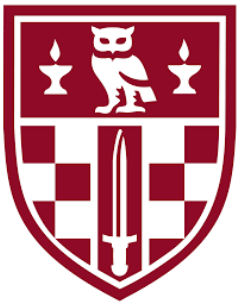
BIRKBECK

UNIVERSITY OF LONDON



MSc. BIOINFORMATICS

Enoyl-Acyl Carrier Protein Reductase (FabI) from Neisseria meningitidis in complex with NAD+ and Triclosan

---

Name: Gabriel López

Professor: Mark Williams

London

January-2023

Part A: Analysis and description of the structure [22 marks]

a) What is the enzymatic reaction that this protein catalyses in vivo? Write the relevant equation

Through the UnitProt web service **(The UniProt Consortium, UniProt.,2023)** the enzymatic reaction of the protein is:

a 2,3-saturated acyl-[ACP] + NAD^+ = a (2E)-enoyl-[ACP] + H^+ + NADH

b) What is the CATH structural classification of your protein? [1 Ma

The CATH **(Cuff et al., 2009)** and SCOP **(Murzin et al., 1995)** depicts the structural classification:

Class: Alpha and beta proteins (a/b)

Architecture: 3-Layer(aba) Sandwich

Topology: Rossmann fold

Homologous Superfamily: NAD(P)-binding Rossmann-like Domain

Protein type: Globular proteins

Fold: SDR-type extended Rossmann fold

Superfamily: Short-chain dehydrogenase/Reductase (SDR-like)

Domain: Enoyl-[acyl-carrier-protein] reductase [NADH]

Species: Neisseria meningitidis FAM18

c) Describe the structure of the biologically active form of the protein, accompanied by relevant illustrations including your own Chimera figures and a plot indicating the secondary structure. [5 Marks]

Enoyl-Acyl Carrier Protein Reductase (FabI) is a dimeric protein that depicts α-Helices, 3_{10}, β-strands, loops, and other motifs which are colored in orange, blue light, and green, respectively **(Figure 1).** Additionally, the protein presents strand (36%), alpha-helix (40,1), 3_{10} (7,4%), and other (38,5%) where the most predominant structure is α-Helix. Likewise, Ramachandran's plot shows up a bunch of spots (red) located on the right-hand side symbolizing the α-Helix **(Figure 2)**

Retrieved from http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=4m89&template=protein.html&o=SUMMARY&l=1&c=1&chain=A on 26 of December 2022

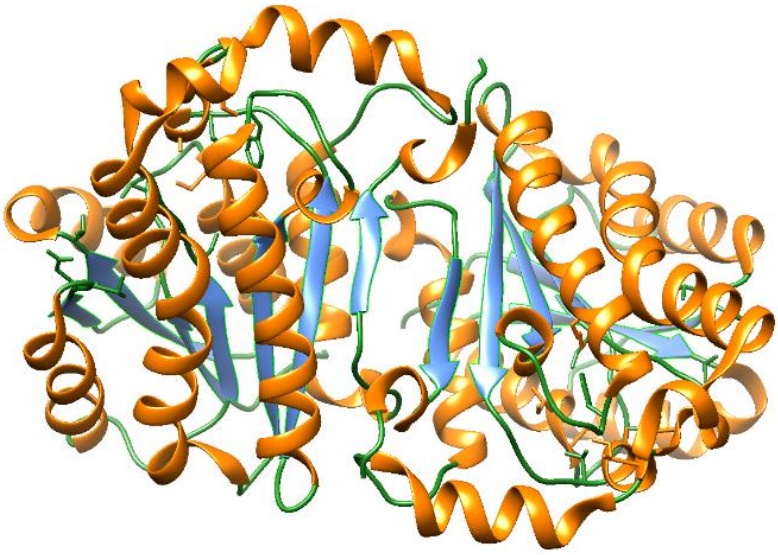


Figure 1. Enoyl-Acyl Carrier Protein Reductase (FabI) from Neisseria meningitidis.

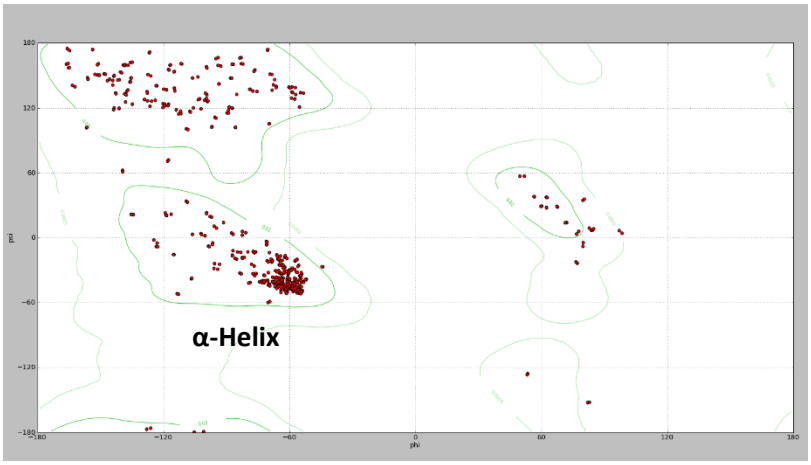


Figure 2. Ramachandran plot of Enoyl-Acyl Carrier Protein Reductase (FabI) from Neisseria meningitidis by Chimera.

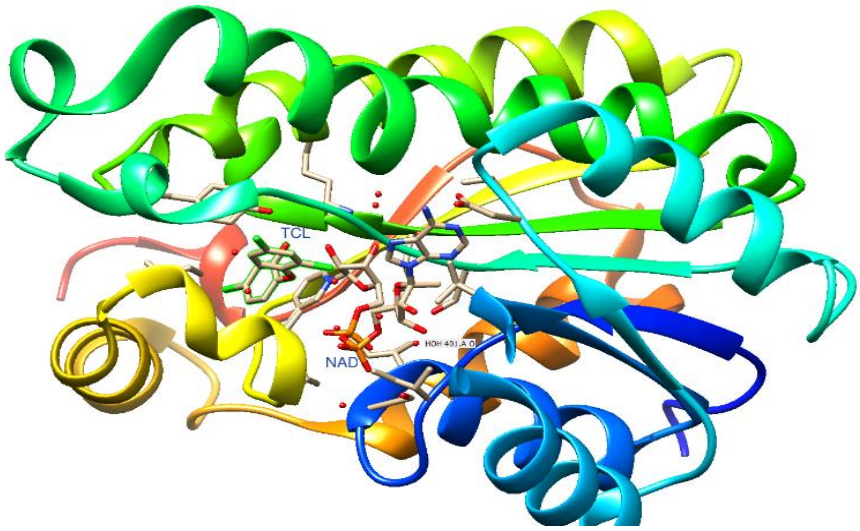


Figure 3. Enoyl-Acyl Carrier Protein Reductase (FabI) in complex with ligand (TCL) and inhibitor (NAD)

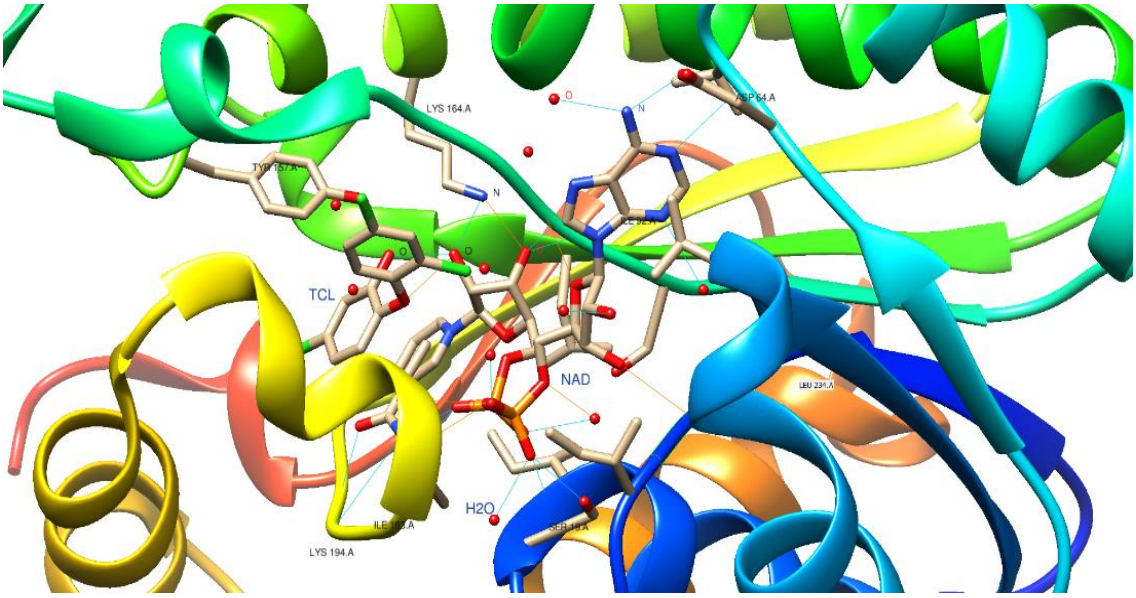


Figure 4. Enoyl-Acyl Carrier Protein Reductase (FabI)

d) Is the structure of good quality? Support your opinion with argument and evidence, including supporting plots. [4 Marks]

The structure shows up several characteristics which present high quality of the structure. Firstly, the resolution displays 1.90 Å obtained by X-RAY method, $R_{observed}$ (0.168) and R_{work} (0.167). However, R is not considered a quality value but R_{free} (0.186) sustains good quality. Besides, Ramachandran plot reveals that 92% fell in favourite regions, 7.1% in additional regions validating the quality of the model **(Figure 5).** Additionally, the server provides G-factor which reveals the plausibility of a stereochemical property **(Aslanzadeh & Ghaderia, 2012)**. The G-factor of dihedral angles and covalent forces is 0.18. Moreover, Prosa was used to obtain Z-score **(figure 6)** which displays a -8.13 value, characteristic of structures determined by NMR and X-ray methods **(Sippl, 1993)**

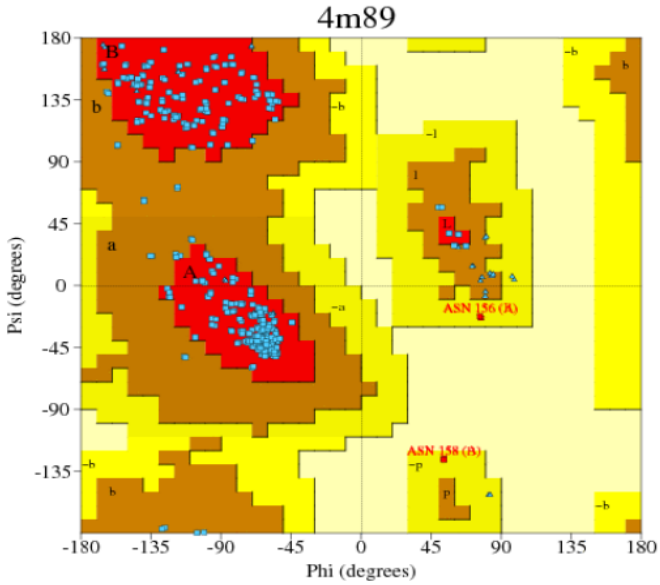


Figure 5. Ramachandran Plot of Enoyl-Acyl Carrier Protein Reductase (FabI) retrieved from:http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=4m89&template=procheck_summary.html

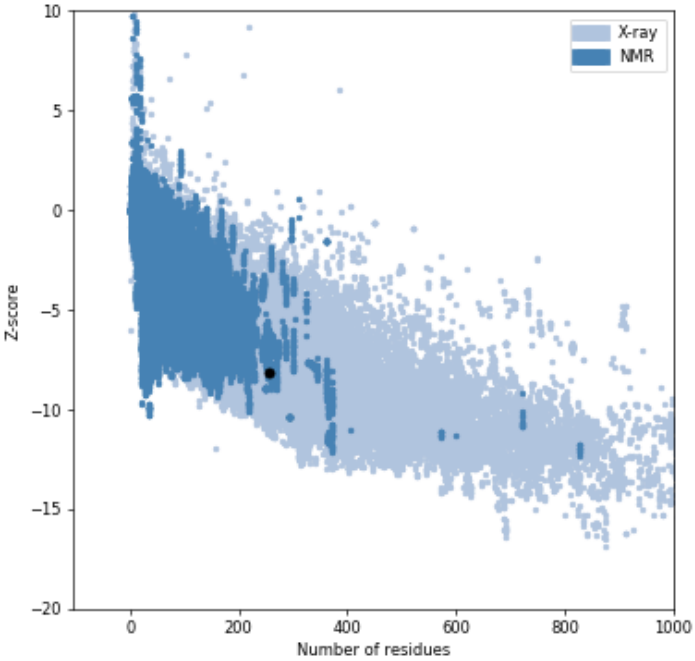


Figure 6. Number of residues in function of the Z-score of Enoyl-Acyl Carrier Protein Reductase (FabI). Retrieved from Prosa Database **(Nanson et al., 2013)**

d) What small molecule ligands are found in your structure? One is the cofactor. One of the others is a small molecule that may have some resemblance with the reaction substrate for this enzyme (NOTE: this resemblance might not be great!) or acts as an inhibitor to substrate binding. What is this substrate/inhibitor-like ligand?

The small molecules found in FabI are Triclosan (TCL) and Nicotinamide-Adenine-Dinucleotide (NAD). **Heath et al., 1998** suggests that TCL acts as an inhibitor of enoyl-ACP reductase, disinhibiting to catalyse the biosynthesis of the bacteria. Thus, NAD is the cofactor that enable a variety of redox reactions **(Shantanam & Mueller, 2018).**

f) Identify the functional groups of your protein that interact with this substrate-like ligand and the type of interaction (hydrogen bonds, hydrophobic interactions etc.). Present this information in a table and as a figure (e.g., using ligplot/Chimera). [5 Marks]

The analysis of functional groups was carried out by Chimera and Ligplot **(figure 7).** It has been found hydrogen bonds, hydrophobic and Van der Waals interaction. The interactions are displayed in the next table:

Table 1. Hydrogen bonds of Enoyl-Acyl Carrier Protein Reductase (FabI).

	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	735	N	ALA	95	A	3925	CL15	TCL	302	A	3.28
2.	1206	OH	TYR	157	A	3927	O17	TCL	302	A	2.51
3.	1206	OH	TYR	157	A	3927	O17	TCL	302	A	2.51

Table 2. Non-bonded contacts of Enoyl-Acyl Carrier Protein Reductase (FabI).

	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	720	N	GLY	93	A	3926	CL16	TCL	302	A	3.59
2.	721	CA	GLY	93	A	3926	CL16	TCL	302	A	3.52
3.	722	C	GLY	93	A	3918	C10	TCL	302	A	3.77
4.	722	C	GLY	93	A	3926	CL16	TCL	302	A	3.61
5.	723	O	GLY	93	A	3918	C10	TCL	302	A	3.60
6.	723	O	GLY	93	A	3926	CL16	TCL	302	A	3.39
7.	735	N	ALA	95	A	3925	CL15	TCL	302	A	3.28
8.	738	O	ALA	95	A	3925	CL15	TCL	302	A	3.16
9.	776	CD1	LEU	100	A	3921	C12	TCL	302	A	3.76
10.	776	CD1	LEU	100	A	3925	CL15	TCL	302	A	3.60
11.	1125	CG	TYR	147	A	3911	C1	TCL	302	A	3.81
12.	1128	CE1	TYR	147	A	3924	CL14	TCL	302	A	3.69
13.	1130	CZ	TYR	147	A	3924	CL14	TCL	302	A	3.71
14.	1203	CE1	TYR	157	A	3911	C1	TCL	302	A	3.41
15.	1203	CE1	TYR	157	A	3913	C6	TCL	302	A	3.62
16.	1203	CE1	TYR	157	A	3927	O17	TCL	302	A	3.49
17.	1205	CZ	TYR	157	A	3911	C1	TCL	302	A	3.87
18.	1205	CZ	TYR	157	A	3927	O17	TCL	302	A	3.42
19.	1206	OH	TYR	157	A	3911	C1	TCL	302	A	3.63
20.	1206	OH	TYR	157	A	3913	C6	TCL	302	A	3.42
21.	1206	OH	TYR	157	A	3927	O17	TCL	302	A	2.51
22.	1254	CE	LYS	164	A	3927	O17	TCL	302	A	3.67
23.	1482	O	ALA	197	A	3921	C12	TCL	302	A	3.58
24.	1482	O	ALA	197	A	3922	C13	TCL	302	A	3.66
25.	1483	CB	ALA	197	A	3918	C10	TCL	302	A	3.76
26.	1483	CB	ALA	197	A	3919	C9	TCL	302	A	3.33
27.	1483	CB	ALA	197	A	3920	C8	TCL	302	A	3.64
28.	1483	CB	ALA	197	A	3926	CL16	TCL	302	A	3.52
29.	1485	CA	ALA	198	A	3915	C4	TCL	302	A	3.71
30.	1506	CD1	ILE	201	A	3912	C2	TCL	302	A	3.55
31.	1506	CD1	ILE	201	A	3916	C3	TCL	302	A	3.54
32.	1528	CE1	PHE	204	A	3916	C3	TCL	302	A	3.87
33.	1528	CE1	PHE	204	A	3924	CL14	TCL	302	A	3.80

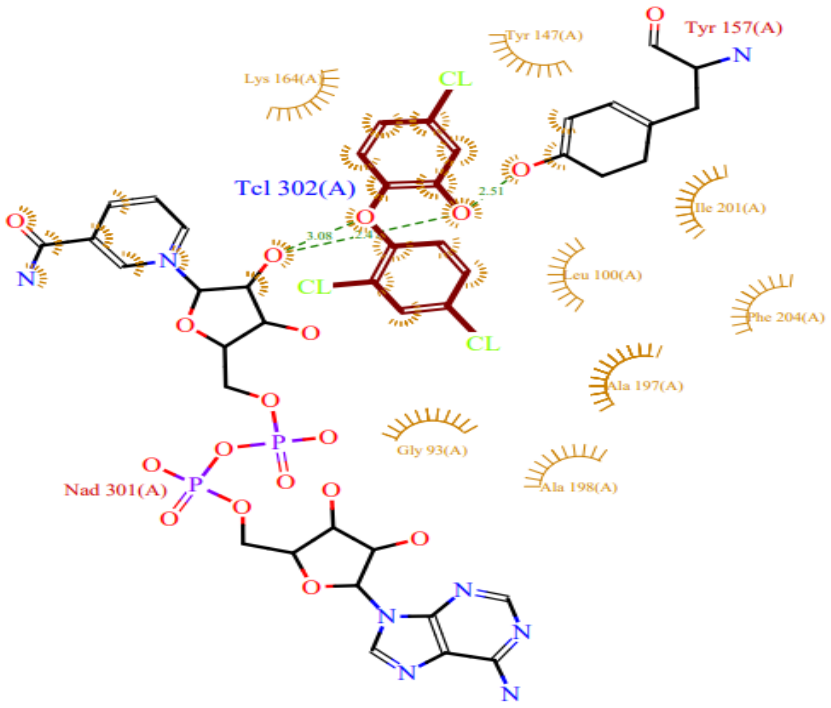


Figure 7. Ligplot analysis results of protein-ligand (FabI-Triclosan 3Fi02)

Table 3. Hydrogen bonds of NAD (Inhibitor)

	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	89	O	GLY	13	A	<--	3875	O3B	NAD	301	A	2.86
2.	137	OG	SER	19	A	-->	3869	O2A	NAD	301	A	2.55
3.	137	OG	SER	19	A	<--	3869	O2A	NAD	301	A	2.55
4.	138	N	ILE	20	A	-->	3892	O2N	NAD	301	A	2.86
5.	503	OD1	ASP	64	A	<--	3884	N6A	NAD	301	A	2.78
6.	505	N	VAL	65	A	-->	3885	N1A	NAD	301	A	3.01
7.	715	O	ILE	92	A	<--	3898	O3D	NAD	301	A	3.16
8.	1255	NZ	LYS	164	A	-->	3898	O3D	NAD	301	A	2.89
9.	1255	NZ	LYS	164	A	-->	3900	O2D	NAD	301	A	2.77
10.	1447	N	ILE	193	A	-->	3906	O7N	NAD	301	A	2.69
11.	1450	O	ILE	193	A	<--	3907	N7N	NAD	301	A	2.89
12.	1469	OG1	THR	195	A	-->	3891	O1N	NAD	301	A	2.69
13.	1469	OG1	THR	195	A	-->	3907	N7N	NAD	301	A	3.22
14.	1479	N	ALA	197	A	-->	3868	O1A	NAD	301	A	2.96

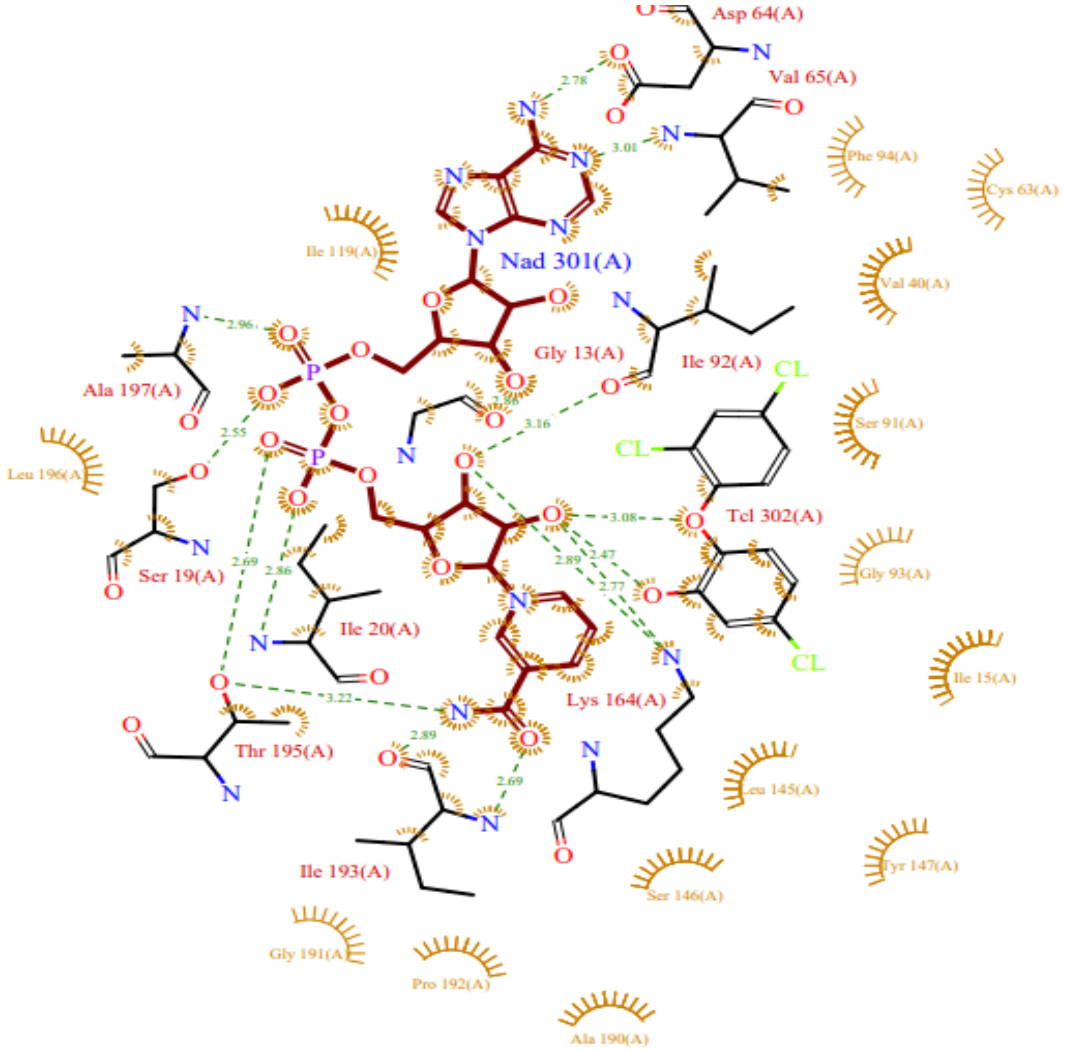


Figure 8. Ligplot analysis results of protein-ligand (FabI- NAD)

g) Which of the interactions identified in (f) are the most important to the binding affinity and which to the catalysed reaction? Support your opinion with argument and evidence. [4 Marks]

It is generally accepted that the affinity between a protein and its ligand is chiefly due to hydrophobic interactions, which are non-directional, whereas specificity of binding is chiefly due to anisotropic or directional, forces such as hydrogen bonding **(Gregory, A. & Dagmar,R., 2014).** In the substrate binding pocket, NAD (cofactor) is bound into the hydrophobic site **(Figure 8)** displaying 14 hydrogen bonds and 120 non-bonded contacts. The adenine site of NAD is binding to Asp 64, forming hydrogen bonds and Val 65 displays covalent interaction due to the triple bonds between two nitrogen molecules of the protein-ligand. Besides, it has found hydrophobic interactions in Ala 197, Ile 20 and the phosphate group of NAD. The interaction between oxygen molecules of Thr 195 and the phosphate group forms a covalent bond

(www.uniprot.org/uniprotkb/A1KVU8/feature-viewer). Finally, amino acids such as Ile 193, Lys 164, and Thr195 display hydrophobic interaction. On the other hand, the interactions of TCL (inhibitor) and NAD depict covalent bonds between oxygen molecules **(figure 9).**

The active site was found in Tyr 147 and Tyr 157 **(figure 10)** where can be found hydrogen bonds between the hydroxyl group and the oxygen of the TCL. Similar studies revealed that the inhibitor binds into a hydrophobic pocket in the active site depicting well conserved molecules **(Seol et al., 2019).**

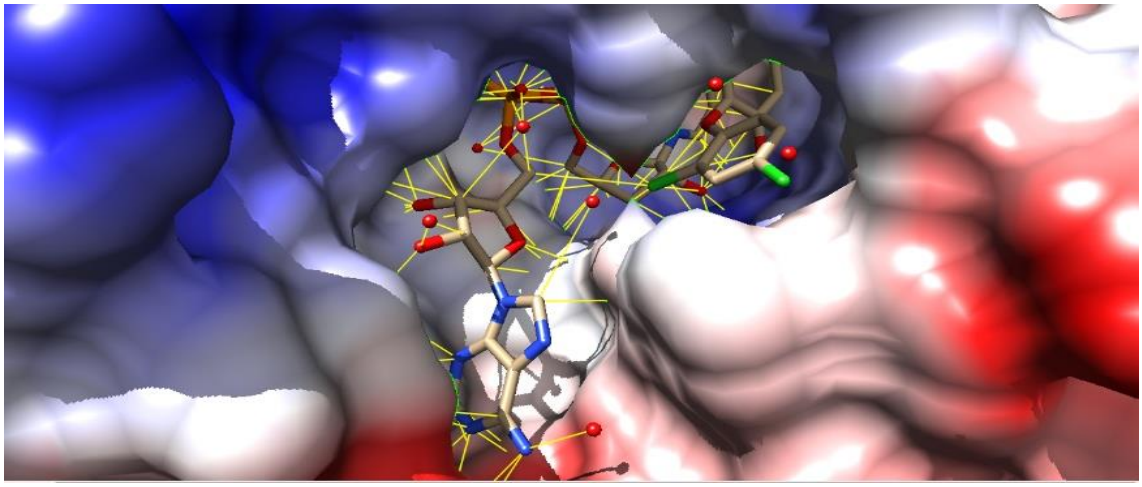


Figure 9. The active site Enoyl-Acyl Carrier Protein Reductase (FabI).

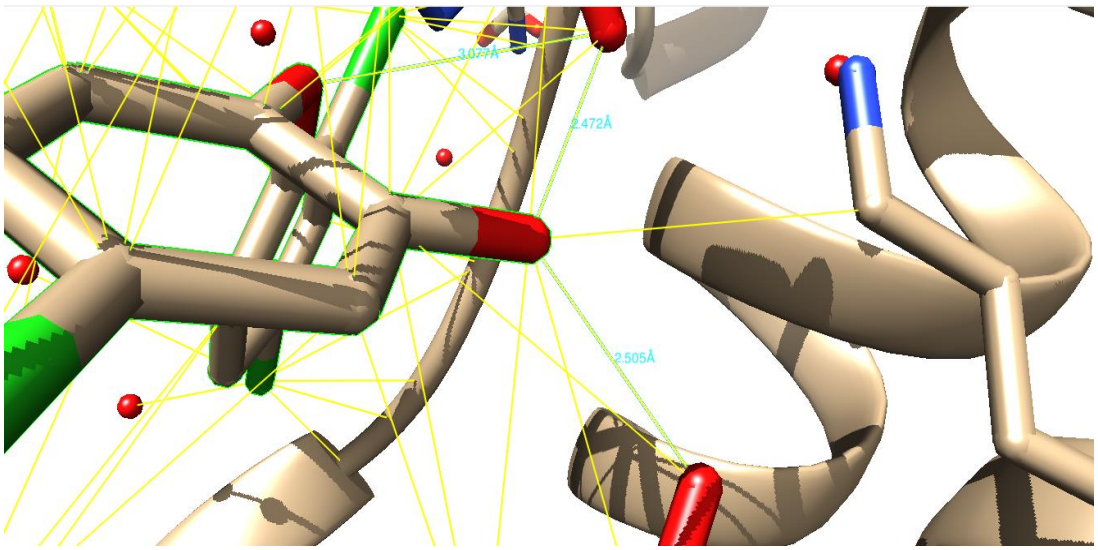


Figure 10. Interaction in the active side (Tyr 147) of Enoyl-Acyl Carrier Protein Reductase (FabI).

Part B: Identifying potential inhibitors

a) Create a connection table or SMILES string for the selected ligand in your complex (do not select the cofactor!). You may find this molecule in .sdf or or .sd or .mol format. (see: http://www.ebi.ac.uk/chebi). Alternatively, you can find or create your own SMILES string for it. You can use any resource on the web, but explain how you got your connection table/SMILES string and copy the file as text in your answer. [2 Marks]

The DrugBank webside (**go.drugbank.com/drugs/DB08604)** provides several small molecules using in protein-ligand interaction. Through the server the next SMILE string was obtained:

OC1=CC(Cl)=CC=C1OC1=C(Cl)C=C(Cl)C=C1

b) Does your selected ligand obey the Lipinski rule of 5? Explain your answer. [2 Marks]

The Triclosan presents a hydrogen bond donor and two hydrogen bond acceptors. Besides, the molecular mass is 289.5 Daltons, and the coefficient log P is 5. In conclusion, the ligand obeys the Lipinski rules. Retrieved from pubchem.ncbi.nlm.nih.gov/compound/Triclosan#section=Computed-Properties.

c) Create a set of drug or drug-like molecules for subsequent searching. Document how and why you chose your 'search set' and how many molecules you have ended up with. You should have either SMILES strings or an .sdf file for these molecules. Do NOT include the whole list in the coursework you hand in - the first page of the list/file will be fine in the appendix. [2 Marks]

Table 4. Drug-like molecules retrieved form ChEMBL,DrugBank and ZINC

N°	Name	Database	Smile string
1	CG-400549	ChEMBL	Cc1c(N)cccc1Cn1ccc(OCCc2cccs2)cc1=O
2	AFN-1252	ChEMBL	Cc1c(CN(C)C(=O)/C=C/c2cnc3c(c2)CCC(=O)N3)oc2ccccc12
3	AFABICIN	ChEMBL	Cc1c(CN(C)C(=O)/C=C/c2cnc3c(c2)CCC(=O)N3COP(=O)(O)O)oc2ccccc12
4	Soneclosan	DrugBank	OC1=C(OC2=CC=C(Cl)C=C2)C=CC(Cl)=C1

5	Indole Naphthyridinone	DrugBank	CN(CC1=CN(C)C2=CC=CC=C12)C(=O)\C=C\C1=CN=C2NC(=O)CCC2=C1
6	3-(6-Aminopyridin-3-Yl)-N-Methyl-N-[(1-Methyl-1h-Indol-2-Yl)Methyl]Acrylamide	DrugBank	CN(CC1=CC2=C(C=CC=C2)N1C)C(=O)\C=C\C1=CN=C(N)C=C1
25	Triclosan Sulfate	ZINC	O=S(=O)(O)Oc1cc(Cl)ccc1Oc1ccc(Cl)cc1Cl

DrugBank (drugbank.com), ChEMBL(www.ebi.ac.uk) and ZINC (zinc.docking) dabases were used to search for information about drugs and the functions they perform obtaining 25 small molecules **(Appendix 1).** The name of the protein (Enoyl-[acyl-carrier-protein] reductase (FabI) was input as target through DrugBank and ChEMBL websites. On the other hand, the name of small molecule (Triclosan) was used to identify similar ligands at ZINC website. Each entry has several data fields, including the trials conducted, chemical properties and the experimental phases in which they are currently in. The small molecules were selected through Lipinski rule of 5 which is related to the molecular properties important and drug's pharmacokinetics in the body: absorption, distribution, metabolism and excretion (ADME) **(Ivanović et al., 2020).**

d) Fingerprint the drugs in the set and calculate the Tanimoto similarity scores for the comparison of your selected ligand against each one of the drugs. You can do this using the OpenBabel ready-made programs (i.e. the binaries you can access as advised in the chemoinformatics tutorials) or by writing your own program. No extra marks will be given for writing code as opposed to using ready-made binaries. Produce a histogram of the scores and submit this together with a list containing the actual commands you used or the source code of your program (in the appendix if the code is longer than a few lines). What are the mean and standard deviation of your scores? [4 Marks]

The fingerprint was obtained by R (**Appendix 2)** which reveals several compounds depict in the **table 5.** The first value was obtained due to the molecule was compared itself and then against each of the drugs stored in the sdfset. The cutoff value is 0.1 displaying 14 molecules, however, 11 of the compounds present a Tanimoto coefficient less than 0.2.

Table 5. Fingerprint of the drugs and the Tanimoto similarity scores

N°	Drugs	Tanimoto coefficient
1	Triclosan	1
2	Soneclosan	0.8333333
3	Triclosan Sulfate	0.6086957
4	CMP10	0.1842105
5	CMP13	0.1842105
6	CMP21	0.1792453

7	CMP18	0.1465517
8 9	CMP23	0.1465517
10	CMP24	0.122807
11	CMP8	0.1092437
12	CMP11	0.1066667
13	CMP19	0.1041667
14	CMP9	0.1024096
Mean		0.2944687
Estandard deviation		0.3082726

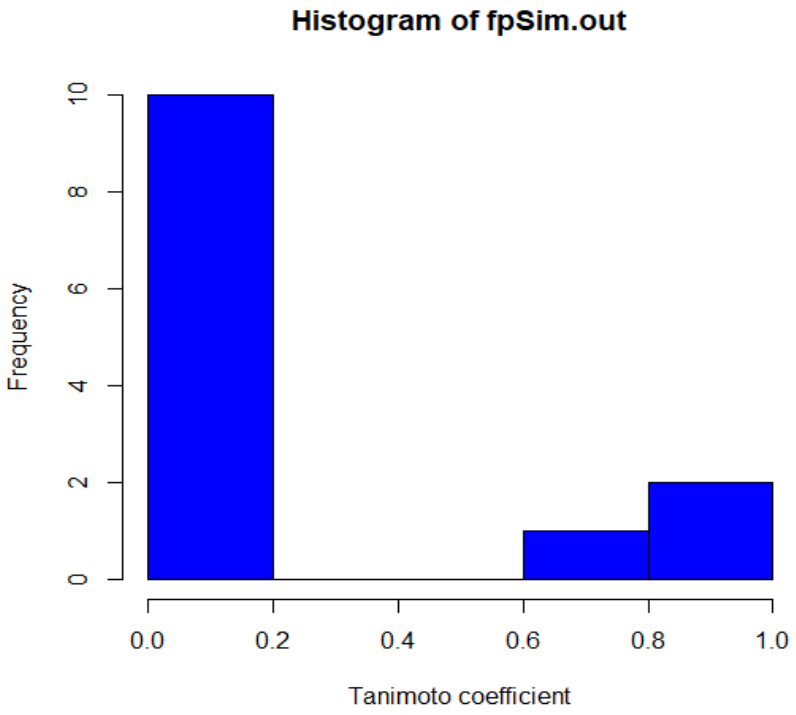


Figure 11. Histogram of Tanimoto coefficient of the druf set

e) Did you find any drug-like molecules that significantly resembles your ligand – if so, give examples. If not why do you think nothing similar was found? [2 marks]

Trough the analysis the main molecules are Soneclosan and Triclosan sulphate. The first drug presents a similar structure due to it has chlorinated aromatic compounds and functional groups such as ether and phenols. On the other hand, the second drug has a sulphuric acid that differs slightly from Triclosan. The histogram **(figure 11)** depicts an unusual data due to the first values belongs to drugs in the range of 0.1-0.2, in contrast with Soneclosan and Triclosan sulphate drugs that display less frequency but high Tanimoto coefficient.

Part C: Docking to a model structure

Using your method of choice (or multiple methods) identify a reliable 'template' protein that is homologous to your target. Explain how you chose the template and what score/s and other criteria you used to reach this decision. [2 marks]

The analysis of the template was carried out by GenTHREADER tool that provided 390 hits. The criteria to choose the protein was the high sequence identity, RMSD (Chimera) and the long alignment length. For this reason, the Fabl protein from *Neisseria meningitidis FAM 18* (4m89) was excluded and the findings depicts that *Pseudomonas aeruginosa Enolyl-Acyl* (4nr0) presents a net score of 196.684, p-value 7e-19, the long alignment is 259 out 260 and RMSD is 0.645. Furthermore, the sequence (4nr0) was compared using Modeller and portrayed 61.3% of identity.

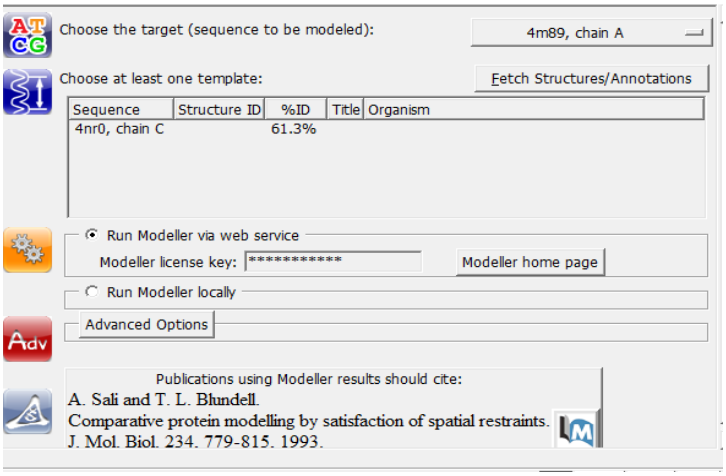


Figure 12. Comparative protein modelling between 4m89 and 4nr0 proteins (Modeller software)(Sali et al., 1993)

Results: http://bioinf.cs.ucl.ac.uk/psipred/&psipred_uuid=f0bace6c-8d10-11ed-a60c-00163e100d53

b) Generate a comparative (homology) model of your target protein using the template structure you've selected in Part C (a). You can use any method you wish, but you need to explain your choice and how you have obtained your model (including model assessment score/s). Be careful to document any settings of the modelling process (especially non-default ones). Give a brief evaluation of the quality of the model structure. Include figures illustrating your results. [4 Marks]

The comparative modelling was performed by GenTHREADER due to uses traditional sequences alignment algorithm to generate fast and reliable alignments which are evaluated by method derived form threading techniques. **(Jones, 1999).** Then, the result displayed a protein is classified as oxidoreductase and has been found in *Burkkilderia pseudomallei* organism. The Fabl in complex with NAD and PT01 dominated as 5i7s in PDB website. In the first stages, two web servers (GenTHREADER and HHpred) were used to generate a comparative homology. However, the best hit of the HHpred website depicted low identity, in contrast with GenTHREADER displayed 61.5 % of identity **(figure 13).** Besides, the p-value is 2e-19, the alignment length is 257 out 273 and the net

score is 201.322. Moreover, the protein was evaluated by statistical potential score using Dope and Prosa servers. The Z-score value (-8.66) is located within the space of protein determined by NMR method **(figure 14).** The value is close to the finding of the most template which recommends that the obtained model is reliable and very close to experimentally determined structure **(Syed et al., 2012).** The local model quality **(figure 15)** shows the energy as a function of amino acid sequence position. The positive values correspond to problematic part of the input structure **(Kumar et al., 2014).** Finally, the quality of the structure was evaluated by the resolution which depicts 1.60 Å, the Ramachandran plot the 92,4 % of the most favoured regions, R-value of 0.153 and Rfree of 0.178. In conclusion, the comparative model is a reliable homology of the template protein.

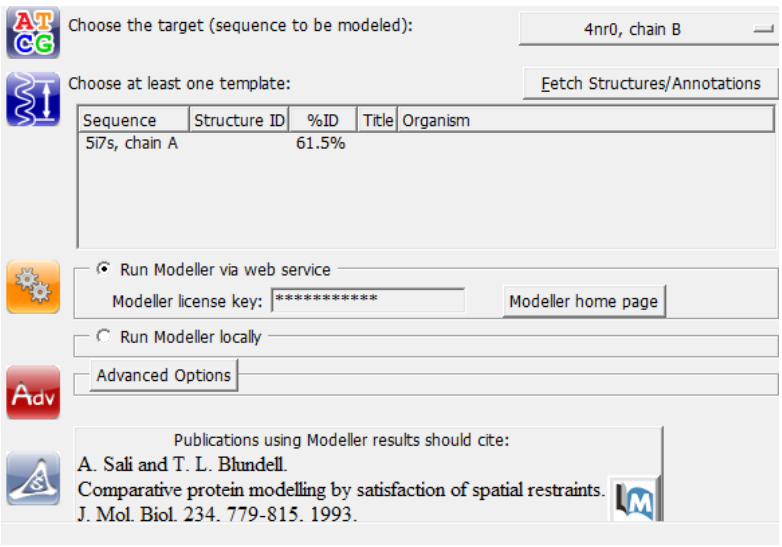


Figure 13. Comparative protein modelling between 4nr0 and 5i7s(Modeller software

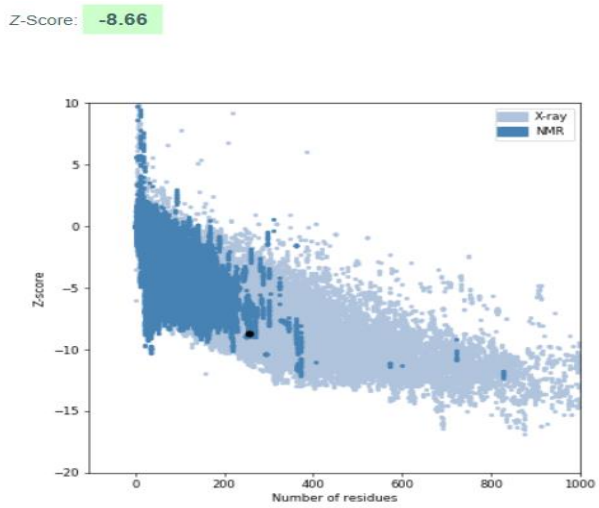


Figure 14. The Z-score value of 5i7s using Prosa database

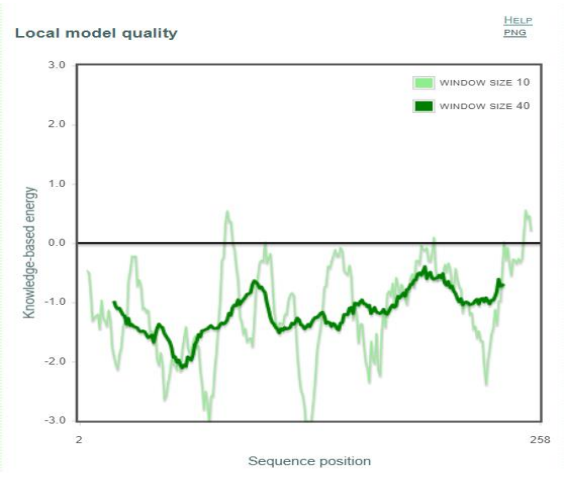


Figure 15. The local model quality value of 5i7s using Prosa database

Results retrieved from http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=5i7s&template=procheck_summary.html

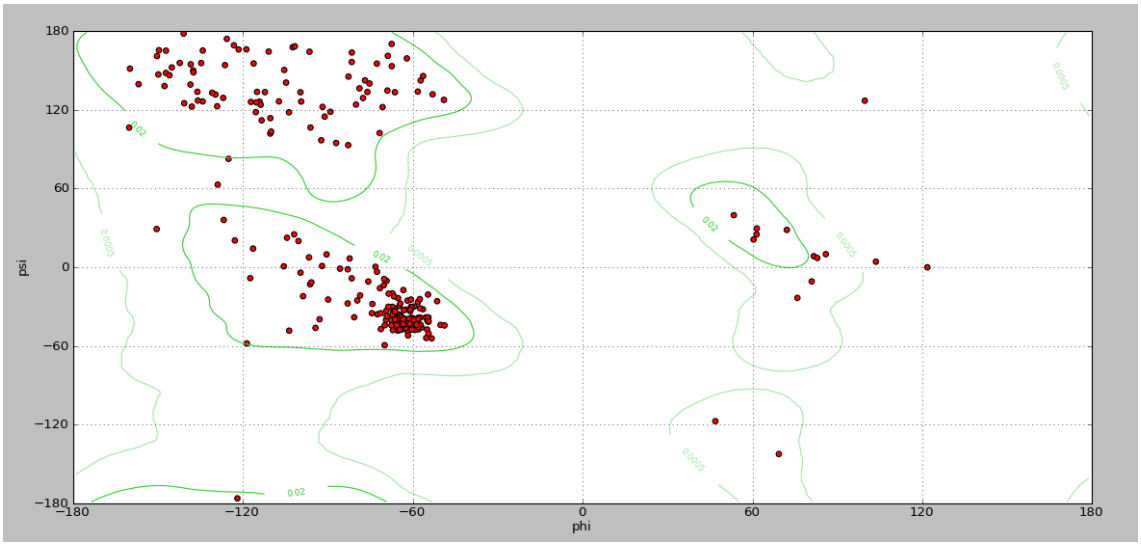


Figure 16. Ramachandran plot of 5i7s by Chimera

Results from http://bioinf.cs.ucl.ac.uk/psipred/&uuid=c3b40708-8d51-11ed-967e-00163e100d53

c) Compare the best scoring model of your model to the actual PDB structure of the target. Describe any differences in the protein structure between model and the actual structure. Use assessment scores (e.g RMSD and TM-score). Illustrate global and/or local structural differences in a figure (or figures) [4 marks].

The *Burkholderia pseudomallei* FabI1 Enoyl-ACP Reductase (5i7s) displays a high-quality resolution (1.60 Å). The protein is made up 257 residues which 16 % represents strand, 37.7 % of alpha helix, 7% of 3-10 helix and other structure displays 39.3 %. In addition, the TM-score is 0.5887 which counts all residue pairs using the Levitt-Gerstein weight and does not need discrete distance cutoffs **(Zhang & Xu, 2010).** In the other hand, the value of RMSD represents how different the obtained docking orientation is from corresponding co-crystallized pose of the same ligand molecule **(Ramírez & Caballero, 2018).** The value obtained is 2.888 Å that deviates from the position of the reference. In the **figure 16** the main difference reveals the dimmer obtained. The 4rn0 shows tetra dimer molecules, in contrast with 5i7s that displays a dimer.

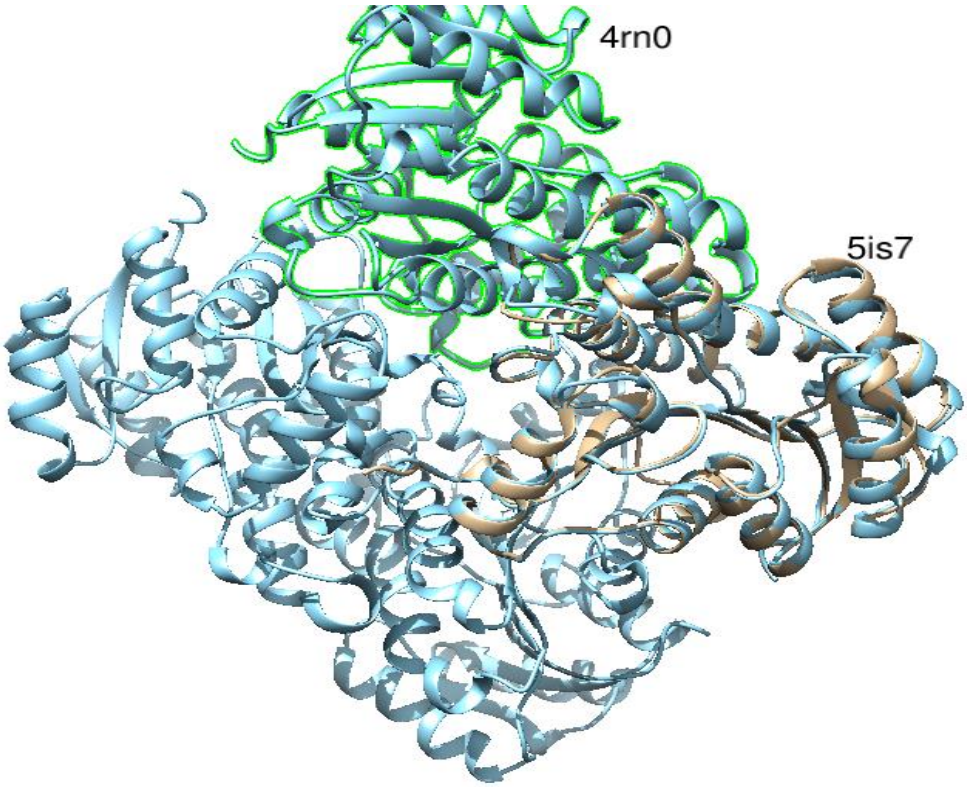


Figure 16. Structure1 colored by blue (4nr0) and Structure colored by light brown (5i7s)

d) Dock the ligand molecule identified in Part A (e) into the comparative model using SwissDock (or any other program/server of your choice, eg. haddock/autodock vina). Present the results in a table/figure [2 marks]

Table 6. Score of molecular docking of 5i7s and Triclosan by Chimera

Score (kcal/mol)	RMSD lower bound	RMSD upper bound	Hydrogen bonds
-7.6	1.591	6.065	1
-7.2	3.185	6.181	0
-7.1	1.997	6.036	0
-7.0	1.906	6.136	0
-6.9	2.05	2.461	0
-6.9	3.799	5.582	0
-6.7	2.223	6.015	0
-6.6	12.261	13.712	0
-6.6	3.88	7.354	0

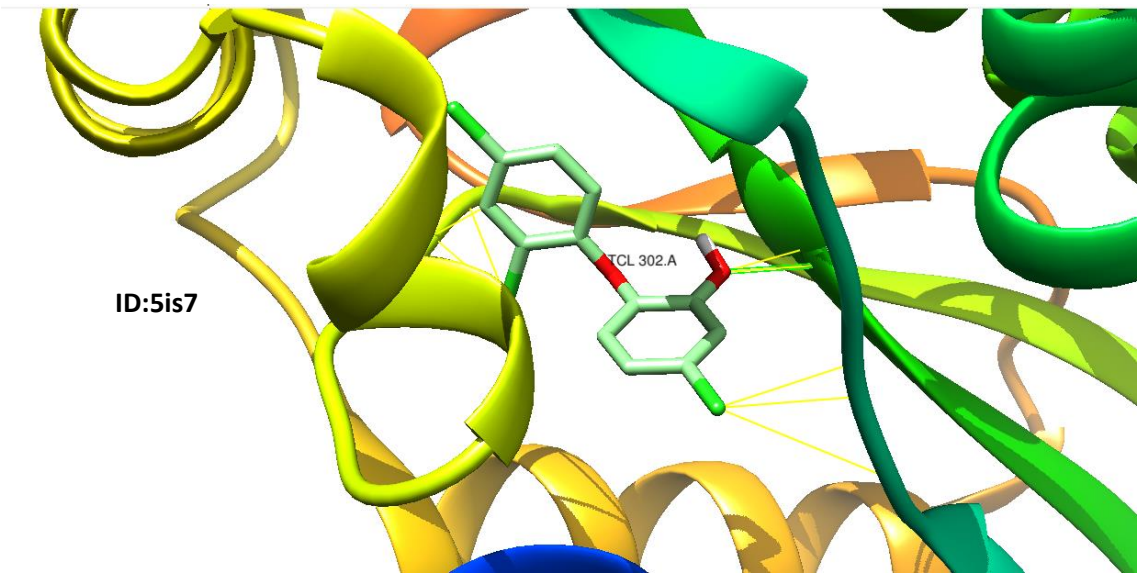


Figure 17. Interaction between 5i7s and Triclosan

e) Describe and illustrate the similarities and differences of the ligand pose and interactions between protein and ligand in the model of the complex and the actual structure from the PDB. Why do you think any differences have arisen? If there was no PDB structure of the target complex, how useful would your model structure be in understanding ligand binding? [4 Marks]

The evaluation revealed slightly difference of the ligand in the new model complex which could be affected in the *in vivo* trials. However, the stabilization of the ligand was found by a hydrogen bond (O17-SER145), in contrast with the actual PDB structure, which shows three hydrogen bonds. Moreover, there are 43 interactions in the new model, displaying ionic interaction and Van der Waals interaction. In addition, the molecule shows interaction in F 206, Y 146, I 92, K163, I 20 **(figure 18).** The **figure 19** shows the docking was performed in the hydrophobic site which is an important factor in molecular recognition and accurate prediction of the binding modes of nonpolar molecules to proteins in aqueous solvent is useful for ligand docking and drug design **(Majeux, et al., 2000).** The table 6 depicts the values obtained by docking that the lowest score represents the energy of binding. In this case, -7.6 kcal/mol shows a good affinity and it is related to Gibbs energy **(Pantsar & Poso, 2018)**. Thus, the difference is slightly due to the model is close to the target protein. Finally, the complex shows a good candidate to carry out an *in vitro* trial.

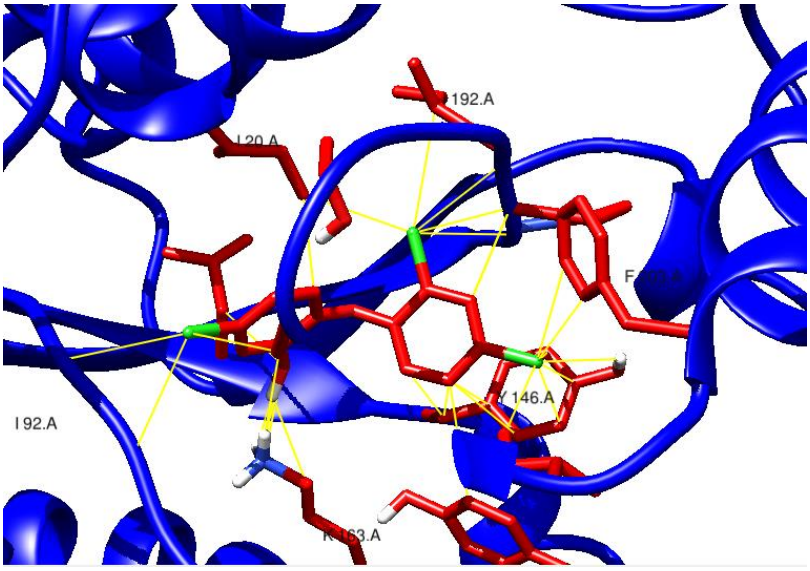


Figure 18. Interaction between 5i7s and Triclosan

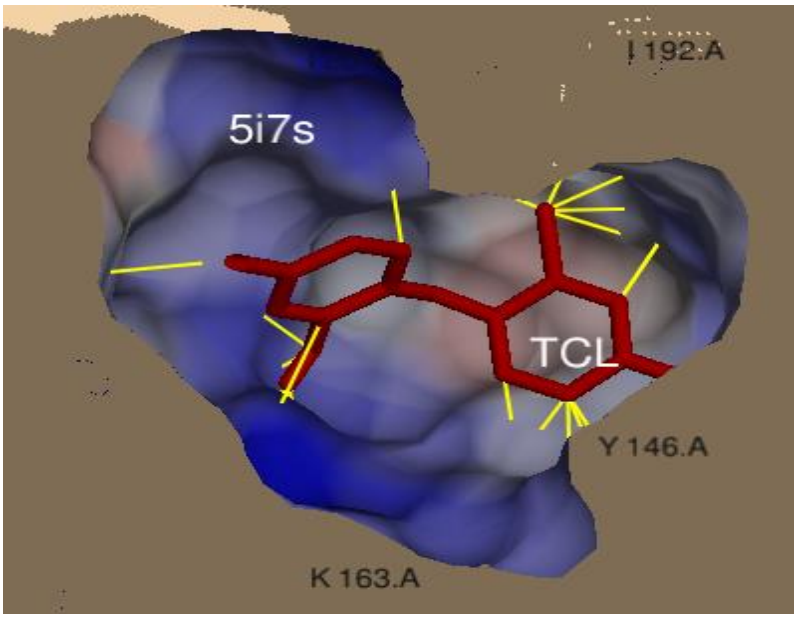


Figure 19. Hydrophobic site finding in 5i7s complex

BIBLIOGRAPHY

Aslanzadeh, V., & Ghaderia, M. (2012). Homology Modeling and Functional Characterization of PR-1a Protein of Hordeum vulgare subsp. Vulgare. *Bioinformation*, *8*(17), 807–811. doi: 10.6026/97320630008807

Cuff AL, Sillitoe I, Lewis T, Redfern OC. et al. (2009). The CATH classification revisited -- Architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucl Acids Res.* **37**:D310–D314. doi: 10.1093/nar/gkn877. Retrieved: http://www.cathdb.info/pdb/4m89

Gregory A Petsko & Dagmar Ringe. (2014). *Protein structure and function-New Science*. Science.

Heath, R. J., Yu, Y. T., Shapiro, M. A., Olson, E., & Rock, C. O. (1998). Broad spectrum antimicrobial biocides target the FabI component of fatty acid synthesis. *Journal of Biological Chemistry*, *273*(46), 30316–30320. doi: 10.1074/jbc.273.46.30316

Ivanović, V., Rančić, M., Arsić, B., & Pavlović, A. (2020). Lipinski's rule of five, famous extensions and famous exceptions. *Popular Scientific Article*, *3*(1), 171–177.

Jones, D. T. (1999). GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology*, *287*(4), 797–815. doi: 10.1006/jmbi.1999.2583

Kumar, S., Govil, S., Singh Ranawat, A., Yadav, M. K., Tiwari, M., & Srivastava, D. (2014). In-silico 3D Structure Modeling of IpLA7 Protein of Borrelia burgdorferi using Homology Modeling. *IJPTB. International Journal of Pharmaceutical Technology and Biotechnology*, *1*(28), 28–33.

Majeux, N., Scarsi, M., Tenette-Souaille, C., & Caflisch, A. (2000). Hydrophobicity maps and docking of molecular fragments with solvation. *Perspectives in Drug Discovery and Design*, *20*, 145–169. doi: 10.1023/A:1008785123705

Murzin AG, Brenner SE, Hubbard TJP, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247:536-540. Retrieved: http://scop.mrc-lmb.cam.ac.uk/term/8063132

Nanson, J. & Forwood, J. (2013).Crystal Structure of Enoyl-Acyl Carrier Protein Reductase (FabI) from Neisseria meningitidis in complex with NAD+ and Triclosan. Retrieved : https://www.rcsb.org/structure/4M89

Pantsar, T., & Poso, A. (2018). Binding affinity via docking: Fact and fiction. *Molecules*, *23*(8), 1DUMMY. doi: 10.3390/molecules23081899

Ramírez, D., & Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules*, *23*(5), 1–17. doi: 10.3390/molecules23051038

Šali, A. & Ther, L. (1993). Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815.

Seol, G., Park, H., Ahn, Y. J., & Kang, L. W. (2019). Crystal Structure of Enoyl-Acyl Carrier Protein Reductase (FabI) from Acinetobacter baumannii as a Target for Broad-Spectrum Antibacterial Drug. *Bulletin of the Korean Chemical Society*, *40*(10), 1039–1042. doi: 10.1002/bkcs.11861

Shantanam, S., & MUELLER. (2018). Biogenesis and Homeostasis of Nicotinamide Adenine Dinucleotide Cofactor. *Physiology & Behavior*, *176*(1), 139–148. doi: 10.1128/ecosalplus.3.6.3.10.Biogenesis

Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, *17*(4), 355–362. doi: 10.1002/prot.340170404

Syed, R., Rani, R., Sabeena, Masoodi, T. A., Shafi, G., & Alharbi, K. (2012). Functional analysis and structure determination of alkaline protease from Aspergillus flavus. *Bioinformation*, *8*(4), 175–180. doi: 10.6026/97320630008175

The UniProt Consortium, UniProt. (2023): the Universal Protein Knowledgebase in 2023, *Nucleic Acids Research*, Volume 51, Issue D1, 6 January 2023, Pages D523–D531. Retrieved : www.uniprot.org/uniprotkb/A1KVU8/entry

Zhang, Y. & Skolnick, J. (2010). Scoring function for automated assessment of protein structure template quality, Proteins, 57: 702-710. Retrieved: https://zhanggroup.org/TM-score/tmp/853568.html

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.J Mol Biol. 2018 Jul 20. S0022-2836(17)30587-9. Retrieved : https://toolkit.tuebingen.mpg.de/jobs/8079215

Appendix 1

N°	Name	Database	Smile string
1	CG-400549	ChEMBL	Cc1c(N)cccc1Cn1ccc(OCCc2cccs2)cc1=O
2	AFN-1252	ChEMBL	Cc1c(CN(C)C(=O)/C=C/c2cnc3c(c2)CCC(=O)N3)oc2ccccc12
3	AFABICIN	ChEMBL	Cc1c(CN(C)C(=O)/C=C/c2cnc3c(c2)CCC(=O)N3COP(=O)(O)O)oc2ccccc12
4	Soneclosan	DrugBank	OC1=C(OC2=CC=C(Cl)C=C2)C=CC(Cl)=C1
5	Indole Naphthyridinone	DrugBank	CN(CC1=CN(C)C2=CC=CC=C12)C(=O)\C=C\C1=CN=C2NC(=O)CCC2=C1
6	3-(6-Aminopyridin-3-Yl)-N-Methyl-N-[(1-Methyl-1h-Indol-2-Yl)Methyl]Acrylamide	DrugBank	CN(CC1=CC2=C(C=CC=C2)N1C)C(=O)\C=C\C1=CN=C(N)C=C1
7	4-(2-Thienyl)-1-(4-Methylbenzyl)-1h-Imidazole	DrugBank	CC1=CC=C(CN2C=NC(=C2)C2=CC=CS2)C=C1

8	3-[(Acetyl-Methyl-Amino)-Methyl]-4-Amino-N-Methyl-N-(1-Methyl-1h-Indol-2-Ylmethyl)-Benzamide	DrugBank	CN(CC1=C(N)C=CC(=C1)C(=O)N(C)CC1=CC2=C(C=CC=C2)N1C)C(C)=O
9	1,3,4,9-Tetrahydro-2-(Hydroxybenzoyl)-9-[(4-Hydroxyphenyl)Methyl]-6-Methoxy-2h-Pyrido[3,4-B]Indole	DrugBank	OC1=CC=C(CN2C3=C(CCN(C3)C(=O)C3=CC=C(O)C=C3)C3=CC=CC=C23)C=C1
10	Beta-D-Glucose	DrugBank	OC[C@H]1O[C@@H](O)[C@H](O)[C@@H](O)[C@@H]1O
11	2-(TOLUENE-4-SULFONYL)-2H-BENZO[D][1,2,3]DIAZABORININ-1-OL	DrugBank	CC1=CC=C(C=C1)S(=O)(=O)N1N=CC2=C(C=CC=C2)B1O
12	6-METHYL-2(PROPANE-1-SULFONYL)-2H-THIENO[3,2-D][1,2,3]DIAZABORININ-1-OL	DrugBank	CCCS(=O)(=O)N1N=CC2=C(C=C(C)S2)B1O
13	Triclocarban	DrugBank	ClC1=CC=C(NC(=O)NC2=CC(Cl)=C(Cl)C=C2)C=C1
14	**Pyrazinamide**	DrugBank	NC(=O)C1=NC=CN=C1
15	**Ethionamide**	DrugBank	CCC1=NC=CC(=C1)C(N)=S
16	**Isoniazid**	DrugBank	NNC(=O)C1=CC=NC=C1
17	Genz-10850	DrugBank	O=C(N1CCN(CC1)C1C2=CC=CC=C2C2=CC=CC=C12)C1=CC2=C(NC=C2)C=C1
18	(3S)-N-(3-CHLORO-2-METHYLPHENYL)-1-CYCLOHEXYL-5-OXOPYRROLIDINE-3-CARBOXAMIDE	DrugBank	[H][C@]1(CN(C2CCCCC2)C(=O)C1)C(=O)NC1=C(C)C(Cl)=CC=C1
19	N-(4-METHYLBENZOYL)-4-BENZYLPIPERIDINE	DrugBank	CC1=CC=C(C=C1)C(=O)N1CCC(CC2=CC=CC=C2)CC1

20	(3S)-1-CYCLOHEXYL-5-OXO-N-PHENYLPYRROLIDINE-3-CARBOXAMIDE	DrugBank	[H][C@]1(CN(C2CCCCC2)C(=O)C1)C(=O)NC1=CC=CC=C1
21	(3S)-1-CYCLOHEXYL-N-(3,5-DICHLOROPHENYL)-5-OXOPYRROLIDINE-3-CARBOXAMIDE	DrugBank	[H][C@]1(CN(C2CCCCC2)C(=O)C1)C(=O)NC1=CC(Cl)=CC(Cl)=C1
22	(3S)-N-(3-BROMOPHENYL)-1-CYCLOHEXYL-5-OXOPYRROLIDINE-3-CARBOXAMIDE	DrugBank	[H][C@]1(CN(C2CCCCC2)C(=O)C1)C(=O)NC1=CC=CC(Br)=C1
23	(3S)-N-(5-CHLORO-2-METHYLPHENYL)-1-CYCLOHEXYL-5-OXOPYRROLIDINE-3-CARBOXAMIDE	DrugBank	[H][C@]1(CN(C2CCCCC2)C(=O)C1)C(=O)NC1=C(C)C=CC(Cl)=C1
24	**Pretomanid**	DrugBank	[O-][N+](=O)C1=CN2C[C@@H](COC2=N1)OCC1=CC=C(OC(F)(F)F)C=C1
25	Triclosan Sulfate	ZINC	O=S(=O)(O)Oc1cc(Cl)ccc1Oc1ccc(Cl)cc1Cl

Appendix 2

Conversion of smiles to sdf format

```
smi_string<- "CCCS(=O)(=O)N1N=CC2=C(C=C(C)S2)B1O"
sdf <- smiles2sdf(smi_string)
write.SDF(sdf[[1]], file="/Final_project/12.sdf")
sdf3D <- generate3DCoords(sdf)
write.SDF(sdf3D[[1]], file="/Final_project/12_3D.sdf")
```

dataset of ligands

```
sdfset<- read.SDFset("/Final_project/drugset.sdf")
length(sdfset)
```

Tanimoto_coefficient

```
sdfset <- read.SDFset("/Final_project/drugset.sdf")
fp2set <- fingerprintOB(sdfset, "FP2")
fpSim.out <- fpSim(fp2set[1], fp2set, method="Tanimoto", cutoff=0.1, top=25)
fpSim.out
```

 # Plot Histogram

```
hist(fpSim.out, xlab= "Tanimoto coefficient", col = "blue", border= "black")
```

#Standar Deviation

```
sd(fpSim.out)
```

#Mean

```
mean(fpSim.out)
```

Appendix 3

Count: 1679