Functional Annotation of Laccase from *Pleurotus eryngii* with Comparative Modelling

Alessandro Caula

Bioinformatics Master Degree Course, University of Bologna, Bologna, Italy

INTRODUCTION

Laccases (EC 1.10.3.2) are enzymes that catalyse oxidoreductase reactions, they belong to the blue multicopper oxidases (bmCuO) family[1]. They oxidise a broad range of substrates, in particular phenolic compounds, by the reduction of molecular oxygen in water molecules that is accompanied with a one-electron oxidation of the substrates[2].
The *T. versicolor* Laccase (TvL) has a monomeric structure, organized in three sequentially arranged domain, composed mainly of β-sheets, α-helix and β-barrel motifs. The active site of TvL is characterized by the presence of four total copper atoms. The copper in the Type-1 (T1) site is coordinated by two histidines (His-395 and His-458) and one cysteine (Cys-453) and it is the primary electron acceptor. The Type-2 and Type-3 coppers, where the reduction of molecular oxygen takes place, are arranged in a trinuclear cluster, the two coppers of the T3 site are coordinated to six histidine while two histidine residues and an oxygen ligand coordinate the T2 copper[2.1] (**Figure 1**)
The goal of this study is to establish the possibility to transfer by homology modelling the GO terms from the highly annotated and well-known *T. versicolor* Laccase to the poorly annotated Laccase produced by the *P. eryngii* fungi.

METHODS

Databases

The sequences, the annotations and other useful feature of our template and target are both found in the UniProt[3] database (release October 15, 2019).
The sequence and the 3D structure of our template is retrievable on the Preotein Data Bank (PDB)[4] database (release 2019_05).

Computational methods

To find the template we run the target sequence in the BLASTp[5] algorithm tool that is present in UniProt. This algorithm retrieves all the similar sequences from the whole UniProt database.
For the global sequence alignment between the target and the template we used Lalign[6] program, a dynamic alignment tool capable of doing both global and local alignment.
The putative structure of the target was obtained with Modeller[7] (version 9.23), a program that implements comparative protein structure modelling by satisfaction of spatial restraint.
Chimera[8] (version 1.14) was used for the graphical visualization and analysis of the three-dimensional structure of the target, template and the superimposition between the two.
To evaluate the quality of the model we used jCE[9,10,11] (version 3.0.8), a Java implemented programme that calculates the pairwise structure alignment.
To evaluate the stereochemical quality of the protein structure Procheck[12] was used to the Ramachandran plot.

TEMPLATE SELECTION

We used the UniProt database and its build in BLASTp tool to find the best template, putting the sequence of our target protein (B0JDP9) as a query.
The following parameters were changed before running BLAST: Target database with

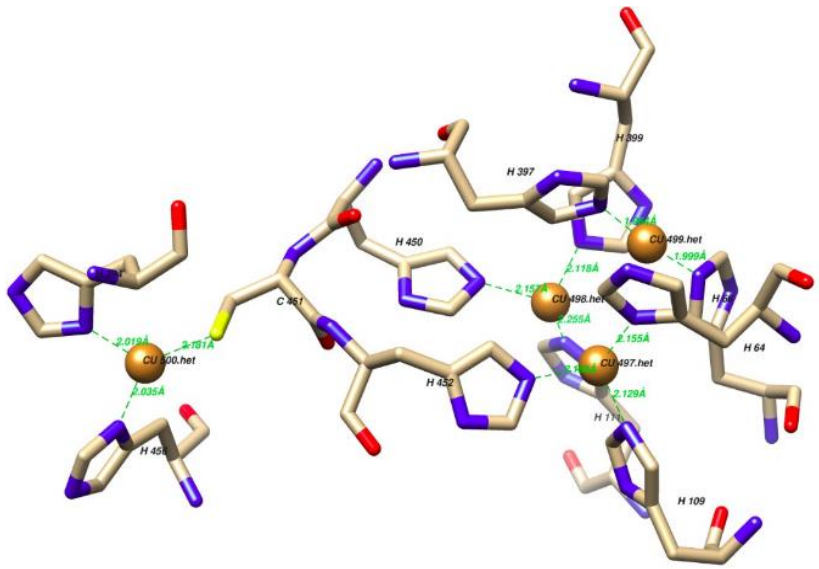

Figure 1. Chimera image of the active site of the model 2 with relative residues and distances.

3D structure (PDB); the E-Threshold set to 0.01, to obtain only the more significant matches; as the matrix score, we choose the BLOSUM-62, that is very good for longest sequences like ours (B0JDP9).
The template must be chosen according to three major criteria: the highest coverage of the target, the highest resolution of the 3D structure of the protein and the high sequence identity between the target and the template.
Once the results of BLAST are shown it's important to further limit the results to the reviewed proteins from Swiss-Prot (version) and of course also to the proteins that have a 3D structure, since we have to do comparative modelling.
Since the high sequence identity of 62.0%, the high resolution (1.9 Å) for the X-ray resolved structure and the presence of all the four copper atoms in the active site, we chose as a template the Laccase-2 from *T.versicolor.*
The sequence of this Laccase used for the model was the one that is retrievable from the PDB database with the entry 1GYC.

SEQUENCE ALIGNMENT

The next step was to obtain a good alignment between the target and the template for Modeller to work with, this was made using the global alignment algorithm present in Lalign program (**Supplementary Figure 1**). The parameter setting for the alignment were: global as the alignment method; BLOSUM62 as the scoring matrix; the opening gap penalty kept as default; and the E-value threshold set to 0.01.

The result of the alignment highlights also that the sequence identity between the two is 56.8%, that is greater than the 30% threshold, and therefore we can proceed with the construction of the model.

MODELLER

Before using Modeller is important to transform the alignment from Lalign, in a .pir format file, modify it and then use it as the input for our modelling procedure.
From the .pir format file manually delete the first 20 residues, known as the signal peptide, from the target sequence and consequently the 20 gaps at the beginning of 1GYC sequence. Then is needed to remove the last 16 gaps plus the Glutamine (Q), that should be in the position 499 of the template sequence and the last 17 residues from B0JDP9. After that at the end of both the sequences add 4 dots "." for the hetero atoms of copper and the "*" that mark the end of the sequence.
The .pdb file of the template was also modified. The records containing the atom coordinates of Gln499 were deleted and the coordinates of the 4 atoms of copper were renumbered (starting from the 499 till the 502 position).

Last thing to do before running modeller was to specify on the model.py script the code of the target and the template, the name of the .pir file where there are the two sequences and the number of models that we wanted to obtain.
Seven different models were generated with modeller. In order to choose the best one, is essential to look at the smallest DOPE score and the smallest molpdf in the result table shown after the Modeller run. Very important is also the RMSD (Root Mean Square Deviation) given by jCE structural superimposition and the quality of the Ramachandran plot.
As shown in **Table 1** the smallest molpdf value and DOPE score, the small RMSD and the good percentage of favoured residues in the favoured regions in the Ramachandran plot led the choice to the model number 2 as the best one.

Model number	molpdf value	DOPE score	% of Favoured	RMSD (Å)
1	3602,29688	-57239,14062	90,0	0,26
2	3328,70459	-57901,96875	91,5	0,26
3	3850,48364	-57815,37109	91,2	0,30
4	3672,39355	-58051,73828	89,3	0,27
5	3412,36499	-57943,47266	91,5	0,26
6	3788,18457	-57628,60547	88,8	0,28
7	3521,69580	-58045,62500	90,7	0,26

Table 1. In this table there are all the values and scores from Modeller, jCE and Procheck that we kept in consideration for the selection of the best model.

TARGET ANNOTATION

jCE was used to do the structure superimposition between the model 2 and the template 1GYC (**Figure 2)**.
The RMSD, the measure of the average distance between the atoms of the superimposed proteins, is very low and that is a visible sign of the goodness of the model.
Also, the z-score that is 8.30 statistically confirms the significance of the structural superimposition of the model.

Figure 2. jCE pairwise sequence alignment from structural superimposition of the template towards the model.

Moreover, the superimposition with jCE also shows the conservation of several important structural residues, like the residues present in the disulfide bridges and the residues that participate in the copper ligands in the active site.Using Chimera, we could confirm once again the good quality of the selected model and the conservation of its active site carefully comparing the structure of the model with the template. (**Figure 3**)

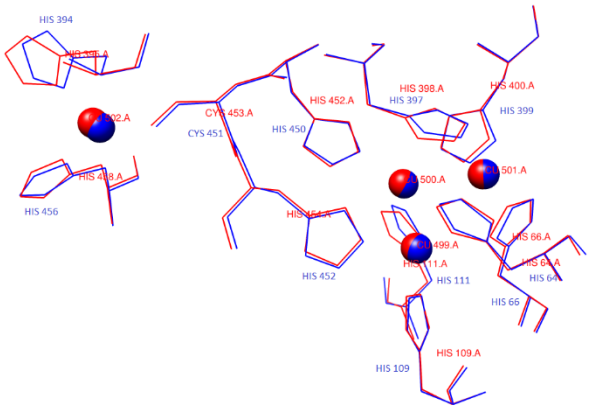

Figure 3. Chimera image of the superimposition of the active site of the model 2 (blue) and the template (red).

More in detail as shown in **Table 2**, all the distances, picked with Chimera, between the histidine involved in the T2 and T3 copper ligands, the His-Cys-His residues coordinating the T1 copper and the Cysteines that participate in the disulfide bridge are well conserved (**Supplementary Figure 2**).

	Position in the Template	Position in the model	Residues	INFO	Distance in Template	Distance in Model
Metal Binding	64	64	His	CU1	2,01	2
	66	66	His	CU2	2,15	2,15
	109	109	His	CU2	2,12	2,13
	111	111	His	CU3	2,23	2,26
	395	394	His	CU4	2,02	2,02
	398	397	His	CU1	1,97	1,96
	400	399	His	CU3	2,12	2,12
	452	450	His	CU3	2,16	2,16
	453	451	Cys	CU4	2,19	2,18
	454	452	His	CU2	2,17	2,19
	458	456	His	CU4	2,04	2,18
S-S	105/508	85/488	85/486	Cys/Cys	1,98	2,04
	137/225	117/205	117/204	Cys/Cys	2,04	2,02

Table 2. Table with all the distances between a specific Cu-Type and the relative coordinate residues

In addition, the analysis of the Ramachandran plot generates with Procheck with the 91.5% of residues in the most favourite region, reinforces the good quality of the model, even if there are three residues in non-favourite regions, but that anyway don't influence the structure and function of the active site (**Supplementary Figure 4**).
Thus, we can confirm that this is a very high-quality model, and that the general oxidoreduction function of the template can be ascribed to this model.

DISCUSSION

The aim of the homology modelling is to functionally annotate a protein for which we only know the sequence through structure comparison with another known entity with high sequence identity. The final goal is to establish whether is legitimate to transfer GO terms, all or part of them, from the template to the target. After the structure analysis with Chimera and with jCE it's possible to conclude that the target and the template are very closely related and for these reasons it's possible to say that the target respect all the parameters to be considered as a member of the Laccase family. Therefore, it is possible to transfer all the following experimentally validated GO terms that are included in the molecular function properties of the protein from the *T. versicolor* Laccase to the one of *P. eryngii:*

- Copper ion bonding (GO:0005507);
- Hydroquinone: oxygen oxidoreductase activity (GO:0052716);
- Lignin catabolic process (GO:0046274);
- Extracellular region (GO:0005576).

The last GO term is transferable because of the presence of the signal peptide and the conservation of the disulfide bridge that's characteristic of the secreted protein.

REFERENCES

1. Messerschmidt, Albrecht. *Multi-copper oxidases*. World Scientific, 1997
2. Piontek, Klaus, Matteo Antorini, and Thomas Choinowski. "Crystal structure of a laccase from the fungusTrametes versicolor at 1.90-Å resolution containing a full complement of coppers." *Journal of Biological Chemistry* 277.40 (2002): 37663-37669.
3. UniProt Consortium. "UniProt: a hub for protein information." *Nucleic acids research* 43.D1 (2014): D204-D212. (https://www.uniprot.org/)
4. Berman, Helen M., et al. "The protein data bank." *Acta Crystallographica Section D: Biological Crystallography* 58.6 (2002): 899-907. (https://www.rcsb.org/)
5. Altschul, Stephen F., et al. "Basic local alignment search tool." Journal of molecular biology 215.3 (1990): 403-410.
6. Huang, Xiaoqiu, and Webb Miller. "A time-efficient, linear-space local similarity algorithm." Advances in Applied Mathematics 12.3 (1991): 337-357. (https://embnet.vital-it.ch/software/LALIGN_form.html)
7. Šali, Andrej, and Tom L. Blundell. "Comparative protein modelling by satisfaction of spatial restraints." Journal of molecular biology 234.3 (1993): 779-815.
8. Pettersen, Eric F., et al. "UCSF Chimera—a visualization system for exploratory research and analysis." Journal of computational chemistry 25.13 (2004): 1605-1612.
9. Prlić, Andreas, et al. "Pre-calculated protein structure alignments at the RCSB PDB website." Bioinformatics 26.23 (2010): 2983-2985.
10. Shindyalov, Ilya N., and Philip E. Bourne. "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path." Protein engineering 11.9 (1998): 739-747.
11. Ye, Yuzhen, and Adam Godzik. "Flexible structure alignment by chaining aligned fragment pairs allowing twists." Bioinformatics 19.suppl_2 (2003): ii246-ii255.

12. Laskowski, Roman A., et al. "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography 26.2 (1993): 283-291.

Figure1. Global alignment of the target (B0JDP9) and the template (1GYC) sequences using Lalign program. Where: ":" means that the residues of the alignment are the same, "." That the residues are similar.

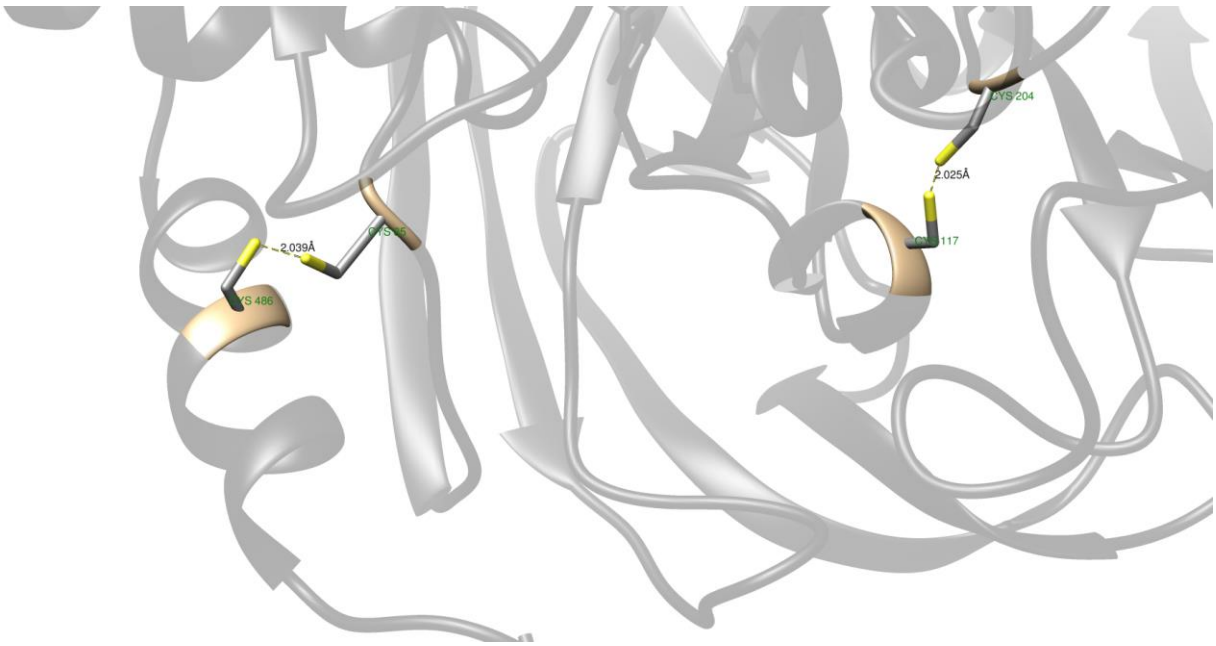

Figure 2 Image showing the disulfide bridges that occurs between the Cys-486 – Cys-85 and the Cys-117 – Cys-205 in the model 2.

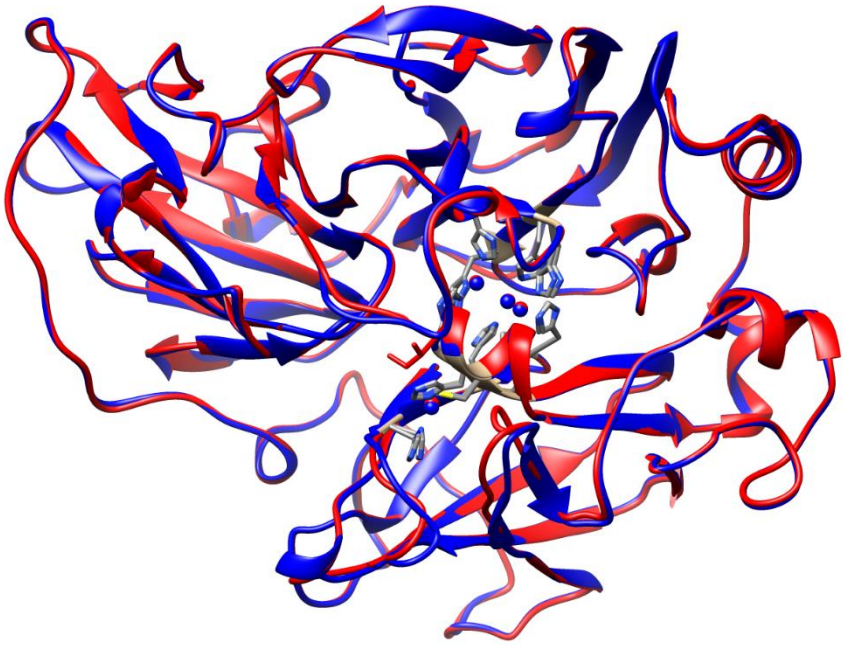

Figure 3. Structural superimposition of the model 2 (blue) with the template (red) (1GYC)

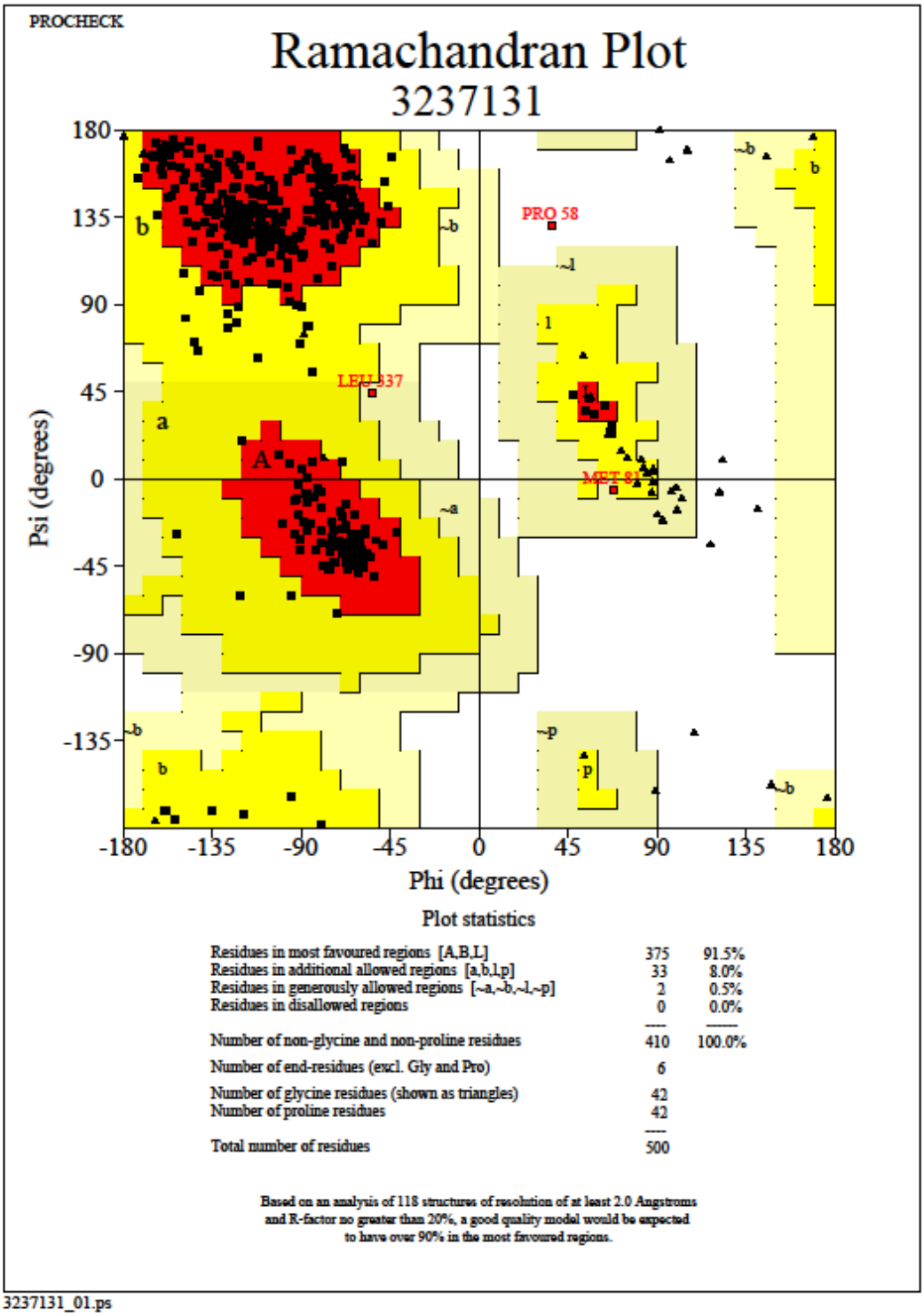

Figure 4. Ramachandran plot of the model 2 executed with Procheck program.