**Caffeine Evolution**

By Brian Sullivan1, Shuchismita Dutta2\*

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

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

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

**Part 3: Biosynthesis of Caffeine in Molecular Detail**

*Motivation 3*: 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.

Ans:

|  |  |  |
| --- | --- | --- |
| 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 respectively.

Q2. Why do you think that the methylated form of the cosubtrate (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 powerpoint 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 UniPort 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 4: Evolution of Caffeine Biosynthesis**

*Motivation 4*: In this part we explore if the same enzymes that make caffeine in coffee plants are 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?

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

**The Molecular Case Study Question:**

**Why and how do plants make caffeine? Do they all make it the same way?**