FUNCTIONAL ANNOTATION OF LACCASE FROM THE FUNGUS *PLEUROTUS ERYNGII* WITH COMPARATIVE MODELLING

Alessia Campo

Bioinformatics Master's degree Course, University of Bologna, Bologna, Italy

INTRODUCTION

Laccases (EC 1.10.3.2) belong to the family of blue copper oxidoreductases. These enzymes catalyse the oxidation of the substrate and the reduction of the catalytic multicopper site in which occurs the reduction of molecular oxygen in water molecules. In plants they are mainly involved in lignin biosynthesis, conversely in bacteria and fungi they could be involved in lignin degradation [1].

The laccase from *T. versicolor* is involved in lignin degradation and detoxification of its derived products [1]. It has a monomeric structure made up of three different domains. The secondary structure of the protein consists mainly of β-sheets, α-helix and β-barrel motifs. The enzymatic reaction led by the Laccase from *T. versicolor* (TvL) on its oxidised form is shown in the Figure 1.

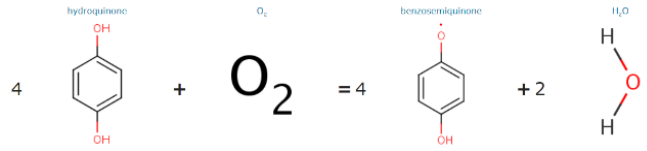


Figure 1. *Chemical reaction led by Laccase from T.versicolor.*

The active site of Laccases generally comprises one copper atom. In the multicopper oxidase TvL (Uniprot code: Q12718), three additional coppers occur in the active site, spatially arranged in a trinuclear cluster in which occurs the reduction of molecular oxygen during the catalytic activity. Six residues of histidine are coordinated to the coppers of the Type-3 (T3) site. Cu1 in the Type-2 (T2) site is coordinated to two histidine residues and there is an oxygen ligand in both positions (T2/T3 site). The mononuclear site of Type-1 (T1) occurs in the third domain and occupies a cavity of the enzyme surface. The copper in this site (Cu4) is the first electron acceptor and it is linked to the trinuclear cluster T2/T3 through a His-Cys-His tripeptide. It is also known that in the T1 site one ligand is supplied by a sulphur atom, given by a cysteine residue. The copper in the T1 site is closer to the binding site of the substrate and the cluster T2/T3, located between the first and the third domain, gives access to the solvent through two channels, allowing the release of water molecules, as provided by the chemical reaction shown previously. This disposition makes the transfer of electrons from the substrate possible.

From post translational modifications, two disulphide bridges concur to the stability of the protein, indeed they are required in many secreted proteins. The first bridge (Cys85-Cys488) stabilizes an α-helix structure in the first domain. The second bridge (Cys117-Cys205) maintains a connection between the first and the second domain. Glycosylation sites that are N-linked asparagine residues, contribute to maintain the stability of the structure [2].

The Laccase from *Pleurotus Eryngii* (Uniprot code: B0JDP9) has not been well functionally annotated but there are important expression data that connote the existence of the protein at the transcript level.

The aim of this study is to functionally annotate the protein of unknown structure towards the one that is functionally and structurally annotated, by using the method of comparative modelling.

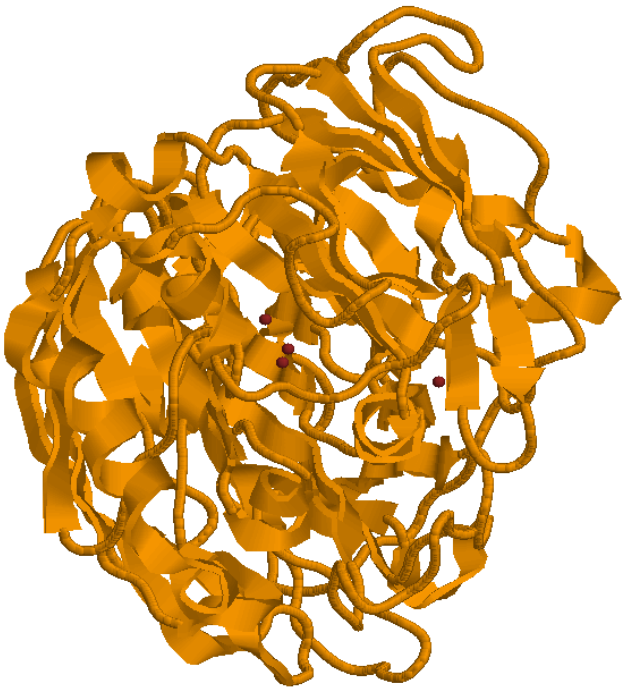


Figure 2. *It is shown the three-dimensional structure of the TvL, (PDB code: 1GYC), where the conformation of the secondary structure is emphasized. The four copper ions of the active site are also displayed in CPK.*

METHODS

Database

We retrieved all the information about the annotation level, the cellular location, the PTM/processing and other useful features for both the target and the template sequence on Uniprot database (release 2019_12) [3].

We found the information about the three-dimensional structure, including the atoms coordinates that are provided in the PDB file on PDB (Protein Data Bank; release 2019_05) [4].

PDBsum provided us an overview about the template 3D structure deposited in the Protein Data Bank, including the molecular components and the enzyme reaction.[5].

Computational methods

We used BLASTp (Basic Local Alignment Search Tool protein-protein) to search for the best candidate template against Uniprot to find out those sequences which have similarities with the sequence of interest above defined thresholds [6].

PDB files can be opened and visualized with Rasmol (2.7.5.2 version) and through a script language, structures have been explored and analysed [7].

In order to obtain a global sequence alignment between the target and the template we used Lalign [8]. The algorithm used for the global alignment is the Needleman-Wunsch algorithm [9].

To perform the homology comparative modelling of the 3D structures of the target and the template we used Modeller (9.23 version) whose working is based on spatial restraints satisfaction [10].

To calculate pairwise structure alignments and evaluate the goodness of the model was used jCE (3.0.8 version) [11][12][13].

To check the stereochemical quality of the model structure and obtain the Ramachandran plot, we used Procheck [14].

We used Swiss-model to perform the automated homology modelling [15].

TEMPLATE SELECTION

We first used BLASTp to find the best template sequence for the homology modelling procedure. Before running BLAST, we set the e-threshold value to 0.01. We used BOSUM62 as substitution matrix. We also filtered the search for sequences manually curated in Swiss-prot and with a 3D-structure known. We found out that the best identity score (sequence identity equal to 62%) corresponds to Laccase from *Trametes versicolor* (Uniprot code: Q12718). We retrieved information about its structure and functions on PDB (PDB code: 1GYC).

In PDBsum we found that the protein structure was obtained by X-ray crystallography with a resolution of 1.90 Å and a coverage of 100%.

Using the tool Lalign we performed a global sequence alignment between the target and the template. We used as default parameters the gap penalty (-12), an e-threshold value of 10 and we selected BLOSUM62 as substitution matrix. The output (Figure 3) included the scores of sequence identity (56.8%) and similarity (76.4 %).

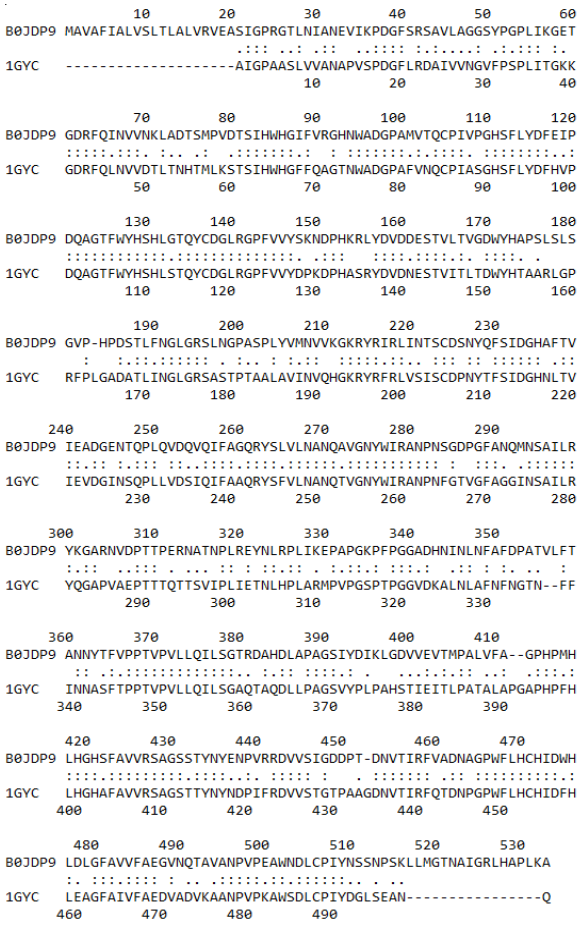


Figure 3. *Global sequence alignment obtained with Lalign between the target (B0JPD9) and the template (1GYC) sequences.* ***":"*** *indicates the alignment of same residues;* ***"."*** *of similar ones.*

Based on the parameters for the selection of the best template, that are a sequence identity above the 30% threshold, a high resolution and a high coverage, we ended up that the template selected was the optimal one for the homology modelling procedure.

MODELLER AT WORK

Input preparation and model generation

First, we converted the alignment file in a PIR file in order to obtain a proper input file for the modelling procedure. Then we removed positions from the target sequence (1-20 a.a) because of the cleavage of the signal peptide from the template sequence before the crystallization procedure [2]. We also removed a Gln499 from the template as well as the previous gaps and the corresponding positions on the target. The deletions were made knowing that they do not influence the stability of the active site. Then, we added 13 dots that denote the heteroatoms, in this case the four copper atoms of the active site and the glycosylation sites, in order to map them in the model. We also edited the PDB file of our template according to the changes made on the PIR file format (*Supplementary; section 1*). Once we have run modeller, we obtained as output six different models and the respective scores (Table 1).

Filename	**molpdf**	**DOPE score**	**GA341 score**
B0JDP9.B99990001	3743.09	-57575.90	1.00
B0JDP9.B99990002	3799.20	-57325.30	1.00
B0JDP9.B99990003	3657.02	-57896.10	1.00
B0JDP9.B99990004	3771.90	-57499.30	1.00
B0JDP9.B99990005	3983.20	-57557.40	1.00
B0JDP9.B99990006	3675.20	-57835.20	1.00

Table 1. *It is shown the Modeller output: the molpdf score is the sum of restraints violations, so the lower it is, the better is the model obtained. The DOPE score is related to the protein stability and must be the lower value. This two scores with the GA341 score are used to assess the goodness of the model.*

Model selection

Using the PDB files of the six models obtained and loading them on jCE, we performed the structural alignment of the models towards the template *(supplementary, section 2*). We scanned all the scores, focusing on the RMSD values. The lowest values belong to the third and the sixth model, with an RMSD value of 0.26 and 0.24 respectively. As regards the third model, the sequence identity and the sequence similarity scores are respectively of about 61% and 71%. The z-score of 8.30 proves the significance of the goodness of the superimposition obtained. Other important parameters are those obtained with Procheck respect to the Ramachandran plot. We found that, for the third model, the 90.5% of the residues occurs in the favoured region. For all these reasons and according to the molpdf and DOPE score, we selected the third model (B0JDP9.B99990003; Table 1) as the best output of the modelling procedure (*Supplementary; section 3 and 4*).

Target annotation

Since we know that the functionality of the active site is related to the presence of copper atoms, we had to verify if the architecture of the active site described previously was maintained. We picked the distances between the copper atoms and residues with which are coordinated both in the

template and in the model and we compared them (Figure 6; Table 2).

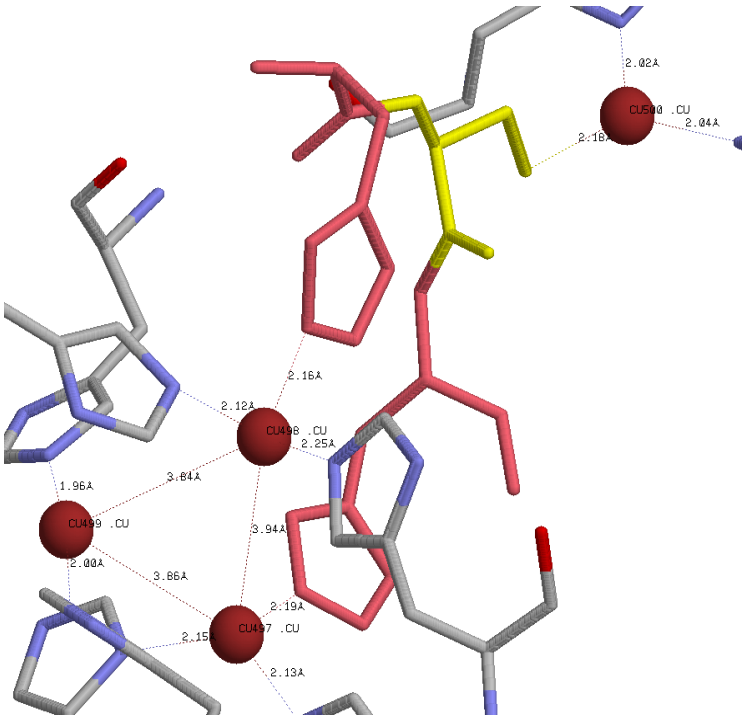


Figure 6. *Visualization of the picked distances between copper ions and amino residues. It is emphasized the tripeptide His-Cys-His displayed in yellow (Cys) and pink (His) that connect the T2/T3 and T1 sites.*

Template position	Residue	Ligand	Template distance (Å)	Model distance (Å)	Model position
64	His	Cu-1	2.01	2.00	64
66	His	Cu-2	2.15	2.15	66
109	His	Cu-2	2.12	2.13	109
111	His	Cu-3	2.23	2.25	111
395	His	Cu-4	2.02	2.02	394
398	His	Cu-1	1.97	1.96	397
400	His	Cu-3	2.12	2.12	399
452	His	Cu-3	2.16	2.16	450
453	Cys	Cu-4	2.19	2.18	451
454	His	Cu-2	2.17	2.19	452
458	His	Cu-4	2.04	2.04	456

Table 2. *Distances between the copper atoms and the residues both in the template and in the model structure. Some positions of the residues have changed due to the introduction of gaps during the sequence alignment procedure.*

According to the distance values reported in the Table 2, no big differences were detected between the template and the model. Moreover, not even the distances between the copper atoms themselves changed. e.g. the distance between the copper atoms Cu4 (T1 site) and Cu3 (T2/T3) is of about 12.30 Å in the model, respect to the distance picked in the template of 12.10 Å.

Having verified if the distances of the disulphide bonds were maintained in the model was an additional way to check if the model could preserve the same function (Table 3). Not big differences in terms of distance were detected.

Template position	Template distance (Å)	Model distance (Å)	Model position
Cys85-Cys488	1.98	2.02	Cys85-Cys486
Cys117-Cys205	2.04	2.02	Cys117-Cys204

Table 3. *Distances of the disulphide bonds in the template and in the model structure. Only two cysteine positions have changed.*

Figure 7. *Visualization of disulphide bridges in the folding of the model.*

As discussed previously, the presence of NAG in the structure is important for the stability of the protein. Looking at the sequence alignment (Supplementary; section 2) it is possible to notice that not all the Asn position has been conserved. In addition, the detection of the Asn residues and of the distances with the glycan groups confirmed the missing of some glycosylation sites in the model (Table 4).

Asn position	Template distance	Model distance
54	1.42	Not conserved
141	/	/
208	/	/
217	1.41	Not conserved
251	1.41	1.43
333	1.39	Not conserved
341	1.43	3.35
436	1.42	1.45

Table 4. *The table shows the distances for the glycosylation sites that occur in the template and in the model.*

Comparison with Swiss- Model

Uploading the sequence of the target on Swiss-Model we obtained the automated generation of the best model. The tool searched for the best template, that was 5mew.1A, with a sequence identity of 62.93% respect to the 63.20% of sequence identity for 1GYC. The GMQE values (Global Model Quality Estimation) were of 0.79 for 1GYC and 0.80 for 5MEW. The QMEAN for 1GYC is a positive value (0.34). This value is an indication of the high quality of the model [16]. Performing the pairwise structural alignment of the model obtained with Swiss-Model towards the template we found that the RMSD value was of 0.74 Å, respect to the one of the model implemented with Modeller (0.26 Å).This result prompted us to claim that the model obtained with modeller is better in terms of structural alignment. The sequence identity and similarity percentages are approximately the same in the two models.

DISCUSSION

Through homology modelling we can functionally annotate a protein of known sequence but unknown structure and function by building a model. This procedure provides a template whose functionality and structure are well known. Once we have verified that all the features of the active site and the distinctive elements of the protein stability that occur in the template were maintained in the model, we claimed that the model was able to execute all the functions and that it has been functionally annotated. The pairwise sequence alignment and the structural superimposition showed a very high sequence identity percentage (higher than the 30% threshold), as well as a very high similarity percentage. Even the RMSD values confirm that the two structure are well superimposable (Figure 8 and Figure 9). All these considerations are necessary to clarify that the GOterms transfer, from the template to the target is possible, meaning that what before was an unknown sequence it is now functionally annotated in relation to a specific protein family.

GOterms that can be transferred are those experimentally validated: copper ion bonding (GO:0005507); hydroquinone: oxygen oxidoreductase activity (GO:0052716); lignin catabolic process (GO:0046274); extracellular region (GO:0005576)

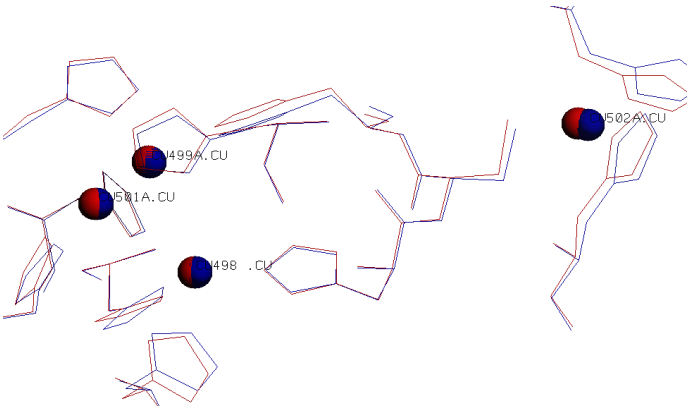


Figure 8. *Visualization of structural superimposition of the active site of the TvL (blue) and the model (red).*

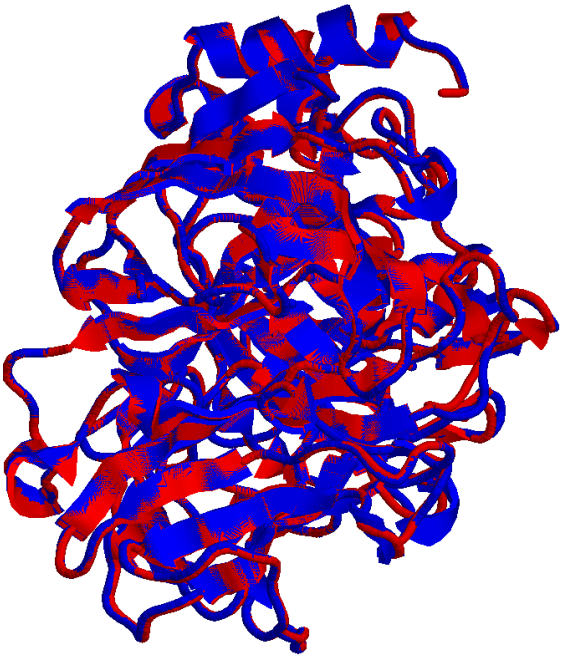


Figure 9. *Whole structural superimposition of the model towards the template.*

REFERENCES

[1] Alfred M. Mayer, Richard C. Staples. Laccase: new functions for an old enzyme. *Phytochemistry*, Vol 60, pp.551-565. 2002.

[2] Klaus Piontek, Matteo Antorini and Choinowski. Crystal Structure of a Laccase from the *Fungus Trametes versicolor* at 1.90-A Resolution Containing a Full Complement of Coppers. *The Journal of Biological Chemistry*, Vol 277, pp 37663-37669. 2002.

[3] The Uniprot Consortium. Uniprot: a worldwide hub of protein knowledge. *Nucleic Acids Research*, Vol 47, pp.506-D515. 2019. (https://www.uniprot.org/)

[4] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, Vol. 28, pp 235-242. 2000. (http://www.rcsb.org/)

[5] Laskowski R A, Jabłońska J, Pravda L, Vařeková R S, Thornton J M. PDBsum: Structural summaries of PDB entries. *ProteinScience* pp.129-134.2018. (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html)

[6] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. "Basic local alignment search tool." *Molecular Biology Journal*. Vol. 215, 403-410. 1990.

[7] www.RasMol.org

[8] X Huang, W Miller. Advances in Applied Mathematics. Vol 12,pp.337-357.1991.(https://embnet.vital-it.ch/software/LALIGN_form.html)

[9] Needleman, Saul B.; and Wunsch, Christian D., A general method applicable to the search for similarities in the amino acid sequence of two proteins, in *Journal of Molecular Biology*, Vol. 48, pp. 443–53. 1970

[10] A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Molecular Biology Journal* pp. 779-815, 1993.

[11] Prlìc A et al. Pre-calculated protein structure alignments at the RCSB DB website Bioinformatics. Vol 26, pp. 2983-2985. 2010.

[12] Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* Vol. 11, pp. 739-747. 1998.

[13] Yuzhen Ye and Adam Godzik. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* Vol. 19, pp.246-255. 2003-

[14] Laskowski R A, MacArthur M W, Moss D S, Thornton J M. PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* Vol. 26, pp.283-291.1993.(https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/)

[15] Waterhouse A et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, Vol 46, pp. 296-303. 2018. (https://swissmodel.expasy.org/)

[16] Benkert, P., Biasini, M., Schwede, T. Toward. The estimation of the absolute quality of individual protein structure models. *Bioinformatics* Vol. 27, pp. 343-350. 2011

SUPPLEMENTARY

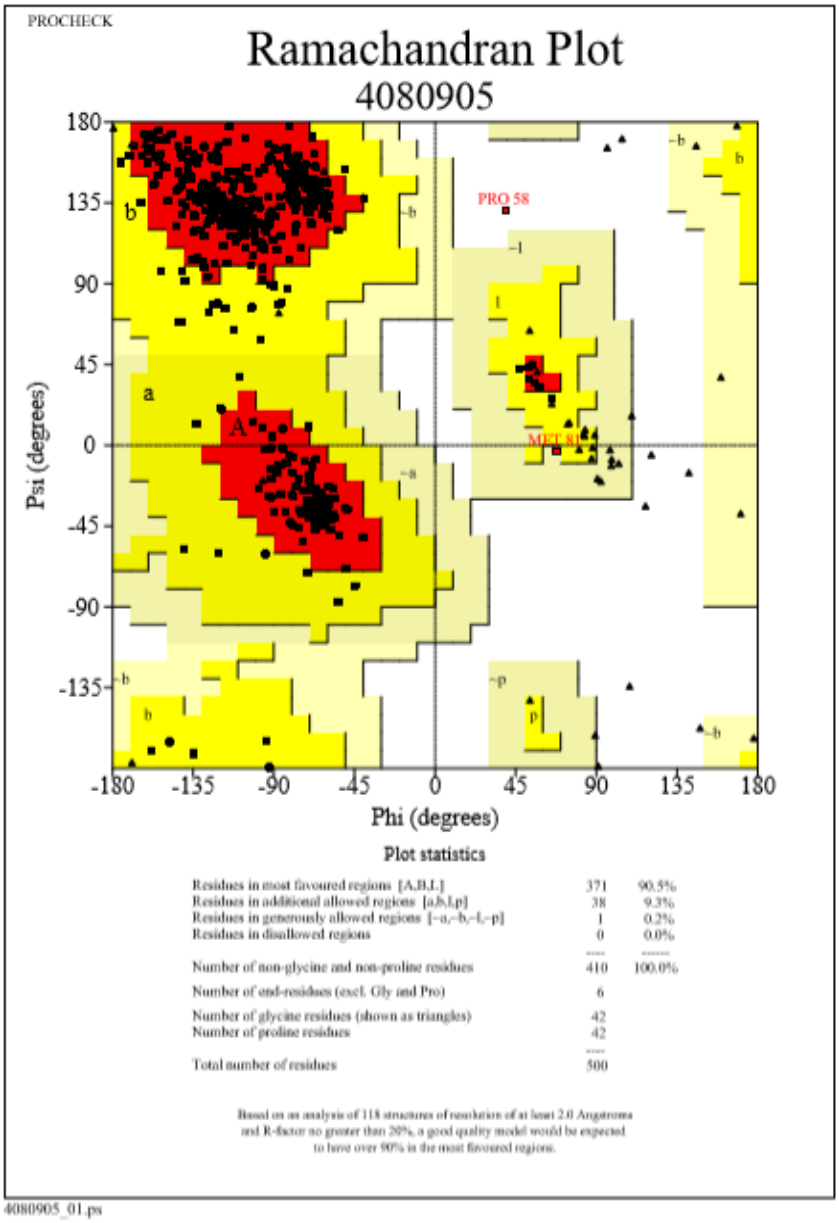
Section 1. PIR file input for Modeller procedure derived by the sequence alignment of the target and the template.

```
>P1;B0JDP9
sequence:B0JDP9:::::::
SIGPRGTLNIANEVIKPDGFSRSAVLAGGSYPGPLIKGET
GDRFQINVVNKLADTSMPVDTSIHWHGIFVRGHNWADGPAMVTQCPIVPGHSFLYDFEIP
DQAGTFWYHSHLGTQYCDGLRGPFVVYSKNDPHKRLYDVDDESTVLTVGDWYHAPSLSLS
GVP-HPDSTLFNGLGRSLNGPASPLYVMNVVKGKRYRIRLINTSCDSNYQFSIDGHAFTV
IEADGENTQPLQVDQVQIFAGQRYSLVLNANQAVGNYWIRANPNSGDPGFANQMNSAILR
YKGARNVDPTTPERNATNPLREYNLRPLIKEPAPGKPFPGGADHNINLNFAFDPATVLFT
ANNYTFVPPTVPVLLQILSGTRDAHDLAPAGSIYDIKLGDVVEVTMPALVFA--GPHPMH
LHGHSFAVVRSAGSSTYNYENPVRRDVVSIGDDPT-DNVTIRFVADNAGPWFLHCHIDWH
LDLGFAVVFAEGVNQTAVANPVPEAWNDLCPIYNSSNPSK.............*

>P1;1GYC
structureX:1gyc.pdb:1:A:511:A::::
AIGPAASLVVANAPVSPDGFLRDAIVVNGVFPSPLITGKK
GDRFQLNVVDTLTNHTMLKSTSIHWHGFFQAGTNWADGPAFVNQCPIASGHSFLYDFHVP
DQAGTFWYHSHLSTQYCDGLRGPFVVYDPKDPHASRYDVDNESTVITLTDWYHTAARLGP
RFPLGADATLINGLGRSASTPTAALAVINVQHGKRYRFRLVSISCDPNYTFSIDGHNLTV
IEVDGINSQPLLVDSIQIFAAQRYSFVLNANQTVGNYWIRANPNFGTVGFAGGINSAILR
YQGAPVAEPTTTQTTSVIPLIETNLHPLARMPVPGSPTPGGVDKALNLAFNFNGTN--FF
INNASFTPPTVPVLLQILSGAQTAQDLLPAGSVYPLPAHSTIEITLPATALAPGAPHPFH
LHGHAFAVVRSAGSTTYNYNDPIFRDVVSTGTPAAGDNVTIRFQTDNPGPWFLHCHIDFH
LEAGFAIVFAEDVADVKAANPVPKAWSDLCPIYDGLSEAN.............*
```

Section 2. jCE sequence alignment derived from pairwise structural alignment

```
                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:    1 AIGPAASLVVANAPVSPDGFLRDAIVVNGVFPSPLITGKKGDRFQLNVVDTLTNHTMLKSTSIHWHGFFQ
              .|||...|.:||..:.||||.|.|:...|.:|.|||.|..|||||:|||..|....|...|||||||.|.
Chain 2:    1 SIGPRGTLNIANEVIKPDGFSRSAVLAGGSYPGPLIKGETGDRFQINVVNKLADTSMPVDTSIHWHGIFV

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:   71 AGTNWADGPAFVNQCPIASGHSFLYDFHVPDQAGTFWYHSHLSTQYCDGLRGPFVVYDPKDPHASRYDVD
              .|.|||||||.|.||||..||||||||.:|||||||||||||.|||||||||||||...|||...||||
Chain 2:   71 RGHNWADGPAMVTQCPIVPGHSFLYDFEIPDQAGTFWYHSHLGTQYCDGLRGPFVVYSKNDPHKRLYDVD

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  141 NESTVITLTDWYHTAARLGPRFPLGADATLINGLGRSASTPTAALAVINVQHGKRYRFRLVSISCDPNYT
              .||||:|..||||.........| ..|.||.||||||...|...|.|.||..|||||.||:..|||.||.
Chain 2:  141 DESTVLTVGDWYHAPSLSLSGVP-HPDSTLFNGLGRSLNGPASPLYVMNVVKGKRYRIRLINTSCDSNYQ

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  211 FSIDGHNLTVIEVDGINSQPLLVDSIQIFAAQRYSFVLNANQTVGNYWIRANPNFGTVGFAGGINSAILR
              ||||||..||||.||.|.|||.||.:|||||.||||.||||||.||||||||||.|..|||...||||||
Chain 2:  210 FSIDGHAFTVIEADGENTQPLQVDQVQIFAGQRYSLVLNANQAVGNYWIRANPNSGDPGFANQMNSAILR

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  281 YQGAPVAEPTTTQTTSVIPLIETNLHPLARMPVPGSPTPGGVDKALNLAFNFNGTN--FFINNASFTPPT
              |.||...:|||.:.....||.|.||.||.:.|.||.|.|||.|..:||.|.|....  |..||..|.|||
Chain 2:  280 YKGARNVDPTTPERNATNPLREYNLRPLIKEPAPGKPFPGGADHNINLNFAFDPATVLFTANNYTFVPPT

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  349 VPVLLQILSGAQTAQDLLPAGSVYPLPAHSTIEITLPATALAPGAPHPFHLHGHAFAVVRSAGSTTYNYN
              ||||||||||...|.||.||||:|.:.....:|:|:||...|  .|||.|||||.|||||||.|||||.
Chain 2:  350 VPVLLQILSGTRDAHDLAPAGSIYDIKLGDVVEVTMPALVFA--GPHPMHLHGHSFAVVRSAGSSTYNYE

                .    :    .    :    .    :    .    :    .    :    .    :    .    :
Chain 1:  419 DPIFRDVVSTGTPAAGDNVTIRFQTDNPGPWFLHCHIDFHLEAGFAIVFAEDVADVKAANPVPKAWSDLC
              .|:.|||||.|... .|||||||..||.||||||||||.||:.|||:||||.|.....|||||.||.|||
Chain 2:  418 NPVRRDVVSIGDDP-TDNVTIRFVADNAGPWFLHCHIDWHLDLGFAVVFAEGVNQTAVANPVPEAWNDLC

                .    :
Chain 1:  489 PIYDGLSEAN
              |||.......
Chain 2:  487 PIYNSSNPSK
```

Section 3. Ramachandran plot of the model selected executed by Procheck



Section 4. Table of the parameters used for the selection of the best model

Model	RMSD (Å)	Favoured regions (%)	Allowed regions (%)	Not allowed regions (%)	Generously allowed (%)
B0JDP9.B99990001	0.25	90.2	9.5	0.0	0.2
B0JDP9.B99990002	0.27	89.8	9.5	0.0	0.7
B0JDP9.B99990003	0.26	90.5	9.3	0.0	0.2
B0JDP9.B99990004	0.25	89.3	9.5	0.2	1.0
B0JDP9.B99990005	0.26	90.2	9.3	0.0	0.5
B0JDP9.B99990006	0.24	90.3	9.3	0.2	0.2