

# Too Hot To Handle: Creating a Spicy Tomato and Assessing the Susceptibility of Common Agricultural Pests to Capsaicin

Group 8: Nikita Sajai, Dominic Matras, Alice Gevorgyan

## Abstract

The purpose of this project was to evaluate the feasibility of capsaicin as an alternative, organic pesticide. With huge annual crop losses to pests and the toxicity of current pest control chemicals, it would be doubly beneficial to use a less toxic, but equally effective, compound. This paper evaluates whether capsaicin could be an organic alternative to conventional pesticides. There were two fronts studied to evaluate (1) capsaicin's mode of actions in insect targets and (2) to understand whether capsaicin production could be engineered in plants other than pepper to allow for innate defenses. This research identified key motifs conserved in capsaicin-metabolizing proteins, namely CYP6B6, found in common agricultural pests. In addition, the feasibility of developing spicy plants in the Solanaceae family was evaluated by looking at the conservation of key residues in capsaicin synthase homologs in crops closely related to pepper. Protein-ligand docking was used in both sections in order to identify binding residues and to identify putative active sites in relevant proteins.

# Introduction

Capsaicin is among the world's most recognizable compounds. From its origins in Bolivia, spicy peppers have spread all over the world and become a staple in millions of diets (Aguilar-Meléndez et al., 2009; Kaiser, Higuera and Goycoolea, 2017). Although "spicy" was not a novel taste to European explorers, the compounds which made peppers spicy were very different from the black pepper previously known to the Old World. Instead of piperine being the spicy component, chili peppers get their pungency from an alkaloid compound called capsaicin (Milenković and Stanojević, 2021; Ogawa et al., 2015). Only plants in the *Capsicum* genus of the family *Solanaceae* are able to produce capsaicin due to unique biosynthetic pathways and enzymes that evolved after their divergence from other *Solanaceae* crops (Kim et al., 2014; Ogawa et al., 2015). Current phylogenetic trees suggest that the closest economically relevant crops to peppers are tomatoes (*Solanum lycopersicum*) and eggplant (*Solanum melongena*) (Särkinen et al., 2013; Figure 1B). It is thus plausible that these closely related species would be good candidates to try and artificially engineer capsaicin biosynthesis in plants outside *Capsicum*.

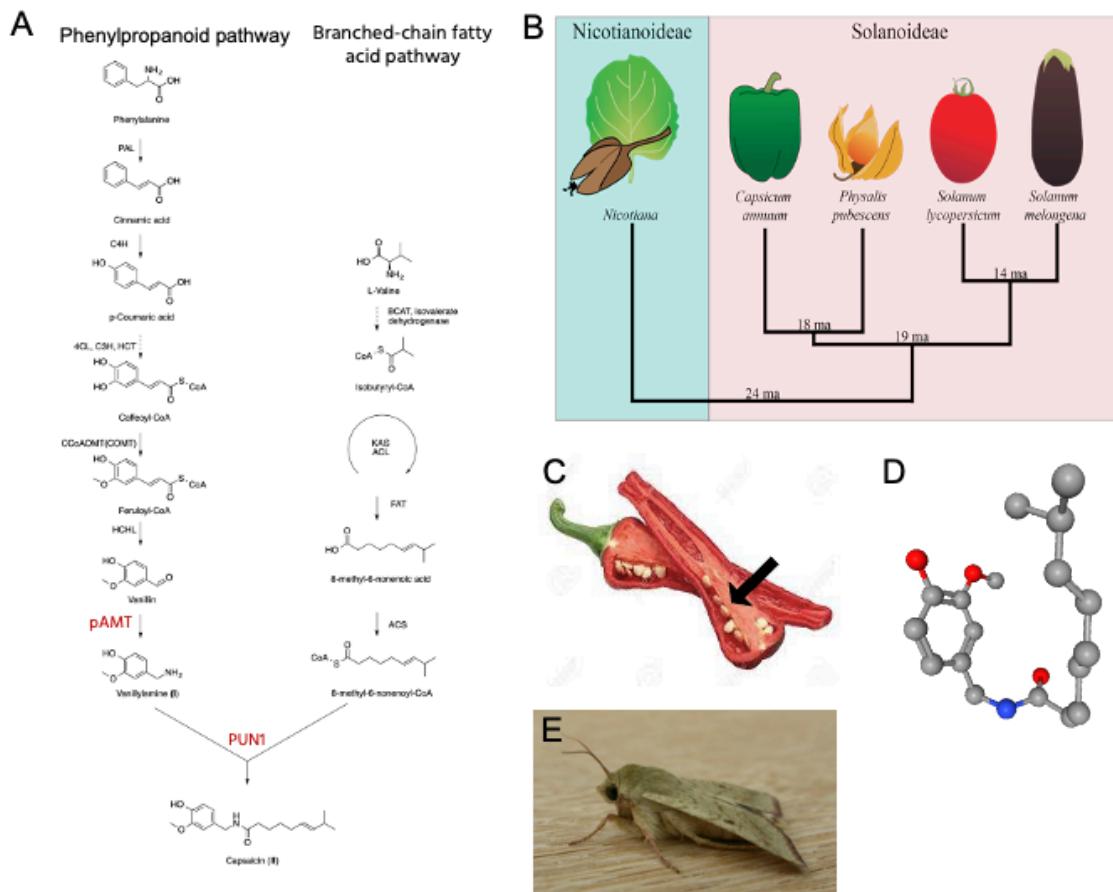
Capsaicin itself is a compound produced in the placenta of pepper fruits in conjunction with other capsaicinoid compounds including dihydrocapsaicin and homocapsaicin (Li et al., 2019; Figure 1C-D). There are two biosynthetic pathways required to produce capsaicin, the phenylpropanoid pathway and the branched-chain fatty acid pathway (Figure 1A). The important products from these pathways are vanillylamine and 8-methyl-6-nonenoic acid, respectively. The two pathways meet when these two compounds are catalyzed by an acetyltransferase to become capsaicin (Kaiser, Higuera and Goycoolea, 2017; Figure 1A). The enzyme that facilitates this last step is capsaicin synthase, now known to be PUN1 (Stewart et al., 2005; Han et al., 2013). In addition to PUN1, a putative aminotransferase (pAMT) has also been identified as a key regulator in capsaicin. pAMT catalyzes the last step in the phenylpropanoid pathway from vanillin to vanillylamine (Kaiser, Higuera and Goycoolea, 2017).

The perception of capsaicin is another interesting aspect of this compound. As humans, we experience the spicy taste of peppers due to the Transient Receptor Potential Vanilloid 1 (TRPV1) ion channel. Upon consumption, capsaicin binds to TRPV1 and an action potential is fired in neurons, evoking the familiar spicy sensation (Shuba, 2021). However, insects lack this receptor. Current research has not elucidated what makes insects susceptible to capsaicin, but it is well-documented that exposure to capsaicin has detrimental effects on pest growth and development. Common agricultural pests including *Drosophila melanogaster* (fruit fly), spider mites (*Tetranychus urticae*), and cotton aphid (*Aphis gossypii*) have been found to have decreased oviposition, increase mortality, and suppressed populations in the presence of capsaicin (Li et al., 2019; Li et al., 2020; Tomita and Endo, 2007). However, *Helicoverpa armigera* (cotton bollworm) and *Helicoverpa assulta* (tobacco budworm), both from the

Noctuidae (moth) family are not susceptible to capsaicin and frequently consume *Capsicum* plants (Tian et al., 2019; Figure 1E). This suggests that some insects have ways of metabolizing, or detoxifying, capsaicin.

One particular enzyme that has been studied in relation to capsaicin is CYP6B6, a member of the heme-thiolate protein superfamily of Cytochrome P450s (CYPs for short) (Tian et al., 2019). Tian et al. (2019) found that CYP6B6 in *H. armigera* was able to metabolize capsaicin and turn it into multiple hydroxylated metabolites. These CYP6 proteins have even been found to help with detoxifying commercial chemicals such as DDT in *Drosophila* (Water et al., 1992; Ranasinghe and Hobbs, 1998). Other studies with *H. armigera*, a generalist herbivore, and *Solanaceae* specialist *Helicoverpa assulta* (tobacco budworm) suggested that increased CYP6B6 activity in *H. assulta* allows the increased detoxification of capsaicin (Jiang et al., 2020; Shi et al., 2021). Thus identifying if CYP6B6 or similar proteins are present in other agricultural pests would indicate if there would be widespread pest resistance to capsaicin.

In this study we evaluated the conservation of CYP6B6 across common agricultural pests. In addition, we modeled the docking of capsaicin onto CYP6B6 to find key residues that might play a critical role in detoxification. We investigated whether plants outside Capsicum could be modified to produce capsaicin, similar to Bt-producing maize. We conducted BLAST searches to identify possible homologs to PUN1 and pAMT to evaluate if other *Solanaceae* crops might have the tools to produce capsaicin. Docking was also done on PUN1 to find key residues that allow capsaicin synthesis.



**Figure 1. Introductory figure.** (A) The two pathways involved in capsaicin biosynthesis, the phenylpropanoid and branched-chain fatty acid pathways. The two key enzymes in this paper are in red, pAMT and PUN1. A modeled capsaicin molecule is shown in (D). (B) A phylogenetic tree to show the close relatives of *C. annuum*. (C) Capsaicin is produced in the placenta of peppers, shown by the arrow. (E) *H. armigera*, the cotton bollworm, is one of the main organisms studied to understand the effect of capsaicin on insects. Images are taken/modified from Kaiser, Higuera, and Goycoolea, 2017; Thomas (unpublished); 123rf.com/photo\_126276269\_kashmiri-mirch-pepper-capsicum-annuum-fruit-isolated; Donald Hobern from Canberra, Australia, CC BY 2.0

## Materials & Methods

### Looking from the Insect Side: Protein-ligand interaction between capsaicin and CYP6

Previous studies have demonstrated the ability of CYP6B6 in *H. armigera* to detoxify capsaicin through hydroxylation (Tian et. al 2019). To examine this adaptive capsaicin metabolism pathway in *H. armigera*, we used protein-ligand docking software to analyze the interaction between capsaicin and CYP6B6. In this analysis, we make use of Chimera, p2Rank, MedusaDock, and PyMol to complete the successive steps of the docking procedure. First, the

predicted protein structure, generated with AlphaFold, was downloaded from Uniprot (UniProt: Q95031). The protein and ligands were then optimized for docking using the DockPrep extension in Chimera. The parameters were set to: (i) delete solvent molecules that are not important in ligand binding, (ii) complete truncated side chains using the Dunbrack 2010 rotamer library, (iii) add hydrogens and generating protonation states at the physiological pH, and, finally, (iv) add charges using the Gasteiger method to ensure that the net charge is zero. The dock-prepped molecule was written into a mol2 file and saved. This process was repeated for the ligand (capsaicin) molecule preparation.

Following this step, the active site was predicted using *p2rank*, a machine-learning based command-line tool that uses a Random Forest algorithm to assess, cluster, and rank the ligandability of near-surface points of the protein surface (Krivak and Hoksza 2018). The program outputs a set of viable binding pockets, each with a score and probability. The residues listed under pocket1 in the output were used in further analyses as the active site. This data was imported into Chimera using the command line tool and the CYP6B6 active site structure was saved as a separate PDB file.

The docking was then completed using MedusaDock 2.0, an online web service that uses sets of discrete rotamers to model ligand and receptor flexibility, allowing us to assess the likelihood of ligand binding under the induced fit model, (Wang and Dokholyan 2019). Their algorithm operates in three stages: (i) generating ligand rotamers, (ii) modeling a coarse docking that rigidly docks the ligand to the receptor side chain, and (iii) fine docking and selecting the ligand rotamer with the lowest BDE (Wang and Dokholyan 2019). The PDB file containing active site residues was imported, and the parameters set were distance from 2.7-3.2 Å and the energy bonus as -1 kcal/mol. The docking data from Medusa Dock 2.0 was then imported into PyMOL to complete ligand-protein interaction analysis. First, the ligand and active sites were identified using PyMOL command line. The residues involved in H-bonding were then parsed apart and labeled, and the hydrogen bond distances were calculated. The predicted active site residues were then cross-verified with the model and the key residues involved in the interaction were recorded.

### **Identifying key residues and motifs among CYP6B6 homologues**

Following the analysis of the protein-ligand interactions, we assessed the conservation of those key residues among homologs of the protein and other closely related proteins in pests. BLAST was used to find homologs of CYP6B6 with a cutoff E-value of 1e-10, and the BLAST output was further filtered using the python script titled “parse2.py” to filter out results with an identity match of less than 80% or coverage similarity of less than 60% (Supplemental Files). The sequences from the BLAST output were extracted from the web using another python script titled “get\_fasta.py” (Supplemental Files). The sequence files were cleaned using awk and sed

commands in linux and the unique sequences were recorded in a separate file; these unique sequences can be found in Table 1. In order to expand the number of sequences in the analysis we conducted a literature search to identify common agricultural pests thought to be resistant to capsaicin via CYP's. Cytochrome P450 sequences for these pests were then downloaded from UniProt and added to the BLAST-result sequences. The details about the BLAST and literature sequences can be found in Table 1.

A sequence file containing the protein sequences of CYP6B6 homologs and additional CYP sequences from other agricultural insect pests was compiled. These sequences were aligned using MUSCLE in MEGA11, using the UPGMA cluster method with the parameters set as -2.9 penalty for gap open and 1.2 hydrophobicity multiplier. A Maximum Likelihood phylogeny was then constructed using IQTree with 1000 bootstrap replicates.

	<b>Sequence/Species Source</b>	<b>Organism</b>	<b>Protein(s)</b>
<b>BLAST</b>	<i>Helicoverpa armigera</i>		<i>CYP6B6</i>
	<i>Helicoverpa zea</i>		<i>CYP6B6, CYP6B27</i>
	<i>Heliothis virescens</i>		<i>CYP6B10</i>
	<i>Spodoptera frugiperda</i>		<i>CYP6B67-like</i>
	<i>Spodoptera littoralis</i>		<i>CYP6B42</i>
	<i>Spodoptera exigua</i>		<i>CYP6B8</i>
	<i>Mythimna separata</i>		<i>CYP103</i>
	<i>Spodoptera litura</i>		<i>CYP6B7-like</i>
<b>Literature Review</b>	<i>Aphis gossypii</i>		<i>CYP6CY22, CYP6CY13</i>
	<i>Aedes aegypti</i>		<i>CYP6BB2, CYP6Z8, CYP9M6</i>
	<i>Myzus persicae</i>		<i>CYP6CY3, CYP4C</i>
	<i>Aphis craccivora</i>		<i>CYP6CY22, CYP6CY13</i>
			<i>CYP9G2, CYP6BF1v1,</i>
	<i>Plutella xylostella</i>		<i>CYP9G4</i>
	<i>Ostrinia furnacalis</i>		<i>CYP4AU1</i>
	<i>Operophtera brumata</i>		<i>CYP4AU1</i>

**Table 1. Species and sequences used to understand CYP6B6 conservation.** Homologs for CYP6B6 in the left column and cyp450 sequences from literature in the right column.

In addition to building a tree, the alignment was used to find conserved motifs across the cytochromes. MEME suite was used to determine the top 10 conserved motifs with a 0th order background model and motifs widths between 15-20. Motifs with shuffling, which shuffles each sequence to randomize them and erase existing patterns, was completed to determine the average cut off E-value for analysis.

### **Looking from the plant side: Examining capsaicin production through protein-ligand docking**

On the plant side, we aim to better understand capsaicin synthesis, particularly the interactions between capsaicin and its biosynthetic precursors and the enzymes pAMT and PUN1. While the reaction pathway is not fully characterized, these enzymes are hypothesized to carry out the final steps to synthesize capsaicin. The former is a putative aminotransferase that catalyzes the reaction from vanillin (PubChem CID: 1183) to vanillylamine (CID: 70966) to complete the phenylpropanoid pathway. PUN1 accepts this product and 8-Methyl-6-nonenoic acid (CID: 5365959) (the final product from the fatty acid pathway) to produce capsaicin (CID: 1548943) (Stellari et. al.). Examining the substrate docking interactions between these organic molecules and the involved enzymes would better characterize the active sites of catalysis. Furthermore, looking at conservation of key residues in these regions, as well as the relative “fit” of possible transition state analogs could help illuminate the underlying mechanism of each step.

In our study, we examine the docking of both vanillin and vanillylamine with pAMT, as well as vanillylamine, 8-Methyl-6-nonenoic acid, and capsaicin with PUN1. In doing so, we searched for the amino acid residues forming polar contacts with these molecules using PyMol’s calculation feature. We then conducted a BLAST search and multiple sequence alignment of PUN1 and pAMT homologs to assess the conservation of these key residues among *Capsicum* species, as well as other species in *Solanaceae*. This gives context regarding the evolutionary development of PUN1, as well as the feasibility of inducing pungency among other species. If induced in a tissue-specific manner, this pungent phenotype could serve as a deterrent for common agricultural pests among *Solanaceae*.

### **Docking Procedure - Adapted from Insect Procedure**

The docking procedure for the plant enzymes is nearly identical to that for the insect enzymes, with some minor changes to address the ambiguous binding pockets in pAMT. For the case of the enzymes PUN1 and pAMT, which catalyze the final steps in the capsaicin biosynthetic pathway, we dock both the substrate(s) and the product of each step with its corresponding enzyme. We first obtained molecular structure files from the AlphaFold database made available via UniProt, and used PUN1 from *Capsicum annuum* (UniProt: D2Y3X2, AlphaFold:

AF-D2Y3X2-F1) and pAMT (UniProt: O82521, AlphaFold: AF-O82521-F1) from *Capsicum chinense*, a hot pepper, to perform ligand docking.

In Chimera, we performed an optimization step for docking using the software's DockPrep feature, applying it to both the enzyme and the ligand. This step used the same settings as with CYP6B6 to remove solvent, add hydrogens and charges, and find possible rotamers. Finally, the newly prepped molecule was saved as a Mol2 file for subsequent use in MedusaDock.

As with CYP6B6, we predicted active sites for PUN1 and pAMT using *p2rank*. For PUN1, we selected the pocket with the highest score (43.79, compared with p2's 6.51). For pAMT, we modeled both pocket 1 and pocket 2, which scored 13.53 and 12.63, respectively. Because these two scored similarly (in comparison with the best two pockets for PUN1), we selected based on the polar contacts (number and distance) as calculated in PyMol.

After prepping the enzymes and ligands, we performed docking with MedusaDock. We used the prepped Mol2 files from Chimera with the docking site predicted by *p2rank* to create the docked Mol2, which could be visualized in PyMol to measure the distances of polar contacts and record the amino acid residues that engaged in non-covalent binding with the ligands.

### **BLAST Searching for Homologs and Alignment/Phylogeny/Motif Finding**

To find homologs of the PUN1 and pAMT genes in *Capsicum Annuum*, we first obtained polypeptide sequence data from NCBI (pAMT identifier: BAU36961.1, PUN1 identifier: NP\_001311698.1). We then performed a BLAST search against the NCBI database, generating a tabular output file, which we could then parse these results to filter out those below 40% identity, using the following scripts: jBlastp.pamt.sh, jBlastp.pun1.sh, and parse2\_blast.py (Supplemental Files). After this, we removed duplicates from the blast output and retrieved the full sequence hits from the Entrez molecular sequence database system using the script remove\_duplicate.py (Supplemental Files).

### **Phylogeny and Motif Finding for PUN1 and pAMT**

Using Muscle (version 3.8.31), we performed a multiple sequence alignment using the jMuscle4.sh script, which we could then visualize with MEGA to find the residues involved in substrate binding, and likely in catalysis (Supplemental Files). We also used the jIqtree4.sh script to perform model testing and bootstrapping (with 1000 replicates) and build phylogenies for both pAMT and PUN1 (Supplemental Files). In addition, we used the aligned sequences to perform a motif finding analysis with MEME suite, searching for motifs of 10-20 residues with zero or one occurrence per sequence within a single dataset (Classic mode). These two procedures were not

as informative as examining the raw alignment, and less time is spent discussing the associated results.

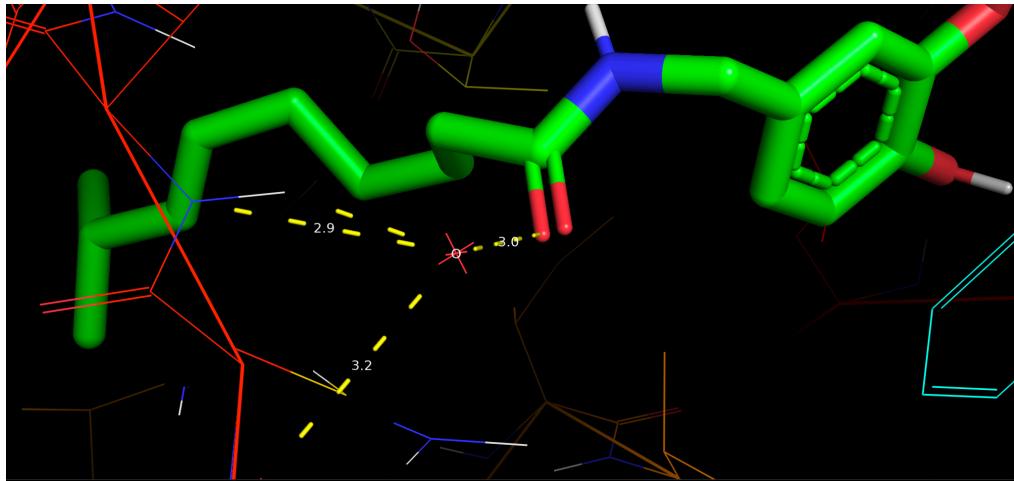
## Results

### Two highly-conserved motifs were identified in capsaicin-detoxifying proteins

Fifty-eight residues were determined to play key roles in the CYP6B6-capsaicin interaction, all of which are listed in Table 2. These residues were calculated using p2rank, and their location and residue ID numbers are listed in the last four columns of the table. Using PyMOL, two of these key residues, Arginine 129 and Asinine 444, were found to be involved in hydrogen bonds between the ligand and the protein. The hydrogen bond distances between these residues were 2.9, 3.2, and 3.0 Å, with an average bond distance of 3.03 Å which are shown in Figure 2.

Feature	Rank	Score	Probability	Residue ID's
Active site (Pocket 1)	1	66.67	0.990	A_104 A_106 A_108 A_117 A_118 A_119 A_125 A_129 A_136 A_183 A_208 A_209 A_211 A_212 A_213 A_214 A_216 A_217 A_220 A_233 A_234 A_300 A_301 A_302 A_304 A_305 A_306 A_309 A_310 A_313 A_361 A_365 A_370 A_371 A_372 A_373 A_374 A_375 A_376 A_437 A_438 A_439 A_443 A_444 A_445 A_446 A_447 A_450 A_451 A_48 A_482 A_483 A_484 A_485 A_486 A_49 A_55 A_76
Key residues in H-bond	n/a	n/a	n/a	ARG 129, ASN 444

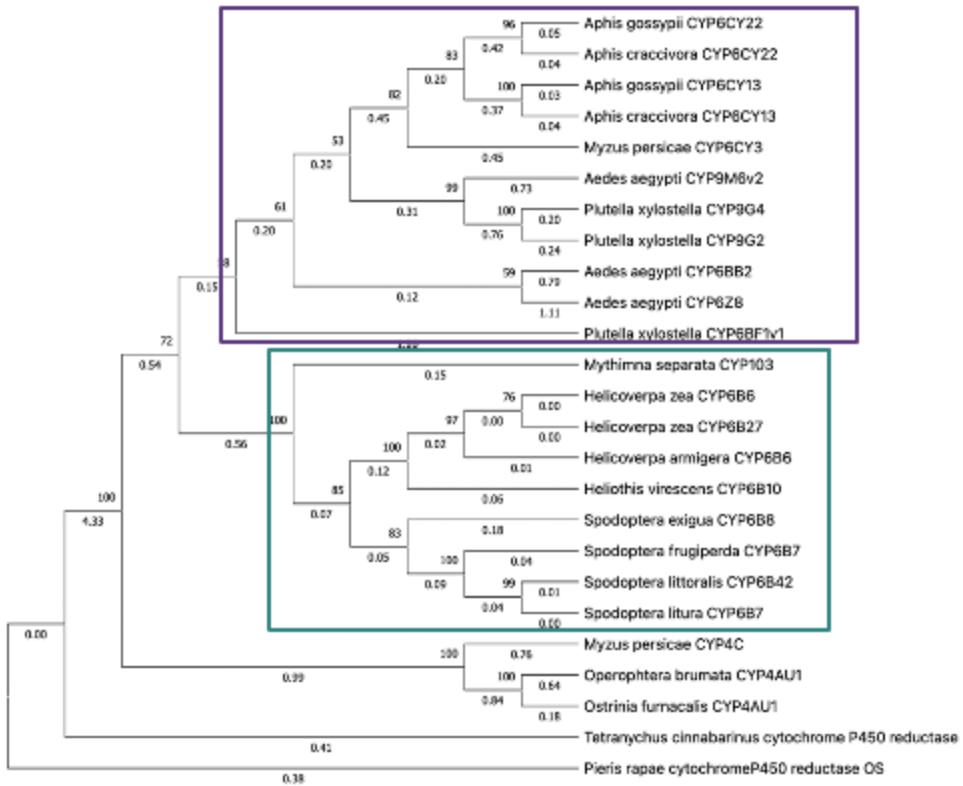
**Table 2. Active site residues from CYP6B6-Capsaicin interaction.** There were 58 residues identified as important to binding in CYP6B6. Two of these key residues, Arginine 129 and Asinine 444, were found to be involved in hydrogen bonds between the ligand and the protein.



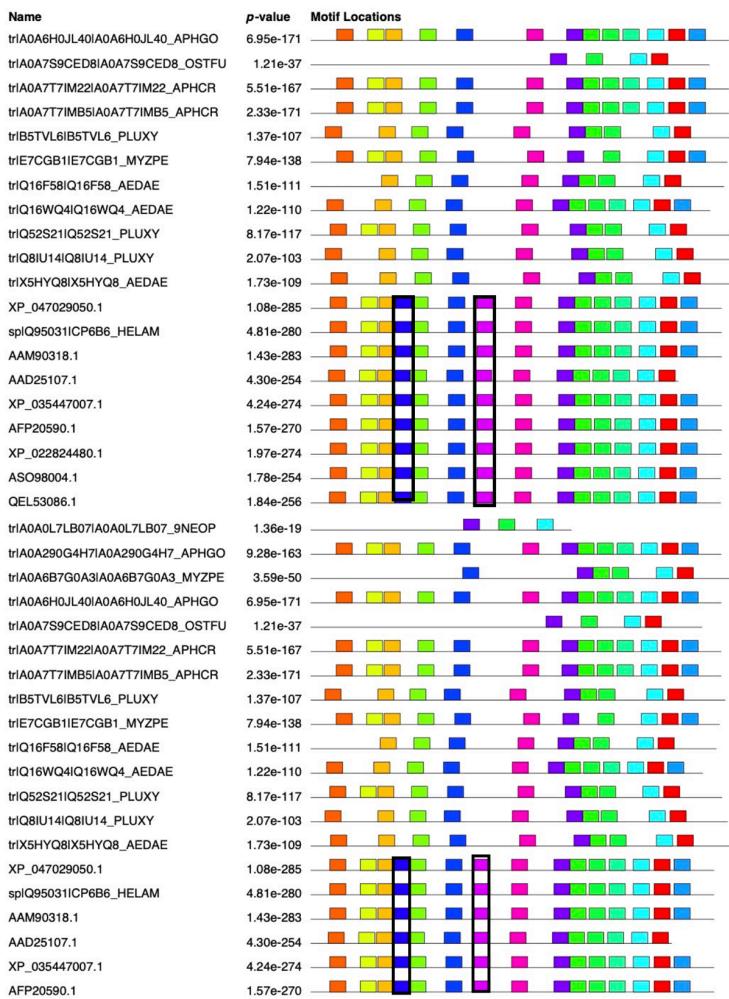
**Figure 2. Hydrogen bond distances between capsaicin and CYP6B6.** Capsaicin can be seen as the ligand (green molecule) while the hydrogen bonds are shown as dotted yellow lines.

From the BLAST query, eight species with homologs to CYP6B6 were found; these homologs included *Helicoverpa zea* CYP6B6 and CYP6B27 proteins, *Heliothis virescens* CYP6B10, *Spodoptera frugiperda* CYP6B67-like, *Spodoptera littoralis* CYP6B42, *Spodoptera exigua* CYP6B8, *Mythimna separata* CYP103, *Spodoptera litura* CYP6B7-like. The phylogeny constructed with these species and the supplemented sequences from literature is shown in Figure 3. The CYP6B6 homologues and the supplemental species grouped into separate clades suggesting their protein sequences contained differences. This is important as it means the supplemented organisms may have alternate methods of detoxifying capsaicin besides cytochromes.

Although the phylogeny did not reveal meaningful structural similarities/differences between the proteins in the two clades, motif finding revealed two unique motifs from residues 104-124 and 205-224. Residues 104-124 and 205-224 contained key active site residues (104, 106, 108, 117, 118, 119 and 208, 209, 211, 212, 213, 214, 216, 217, 220, respectively) that are only present among homologs, distinguishing these homologs from other common agricultural pests (Figure 4). Shuffling the motifs helped establish an average E-value cutoff for motifs in the unshuffled primary sequence dataset, which was 4.6e+002. The E-values for the unique motifs discovered were far lower than this cutoff, respectively 8.6e-.098 and 1.2e-.089, suggesting that these motifs may be used to identify other proteins that could detoxify capsaicin.



**Figure 3. Phylogeny constructed with CYP6B6 (*H. armigera*) homologs and cytochrome p450s predicted to metabolize pesticides in other agricultural pests.** The CYP6B6 homologues are in the teal box while the supplemented sequences are in the purple box. There is a clear clustering of each group, suggesting that there are differences between these two groups of CYP proteins.

**A****B**

Motif	Symbol	Motif Consensus
1.		RHPCAYLPFGPRLPRNCIGMR
2.		IHYDPKYYPBPDQFBPDRFD
3.		WRALRNRFSPIFTSGKLKNM
4.		AGYETSATTMAYLLYZLALN
5.		RDPEJIKHIMIKDFEVFSDR
6.		DTLKEMKYLKVFDETLRLMY
7.		EVHSLLQTYTMDTIACAFG
8.		QRNGKPSGRNDFMDLJLELR
9.		FNYWKRNVPGPPEPVFFGN
10.		YNMFPNEKVVGIVYRMTSPCL
11.		PDIQBKLIAEIDEVLKKNDG
12.		LQRKATRDYKVPGTDLVIEK
13.		SKLCJVKILSKFRVEPSKNT
14.		GVEFSKEGLGQNLFHADGET
15.		VDKIISAPSYANELDDMMYPG

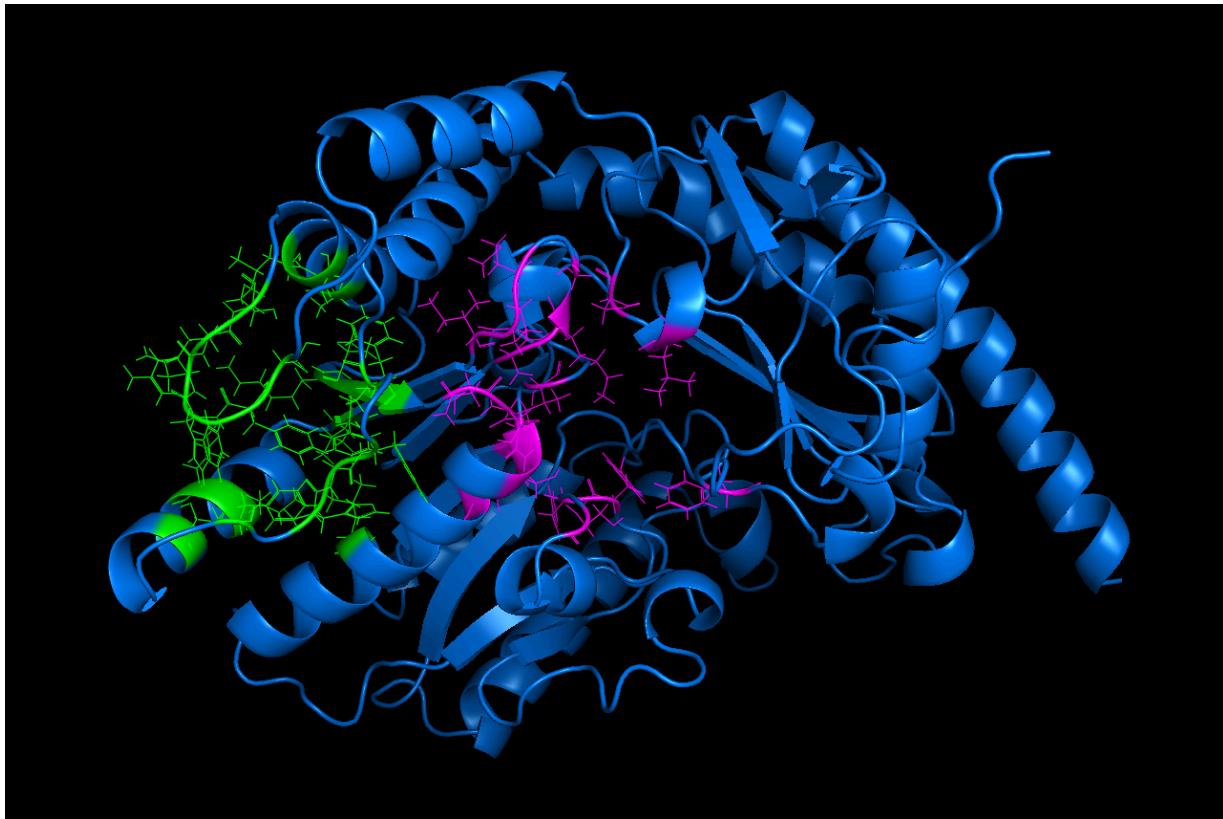
**Figure 4. Unique motifs discovered in CYP6B6 homologs and additional putative detoxifying proteins.** (A) Notable motifs are colored dark blue (105-124) and magenta (204-224) and have been boxed. (B) The motif sequence represented by each color.

### Characterizing pAMT and its active sites during binding

The docking procedure predicted two binding pockets with comparable *p2rank* scores for pAMT. The higher scoring pocket (called Pocket 1) achieved a score of 13.53 and a probability of 0.69, while the second pocket (2) scored 12.63 with a probability of 0.657 (Table 3, Figure 5).

In order to assess the relative binding affinity for the ligands in each binding pocket, we performed docking with both, and used PyMol to compute the number and lengths of polar contacts (including hydrogen bonds) for each ligand in each predicted binding site. The binding of vanillin in each pocket is shown in Figure 6. The non-covalent interactions are of comparable

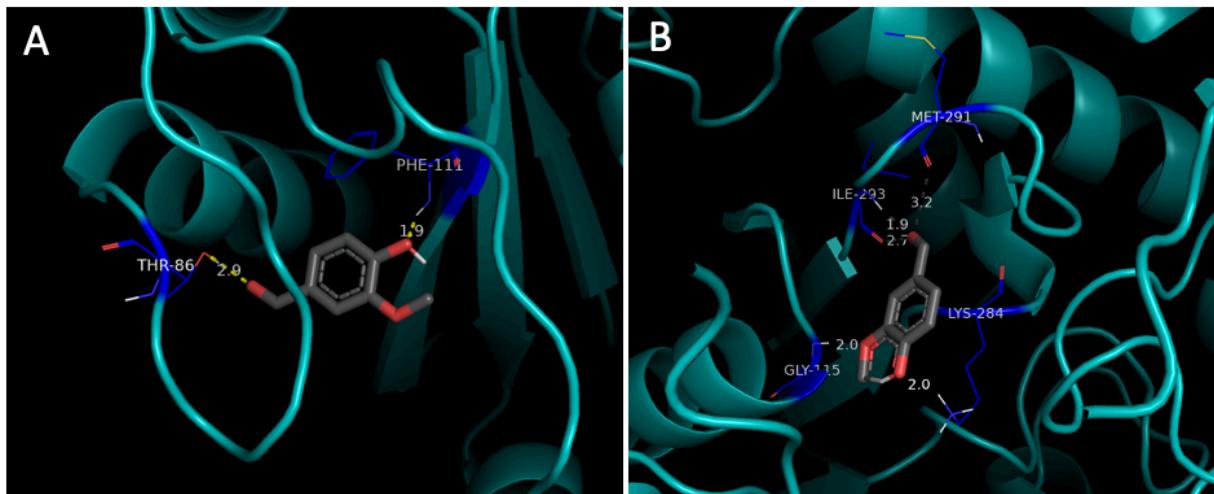
lengths, but pocket 2 shows a greater number of contacts, indicating a binding site that is likely more energetically favorable (Figure 7). For vanillylamine, we observe similar average contact lengths (1.9 Å for pocket 1, and 2.0 Å for pocket 2), but with three binding contacts in pocket 2, compared with only 2 in pocket 1 (Table 3). Therefore, we propose that pocket 2 is more likely to be the active site of the enzyme, considering the relative predicted binding of both vanillin and vanillylamine.



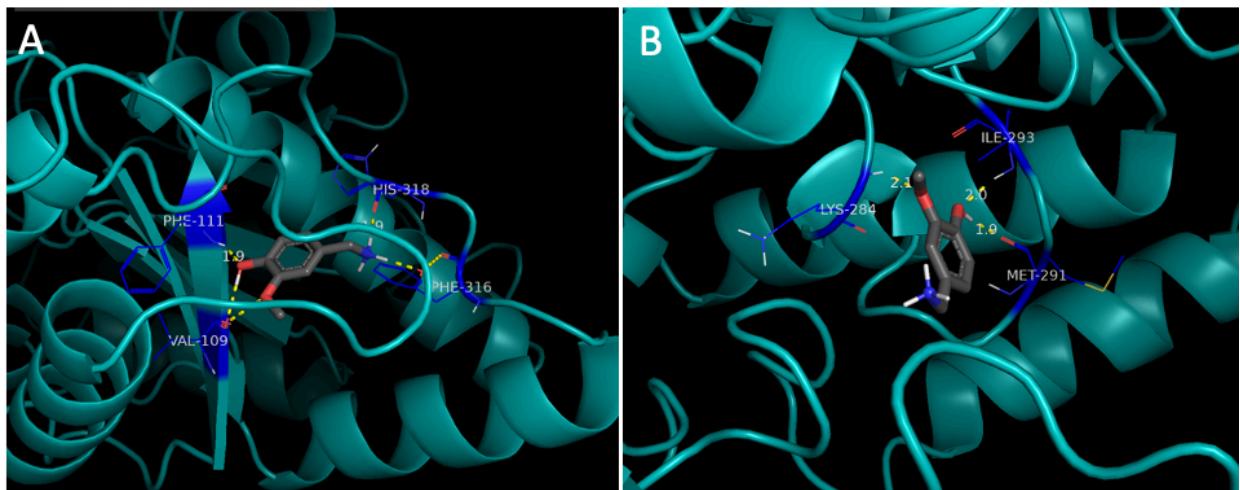
**Figure 5. Predicted active sites in pAMT.** There were two pockets identified, Pocket 1 is shown in green while Pocket 2 is shown in magenta. The pockets have comparable *p2rank* scores, although Pocket 1 has the higher score.

<b>Pocket</b>	<b>Rank</b>	<b>Score</b>	<b>Probability</b>	<b>Residue ID's</b>
pocket1*	1	13.53	0.69	A_108 A_109 A_110 A_111 A_125 A_298 A_306 A_307 A_310 A_316 A_318 A_319 A_320 A_323 A_81 A_82 A_83 A_84 A_85 A_86 A_90 A_91 A_94
pocket2*	2	12.63	0.657	A_113 A_114 A_115 A_116 A_119 A_148 A_149 A_151 A_17 A_255 A_257 A_283 A_284 A_289 A_291 A_292 A_293 A_53 A_55
pocket3	3	9.31	0.499	A_28 A_289 A_290 A_30 A_33 A_35 A_36 A_37 A_38 A_412 A_425 A_48 A_50 A_52 A_55 A_56 A_59 A_60 A_66
pocket4	4	8.46	0.452	A_148 A_161 A_164 A_17 A_226 A_227 A_258 A_284 A_414 A_416 A_421 A_53 A_54
pocket5	5	4.31	0.179	A_146 A_161 A_184 A_185 A_186 A_228 A_381 A_383 A_388 A_391 A_416 A_418 A_419
pocket6	6	3.83	0.149	A_104 A_242 A_245 A_246 A_249 A_250 A_252 A_276 A_278 A_301
pocket7	7	2.28	0.057	A_145 A_146 A_160 A_162 A_165 A_171 A_177
pocket8	8	1.87	0.037	A_330 A_333 A_334 A_337 A_68 A_71 A_72 A_88 A_89 A_92
pocket9	9	1.84	0.035	A_183 A_193 A_232 A_233 A_370 A_372
pocket10	10	1.57	0.024	A_264 A_265 A_272 A_349 A_352 A_353 A_356 A_360 A_373 A_374 A_375
pocket11	11	1.57	0.024	A_124 A_128 A_131 A_313 A_316 A_317 A_318

**Table 3. Details of the predicted pockets of pAMT.** There were 11 predicted pockets, but the first two (indicated with an asterisk) had significantly better scores than the other pockets.



**Figure 6. Binding of vanillin to pAMT.** (A) Binding in Pocket 1 revealed 2 polar contacts with an average bond length of 2.4 Å. (B) Binding in Pocket 2 revealed 5 polar contacts with an average bond length of 2.36 Å. The higher number of polar contacts in pocket 2 suggest it is the superior binding site.



**Figure 7. Binding of vanillylamine to pAMT.** (A) Binding in Pocket 1 revealed 2 polar contacts with an average bond length of 1.9 Å. (B) Binding in Pocket 2 revealed 3 polar contacts with an average bond length of 2.0 Å. The higher number of polar contacts in pocket 2 suggest it is the superior binding site.

Looking at the pAMT more closely, several key residues were identified for the binding of vanillin and vanillylamine to pAMT: Met-291, Lys-284, and Ile-293. Although it was also identified, we neglect Gly-115, as it has no characteristic side chain that would form a polar interaction with the ligand. As a result it could likely be swapped for another, non-bulky residue

(alanine, for example) without much disruption of the binding site. As a note, this residue is conserved among nearly all sequences in the alignment, and is not specific to *Capsicum*.

After completing the BLAST search, we found similar sequences in several species within *Solanales*, including those from the genera *Solanum*, *Capsicum*, *Nicotiana*, and the closely related *Ipomoea* genus in the *Convolvulaceae* family. The full list of BLAST hits was parsed to remove those below 40% identity, and passed to a script which removed duplicate sequences and fetched the sequence data and concatenated it into a single FASTA file. After aligning in Muscle, the key residues identified in the preceding step were located in the raw alignment data, to assess their conservation among the species present in the BLAST hits.

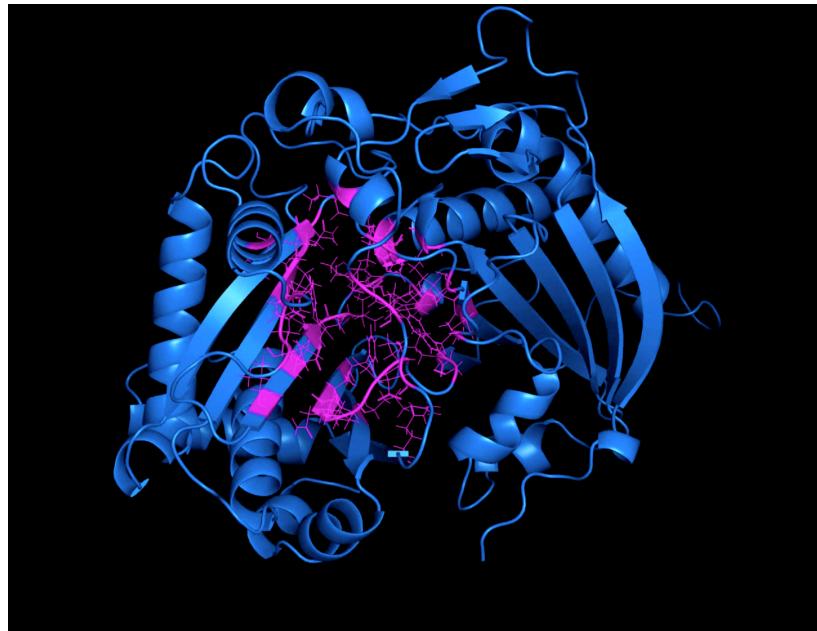
In the cases of all three residues, conservation was observed in sequences outside the *Capsicum* genus, with Met-291 conserved in >70% of all sequences in the alignment, and both Lys-284 and Ile-293 showing >90% conservation. This is likely the result of pAMT having a function less specific to capsaicin biosynthesis than PUN1. Generally, the BLAST search with a pAMT query produced a greater number of hits in a greater diversity of species, indicating that PUN1 is a more likely indicator of pungency in peppers, and a better candidate for inducing this trait in other species.

### **Characterizing PUN1 active sites and phylogeny**

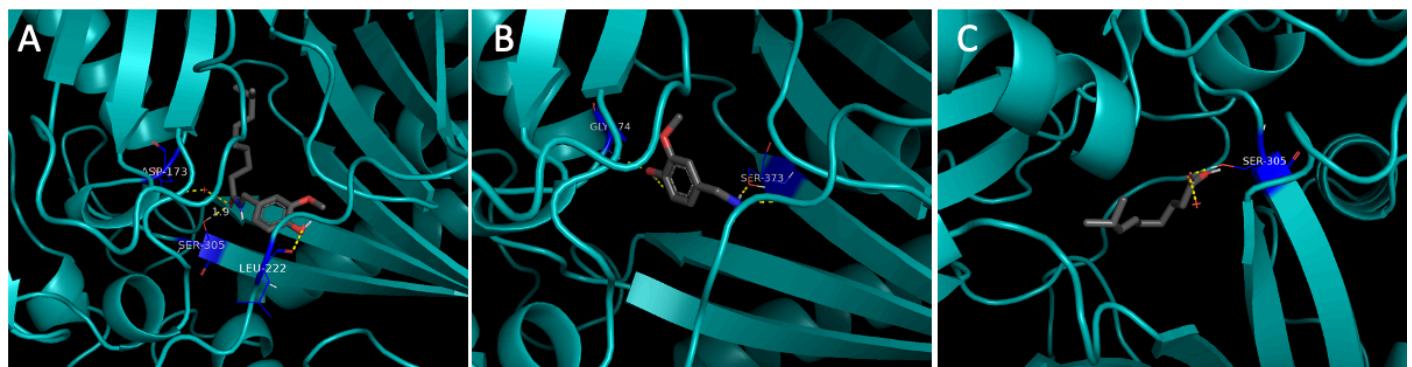
The same analysis was conducted with PUN1 and its associated substrates. Because *p2rank* predicted only one high-scoring putative binding site (score 56.09, compared with pocket 2's 17.39), we skip the step of comparing multiple pockets by assessing their bond lengths (Table 4). The predicted binding site for PUN1 is shown in Figure 8. The docking procedure was carried out for each ligand in the capsaicin biosynthetic pathway, and the visualized results (with binding residues labeled) are shown in Figure 9.

Pocket	Rank	Score	Probability	Residue ID's
pocket1*	1	56.09	0.984	A_139 A_169 A_173 A_174 A_175 A_221 A_222 A_224 A_230 A_285 A_286 A_287 A_305 A_345 A_359 A_36 A_360 A_361 A_369 A_371 A_373 A_374 A_375 A_378 A_380 A_395 A_396 A_399 A_40 A_400 A_401 A_402 A_403 A_405 A_418 A_43 A_45 A_46
pocket2	2	17.39	0.801	A_241 A_244 A_257 A_258 A_259 A_260 A_284 A_285 A_286 A_287 A_290 A_291 A_308 A_331 A_335 A_374 A_376 A_377 A_378
pocket3	3	7.17	0.373	A_281 A_307 A_308 A_309 A_311 A_330 A_333 A_337 A_344 A_347 A_348 A_351 A_352 A_355 A_357 A_361 A_363 A_364
pocket4	4	5.53	0.264	A_159 A_161 A_191 A_385 A_386 A_387 A_52 A_53 A_54 A_56 A_57 A_58 A_66 A_69 A_70 A_73
pocket5	5	5.53	0.264	A_274 A_275 A_276 A_360 A_362 A_367 A_368 A_369 A_401 A_402 A_422 A_424 A_425
pocket6	6	3.96	0.157	A_127 A_130 A_136 A_138 A_225 A_228 A_230 A_397 A_398 A_420
pocket7	7	3.17	0.108	A_175 A_179 A_182 A_287 A_289 A_290 A_298 A_377 A_378 A_379
pocket8	8	2.46	0.067	A_106 A_137 A_142 A_146 A_147 A_15
pocket9	9	2.16	0.051	A_142 A_143 A_145 A_217 A_219 A_42 A_92 A_95
pocket10	10	2.03	0.044	A_123 A_124 A_125 A_126 A_127 A_2 A_229 A_230 A_231 A_232
pocket11	11	2.01	0.043	A_109 A_13 A_14 A_65 A_67 A_68 A_71 A_72 A_75
pocket12	12	1.75	0.032	A_171 A_202 A_204 A_205 A_301 A_302 A_31 A_32 A_37
pocket13	13	1.47	0.02	A_115 A_116 A_117 A_158 A_160 A_391 A_52 A_53

**Table 4. The predicted active pockets of *C. annuum* PUN1.** There were 13 predicted pockets, but pocket 1 was the best based on its high p2rank score (indicated by an asterisk).



**Figure 8. Predicted binding site in *C. annuum* PUN1.** The best binding pocket predicted by p2rank is shown in magenta.

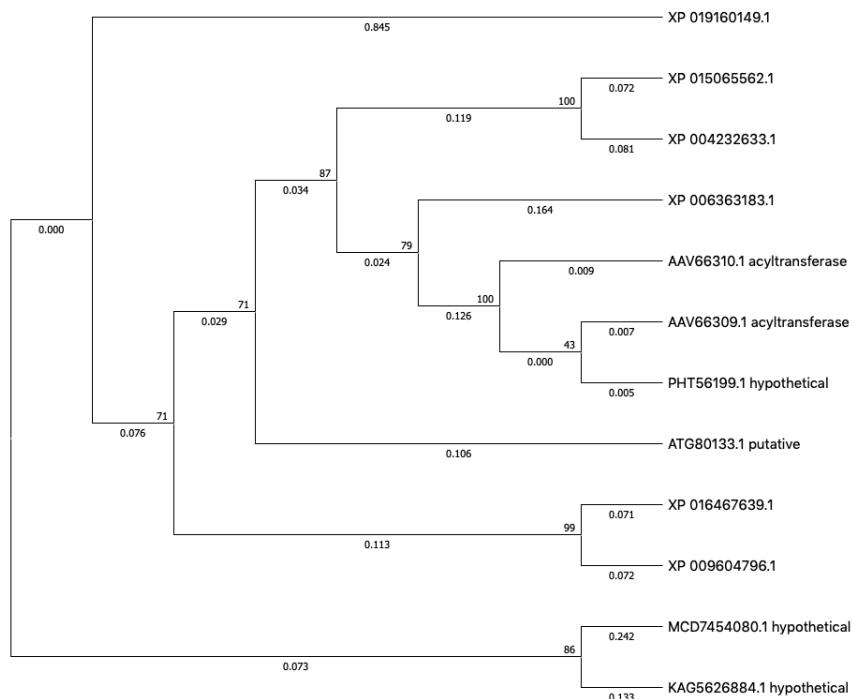


**Figure 9. Docking modeled on *C. annuum* PUN1 with capsaicin precursors and capsaicin.** (A) Capsaicin, Bonding residues: D173, S305, L222. (B) Vanillylamine, Bonding residues: G174, S373. (C)8-Methyl-6-nonenoic acid, Bonding residue: S305. Note: MedusaDock only permits binding of one ligand at a time, so each of the substrates is shown separately.

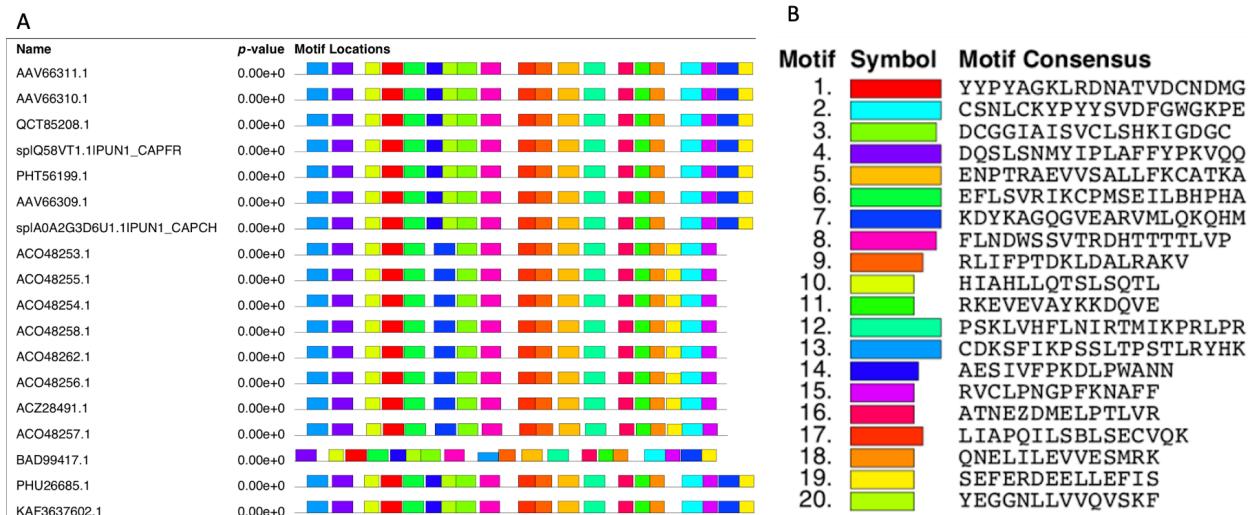
From these results, we identify four key residues involved in binding: Asp-173, Ser-305, Leu-222, Ser-373. We neglect the glycine residue in this case for the same reason as that in the pAMT binding site, outlined above. As a note, this residue is conserved among all sequences in the alignment and is not specific to *Capsicum*, an unsurprising result given its role in catalysis.

Like pAMT, we conducted a BLAST search against the NCBI database and a multiple sequence alignment in Muscle for PUN1 homologs, and used this to study the conservation of key binding residues among species. The BLAST hits included sequences from *Capsicum*, as well as others among species in Solanales. Among these sequences, the Asp-173 residue was conserved most broadly, in several *Capsicum* sequences, as well as more broadly among Solanaceae. The other three key residues all displayed similar patterns of conservation. Ser-305, Ser-373, and Leu-222 were all conserved among *Capsicum*, and found in few species outside this genus, with a few notable exceptions, including *S. lycopersicum*, *S. pennellii*, and *S. melongena*.

In addition to examining the raw alignment output from muscle, we also attempted to build phylogenies for PUN1 in IQTree, and perform motif finding in MEME suite (Figure 11). These results did not prove to be particularly informative beyond the conclusions drawn from the alignment itself. The phylogeny predicted a closer evolutionary relationship between *Capsicum* species and some *Solanaceae*, notably *S. lycopersicum*, *S. pennellii*, and *S. tuberosum*, and a more distant relation to *Nicotiana* and *Ipomoea* (Figure 10). The predicted tree also places *Datura stramonium* and *Solanum commersonii* much further from *Capsicum* than other members of *Solanaceae*, though these sequences are hypothetical proteins, which may contribute to this unexpected result. Motif finding failed to show any motifs unique to *Capsicum* or a similarly narrow subgroup of aligned sequences. Most of the motifs were well conserved among the PUN1 homologs, and did not provide much insight into structural differences between enzymes in different species.



**Figure 10. PUN1 phylogeny.**



**Figure 11. Conserved motifs among PUN1 homologs.**

Due to the limitations of the phylogeny and motif finding for PUN1, these procedures were not carried out for pAMT, and the raw alignment data was analyzed instead. Conclusions drawn from the alignment data are summarized in the discussion below.

## Discussion

To sum, the insect phylogeny revealed the distinct groupings between CYP6B6 and its homologs, and their relationship to other cytochrome p450s in common agricultural pests. However, motif finding revealed the structural similarities and differences that conferred insects with CYP6B6 resistance to capsaicin. Based on the results from the *H. armigera* CYP6B6 investigation, it can be concluded that two unique motifs discovered are involved in capsaicin metabolism and enable the organism to successfully detoxify the compound and resist any of its effects. This also tells us that insect pests with related cytochrome p450 proteins that *do not* contain these unique motifs are likely susceptible to capsaicin, and capsaicin can be used as an insecticide against them. Further testing, especially field testing, will reveal capsaicin's true capabilities.

The findings in this paper complement previous publications well. The unique motifs we found were previously found to be conserved in the CYP6 in a paper by Ranasinghe and Hobbs (1998). In addition, other CYP6B genes have been found to detoxify other plant compounds, such as furanocoumarins (Heckel, 2018). This research came to the conclusion that the CYP6B enzymes expressed in specialist herbivores were more specific than those in generalist herbivores (Li, Schuler and Berenbaum, 2003; Heckel, 2018). Li et al. (2004) concluded that generalist defense

proteins are able to detoxify a wider range of chemicals due to their less-specific binding patterns. A particularly interesting example of generalist versus specialist enzymes occurs in the *Papilio* genus where a two-residue mutation in the generalist prevents the metabolism of furanocoumarins (Li, Schuler and Berenbaum, 2003). This relates closely to our finding since these key residues could be the determining factor that allows capsaicin metabolism. Looking more closely at the conservation of these residues, particularly at the molecular level, is an important step to continue this research.

An additional avenue of research involves UDP-glycosyltransferases (UGT). Besides CYPs, UGTs have been cited in many insect studies as possible proteins involved in detoxifying plant compounds. It is thought that UGT is able to metabolize capsaicin and make it easier for insects, particularly *Helicoverpa* species, to sequester and/or excrete the compound (Ahn et al., 2011; Krempl et al., 2016). The same procedure used for CYP8B6 in this paper could be used to examine UGT as well.

Overall, the analysis of pAMT and its homologs across species proved less informative than PUN1. Key active site residues identified via docking were found to be highly conserved among species outside *Capsicum*. Therefore *pAMT* is then a less definite genetic marker of a pungent phenotype, and less useful in hopes of creating a modified species that could synthesize its own capsaicin as a pest deterrent. In future iterations of this project, pursuing a motif-finding approach may prove more illuminating. At minimum, we predict that, due to the number of BLAST hits among a greater number of species, pAMT should not prove a limiting factor in inducing capsaicin synthesis, and propose PUN1 as a more appropriate target.

In light of the docking and alignment results, we determined that three residues of PUN1 (Ser-305, Ser-373, and Leu-222) were all highly conserved among *Capsicum* species, with limited presence in homologs outside of this genus. Three notable exceptions were found in *Solanum lycopersicum*, *S. pennellii*, and *S. melongena*. As a result, we postulate that these three residues are likely responsible for the function of PUN1 as a catalyst for the last step of capsaicin synthesis, since they are relatively limited to *Capsicum* compared with homologs, and were all determined to interact with the substrates by the docking procedure. Additionally, we predict that the three species of *Solanum* which share these residues would be the best candidates for induced capsaicin synthesis as a method of deterring pests, either by mutagenesis or genetic modification to insert a *Capsicum* homolog into the genome.

Other research groups have also looked into the prospect of spicy crops, particularly tomato. Although PUN1 is not found in tomato, the phenylpropanoid pathway and the majority of other enzymes are present in *S. lycopersicum* (Kim et al., 2014; Naves et al., 2019). This suggests that genetic engineering, particularly to upregulate genes involved in the phenylpropanoid and

branched-chain fatty acid pathways, could result in capsaicin production outside of *Capsicum* (Naves et al., 2019).

Beyond our study, several approaches would be useful in further characterizing the role of these amino acids in the enzymatic function of PUN1 as capsaicin synthase. Mutating the conserved residues in *Capsicum PUN1* and *pAMT* genes would provide greater insight into the importance of each residue in catalysis. Previous studies have attributed a loss of pungency in pepper to mutations in these two genes, but it is not clear if those mutations were at the residues we have identified (Tanaka et al., 2010; Reddy et al., 2014).

Furthermore, visualization of protein structures of PUN1 and docking with these models in other species would better characterize the differences in the active sites of homologous proteins found via BLAST. An immediate next step could involve using the pepper PUN1 to homology model homologous *Solanaceae* sequences and observe how capsaicin and the precursors dock. Studies of different alkaloids, particularly from the vanilloid family which capsaicin is a part of, or other compounds known to act as evolutionary defenses to herbivory would serve to provide greater context for these mechanisms more broadly and narrow down better candidates for deterrents in other species (Kaiser, Higuera and Goycoolea, 2017).

More general questions surrounding this project include the effect capsaicin could have on beneficial insects, as well as both beneficial and harmful microbes. In order for capsaicin to be used widely, it would be best for beneficial organisms to be resistant.

A perfect result from this study would be a tomato that produces capsaicin in its leaves that herbivores eat and die. There is a long way to go, but this work has confirmed that there is a way to identify insects that are likely resistant to capsaicin (via protein motifs) and that some crops may be prime candidates for inducing capsaicin production as a pest deterrent.

## Acknowledgments

We would like to thank the course staff for teaching us many of the techniques used in this project and for their guidance during the project. We would also like to thank our peers in the class for their feedback and questions at various stages of this paper.

## References

- Aguilar-Meléndez, A., Morrell, P. L., Roose, M. L., & Kim, S. C. (2009). Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *American Journal of Botany*, 96(6), 1190-1202.
- Ahn, S. J., Badenes-Pérez, F. R., & Heckel, D. G. (2011). A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. *Journal of insect physiology*, 57(9), 1212-1219.
- Ahn, S. J., Badenes-Pérez, F. R., Reichelt, M., Svatoš, A., Schneider, B., Gershenson, J., & Heckel, D. G. (2011). Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three *Helicoverpa* species. *Archives of Insect Biochemistry and Physiology*, 78(2), 104-118.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *J. Mol. Biol.* 215:403-410.
- B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37:1530-1534.  
<https://doi.org/10.1093/molbev/msaa015>
- Donald Hobern from Canberra, Australia, CC BY 2.0  
<https://creativecommons.org/licenses/by/2.0/>, via Wikimedia Commons
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797.
- Glen Stecher, Koichiro Tamura, and Sudhir Kumar (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* (<https://doi.org/10.1093/molbev/msz312>)
- Han, K., Jeong, H. J., Sung, J., Keum, Y. S., Cho, M. C., Kim, J. H., ... & Kang, B. C. (2013). Biosynthesis of capsinoid is controlled by the Pun1 locus in pepper. *Molecular breeding*, 31(3), 537-548.
- Heckel, D. G. (2018). Insect detoxification and sequestration strategies. *Annual Plant Reviews Online*, 77-114.
- Jiang, Z., Kai, T., & Reilly Christopher, A. (2020). Capsaicinoid metabolism by the generalist *Helicoverpa armigera* and specialist *H. assulta*: Species and tissue differences.
- Kaiser, M., Higuera, I., & Goycoolea, F. M. (2017). Capsaicinoids: occurrence, chemistry, biosynthesis, and biological effects. *Fruit and Vegetable Phytochemicals: Chemistry and Human Health*, 2nd Edition, 499-514.
- Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., ... & Choi, D. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature genetics*, 46(3), 270-278.
- Krempl, C., Sporer, T., Reichelt, M., Ahn, S. J., Heidel-Fischer, H., Vogel, H., ... & Joußen, N.

- (2016). Potential detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth species *Helicoverpa armigera* and *Heliothis virescens*. *Insect biochemistry and molecular biology*, 71, 49-57.
- Krivák, R., Hoksza, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminform* 10, 39 (2018).  
<https://doi.org/10.1186/s13321-018-0285-8>
- Li, W., Schuler, M. A., & Berenbaum, M. R. (2003). Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: specificity and substrate encounter rate. *Proceedings of the National Academy of Sciences*, 100(suppl 2), 14593-14598.
- Li, X., Baudry, J., Berenbaum, M. R., & Schuler, M. A. (2004). Structural and functional divergence of insect CYP6B proteins: from specialist to generalist cytochrome P450. *Proceedings of the National Academy of Sciences*, 101(9), 2939-2944.
- Li, B., Yang, M., Shi, R., & Ye, M. (2019). Insecticidal activity of natural capsaicinoids against several agricultural insects. *Natural Product Communications*, 14(7), 1934578X19862695.
- Li, Y., Bai, P., Wei, L., Kang, R., Chen, L., Zhang, M., ... & Liu, W. (2020). Capsaicin functions as *Drosophila* ovipositional repellent and causes intestinal dysplasia. *Scientific Reports*, 10(1), 1-11.
- Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. *Nucleic acids research*, 33(suppl\_1), D54-D58.
- Maxim Tatarinov. [https://www.123rf.com/photo\\_126276269\\_kashmiri-mirch-pepper-capsicum-anuum-fruit-isolated.html](https://www.123rf.com/photo_126276269_kashmiri-mirch-pepper-capsicum-anuum-fruit-isolated.html)
- McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).
- Milenković, A. N., & Stanojević, L. P. (2021). Black pepper: chemical composition and biological activities. *Advanced Technologies*, 10(2), 40-50.
- Naves, E. R., de Ávila Silva, L., Sulpice, R., Araújo, W. L., Nunes-Nesi, A., Peres, L. E., & Zsögön, A. (2019). Capsaicinoids: pungency beyond Capsicum. *Trends in plant science*, 24(2), 109-120.
- Ogawa, K., Murota, K., Shimura, H., Furuya, M., Togawa, Y., Matsumura, T., & Masuta, C. (2015). Evidence of capsaicin synthase activity of the Pun1-encoded protein and its role as a determinant of capsaicinoid accumulation in pepper. *BMC plant biology*, 15(1), 1-10.
- Prasad, B. N., Kumar, V., Gururaj, H. B., Parimalan, R., Giridhar, P., & Ravishankar, G. A. (2006). Characterization of capsaicin synthase and identification of its gene (csy1) for pungency factor capsaicin in pepper (*Capsicum* sp.). *Proceedings of the National Academy of Sciences*, 103(36), 13315-13320.
- Ranasinghe, C., & Hobbs, A. A. (1998). Isolation and characterization of two cytochrome P450

cDNA clones for CYP6B6 and CYP6B7 from *Helicoverpa armigera* (Hubner): possible involvement of CYP6B7 in pyrethroid resistance. Insect biochemistry and molecular biology, 28(8), 571-580

Reddy, U. K., Almeida, A., Abburi, V. L., Alaparthi, S. B., Unsel, D., Hankins, G., ... & Nimmakayala, P. (2014). Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in *Capsicum annuum* L. collections. PloS one, 9(1), e86393.

Särkinen, T., Bohs, L., Olmstead, R. G., & Knapp, S. (2013). A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC evolutionary biology, 13(1), 1-15.

Shi, Y., Jiang, Q., Yang, Y., Feyereisen, R., & Wu, Y. (2021). Pyrethroid metabolism by eleven *Helicoverpa armigera* P450s from the CYP6B and CYP9A subfamilies. Insect Biochemistry and Molecular Biology, 135, 103597.

Stellari, G., Mazourek, M. & Jahn, M. Contrasting modes for loss of pungency between cultivated and wild species of Capsicum. Heredity 104, 460–471 (2010).

<https://doi.org/10.1038/hdy.2009.131>

Stewart Jr, C., Kang, B. C., Liu, K., Mazourek, M., Moore, S. L., Yoo, E. Y., ... & Jahn, M. M. (2005). The Pun1 gene for pungency in pepper encodes a putative acyltransferase. The Plant Journal, 42(5), 675-688.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T., & Yazawa, S. (2010). Novel loss-of-function putative aminotransferase alleles cause biosynthesis of capsinoids, nonpungent capsaicinoid analogues, in mildly pungent chili peppers (*Capsicum chinense*). Journal of agricultural and food chemistry, 58(22), 11762-11767.

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Tian K, Zhu J, Li M, Qiu X. Capsaicin is efficiently transformed by multiple cytochrome P450s from Capsicum fruit-feeding *Helicoverpa armigera*. Pestic Biochem Physiol. 2019 May;156:145-151. doi: 10.1016/j.pestbp.2019.02.015.

Timothy L. Bailey, James Johnson, Charles E. Grant, William S. Noble, "The MEME Suite", Nucleic Acids Research, 43(W1):W39-W49, 2015.

Tomita, M., & Endo, H. (2007). Using capsaicin as a less toxic insecticide. In Combined Proceedings International Plant Propagators' Society (Vol. 57, pp. 728-732).

UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.

Wang J, Dokholyan NV. MedusaDock 2.0: Efficient and Accurate Protein-Ligand Docking With Constraints. J Chem Inf Model. 2019 Jun 24;59(6):2509-2515. doi: 10.1021/acs.jcim.8b00905

Waters, L. C., Zelhof, A. C., Shaw, B. J., & Ch'ang, L. Y. (1992). Possible involvement of the

long terminal repeat of transposable element 17.6 in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. *Proceedings of the National Academy of Sciences*, 89(11), 4855-4859.