

GENERAL BIOLOGY LABORATORY EXERCISE

Amino Acid Sequence Analysis of Cytochrome C in Bacteria and Eukarya Using Bioinformatics

INTRODUCTION:

All life forms undergo metabolic processes to obtain energy. The way organisms metabolize substances, however, vary depending upon their ancestry and the environmental conditions they live in. For example, many eukaryotes including humans generate energy from sugars mainly through aerobic cellular respiration. Bacteria, on the other hand, can metabolize through various mechanisms such as nitrogen fixation, photosynthesis, aerobic and anaerobic respiration. They can use different sources for energy such as single carbons, nitrogen and sulfur.

A molecule found common to bacteria and eukaryotes is the protein cytochrome C. It aids in metabolism by helping transfer electrons from one electron carrier to another. Cytochrome C consists of about 100 amino acids. In different species, the amino acid sequence for cytochrome C are similar but not identical.

In this lab, we will use bioinformatics to analyze similarities and differences in amino acid sequences for cytochrome C in two species of bacteria and four species of eukaryotes (humans, chimpanzees, bottle-nose dolphins and honey bees) to understand their ancestry and cytochrome C function.

Purpose:

1. Understand and use bioinformatics with the programs BLAST and Seaview.
2. Compare and contrast amino acid sequences of cytochrome C in humans, other animals and bacteria.
3. Relate amino acid sequences of cytochrome C from various species to genome evolution and protein function.

Hypotheses:

1. Organisms that are more closely related will share more common amino acid sequences in the cytochrome C protein.
2. Organisms with similar types of metabolism will share more common amino acid sequences in the cytochrome C protein.

MATERIALS AND METHODS:

Materials: (See appendix)

Computer with internet connection

BLAST (Basic Local Alignment Search Tool) website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Seaview program

Amino acid sequences for subjects

Description of subjects

Methods:

Exercise 1: Use BLAST to identify accuracy of given amino acid sequences. Fill in Table 1.

Video instructions: <http://youtu.be/yv7Tbo3XyOw>

1. Use the BLAST website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. Under “Basic BLAST”, choose the “protein blast” program to run because we are using amino acid sequences.
3. Enter your amino acid sequence for search
In the “enter query sequence” box on the upper left corner, copy and paste the amino acid sequence for the organism you want to test. For example, for the *Unknown*, enter:

QTAYAGDAANGMRVFNTYCSDCHSITEGKNKIGPSLWNVVGRKPASISDFNYSDAMRKNDII
WTEDRISTYITNPQGLLPGVKMAFPGLKDPQKCADVIQFLSQQH

- a. Scroll down to the bottom of the page and click “BLAST”. Wait for your results to show.
Note: Results show the species and protein with the best match at the top of list.
4. Fill in information in Table 1 using the first match shown under the “description” tab.
 - a. Fill in the species identified located in [brackets].
 - b. The protein is listed before the species.
 - c. For “fraction of AA sequences matched” and “identity value”
 - i. click on the first match.
 - ii. look at the first set of data on the top of page. Find the 4th column labeled “identities”. Here is the fraction of AA matched (e.g. 105/105) and percentage for the identity value (e.g. 100%).
5. Repeat for the remaining organisms. Check for species match, protein, fraction of AA match and identity match.

Exercise 2: Use BLAST to compare amino acid sequences for cytochrome C between species. Fill in Table 2.

Video instructions: http://youtu.be/xY_FOkSfhu0

1. Use the BLAST website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. In your protein BLAST search page under “enter query search”, click the box titled “Align two or more sequences.” A second search box will appear underneath.
3. Enter the sequence of *Homo sapiens* in the second box titled “Subject”. Enter all the sequences of the organisms that you want to compare it to in the first “Query” box. Be sure to hit return after each sequence when entering multiple organisms.
4. Click “BLAST”
5. Scroll down and find the “identities” match percentage between *Homo sapiens* and each other organism. Enter in the appropriate box in Table 2. Note: The two sequences are aligned next to each other. The differences in amino acids are identified with a blank or a “+” sign.
6. Repeat for other organisms as necessary.

Exercise 3: Use Seaview to visually align cytochrome C amino acid sequences for all 6 organisms. Answer questions based on alignment.

Video instructions: <http://youtu.be/-XfF8sJaZ-c>

1. Open Seaview.
2. Enter the sequences to be aligned in Seaview
 - a. Go to Appendix 1. Amino Acid Sequences for Species
 - b. Highlight ALL the text below the heading including the species name.
 - c. Copy (CTRL+C) and paste (CTRL+V) into the window called “alignment drag zone”. Note: you must paste by using CTRL + V on the keyboard, not with the mouse.
 - d. Click “confirm”
3. Align the sequences
 - a. Select the drop down menu called “align”
 - b. Select “align all”
 - c. Click “OK”
4. Answer corresponding questions in the results section.

RESULTS AND GUIDE QUESTIONS:

Exercise 1: Use BLAST to identify accuracy of given amino acid sequences.

Table 1. Protein Identification for Different Species

	Species	Protein	Fraction of AA sequences matched	Identity match (%)
Bacteria	1. <i>Methylothera versatilis</i>	<i>cytochrome C</i>	<i>106/106</i>	<i>100%</i>
	2. <i>Sphingomonas</i>	<i>cytochrome C</i>	<i>106/106</i>	<i>100%</i>
	3. <i>Caldithrix abyssi</i>	<u><i>cytochrome C oxidase subunit II</i></u>	<i>106/106</i>	<i>100%</i>
Eukarya	4. <i>Homo sapiens</i>	<i>cytochrome C</i>	<i>105/105</i>	<i>100%</i>
	5. <i>Pan troglodytes</i>	<i>(predicted) cytochrome C-like (predicted = based on hypothetical data)</i>	<i>105/105</i>	<i>100%</i>
	6. <i>Apis mellifera</i>	<i>Cytochrome C</i>	<i>108/108</i>	<i>100%</i>

1. Based on Species #1

a.. How many seconds did BLAST take to identify the protein for this amino acid sequence? *Faster s*

b. How many seconds did BLAST take to identify the organisms? *Slower s*
Note: exact times depends on the computer processing.

c. Was the protein or organism identified first? What does this imply about genome evolution?

Gene for cytochrome C is conserved across organisms (bacteria, eukarya). There is a common conserved region which was located quickly. It took longer time to find the species because BLAST is searching for the differences in the sequences for each species.

d. Do your given sequences correctly identify the protein and organism? Explain using the data in table.
Yes, all searches matched cytochrome C protein with 100% accuracy. The sequences also matched with the correct organism identified with 100% identity value.

Exercise 2: Use BLAST to compare amino acid sequences for cytochrome C between species.
 Fill in the table below with the “identity values” for each pair of organisms.

Table 2. Amino Acid Identity Table for Cytochrome C

Cytochrome C	<i>Homo Sapiens</i>	<i>Pan troglodytes</i>	<i>Apis mellifera</i>	<i>Sphingo- monas</i>	<i>Caldithrix abyssi</i>
<i>Homo Sapiens</i>	-	94	75	52	30
<i>Pan troglodytes (Clint)</i>		-	72	51	27
<i>Apis mellifera</i>			-	54	29
<i>Sphingomonas</i>				-	27
<i>Caldithrix abyssi</i>					-

2. a. What is the identity value that compares the two bacterial species? 27
- b. What are the identity values for 3 pairwise comparisons among the 4 eukaryotic species?
 94, 75, 72
 average of the identity values = 80
- c. Contrast the identity matches between bacteria and the average for eukaryotes.
 85 vs 27
- d. What do these results imply about the evolution (diversification) of bacterial and eukaryotic genomes?
Bacteria is more diverse. These two bacteria are only 27% match in cytochrome C. Eukaryotes on the other hand are more similar with each other as there is a high match at 85%
- e. List in order the species that have cytochrome C that is the most similar to least similar to humans. What does this result imply about our evolutionary relationships with bacteria and other eukaryotes?
Chimp, dolphin, bee, Sphingomonas, Caldithrix abyssi
We do have a shared ancestor with bacteria because there is some ID match, but this ancestor is in the distant past because there is a lower ID match.
Our shared ancestors are more recent with other Eukarya because there is a higher ID match.

- f. What amino acid(s) are different between *Homo sapiens* and *Pan troglodytes*? Do you think this difference has any effect on cytochrome C function? Briefly justify your answer.

Homo sapiens and *Pan troglodytes* have 99/105 amino acids similar, meaning that there are 6 amino acids that are different.

Homo sapiens → *Pan troglodytes*

P → A (proline in *Homo Sapiens*, alanine in *Pan Troglodytes*)

A → V (alanine to valine)

I → T (isoleucine to threonine)

Y → H (tyrosine to histidine)

G → E (glycine to glutamic acid/glutamate)

Y → F (tyrosine to phenylalanine)

These 6 differences do not likely have an effect on cytochrome C function because humans and chimps have similar metabolic processes (use oxygen in similar ways). These differences are likely nonsubstantial changes (silent mutations) that are not critical to cytochrome C function.

Exercise 3: Use Seaview to visually align cytochrome C amino acid sequences for all 6 organisms.

Comparison of Cytochrome C sequence in 5 species Using SeaView

Use your alignment of cytochrome C sequences in SeaView to answer the following questions.

1. Which domain of life looks more diverse, bacteria or eukarya? Briefly explain your answer by pointing out patterns from the data.

Bacteria look more diverse because there are less aligned amino acids between the two bacteria than between the 4 eukaryotes.

2. Which species has the sequence that is the MOST different from the other species? Suggest a reason (based on protein function) why this species' cytochrome C sequence might differ the most.

Calthidrix abyssis is the most different in amino acid sequence. This difference could be due to the different metabolic process that this bacteria uses. Instead of using oxygen as the electron acceptor, it uses nitrogen. All the eukaryotes and Sphingomonas all do aerobic respiration with oxygen.

3. a. List the name of amino acids that were identical in position in the cytochrome C sequence across all species investigated. The part of the genome that codes for these amino acids are referred to as "conserved regions" of the gene.

G (glycine) position 15, 39, 44, 47, 55

C (cysteine) position 24

H (histidine), position 28

P (proline), position 40, 81

b. What might be the function of these conserved regions for the cytochrome C protein?

All are involved in metabolism. We know cytochrome C in humans carries the electron in the electron transport chain to the final electron acceptor. These amino acids may be the electron carrier to the electron acceptor, the oxidizer (ie. oxygen, iron, sulfur)

CONCLUSION:

Evaluate the two hypotheses based on patterns derived from the data.

*The data will likely support the two hypotheses. 1. The more closely related the species, the more similar their cytochrome C sequences were. For example, humans are closest to chimps and we had 94% identity value. We are least related to *Calthidrix abyssi* and we had only a 30% identity value. 2. All the aerobic respiring organisms that use oxygen had more similar amino acid sequences for cytochrome C than *Calthidrix abyssi* which had a functionally different metabolism that used nitrogen instead of oxygen. For example, we had from 52 – 94% identity values with all the aerobic organisms, but only a 30% identity value with the anaerobic organism, *Calthidrix abyssi*.*

TOOLKIT CREDITS:

Developed by Angela Gee, (Los Angeles Trade Tech Community College, CA) and Marissa Pantastico-Caldas, Los Angeles Trade Tech Community College (CA) with data and guidance by John Kirkpatrick (University of Rhode Island, RI) and the support by the rest of the C-DEBI Collaborative Toolkit Team.

WEBSITE:

http://www.coexploration.org/C-DEBI/toolkits_biology.html