

Metabolism: Cytochrome C in Humans Compared to Other Species Using Bioinformatics

INTRODUCTION: We live in a human-centered world, but there is a whole range of life beyond what we can see or even imagine. 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 and live on land. Bacteria, on the other hand, can metabolize through various mechanisms such as nitrogen fixation, photosynthesis, aerobic and anaerobic respiration and may live in water. 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. In humans, cytochrome C is one of the proteins in the electron transport chain in the inner mitochondrial membrane that transports the electron to oxygen, the final electron acceptor. Cytochrome C consists of about 100 amino acids. In different species, the amino acid sequences 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 (2 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 & protein function.

Hypothesis:

1. If humans share common ancestry, then organisms more closely related to humans will share more common amino acid sequences in the cytochrome C protein.
2. If humans share similar metabolism function with other organisms, then these organisms will have a more common amino acid sequence in the cytochrome C protein as humans.
3. If an individual has a genetic disease with their cytochrome C protein, then this individual's amino acid sequence for cytochrome C will differ than that of a healthy individual.

MATERIALS AND METHODS:

Materials

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.

1. Use the BLAST website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and instructional video: <http://youtu.be/yv7Tbo3XyOw>
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
 - a. 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:
 MGDIEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAVGFSYTDANKNK
 GITWGEETLMEYLENPKKYIPGTKMIFAGIKKKSERADLIAYLKKATNE
 - b. 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”
1. click on the first match.
 2. 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.

1. Use the BLAST website: <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and instructional video: http://youtu.be/xY_FOkSfhu0
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 1* 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 1* 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.

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

1. Use instructional video: <http://youtu.be/-Xff8sJaZ-c>
2. Open Seaview.
3. 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”
4. Align the sequences
 - a. Select the drop down menu called “align”
 - b. Select “align all”
 - c. Click “OK”
5. Answer corresponding questions in the results section.

RESULTS:

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 values (%)
<i>Eukarya</i>	1. <i>Methylothera versatilis</i>	<i>cytochrome C</i>	106/106	100%
	<i>Homo sapiens 1: healthy</i>	<i>cytochrome C</i>	105/105	100%
	<i>Homo sapiens 2: disease</i>	<i>cytochrome C</i>	105/105	100%
	<i>Pan troglodytes</i>	<i>cytochrome C-like</i>	105/105	100%
	<i>Apis mellifera</i>	<i>cytochrome C</i>	108/108	100%
<i>Bacteria</i>	<i>Pseudomonas</i>	<i>cytochrome C</i>	110/110	100%
	<i>Caldithrix abyssi</i>	<i>cytochrome C oxidase subunit II</i>	106/106	100%

1.A. What is the range of identity values in table 1 above? *100%*

B. Do all your organisms above have the protein cytochrome C? Yes or No *Yes*

C. Based on your answers to A and B above, do your given sequences correctly identify the protein and organism? Explain using the data in table 1.

Yes. All of the proteins are supposed to be Cytochrome C and the BLAST findings confirm that this is the protein that matched with 100% to the particular amino acid sequence. These cytochrome C proteins from the given sequences also match the given organism listed in the appendix.

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

Fill in the table below with the “identity value” for each pair of organisms.

Table 2. Amino Acid Identity Table for Cytochrome C

Cytochrome C	<i>Homo Sapiens 2</i>	<i>Pan troglodytes</i>	<i>Apis mellifera</i>	<i>Pseudomonas</i>	<i>Caldithrix abyssi</i>
<i>Homo Sapiens 1</i>	100	94	75	43	30

2. A. List in order the species that have cytochrome C that is the most similar to least similar to humans.

Chimps (Pan troglodytes), bees (Apis mellifera), pseudomonas, caldithrix abyssi

B. What does this result imply about our evolutionary relationships and common ancestry with bacteria and other eukaryotes?

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.

3. Based on the identity value of *Homo sapiens 1* vs 2, do you think *Homo sapiens 2* has a genetic defect in its cytochrome C protein? Please explain.

No, the cytochrome C protein sequence is identical in both humans. There is a 100% identify value showing that there is no difference in Cytochrome C between the healthy individual and the one with a genetic defect.

4. What amino acid(s) are different between *Homo sapiens1* and *Pan troglodytes*? Do you think this difference has any effect on cytochrome C function? Briefly justify your answer.

Homo sapiens1 and Pan troglodytes have 99/105 amino acids similar, meaning that there are 6 amino acids that are different.

Homo sapiens1 → 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 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 six species Using SeaView

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

5. A. Which domain of life looks more diverse, bacteria or eukarya? *bacteria*
- B. Who is more related: the two bacteria or humans and bees? *Humans and bees*
- C. Briefly explain your answers to 5A and B by pointing out patterns from the data.

The two bacteria have amino acid sequences that don't align as well with each other as you can tell by the differences in color pattern. The All 4 eukarya sequences align better with each other than the two bacteria sequences align with one another.

6. 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. (*Hint: look at description of organisms.)

Caldythrix Abyssii has the sequence that is most different than all the other species. Calduthrix abyssii is the only anaerobic organism in this group. It uses nitrogen instead of oxygen as the final electron acceptor. This different in function can reflect the different amino acid sequence for cytochrome C.

7. A. List the name and position of the amino acids that are 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)

K = Lysine (17)

C = cysteine (24, 27)

H = histidine (29)

P = proline (40, 81)

L = leucine (74)

- 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 in these positions may contribute the electron carrier to the electron acceptor, the oxidizer (ie. oxygen, iron, sulfur), for all organisms.

DISCUSSION/CONCLUSION:

Does your data support hypothesis #2? Evaluate based on patterns derived from the data.

Yes, humans are aerobic (use oxygen) to metabolize. Therefore, the other organisms that also use aerobic metabolism should have the most similar amino acid sequences for cytochrome C. This is shown in that humans are closest to each other in metabolism function, then chimps, bees and Pseudomonas which all have aerobic metabolism. Therefore Homo sapiens I share 100%, 94%, 75% and 40% identity value respectively. Caldithrix abyssi has the least common metabolism function since it is anaerobic and indeed they share the least common sequence with humans with a 30% identity value. Therefore metabolism function (aerobic vs anaerobic) is reflected in the similarity of amino acid sequences.

Note: if group states “No” then that is acceptable as long as the data is used to back up the claim. Also, students can do all 3 hypotheses if there is time.

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) with and support by the rest of the C-DEBI Collaborative Toolkit Team.

WEBSITE:

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