ActionBioscience.org lesson

To accompany the article by T. Ryan Gregory:
"Genomic Puzzles Old and New" (August 2006)
http://www.actionbioscience.org/genomic/gregory.html

Bioinformatics (February 2007)

By Brendan Kelley, CSIP Graduate Fellow, Cornell University and Irka Elsevier, AP Biology Teacher, Penn Yan High School

Grades & Levels:
High school (honors/AP) – undergraduate (year 1)

Time Recommendations:

- One class period and/or assignment for article review and discussion questions
- Depending on depth and intensity of work on the activities in the handouts, 1 to 6 full periods (~50 minutes each), allowing for extensive computer use by individual students or groups.

NSES (USA) Content Standards, 9–12:

- NSES 1.2. Unifying Concepts and Processes: Evidence, models, and explanation
- NSES 1.3. Unifying Concepts and Processes: Change, constancy, and measurement
- NSES 2.1. Science as Inquiry: Abilities necessary to do scientific inquiry
- NSES 2.2. Science as Inquiry: Understanding about scientific inquiry
- NSES 6.2. Science and Technology: Understanding about science and technology
- NSES 8.2. History and Nature of Science: Nature of scientific knowledge

Note: View the NSES content standards on this site to choose other curricular applications at:
www.actionbioscience.org/educators/correlationcharts.html

Lesson Objectives: Students will…

- discuss the C-value and G-value enigmas in genomics
- distinguish between nucleotide (gene) and amino acid (protein) sequences
- understand the significance of gene sequence to the resulting protein and the impacts of various types of gene mutations
- gain skill in the use of publicly available bioinformatics programs on the internet and explain the meaning of each program's output
- understand the biological relevance of bioinformatic analysis and be confident in making inferences or defending conclusions based on such analyses
- apply the above technical skills and understanding of biology in an original sequence analysis project devised by the student
- record all intermediate and final results of analyses accordingly and explain the importance of documentation in science
- scrutinize the scientific results of peers and value collaboration

Key Words include: *Alu* sequences, amino acid, bioinformatics, chromosome, complexity, DNA, eukaryote, genes, gene mutation, genome, hypothesis, intron, nucleotide, peptide, phenotype, phylogeny, protein, RNA, taxonomy, transposons

Preparation

Note: Educators may want to read the article by Kathleen M. Gabric.: "Bioinformatics in the Biology Classroom" in preparation for the lesson.
http://www.actionbioscience.org/education/gabric.html

Article Discussion: Several approaches are possible for the article discussion questions below:

- Have students read the article on their own or in groups.
- Give students copies of the questions and have them do the reading and complete the content questions on their own, perhaps as a short-answer writing assignment. Have them discuss their answers and the more complex questions either as a large group or in small groups.

Handouts:

- Refer students to “useful links for student research” at the end of the Gregory article. These links help students with their activities and provide a source for research information.
- See page 3 for a teacher's guide to using the Bioinformatics handouts.

For Educators: Article Discussion

About the article by T. Ryan Gregory.: "Genomic Puzzles Old and New"
http://www.actionbioscience.org/genomic/gregory.html

Article Content Questions

1. What implications does the author see for the notion of a "human genome"?
2. What is one difference between human and chimp genomes?
3. What was the original thinking for the DNA Constancy hypothesis?
4. What paradox did it create? Explain this paradox.
5. What solution did scientists find for this C-value paradox?
6. How does the author distinguish an enigma from a paradox?
7. What are some of the questions posed by the C-value enigma?
8. What are the notable findings to date about non-coding DNA?
9. How can DNA be added or removed from a genome?
10. What is the G-value enigma? How is it the same or different from the C-value enigma?
11. Sum up why gene size and gene numbers do not necessarily indicate complexity.

Article Extension Questions

1. What do you think about the genome size data, for example that the human genome is smaller in size than that of a protozoa? Speculate why this is so?
2. How do you feel that the salamander has 26 times more DNA than humans?
3. What do you know about the Human Genome Project?
4. Review the elements in a copy of the human genome listed in the article (as provided by the International Human Genome Sequencing Consortium). Discuss each element.
5. Discuss the view by W. Ford Doolittle that our genomes “might be ironically viewed as vehicles for the replication of *Alu* sequences.”
6. What concluding message does the author give for anyone doing genomic research? Do you agree? Why or why not?

Bioinformatics
Teacher's Guide

General Purpose: Students will use inquiry skills to make and test predictions about genes and their corresponding proteins, understand the use of bioinformatics programs, and pursue their own studies of genes and proteins of interest to them.

Overview: A working draft of the human genome was released in 2000. Numerous genome sequencing projects are currently underway, so biologists need powerful tools to manage and interpret the vast data sets that these studies produce. In this lesson, students will learn to use some basic tools for gene and protein sequence analysis, consider the biological concepts underlying the programs, and work to analyze their personally chosen sequences. By integrating human ingenuity and reasoning with computational analysis, students can begin to answer some of the following genome-scale questions…and more:

- How can we identify and annotate or describe the protein-coding sequences apart from the rest of the DNA in an organism's genome?
- Is the sequence of a newly discovered gene similar to that of another gene that is better understood, and can we use that information as an experimental starting point?
- If a new gene is unlike any previously studied genes, does the protein that it encodes have identifiable characteristics that give clues as to its possible function? For example, signal peptides (for protein secretion/trafficking), post-translational modifications, and functional motifs can be predicted by various algorithms, given only an amino acid sequence (explore the ExPASy website listed below).

Tentative Scheduling

We strongly recommend that students be familiar with basic molecular biology **before working on this lesson. If all students have not had previous course work in these concepts, then introduce the module during or after presenting these topics in your own curriculum.

1-period Introduction: To expose students to basic bioinformatic tools in the context of problem solving, use **Handout I** alone as a "cook-book" activity. This worksheet will help students gain technical proficiency with the websites and prediction programs. Students will also see how these tools can be used to identify and answer a relevant question that is posed in the worksheet.

3-period Sequence: For motivated students with a strong understanding of basic molecular biology, use **Handout I** and **Handout 2** during two sequential class periods and initiate students' original explorations during the third period.

4- to 6-period Sequences: Extend the 3-period sequence to address students' needs and progress with **Bioinformatics** handouts. Incorporate class discussions and brief student presentations, then initiate students' analyses of their own genes/proteins of interest.

Learning Objectives (handouts)

1) To distinguish between nucleotide (gene) and amino acid (protein) sequences, and understand that genes and proteins have abbreviated names for communication in science.
2) To understand the significance of gene sequence to the resulting protein, and appreciate the impacts of various types of gene mutations.

3) To gain skill in the use of publicly available bioinformatics programs on the internet and explain the meaning of each program's output. For example, SignalP and BLAST give results in terms of probability, so what do the numerical and graphical results mean?
4) To understand the biological relevance of bioinformatic analysis, and be confident in making inferences or defending conclusions based on such analyses. For example, "This protein is highly similar to one that performs a certain function in (organism 1), so it might be involved in a similar process in (organism 2)," and "The probability that this protein has a signal peptide is high/low, so it is likely that it is/is not secreted outside the cell. This protein localization is/is not consistent with the function(s) of similar proteins."
5) To apply the above technical skills and understanding of biology in an original sequence analysis project devised by the student.
6) To record all intermediate and final results of analyses accordingly and explain the importance of documentation in science.
7) To scrutinize the scientific results of peers, and to value collaboration.

Background

What does it mean that biology is entrenched in a genomics era? It means that many new computational tools are being developed and employed to analyze gene and protein sequence data. Bioinformatic tools can facilitate lab-based experiments, which in turn validate or challenge the initial computational analyses. It is important to recognize that bioinformatic tools can only **predict** results (e.g. true starting point of a protein-coding sequence, characteristics or function of a mature protein, how closely related two protein sequences are). Lab experiments can therefore be used to improve the bioinformatic programs that make such predictions. Although these tools cannot stand alone in advancing scientific understanding, they can be a powerful means of generating new hypotheses and efficiently informing new experiments.

Numerous sequence databases and bioinformatic tools are freely available on the internet. Thus, it allows anyone with internet access to pursue their own questions on molecular biology, protein function, evolution, etc. virtually without limit. With the availability of such bioinformatics tools, it sets the stage for students to apply what they know, learn new concepts and techniques, and perform real research.

Equipment Requirements: In order to access the program servers described in this module, be sure that your school has access to internet-ready computers (the faster the better) with browsers such as Firefox and Explorer. Also, some school internet setups prohibit browsing to certain types of web pages, so check in advance that your machines can load the following three pages:
http://www.ncbi.nlm.nih.gov (National Center for Biotechnology Information)
http://us.expasy.org/tools/dna.html (translator, Expert Protein Analysis System proteomics)
http://www.cbs.dtu.dk/services/SignalP (SignalP, Center for Biological Sequence Analysis)

**additional sites of interest that might be useful in an extended project:
http://us.expasy.org (ExPASy link index page, Expert Protein Analysis System proteomics)
http://myhits.isb-sib.ch/cgi-bin/motif_scan (protein motif scan algorithm)
http://www.jgi.doe.gov/ (Joint Genome Initiative database)
http://sgn.cornell.edu/ (Solanaceae Genomics Network)
http://pfgd.org/ (*Phytophthora* Functional Genomics Database)

To encourage students to document their work as they go, and to make assessment easier, check to see that the computers have word processing programs that can be used to compile and print sequence data and subsequent analyses.

Teaching Tips

1) Before the "Bioinformatics" lesson: Although inexperienced students will likely be able to perform the initial tasks on the handouts, they will get the most benefit out of this activity if they already have a basic to advanced understanding of molecular biology and its relationship with cellular and organismal biology. Ahead of time, reinforce the concepts of what a gene is, how it is transcribed and translated, and what happens to the nascent protein if it is trafficked through the endoplasmic reticulum and Golgi network. Also consider discussing examples of genes and proteins that are important in current events or otherwise relevant to students' lives, and how these molecules are being studied. Review the scientific method with students and emphasize the development of a question in order to sharpen the hypothesis about a phenomenon.

2) During the "Bioinformatics" lesson: Instruct students on keeping a thorough record of what they do in their own words. By rewording the handout instructions, students will synthesize the process and be better equipped to consider the biology behind it. Additionally, print-outs of sequences (gene, protein, alignments) and graphical output might be useful to you and your students. Since students will be working mostly at computers, be sure that they stay focused, and arrange them into self-monitoring groups if this would help them to stay on task. Once students have completed the structured work that is outlined on the worksheets, gauge their interest and motivation to choose their own genes and proteins to study, and proceed accordingly.

3) Evaluation Strategy: Students' documentation of their process, along with completion of worksheet questions, will provide an individual measure of comprehension and performance. If students work in groups, consider a way to determine individual knowledge and to maintain accountability for the students' work. Between activities in handouts 1 and 2, it has been informative to teachers to give students the chance to present drawings, text, and flowcharts describing their understanding of the module to their peers. This public speaking experience has multiple benefits to the students in addition to giving them incentive to perform well in front of the class. For example, they can productively critique each other's work and raise new questions for future study. Assign students to read and summarize the online descriptions of the tools that they are using (BLAST, ExPASy, SignalP). Final assessment of students' original analyses can be done using oral or written assignments, poster presentations, or preparing students to mentor another class in what they themselves have learned through this project.

Bioinformatics
Teacher's Guide: Sample Responses to "Bioinformatics II" Questions

Problem Set 1:

*Students can mark triplets/codons using vertical bars

```
                mRNA #1:AUG| ACC| CAC| AGG| UCA| GAC| GCA| UAC| UAA
5'3' Frame 1 Transl.: START    T    H    R    S    D    A    Y   STOP


                mRNA #2:AUG| A AC|C CA|C AG|G UC|A GA|C GC|A UA|C UA|A
5'3' Frame 1  Transl.: START   N    P    Q    V    R    R    I   L
Best Translation #1 Frame:
```

*Reading the codons in frame 2 will give the original "THRSDAY" translation that is disrupted by the **A** insertion.*

Question 1a – If the first mRNA is correct and encodes a peptide (small protein) that functions normally in a cell, what do you think would happen if any of the other mRNAs were translated instead in 5'3' Frame 1?

If the first mRNA is correct and encodes a small protein that functions properly in cells, then the proteins that are encoded by the other (mutated) mRNAs might not have the same function and might harm the cells in which they are synthesized. Those mRNAs that lack correct STOP codons will result in a "read-through" past the original stop site (UAA) and will probably make unexpectedly large proteins. Other mRNAs might have inappropriately early STOP codons in the translation, which result in shorter protein sequences than expected for the correct mRNA.

Question 1b – How are the other mRNAs and translations different from the first example? What would you call these differences? How else could you make variations on the first mRNA? Write out some predicted mRNAs and record your translations. Do all of your variations change the original translation?

The gene that encodes the three mRNAs has different point mutations at the same position in its DNA (2, insertion; 3, deletion; 4, substitution). Other variations on the gene for the first mRNA include sequence inversions and nucleotide substitutions that do not change the encoded amino acid (silent mutation).

Question 1c – Explain why changing a single amino acid or creating an inappropriate stop codon in the middle of a protein might affect the protein's function. How might this happen in nature? Which of these alterations do you think would have a more profound effect on the protein's function and why?

If a normally functioning protein is "cut in half" by a gene mutation that introduces a new stop codon, then maybe this protein will stop performing its normal function or perform it in a different way (increased/decreased activity, etc.). Changing a single amino acid might seem trivial to the function of a protein, but a mutation in the hemoglobin gene that causes a single amino acid change (glutamic acid to valine) results in sickle cell anemia. Other examples include the mutations that are involved in promoting cancer: loss of function of tumor-suppressor proteins, gain of function of onco-proteins. There is no "right" answer to this question, so it would be a good topic for class debate or discussion.

Question 2b – What process do you think SLE18 is involved in and why do you think there are so many BLAST hits matching it?

The SLE18 protein sequence matches "pathogenesis-related protein P2" from Lycopersicon esculentum *(tomato), so it is probably involved in the plant's defense system against invading pathogens. The BLAST result lists highly similar proteins in tobacco, pepper, elderberry, grape, and pea, among other plants, so these proteins probably have similar or conserved function to that of SlE18 / P2.*

Question 2c – Perform the same analysis for "PIE19" as you did for SLE18 and answer the questions posed in 2a. What organism contains all of the best BLAST hits for PIE19? Are there any hits that don't fit with the others?

The PiE19 translation matches "phytotoxin-like SCR74" from Phytophthora infestans *and other closely related proteins from the same organism. Low-probability BLAST hits at the end of the list include proteins from dog and mouse. Such unrelated proteins seem to be unrelated to the rest of the SCR74 hits. In addition, their high E-values indicate that these hits are more likely to come up due to chance than due to an actual biological similarity. These weak or low-probability BLAST hits demonstrate the fact that computers are very good at using mathematical algorithms to make predictions, but they are still not as good as humans are at making conceptual distinctions.*

This material was developed through the **Cornell Science Inquiry Partnership** program (http://csip.cornell.edu), with support from the National Science Foundation's Graduate Teaching Fellows in K-12 Education (GK-12) program (DGE # 0231913 and # 9979516) and Cornell University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Name, Class Period:

Date:

Computer programs are important tools for making predictions based on observations of the natural world. For example, meteorologists use programs to predict weather based on temperature, pressure, wind direction/speed and other factors. In the same way, biologists use programs to help them make predictions based on information that they gather on topics such as animal communication, blood flow in the human brain, plant cell wall properties, or structure and function of proteins.

In this activity, you will learn to use internet-based programs to **extract** gene (nucleotide) sequence data from a public database; **translate** gene sequences into their corresponding protein (amino acid) sequences; **predict** whether these proteins stay inside the cell in which they are made or are shuttled (secreted) outside the cell; and **identify** similar protein sequences to the one you have selected.

Problem: Your friend is reading the list of ingredients on her snack food wrapper and she doesn't know what some of them are, especially a protein called **ara h2**. She knows that it's a protein (gene name is ara h2). Because of her allergy to peanuts, she must be careful about what she eats. Your task is to learn more about this protein and give your friend advice on whether her snack would be safe for her to eat!

Sequence Extraction: Open a new internet browser window (e.g. Firefox, Explorer) and type http://www.ncbi.nlm.nih.gov/ to go to the website for the National Center for Biotechnology Information (**NCBI**). At the top of this page, use the pull-down menu next to "Search," select "Nucleotide" and type the gene name **ara h2** in the box labeled "for." Wow! You brought up 54 possible hits, but scroll down to the fifth one that begins with "Arachis hypogaea" and open the AY158467 link. This new page contains the cDNA coding sequence for the ara h2 protein at the bottom, below "Origin," so click and drag the cursor to highlight the entire sequence from "1atggc…" to "…tactaa." Now, right click the highlighted sequence and select "copy" to store it.

DNA-Amino Acid Translation: Leave your first window open, but start a new internet window and type http://us.expasy.org/tools/dna.html to open the translation tool. Right click the cursor in the box below "Please enter DNA…" and select "paste" to enter your ara h2 gene sequence. To the right of "Output format," select "Includes nucleotide sequence" from the pull-down menu and click "Translate Sequence." Now your results in 5'3' Frame 1 should show the amino acid sequence of the ara h2 protein in capital letters below the corresponding nucleotides of the gene (read from left to right, top to bottom). Notice that the gene starts with "atg" and the corresponding amino acid below this triplet/codon is "M" for methionine. Later, you will learn that all protein-coding DNA sequences begin with "atg" as a start codon or initiator methionine. Similarly, all protein-coding DNA sequences end with a "taa" or "tag" stop codon. How many additional atg/M codons can you find in the ara h2 sequence?_____Now, click on 5'3' Frame 1 and click the link for the initiator methionine that you just discovered (how many other "M" links do you see?). Just as you highlighted the ara h2 gene sequence, highlight the ara h2 protein sequence from "MAKL…" to "…RY" and right click to "copy" the sequence.

Signal Peptide Prediction, Inside or Outside?: Leaving the first two windows open, start a new window and type http://www.cbs.dtu.dk/services/SignalP/ to open the SignalP tool. "Paste" your ara h2 protein sequence in the box under "Submission" and click "Submit" below. The graphical output should show a series of lines and peaks, with the tallest blue peak representing the most likely site for a signal peptide to be cut off from the end of the protein. Under the graph, what is the "mean D" value?_____If a mean D value is close to 1.0, then the protein probably has a signal peptide that directs it through the endoplasmic reticulum, Golgi complex and into a vesicle for secretion outside the cell where it was synthesized. Do you think that ara h2 is secreted? What does the SignalP program predict?

Identifying Similar Proteins, BLASTing! Sequences: Now that you have some information about the ara h2 protein sequence, you can open a fourth internet window (meanwhile, your friend is getting hungry and wants to know if she can eat this snack with ara h2 in it). Type http://www.ncbi.nlm.nih.gov/BLAST/ and read the BLAST explanation at the top. After reading this summary, what does BLAST mean to you? Now, click "Protein-protein BLAST (BLASTp)" under "Protein" and right click to "paste" your ara h2 protein sequence into the box next to "Search." If you can't paste your protein sequence, go back to your Translation window, select and copy the sequence again, and paste it into the box in the new BLAST window. With your sequence entered, click "BLAST!" and click "Format!" in the next page. You have just asked the BLAST program to search the entire NCBI protein database for matches to your ara h2 sequence…amazing! The BLAST results page can be overwhelming, but the color-coded graph at the top shows the most similar sequences in red and other sequences that are less similar in magenta, green, blue and black. Now, scroll down the list to see the alignments of your ara h2 sequence ("Query") with each sequence from the database ("Subject"). At positions where the two sequences have an identical amino acid, that letter appears between the two lines, but less similar matches have more spaces where there are fewer identical amino acids…look towards the bottom of the list to see these less similar alignments.

Finally, go back to the very top of the list just below the graph and select the third link for "allergen Ara h 2 isoform [Arachis hypogaea]" by clicking gi|31322017|gb|AAM78596.1|. Under authors "Becker, W.-M." and others, read the title describing the protein sequence entry. What does it say? Since your friend is still waiting to eat her food that contains the **ara h2** protein, what question would you ask her to determine if this protein is safe for her to eat? (hint…allergens can cause allergic reactions in some people who are sensitive to them)

(Optional)

Try going through the steps above by extracting any of the following nucleotide sequences from the NCBI database: SLE3, PIE1, PIE16 or PIE18

**Note: When you translate these sequences, be sure to inspect the protein translations for an initiator methionine "M" followed by the longest amino acid chain that lacks internal stop codons "—." The correct translation won't necessarily be in 5'3' Frame 1, so check carefully.

This material was developed through the **Cornell Science Inquiry Partnership** program (http://csip.cornell.edu), with support from the National Science Foundation's Graduate Teaching Fellows in K-12 Education (GK-12) program (DGE # 0231913 and # 9979516) and Cornell University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

Student Handout: Bioinformatics II
Understanding The Biology Behind The Programs

Name, Class Period:

Date:

In **Bioinformatics I**, you practiced with some of the tools that researchers use to analyze DNA (nucleotide) and protein (amino acid) sequences. Sometimes, computer-based research tools can be very tedious, but they can also be very powerful in helping you address specific biological questions and hypotheses. For example, you were given the gene name ara h2 for the corresponding protein, and you were able to identify it as a peanut allergen as well as predict that it is a protein that is synthesized inside cells of the peanut plant and then secreted from them. What genes and proteins interest you? Think about that question during this activity, and consider the following concepts as you go through it: Reading frame, types of mutations, protein primary sequence.

1) Nucleotide triplets in mRNA encode amino acids that make up a peptide chain or protein. Recall your translation skills, type the hypothetical mRNAs below into the tool at http://us.expasy.org/tools/dna.html and remember to choose "Includes nucleotide sequence" for your output format. Record the 5'3' Frame 1 translation (substitute START for "M" and STOP for dashes) for each mRNA and note the "frame" that **reads almost the same as the first translation**.

mRNA #1: AUG ACC CAC AGG UCA GAC GCA UAC UAA
5'3' Frame 1 Translation:

mRNA #2: AUG **A** ACC CAC AGG UCA GAC GCA UAC UAA
5'3' Frame 1 Translation:
Best Translation #1 Frame:

mRNA #3: AUG CC CAC AGG UCA GAC GCA UAC UAA
5'3' Frame 1 Translation:
Best Translation #1 Frame:

mRNA #4: AUG **C**CC CAC AGG UCA GAC GCA UAC UAA
5'3' Frame 1 Translation:
Best Translation #1 Frame:

Bioinformatics II cont'd…

Answer questions on reverse or on a separate sheet, as needed:

1a) If the first mRNA is correct and encodes a peptide (small protein) that functions normally in a cell, what do you think would happen if any of the other mRNAs were translated instead in 5'3' Frame 1?

1b) How are the other mRNAs and translations different from the first example? What would you call these differences? How else could you make variations on the first mRNA? Write out some predicted mRNAs and record your translations. Do all of your variations change the original translation?

1c) Explain why changing a single amino acid or creating an inappropriate stop codon in the middle of a protein might affect the protein's function. How might this happen in nature? Which of these alterations do you think would have a more profound effect on the protein's function and why?

2) If you identify a new protein and you don't know what its biological function is, BLAST searching in the NCBI database for similar proteins that are better understood can give you some important clues about your protein of interest. Since some organisms are more thoroughly studied than others, we can sometimes make inferences about their functional similarities when we compare genes or proteins at the nucleotide or amino acid (primary structure) sequence level.

Refer to your **Bioinformatics I** sheet to help you extract the cDNA for "SLE18" then **translate** it to a protein sequence and select the correct reading frame (longest amino acid chain that begins with an initiator methionine "M" and lacks internal stop codons "—"). With the correct SLE18 protein sequence, perform a BLASTp search to identify similar proteins in the NCBI database.
**Remember to use separate browser windows for each action (extraction, translation, BLAST) so that you can go back to them if you have difficulties.

Look at the list of "sequences producing significant alignments" with SLE18 and scroll down to the section labeled "alignments." For each alignment, if your SLE18 translation is the **query** and the **subject** is another protein being compared to it, consider how well the aligned sequences match and what this might mean.

2a) When you click on the link for the best result (first hit) to bring up its database entry, what is the name of the organism (binomial and common) that makes the protein? Is this the same organism as for SLE18? Do you think SLE18 and this first hit are the same protein? Why?

2b) What is the definition/name of the protein in this first hit? Scroll down the list and select three other BLAST hits, each from a different organism, and record the binomial/common names and corresponding protein names. What process do you think SLE18 is involved in? Why do you think there are so many BLAST hits matching it?

2c) Perform the same analysis for "PIE19" as you did for SLE18 and answer the questions posed in 2a. What organism contains all of the best BLAST hits for PIE19? Are there any hits that don't fit with the others?

2d) What protein have you decided to study using these bioinformatics tools? Why? What do you know about it already and what do you predict about its subcellular localization (internal, membrane-associated, secreted) or its similarity to other proteins?

This material was developed through the **Cornell Science Inquiry Partnership** program (http://csip.cornell.edu), with support from the National Science Foundation's Graduate Teaching Fellows in K-12 Education (GK-12) program (DGE # 0231913 and # 9979516) and Cornell University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.