist file you have selected.
 - **Clear:** this deletes all the entries in the History List. This is not un-doable.
 - **Save as Web Page...** This saves all the entries in the **History List** as a set of files in a folder you name; this set of files can be viewed as a web page with a picture and the amino acid sequence of each item.
 - **Information:** this has 2 items:
 - **Help:** the help file with instructions on how to use ProtInv.
 - **About:** some information from the authors of ProtInv.
- **History List Items:** if you double-click on one of the proteins in the **History List**, you will get a menu with these choices:
 - **Send to Upper Panel:** this moves the selected protein from the **History List** to the **Upper Folding Window** so that you can examine it in detail, edit it, etc.
 - **Send to Lower Panel:** this moves the selected protein from the **History List** to the **Lower Folding Window** so that you can examine it in detail, edit it, etc.
 - **Add Notes...:** when you leave the cursor over a protein in the **History List**, a window pops up showing its amino acid sequence; this item allows you to edit this information and add notes.
 - **Delete from History List:** this removes the item from the **History List**; it is not un-doable.
- **Disulfide Bond Controls:** each **Folding Window** has a menu to select whether or not disulfide bonds will form or not. Disulfide bonds form between the side chains of cysteines only under certain conditions. Typically, they do not form inside cells and only form in fluids like blood plasma, etc. These disulfide bonds are covalent bonds and are typically stronger than ionic bonds; if present, they are shown in yellow on the folded protein. We will not be using these in Bio 111; you should always leave **Disulfide Bonds OFF**. (You can play with this feature if you are curious).

Note: this program is still under development and it may tend to crash. When it does, you will lose all the proteins in your **History List**. To protect yourself against this, you should:

- take notes on the proteins you've folded (write down the amino acid sequences and sketch their shapes)
- regularly save your history list to a file on the desktop; here's how:
 1. Choose **Save** from the **History List** menu.
 2. Select the Desktop from the list of files/folders that appears.
 3. Type in a descriptive name for your history list. If you forget to add the .histlist extension, the program will add it for you.
 4. Click "Save". Your **History List** will now be saved to this file whenever you select **Save** from the **History List** menu.
- you can re-load a saved **History List** like this:
 1. Choose **Load** from the **History List** menu.
 2. Select the Desktop (or wherever you saved the .histlist file)
 3. Select your .histlist file.
 4. Click "Open". The **History List** will now be replaced by the contents of the file you selected.

In most of the next parts, we will ask you to predict what you think will happen before you actually try it. Although you can easily cheat and skip this step, we *strongly* suggest that you take the time to make a prediction. This will give you a better chance to apply and test your knowledge and will prepare you better for the exams on this material.

The One-Letter Amino Acid Code

So far in Bio 111, we have been using the full names of amino acids as well as three-letter abbreviations. There is an even shorter abbreviation for amino acids that uses only one letter for each. Because of its simplicity, you will use this code to *type* protein sequences into ProtInv. This one letter code is used in more advanced biology classes, so it is worth familiarizing yourself with it. Because several amino acids start with the same letter, some of the single letter codes are a little obscure; the mnemonics in the table below will help you learn them. In ProtInv, you *type* the single letter code but the program *shows* you the three-letter code for easier reading. You can use this table or the **Amino Acid Table** in ProtInv for reference.

<u>Amino Acid</u>	<u>3-letter code</u>	<u>1-letter code</u>	<u>Mnemonic</u>
Alanine	Ala	A	<u>A</u> lanine
Arginine	Arg	R	a <u>R</u> ginine
Asparagine	Asn	N	asparagi <u>N</u> e
Aspartic acid	Asp	D	aspar <u>D</u> ic acid
Cystine	Cys	C	<u>C</u> ystine
Glutamine	Gln	Q	<u>Q</u> -tamine
Glutamic Acid	Glu	E	glu-t <u>E</u> -amic acid
Glycine	Gly	G	<u>G</u> lycine
Histidine	His	H	<u>H</u> istidine
Isoleucine	Ile	I	<u>I</u> soleucine
Leucine	Leu	L	<u>L</u> eucine
Lysine	Lys	K	lysin <u>K</u>
Methionine	Met	M	<u>M</u> ethionine
Phenylalanine	Phe	F	<u>F</u> enylalanine
Proline	Pro	P	<u>P</u> roline
Serine	Ser	S	<u>S</u> erine
Threonine	Thr	T	<u>T</u> hreonine
Tryptophan	Trp	W	t <u>W</u> ptophan
Tyrosine	Tyr	Y	t <u>Y</u> rosine
Valine	Val	V	<u>V</u> aline

The Amino Acid Table

In addition to showing the one-letter and three-letter codes for each amino acid, the Amino Acid Table in ProtInv shows more information about each amino acid:

- The circles are colored to indicate the relative hydrophobicities of the side chains. The most hydrophilic are white; the most hydrophobic are black; gray amino acids are intermediate.
- Amino acids with (+)-charged side chains have a small blue "+" sign on their symbol; those with (-)-charged side chains have a small red "-" sign.
- Amino acids with uncharged side chains that can form hydrogen bonds have a "*" on their symbol.

3) Consider the protein arg-arg-arg-arg.

a) Based on the simplifications described previously, predict how this protein would fold. Sketch your predicted folded protein below.

b) Construct the protein.

- Click in the **Amino Acid Sequence Box** of the **Upper Folding Window** to activate it. You should see a blinking cursor appear.
- Type "RRRR" (without quotes). This is the single-letter code for four arginines.
- You should see "Arg Arg Arg Arg" in the **Amino Acid Sequence Box**. Note that ProtInv shows the three letter code for easy reading. Note also that the border of the **Upper Folding Window** turns pink - this indicates that the protein sequence in the **Sequence Box** has not yet been folded and therefore, the sequences in the **Sequence Box** and the **Folding Window** are no longer the same.
- Click the "**FOLD**" button at the bottom of the **Upper Folding Window**. Wait a few seconds for the computer to do the calculations.
- You will see the folded protein in the **Folded Protein Window**. You will also see a smaller version of the protein appear in a little box in the **History List**. The border of the **Upper Folding Window** will turn gray to indicate that the sequences in the **Sequence Box** and **Folded Protein** are now the same.

c) Does the actual folded shape match your prediction? If not, make sure you understand why it didn't and draw it correctly.

4) Consider the protein phe-phe-phe-phe.

a) Based on the simplifications described previously, predict how this protein would fold. Sketch your predicted folded protein below.

b) Construct the protein.

- Click in the **Amino Acid Sequence Box** of the **Lower Folding Window** to activate it. You should see a blinking cursor appear.
- Type "FFFF" (without quotes); this is the one-letter code for four phenylalanines.
- you should see "Phe Phe Phe Phe" in the **Amino Acid Sequence Box**.
- Click the "**FOLD**" button in the **Lower Folding Window** and wait a few seconds.

c) Does the actual folded shape match your prediction? If not, make sure you understand why it didn't, and draw it correctly.

5) Consider the protein phe-arg-arg-arg-arg-phe.

a) Based on the simplifications described previously, predict how this protein would fold. Sketch your predicted folded protein below.

b) Construct the protein.

- Click in the **Amino Acid Sequence Box** of the **Lower Folding Window** to activate it. You should see a blinking cursor appear.
- You can build this protein by selecting the entire protein sequence, hitting delete or backspace to clear the text, and then typing "FRRRRF".
- Alternatively, you can edit the existing sequence as described below:
 - Click to put the cursor anywhere in the third "Phe".
 - Hit delete or backspace twice; it should delete two of the Phe's to give you "Phe Phe" with the cursor in the middle. Notice that hitting the delete key deletes an entire amino acid (not just one letter of its name as you might expect).
 - Type "RRRR".
 - The sequence should now read "Phe Arg Arg Arg Arg Phe"
- Click "**FOLD**" and wait a few seconds.

c) Does the actual folded shape match your prediction? If not, make sure you understand why it didn't and draw it correctly.

6) Consider the protein ile-phe-met-gln-ser-arg-thr-asp-ala-ala. (IFMQSRTDAA)

a) Based on the simplifications described previously, predict how this protein would fold. Roughly sketch your predicted folded protein below.

b) Construct the protein. You should clear one of the **Polypeptide Windows** and start over.

c) Does the actual folded shape match your prediction? If not, make sure you understand why it didn't and draw it correctly.

d) Suppose a mutation in the gene for this protein caused the Arg in the middle (amino acid #6) to be replaced by a Glu. How might this effect the shape of the protein? Sketch your prediction below.

e) Make the mutant protein, either by editing the sequence (you can move the cursor to the right of any amino acid to delete it with the backspace key and then type in the new one) or by clearing and starting over. The mutant protein is: ile-phe-met-gln-ser-glu-thr-asp-ala-ala (IFMQSETDAA). Click "FOLD". Does the actual folded shape match your prediction? If not, make sure that you understand why it didn't.

You can easily compare two different proteins by clicking on the inactive **Folding Window** then double-clicking on the protein you want to see in the **History List**. It will appear in the **Folding Window** that you just activated.

f) Consider a different mutation, one that causes the same Arg (#6) to be replaced by an Ile. Follow the same procedure as for the mutation in part (e). The mutant protein is: ile-phe-met-gln-ser-ile-thr-asp-ala-ala (IFMQSITDAA). You may find it useful to have the normal protein in one **Folding Window** and the mutant protein in the other **Folding Window** for comparison purposes (select it in the **History List** and click one of the ">Upper" or ">Lower" buttons as appropriate).

g) Starting from the original protein from step (6a), make a mutation that has **little or no effect** on the shape of the protein. Explain why this mutation has such a small effect.

h) Starting from the original protein from step (6a), make a mutation that has **a large effect** on the shape of the protein. Explain why this mutation has such a large effect.

7) Design a protein where an **ionic bond** is critical its shape. Demonstrate the importance of this ionic bond by mutating one of the amino acids. Have your TA check off that you have shown and explained the role of this ionic bond in your protein. Write the amino acid sequences below:

Normal Protein_____

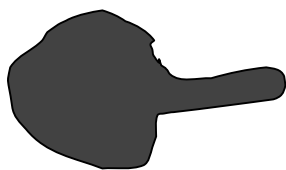
Mutant Protein_____

8) Design a protein where a **hydrogen bond** is critical its shape. Demonstrate the importance of this ionic bond by mutating one of the amino acids. Have your TA check off that you have shown and explained the role of this hydrogen bond in your protein. Write the amino acid sequences below:

Normal Protein_____

Mutant Protein_____

9) Design and construct a protein of at least 6 amino acids that has approximately the following shape:



Amino acid sequence _____

You will need to explain to your TA why your protein has this shape for a check-off.

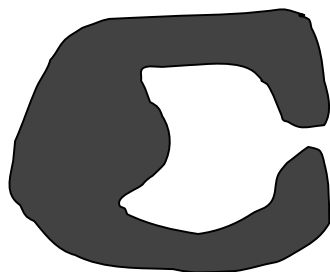
10) Design and construct a protein of at least 6 amino acids that has approximately the following shape:



Amino acid sequence _____

You will need to explain to your TA why your protein has this shape for a check-off.

11) Design and construct a protein with approximately the following shape:



These ends must touch and be joined by a non-covalent interaction, not a covalent bond.

Amino acid sequence _____

You will need to explain to your TA why your protein has this shape for a check-off.

Lab Report

There is no lab report for this lab.