Docking Methods
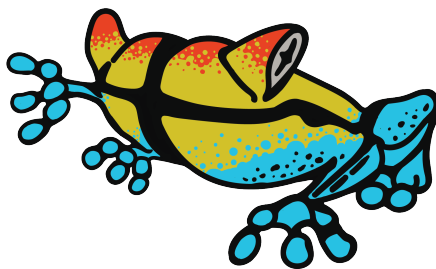
Synthetic Biobots

Gerardo Cendejas Mendoza

iGEM Design League 2022

Introduction

In this document is shown the docking methodology followed by the team Synthetic Biobots during the iGEM Design League 2022 season. Molecular docking is a process to couple proteins with their ligands; for this project molecular docking was performed in order to support the decision of using the proteins that were selected as part of the biosynthetic pathway of piperamide. The different steps necessary for its production are shown in the figure shown in the next page; in this figure all the steps required are marked with a number, this number is an identifier for the enzyme required to perform that step.

From all the enzymes used in the pathway, only 8 require a docking step for different reasons, these reasons can be that it is an enzyme that has not been isolated and even though a previous annotation step indicates it performs that reaction, molecular docking can provide better support for their selection and activity. Another reason to perform molecular docking is when the protein is known and has been isolated but we are using a molecule that is not the original ligand of the protein, in this case molecular docking can help to elucidate if the protein would be able to perform the reaction with a molecule with similar structure to its' original ligand. The enzymes that are subject of molecular docking and the genes that code for these enzymes and will be studied are: 1 (Pn8.2617), 2 (Pn2.84), 4 (Pn1.1317), 5 (Pn3.4770, Pn7.1626), 6 (Pn16.1237, Pn16.1198), 8 (Pn4.3222), 9 (Pn2.2377) and 10 (2cwh).

For each of these genes it was performed a differential exon usage analysis between four different *Piper nigrum* tissues: fruit at 20 days stage, fruit at 40 days stage, leaf and panicle. This analysis can help us to visualize the expression level for each exon of the genes, with this information we can select the exons of the gene to include in the modeling and in the next steps. This analysis was performed using the Bioconductor library "DEXSeq", to see the code of this methodology please visit the GitHub repository of the team for iGEM Design League 2022 season, the Exon_usage directory.[6;1;9] In order to carry out the analysis it was necessary to map the RNAseq reads to the reference genome of *Piper nigrum*, this was done using the STAR ultrafast universal RNA-seq aligner on the Galaxy project platform.[7;5;4] Only results of relevance for the modeling and docking methods are shown in this document.
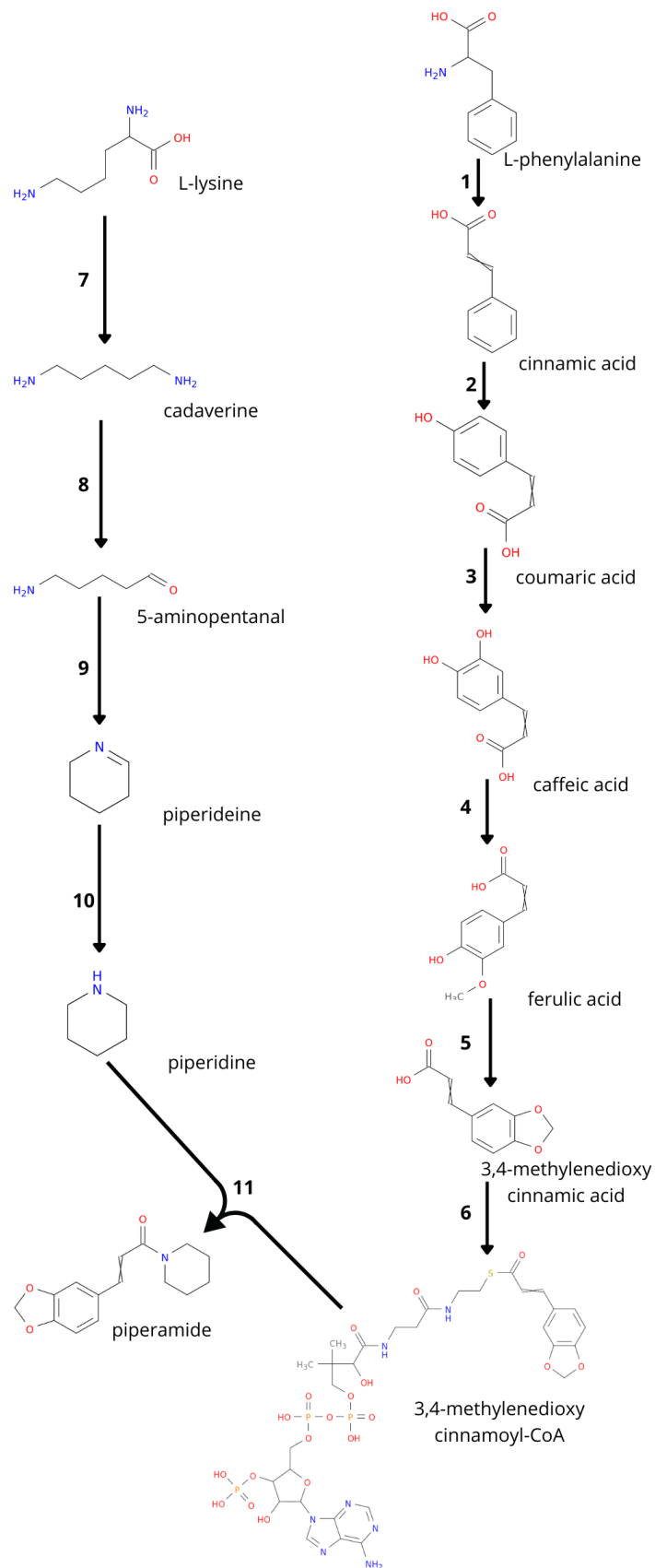
Biosynthetic pathway of piperamide.

1 Docking

1.1 Pn8.2617

The first reaction in the biosynthetic pathway of piperamide is performed by the enzyme phenylalanine ammonia-lyase, wich is part of the phenylpropanoid biosynthesis pathway and catalyzes the reaction shown in Eq. 1.1. The gene Pn8.2617 was selected as the best candidate for enzyme 1, this selection was based on the fact that it obtained the highest score (1170.9) over the threshold (524.93) for that KEGG Orthology ID K10775, with E-value $\approxeq 0$. This annotation was made with KofamKOALA.[2] With the Exon usage analysis no relevant results were found.

$$\text{Phenylalanine} \xrightarrow{\quad 1 \quad} \text{Cinnamic acid} + NH_3 \text{ (Ammonia)} \tag{1.1}$$
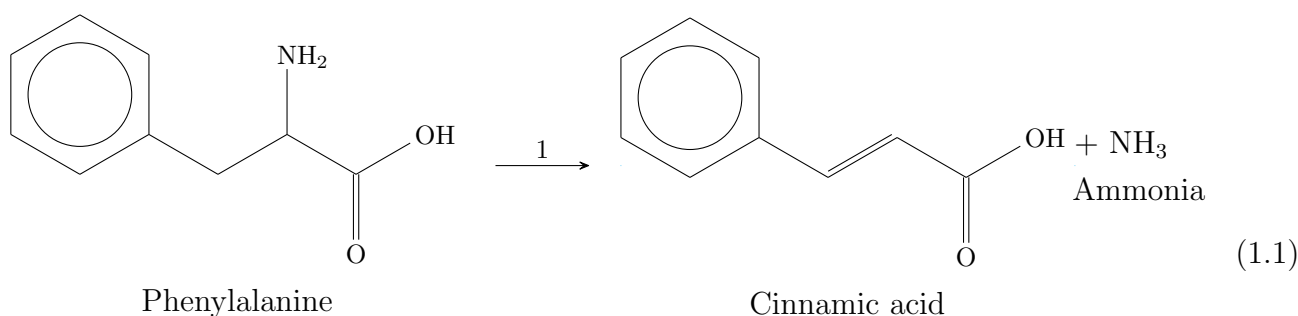
The gene contains two exons that were considered when modeling the three-dimensional structure of the protein. The structure modeling was made using an homology based methodology "SWISS-MODEL".[11;3] This methodology is template based, in this case the template selected was a phenylalanine ammonia-lyase from *Petroselinum crispum* with PDB ID: 1w27; this protein has a 79.97% sequence identity to our query protein Pn8.2617 which make it an appropiate template. The template was found using HHblits, a protein sequence searching algorithm that uses hmm-hmm alignment.[8] The model (Fig. 1) is found in a homo-tetrameric state, as the template is; it has a GMQE value of 0.84 and a QMEANDisCo Global value of 0.81 ± 0.05 (Fig. 2). The QMEANDisCo value is a composite score value for single model quality estimation, this value is a function of underlying single model scores employing statistical potentials of mean force and a distance constraint based on the known structure of homologous proteins. QMEANDisCo value is calculated for every residue of the protein and the QMEANDisCo Global value is the average of these per residue scores, wich aims to be a good approximation of the global overall quality of the model.[10]
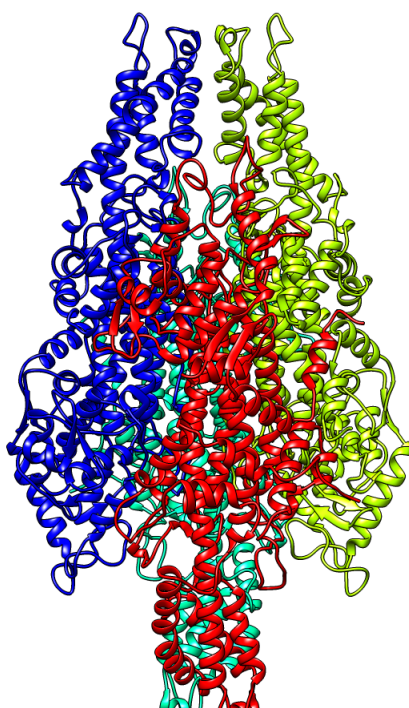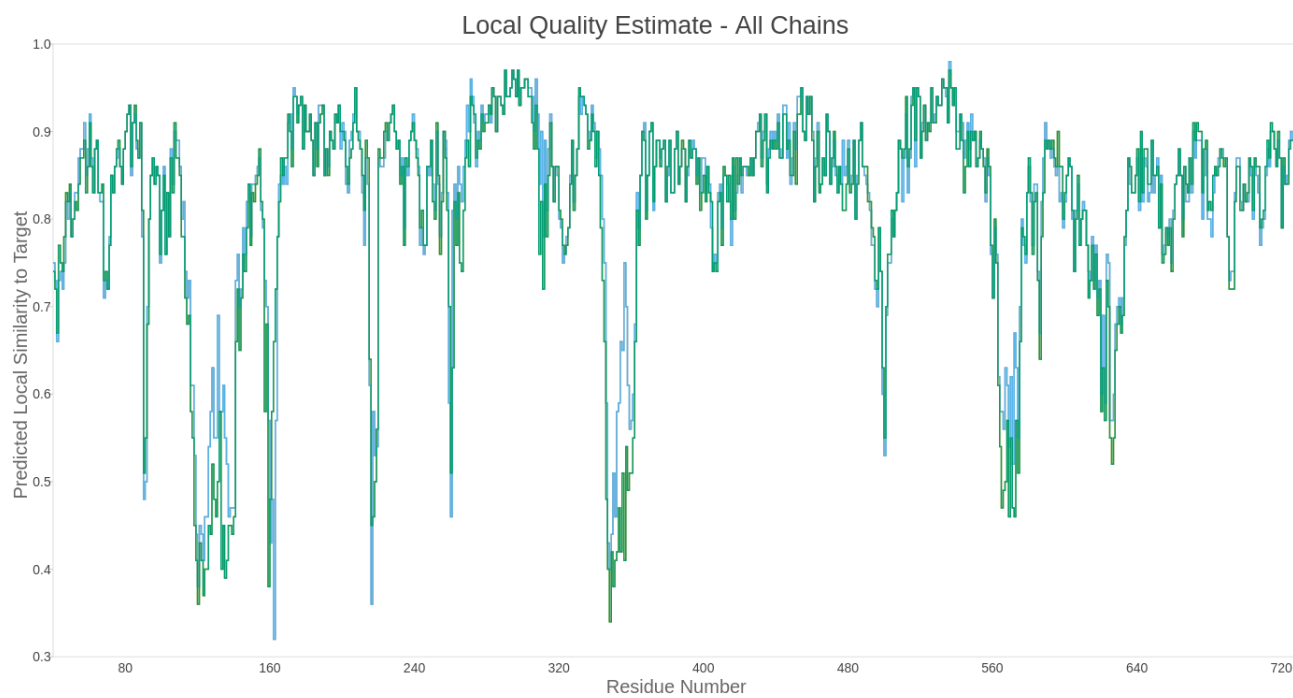
Figure 1: Pn8.2617 three-dimensional model

Figure 2: Pn8.2617 three-dimensional model

References

[1] Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from rna-seq data. *Nature Precedings*, pages 1–1.

[2] Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). Kofamkoala: Kegg ortholog assignment based on profile hmm and adaptive score threshold. *Bioinformatics*, 36(7):2251–2252.

[3] Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., and Schwede, T. (2017). Modeling protein quaternary structure of homo-and hetero-oligomers beyond binary interactions by homology. *Scientific reports*, 7(1):1–15.

[4] Community, T. G. (2022). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1):W345.

[5] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21.

[6] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):1–16.

[7] Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., Wu, H., Qin, X., Yan, L., Tan, L., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nature communications*, 10(1):1–11.

[8] Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2012). Hhblits: lightning-fast iterative protein sequence searching by hmm-hmm alignment. *Nature methods*, 9(2):173–175.

[9] Reyes, A., Anders, S., Weatheritt, R. J., Gibson, T. J., Steinmetz, L. M., and Huber, W. (2013). Drift and conservation of differential exon usage across tissues in primate species. *Proceedings of the National Academy of Sciences*, 110(38):15377–15382.

[10] Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., and Schwede, T. (2020). Qmeandisco—distance constraints applied on model quality estimation. *Bioinformatics*, 36(6):1765–1771.

[11] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., et al. (2018). Swiss-model: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1):W296–W303.