**Evolution of Caffeine Biosynthesis Enzymes**

By Brian Sullivan1, Shuchismita Dutta2\*

1Department of Plant and Microbial Biology, North Carolina State University, Raleigh NJ 27695

2Institute for Quantitative Biomedicine, Rutgers University, Piscataway NJ 08854

\*contact author: sdutta@rcsb.rutgers.edu

**Preparation: The Coffee Genome**

As homework and prior to the case discussion in class, get acquainted with the case.

* Read the New York Times Science article from Sep 2014, authored by Carl Zimmer (<https://www.nytimes.com/2014/09/04/science/how-caffeine-evolved-to-help-plants-survive-and-help-people-wake-up.html>)
* Also read the abstract of the peer-reviewed scientific article mentioned in the newspaper article above (Science, 2014, 345, 1181-1184; doi:10.1126/science.1255274)

These articles set the stage for the case. The first part of this section focuses on understanding information sources, the second introduces you a biological data resource to learn about the chemical nature and use of caffeine (for humans), and finally the third section discusses why plants make caffeine.

* Review these sections and answer the questions in each of the sections.

*A. Understanding sources*

Both the NY Times and Science articles report sequencing of the coffee genome. Here is an opportunity to reflection on what can be learned from reading these articles. Answer the following questions.

Q1. What kinds of information did you learn from the NYT article vs the Science article abstract?

Q2. What is the main difference between the news report and peer-reviewed scientific article?

*B. What is caffeine?*

When you think about foods and drinks that contain caffeine perhaps you think about coffee, tea, and cocoa or chocolate. Here we will learn the chemical nature of Caffeine.

1. The newspaper article describes caffeine as a drug that is “toxic at high doses but enhancing our brains at low doses.” To learn a little more about its chemical structure, search for Caffeine in DrugBank (a curated resource that provides a wide variety of information of drugs and drug-like molecules). Look for information about this molecule in DrugBank (<https://www.drugbank.ca/drugs/DB00201>), then refer to it to answer the following questions?

Q3. What is the chemical structure of caffeine? Draw or paste a picture of this molecule below.

Q4. Describe the function(s) of Caffeine (as listed in DrugBank).

*C. Why do Plants make Caffeine?*

Worldwide humans commonly use caffeine as a stimulant. In this case study we will think about the value of caffeine from the plants’ perspective. However, have you ever thought about why plants (e.g. the Coffee, tea, cocoa, mate plants) make caffeine?

Watch a short video (<https://youtu.be/ohhMF1UlqR4>) to learn about plant growth strategies and allelopathy; review the Box-1 contents and then answer the following questions.

*Box 1: Vocabulary*

**Allelopathy** is a biological phenomenon in which plants produces biochemicals to influence germination, growth, survival, and reproduction of another plant or organism.

**Allelochemicals** are often secondary metabolites and not required for its growth, development, and reproduction.

Caffeine has been described as a germination and growth inhibitor based on various experiments.

Q5. How would you design an experiment to test this function of caffeine? Describe the experiment and expected results in 2-4 sentences.

In the early 1900s Fred Ransom published the results of an experiment in Biochem J. 1912, 6: 156–161; doi: 10.1042/bj0060156, where he compared the germination of a large number of different seeds in tap water and when pre-treated with caffeine. A small portion of the results are included below.



Note that in the above table the date is listed using the convention DDMMYY

Q6. Based on the data shown fill in the table below and describe the impact of caffeine treatment on the germination of the seeds. Does this confirm or disprove the claim that caffeine is an allelochemical?

|  |  |
| --- | --- |
| Name of Seed | # days for germination  |
|  Tap water |  pre-treated with Caffeine |
| Endive  |  |  |
| Onion |  |  |
| Lettuce |  |  |
| Cress |  |  |
| Turnip |  |  |

**Part 1: How do plants make caffeine?**

Coffee, tea, cocoa, and other caffeine producers make caffeine (1,3,7-trimethixanthine) from the nucleoside metabolite xanthosine through a series of enzymatic reactions. Here we will learn what these enzymes are and how they facilitate caffeine biosynthesis.

The Denoeud et al., 2014 (Science paper) explains that coffee plants have a set of 3 enzymes that convert a xanthosine into caffeine. All three enzymes [*i.e.*, xanthosine methyltransferase (XMT); 7-methylxanthine methyltransferase (MXMT); and 3,7-dimethylxanthine methyltransferase (DXMT)] are N-methyl transferases that use the cofactor S-adenosylmethionine (SAM) as the methyl group donor.

Review the chemical structures of xanthosine, caffeine, and various intermediates in the biosynthetic pathway.



Adapted from Figure 2 of Science (2014): Vol. 345, Issue 6201, pp. 1181-1184

Q3. The cofactor/co-substrate SAM is used by all the enzymes (XMT, MXMT, and DXMT). In each case at the end of the enzyme reaction SAM is converted to SAH. Can you figure out what SAH is?

*Box 2: Concept*

Enzymes undergo conformational changes in order to catalyze a reaction. Sometimes these changes include movement of loops and also interacting with other molecules (such as cofactors, ions). **Apoenzyme**s – i.e. the enzyme without any substrate, cofactor etc. bound to it, is often **inactive**. Binding to substrate(s) and cofactor(s) can **activate the enzyme** and prepare it to perform the enzyme reaction.

Q4. If you had to assemble an active complex of the XMT enzyme, that is about to perform the catalysis, what molecule(s) would you include in this complex?

**Part 2: Biosynthesis of Caffeine in Molecular Detail**

Exploring the structures of the N-methyltransferases involved in caffeine biosynthesis can shed light on how they facilitate the methylation reactions. This will also provide an opportunity to examine if the enzymes that use the same cofactor have any structural similarities.

To examine the structures of the N-methyltransferases involved in caffeine biosynthesis we will search the Protein Data Bank.

*Box 3: Resource*

RCSB Protein Data Bank (**RCSB PDB**, [www.rcsb.org](http://www.rcsb.org)) provides access to 3D structural data of biological macromolecules (proteins, nucleic acids, carbohydrates and their various complexes). In addition, it provides information about the experiment used to derive the data, details about the molecules included in the experiment, and links to various bioinformatics resources that can provide additional information about the protein/molecule of interest. Each structure in the PDB is identified by a unique identifier (called PDB ID). Atomic coordinates form the PDB can be visualized and analyzed using various visualization software (some available from RCSB PDB).

A query for “xanthosine methyltransferase” in the RCSB Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) reveals that there is a structure of this protein in the PDB (PDB ID 2eg5). Open the structure summary page (<https://www.rcsb.org/structure/2EG5>) to learn more about this structure. Note that there is a related structure mentioned on this page (PDB ID 2efj). This is the structure of 3,7-dimethylxanthine methyltransferase.

Let us explore the structure summary pages of these structures to learn a little more about them.

*Box 4:* ***What can you find on the structure summary page?***

1. **Title** - that tells you what the structure is about

2. **Snapshot** - of what the structure of the molecule/complex looks like.

3. **Authors** – who solved the structure

4. **Literature** – for access the article that describes the structure.

5. **Macromolecules** – All proteins and nucleic acids present in the structure are listed here. Each unique type of macromolecule or molecular chains is listed as a separate entity. There may be multiple copies of a molecule in the structure.

6. **Small molecules** – All ligands, ions, cofactors, inhibitors that are present in the structure are listed here.

7. **Experimental details** – describe details about the structure determination

8. **Structure quality** – shows a slider that provides insights about the quality of the structure and its agreement with the experimental data and geometric standards.

See <http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction> for details

Q1. Use information from the structure summary pages of both structures to complete the following table.

|  |  |  |
| --- | --- | --- |
| PDB ID: 2eg5 | Authors of entry |  |
| Macromolecules (# and Name) |  |
| Small molecules (# and Name) |  |
| Related structure PDB ID: 2efj | Authors of entry |  |
| Macromolecules (# and Name) |  |
| Small molecules (# and Name) |  |

Scroll to the “Literature” section of the structure summary page of PDB ID 2eg5, with the heading shown below and read the abstract.



The XMT and DXMT enzymes were both co-crystallized with the demethylated cofactor, S-adenosyl-L-cysteine (or SAH), and substrates, xanthosine or theobromine.

Q2. Why do you think that the methylated form of the co-substrate (SAM) was not used? (Hint: what do you think would happen if SAM was included in the structure?)

In the following section you will explore the structures of the 2 enzymes that you have identified in the PDB. The structure exploration is set up as a worksheet. Follow your instructor’s directions regarding completing these worksheets. After completing the worksheet(s), the responses will be reviewed in class.

*Structure exploration worksheet*:

For the structure you are exploring (either PDB ID 2eg5 or 2efj)

*A. Quick glimpse of structure*:

* Click on the 3D view tab on the top of the RCSB PDB structure summary page to see a ribbon model of the structure. Explore the structure interactively by rotating it, changing the colors, representations etc. using the various pull-down menus on the right hand side of the page. Remember to select the asymmetric unit option in the Assembly pulldown menu. This shows you the actual coordinates submitted to the PDB.

Q3. What is the overall shape of the protein(s) in the asymmetric unit of the complex?

*B. Explore closely the location of all components in the complex*:

* For this section we will use a web-based tool called iCn3D:

*Box 5: Resource*

**iCn3D** is a web-based visualization tool that allows users to directly open any structure from the Protein Data Bank (PDB) and visualize the structure. Users can interactively rotate the molecule/complex, select specific regions and represent them in different ways, compare structures, analyze interactions and make simple distance measurements.

* Open the file:
	+ Go to iCn3D (<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>)
	+ Click on the button called File >> Retrieve by ID >> PDB ID so that a new window opens. Input the PDB ID of the structure you wish to visualize and click on Load.

Spend a few minutes playing around with the different pull-down menus to see some of the different ways this protein view can be adjusted.

* Display one copy of the enzyme and molecules (cofactors/substrates) bound to it
	+ If more than one chain is present, visualize the chain of interest (e.g. chain A)
		- Click on Select >> Defined Sets.
		- In the new window that opens on the right select chains C, E, and G simultaneously. In the graphics window all these chains should be highlighted with a yellow halo.
		- Click on Style >> Proteins >> Hide and Style >> Chemicals >> Hide to hide these chains.
* Orient molecule so that you can clearly see the enzyme and all bound components.
* Save an image by clicking on File >> Save files >> iCn3D PNG image.
* Import the image saved above to power point or any other graphics software and label in the image appropriately.

Q4. Where are the substrate and cofactor molecules bound in this structure? Include a suitable image to support your answer.

*C. Explore the neighborhood of the bound substrate*:

* Display the amino acid side chains in the neighborhood of the substrate:
	+ Select the substrate - Click on Windows >> View Sequence and Annotations >> scroll down to select the first listed substrate (e.g. bound to chain A)
	+ Select the neighborhood - Click on Select >> by Distance >> use the default options and click on Display
	+ Display amino acids in the neighborhood - In the graphics window click on Style >> Side chains >> Sticks. This shows the amino acid side chains selected near the substrate molecule.
	+ Focus on selected residues - Click on View >> Zoom in selection to see a closeup of these residues and visualize the nature of interactions.

Read the box below to review some of the interactions in you may see in the structure:

*Box 6: Concept*

Biomolecular structural stability, interactions and functions are dependent on various non-covalent interactions. Some key interactions in molecular structures are

**Hydrogen bonds** - formed between two partially negatively charged atoms with a hydrogen atom between and covalently linked to one of them. e.g. in structures look for examples of O/N … H\_\_O/N, where … denotes hydrogen bond and \_\_ denotes a covalent bond

**Salt bridges** or **ionic interactions** - formed between oppositely charged amino acid side chains and/or charged ligands/ions. e.g. in structures look for interactions between Lys/Arg/His and Glu/Asp. These interactions may also involve phosphate groups and ions such as K+, Na+, Cl- etc.

**Hydrophobic interactions** - formed between hydrophobic amino acid side chains positioned away from the aqueous environment. e.g. look for regions with large numbers of carbon and hydrogen atoms in close proximity. Aliphatic amino acids such as Ala, Leu, Val, Ile participate in hydrophobic interactions.

**Pi stacking** - seen between amino acids with aromatic side chains. Pi clouds of aromatic rings interact with each other in staggered stacks, face to edge interactions. Alternatively, they may also interact with positively charged amino acid side chains (pi-cation interaction).

Learn more at <https://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmolnew/bonding/chymo.html#Topic2>

* Examine the interactions stabilizing the substrate:
	+ First clear all selections (following the neighborhood amino acid display) – Click on Select >> Clear Selections
	+ Now select the substrate again (as before) and click on View >> H-bonds and interactions >> Display to see H-bonds and/or Salt bridge interactions.
	+ Examine the structure to see if you can identify any other interactions stabilizing the substrate (e.g. hydrophobic interactions, pi stacking.)
	+ Save an image and import to powerpoint or any other graphics program to label the figure.

Q5. Describe 2-3 interactions stabilizing the substrate. Support your answer with a suitably labeled image also showing the methylation site.

*D. Explore the neighborhood of the bound cofactor*:

Use the same steps used to examine the substrate’s neighborhood and interactions here, but this time to explore the cofactor.

Q6. Describe any 2 interactions stabilizing the cofactor. Support your answer with a suitably labeled image also showing the methylation site.

*E. Compare the relationship between the 2 coffee N-methyl transferases*:

Although the 2 caffeine biosynthesis enzymes from coffee have similar functions and binds the same cofactor, its substrate specificities are different. Here we will compare the sequences and structures of the two coffee proteins (XMT and DXMT) to see if they are similar or different?

* Compare sequences of the 2 enzymes (PDB IDs 2eg5 and 2efj)
	+ Download the sequences of both these proteins from the top of the structure summary page >> Download Files >> FASTA sequence save as text file.
	+ Go to UniProt Align tool (<https://www.uniprot.org/align/>) to compare these two sequences.
	+ Paste the sequences downloaded or upload the files to run the alignment. This may take a few minutes to complete.
	+ Examine the sequence alignment and answer the following questions. Keep this window open for use later.

Q7. What is the sequence identity between these two proteins? Where is/are the differences in the sequences (if any)?

* Compare structures of the two caffeine biosynthesis enzymes (PDB IDs 2eg5 and 2efj)
	+ Start a fresh iCn3D session (<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>)
	+ Click on the button called File >> Align >> Structure to Structure
	+ In the new box that opens type in the PDB IDs of the 2 caffeine biosynthetic enzymes (PDB ID 2eg5 and 2efj) >> Click on All Matching Molecules Superposed.

In the C-alpha backbone image seen - red regions are conserved, and blue regions are not.

Q8. Make a labeled image of the superposed structures and mark the neighborhoods of the Xanthine derivative and SAH with comments about sequence and structure conservation in these neighborhoods. Did you expect this pattern? Explain your answer.

Q9. What can you conclude about the origin of these coffee enzymes that make caffeine? (Hint: use information from the sequence and structure comparisons to explain the relationship between these enzymes.

**Part 3: Evolution of Caffeine Biosynthesis**

In this part we explore if the same enzymes that make caffeine in coffee plants are also involved in producing caffeine in tea and cocoa plants too.

Other plants that make caffeine (e.g., Tea and Cocoa) also use Xanthosine as a starting metabolite. Currently there are no structures of these enzymes in the PDB archive. To determine if the tea and cocoa enzymes are related to each other and/or to the coffee plants, we will do some sequence comparisons and draw phylogenetic trees using the interface and tools available from UniProt.

*Box 7: Resource*

**UniProt** (<https://www.uniprot.org/>) is a bioinformatics data resource that provides comprehensive, high-quality, freely accessible protein sequences, and their functional information. This information comes from research that has been published by others. For eukaryotic proteins it also lists information about specific domains, post-translational processing and modifications, and pathology resulting from mutations in the protein. UniProt provides links to other biological data resources to access other relevant information about the protein, such as gene sequence, protein structures, functional annotations etc.

* Find and download protein sequences
	+ Go to UniProt (<https://www.uniprot.org/>), search for the following protein sequences and download the FASTA sequences.
		- In tea (organism: *Camellia sinensis*) UniProt ID TCS1\_CAMSI
		- In cocoa (organism: *Theobroma cacao*) UniProt ID Q2HXL8\_THECC
	+ Go to the UniProt Align tool (<https://www.uniprot.org/align/>) to compare the sequences of the 2 Coffee N-methyltransferases and the 2 caffeine synthase enzymes from tea and cocoa.

(If your previous alignment result page is still open add the 2 FASTA sequences to the box called Add and align and rerun the alignment.

* + Save the sequence alignment (by clicking on Download alignment, then copying and pasting the alignment to this document using Courier font size 8)
	+ Also take a screen shot of the phylogenetic tree.

Q1. Based on the sequence comparison results, are the tea, coffee, and cocoa enzymes related?

Q2. What does the phylogenetic tree tell you about the evolution of the caffeine biosynthetic enzymes?

An overview of the Caffeine Metabolism in the KEGG pathways shows how multiple enzymes and combinations of enzymes that can lead to caffeine synthesis (<https://www.genome.jp/kegg-bin/show_pathway?map00232>).



Based on all the things that you learned in your explorations answer the questions asked.

**The Molecular Case Study Question:**

**How do plants make caffeine? Do they all make it the same way?**