CHAPTER 4

RESULTS & DISCUSSION

4.1 Target sequence

In this research project, the target sequence is originated from *Drosophila melanogaster*. Searching through NCBI Protein database, it yields result of glutathione S transferase D3 from *D. melanogaster* with the length of 199 amino acids and accession number AA041561 whereas the NCBI reference sequence is NP_788656.1 (Figure 4.1). The length of DmGSTD3 is only 199 amino acids which is shorter than the other members of Delta GST from the same ancestor (Table 4.1). As shown in the table below, the length of other members in Delta class exceed 209 amino acids whereas for GSTD3, its length is only 199 amino acids.

From the query sequence and NCBI Protein graphics results, the GST N-terminal region can be found at position 1-58 residues whereas GST C-terminal region can be found at position 72-188 residues , which differ from positions reported in GSTD1 which are 1-79 and 86-208, respectively (Figure 4.2) (Low et al. 2010). Albeit they have different positions of terminal domain, mainly the N-terminal domain spanning at the first 79 residues whereas the C-terminal domain spanning at the next 86 residues. From the NCBI Protein database, I obtained that the position of glutathione binding occurred at position 34-49 which codes for protein histidine, serine, and isoleucine as well as glutamic acid and others whereas the substrate binding pocket is located at positions 85-187 which differ from the members of Delta that have been reported such as in GSTD1 whereby its GSH binding site occur at position 10-66 and substrate binding pocket occur at position 102-204 (Low et al. 2010). The substrate binding pocket which is the H-site is important in the binding of xenobiotics

alkylating agents such as carcinogens, therapeutic drugs, environmental toxins, and products of oxidative stress. GSH binding site is common to all GSTs; however the substrate binding site is different between classes and isoforms. This speculation was based on similarity between target sequence and sequences with highest similarity.

The conserved residue which is tyrosine is present in other members of Delta class exactly at positions 5 and 6 but absence in Delta isoform 3. Also, from the sequence alignment among the members of Delta, it shows that in D3, there are truncations of several amino acids at the N-terminal of the sequence (Figure 4.3). This truncation caused conserved tyrosine at positions 5 and 6 to be missing in which it appears in other members. Tyrosine residue is important since it determines the catalytic activity of GSTs enzymes.

```
>gi|28571670|ref|NP_788656.1| glutathione S transferase D3
[Drosophila melanogaster]
MVGKALGLEFNKKIINTLKGEQMNPDFIKINPQHSIPTLVDNGFTIWESR
AILVYLVEKYGKDDALYPKDIQKQAVINQRLYFDMALMYPTLANYYYKAF
TTGQFGSEEDYKKVQETFDFLNTFLEGQDYVAGDQYTVADIAILANVSNF
DVVGFDISKYPNVARWYDHVKKITPGWEENWAGALDVKKRIEEKQNAAK
```

Figure 4.1: GST isoform 3 from fruit fly (NCBI Protein).

Table 4.1: Amino acid length of Delta members

Delta Member	Length of Amino Acid
1	209
2	215
3	199
4	215
5	216
6	215
7	224
8	212
9	218
10	210

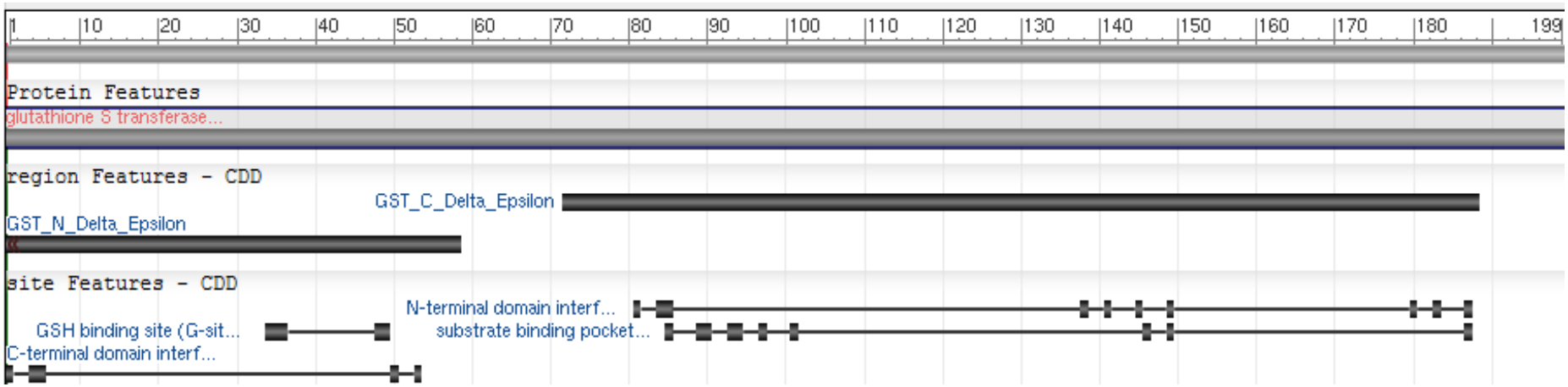

Figure 4.2: Positions of N-terminal, C-terminal domain, GSH binding site, and substrate binding pocket in GSTD3 (NCBI Protein).

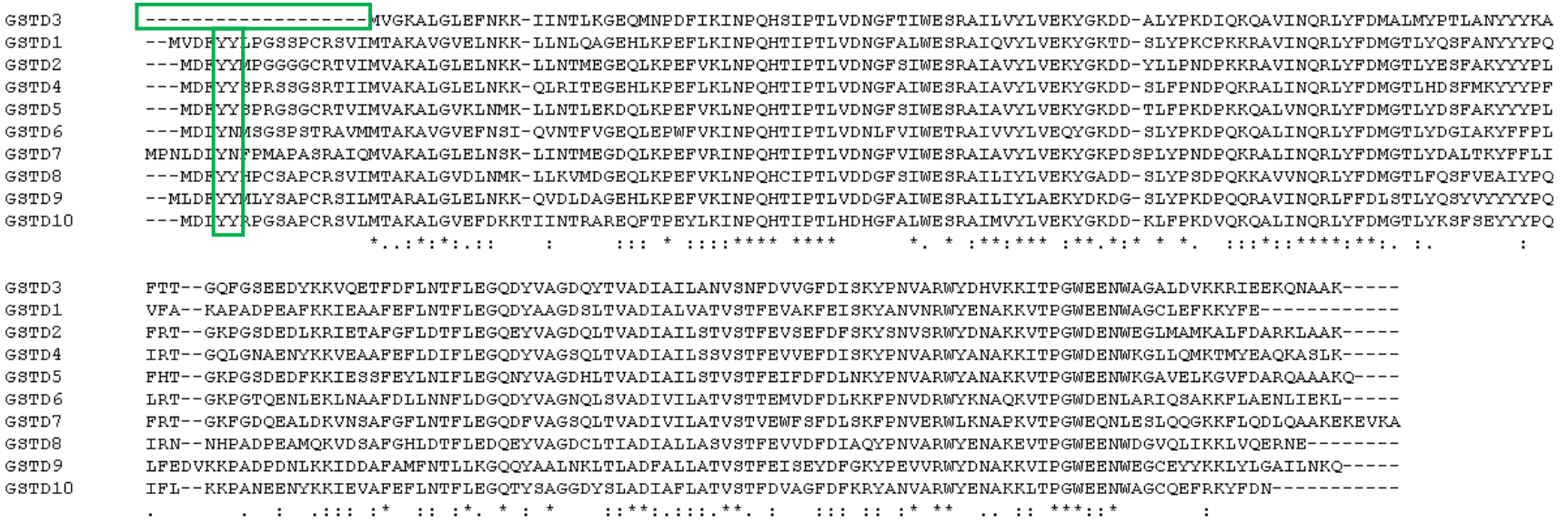

Figure 4.3: Multiple sequence alignment of D1 until D10 of *D. melanogaster*. Tyrosine residue in green box is the catalytic residue which is conserved among D1 until D2 and D4 until D10. The missing of tyrosine residue in D3 is due to truncation of 15 amino acid at the N-terminal of the protein sequences.

4.2 Template sequences

The template sequences were searched by BLAST-ing the protein sequences of GSTD3 from *Drosophila melanogaster* against the sequences in UniProt database. From BLAST program in UniProt database, the result produced numbers of sequences in which the sequence similarities between target and template sequence were recorded between 95% to 39%. Although GD18816 produced highest similarity with target sequence, however, GD18816 did not have experimentally determined crystallographic 3D structure. Therefore, in order to select the suitable templates, I carried out a thorough search along the database. Fortunately, two sequences with 63% similarity with the target sequence and both of them were having the experimentally determined 3D structure, which are GSTD10 and GSTD1. Remarkably, both of them originated from the common ancestor with the target sequence which is *Drosophila melanogaster*. Moreover, another two template sequence with predetermined 3D structures also being discovered (Table 4.2).

In making the decision to choose and select an appropriate template among four template sequences obtained from UniProt database, those four protein sequences were run through Modeller program. The structural and sequence similarity between four templates were assessed. Basically, the program will read each of the six pdb files (3EIN, 3MAK, 3GH6, 3F6F, 1JLV, and 1PN9) and aligned them along to each other (as in multiple sequence alignments). Then, the alignment were compared based on the structures by comparing their atomic positions and distances, distances of dihedral angles between mainchain and sidechain, percentage of sequence identities, as well as other properties (Table 4.3). The comparisons of those six templates yield result of distances between two template sequences which later been used in developing the clustering tree (Figure 4.4).

Table 4.2: Information about template sequences obtained from UniProt database.

Entry	PDB ID	Protein Names	Organism	Length	Identity	Score	E-value	Query coverage
Q9VGA1	3F6F (1.60 Å) 3GH6 (1.65 Å)	Glutathione S transferase D10	*Drosophila melanogaster*	210	63%	665	2×10^{-86}	96%
P20432	3EIN (1.13Å) 3MAK (1.80 Å)	Glutathione S-transferase 1-1	*Drosophila melanogaster*	209	63%	668	6×10^{-87}	94%
Q9GNE9	1JLV (1.75 Å)	Glutathione transferase GST1-3	*Anopheles dirus*	209	55%	589	6×10^{-75}	96%
Q93113	1PN9 (2.00 Å)	Glutathione S-transferase 1, isoform D	*Anopheles gambiae*	209	52%	561	1×10^{-70}	95%

4.2.1 Selecting the appropriate templates

The comparison obtained from Modeller showed that 3F6F and 3GH6 as well as 3MAK and 3EIN were almost identical in terms of their structural and sequences (Table 4.3). This is because the sequence identity and identical residues between 3F6F and 3GH6 record 100 and 209 correspondingly. Meanwhile, the sequence identity and identical residues between 3MAK and 3EIN were 100 and 207 respectively. However, 3EIN has the highest crystallographic resolutions compared with the crystallographic resolutions of 3F6F, 3GH6, and 3MAK (1.13 Å, 1.6 Å, 1.65 Å, and 1.8 Å respectively) which eliminating 3MAK. A second group of structures which are 1JLV and 1PN9 share same similarities among each other as their sequence identity and identical residues were 81 and 168 respectively. From the second group, 1PN9 has the poorest resolution which is 2.0 Å compared to 1JLV which is 1.75 Å, therefore eliminating 1PN9 (Figure 4.4).

Regarding the sequence identity between template sequences and query sequence, all of these three structures, 3EIN, 3GH6, and 3F6F has highest identities with the target sequence which are 63%. However, for 1JLV, the sequence identity is 55% which resulted in eliminating the 1JLV. This leaves only 3EIN, 3GH6, and 3F6F. 3GH6 and 3F6F are same sequences with different properties, since 3GH6 is the three dimensional structure in complex with glutathione and 3F6F is the 3D structure without glutathione. Referring to their pairwise sequence alignment between query sequence, Delta isoform 3, and target sequence, Delta isoform 1 and 10, the results produced are not too varied (Figure 4.5 and Figure 4.6). From the pairwise alignment, D3 and D1 have more sequence identity than D3 and D10, 121 amino acid and 120 amino acid respectively. Since 3EIN has the highest crystallographic resolution which is 1.13 Å, 121 of sequence identity with query, and its score is 668, therefore, 3EIN is a suitable template for constructing the three dimensional

structure of glutathione S-transferase Delta isoform 3. Also, during the pairwise alignment between D10 and D3, a gap is introduced into the protein sequence of D3 which affect the score. On the other hand, during the pairwise alignment between D3 and D1, there is no gap introduced inside the sequence of D3 itself.

Table 4.3: Sequence identity among templates using Modeller. Red color represents number of residues, green color depicts number of identical residues, and blue color represents percentage of sequence identity. The bold numbers show the highest score. Target sequence is not included.

	1JLV	**1PN9**	**3F6F**	**3EIN**	**3MAK**	**3GH6**
1JLV	207	168	133	146	146	133
1PN9	81	209	125	140	140	125
3F6F	64	60	210	148	148	**209**
3EIN	71	68	71	209	**207**	148
3MAK	71	67	71	**100**	209	148
3GH6	64	60	**100**	71	71	210

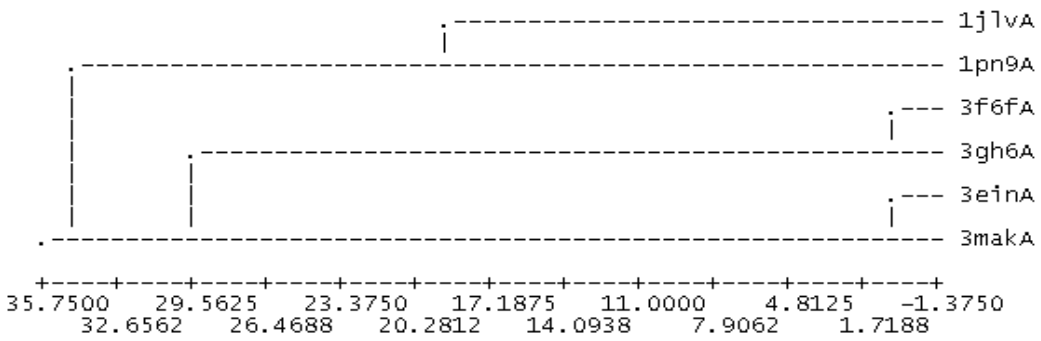

Figure 4.4: Weighted pair-group average clustering based on distance matrix

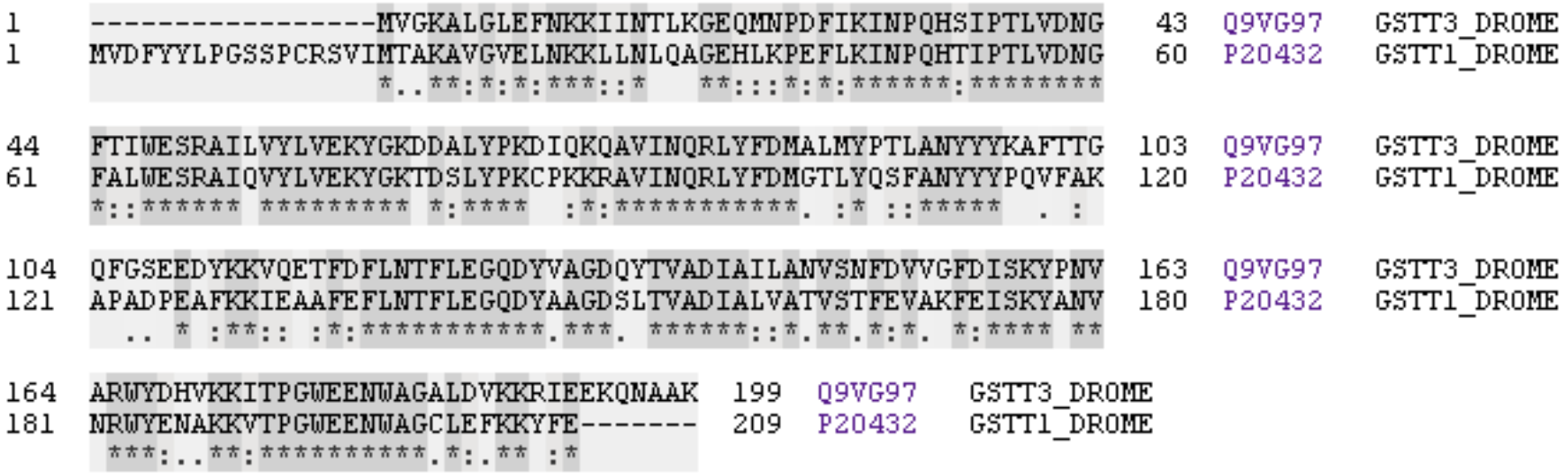

Figure 4.5: Pairwise sequence alignment between GSTD3 (Q9VG97) and GSTD1 (P20432). Sequences highlighted in grey color and asterisk sig '*' represent similarity.

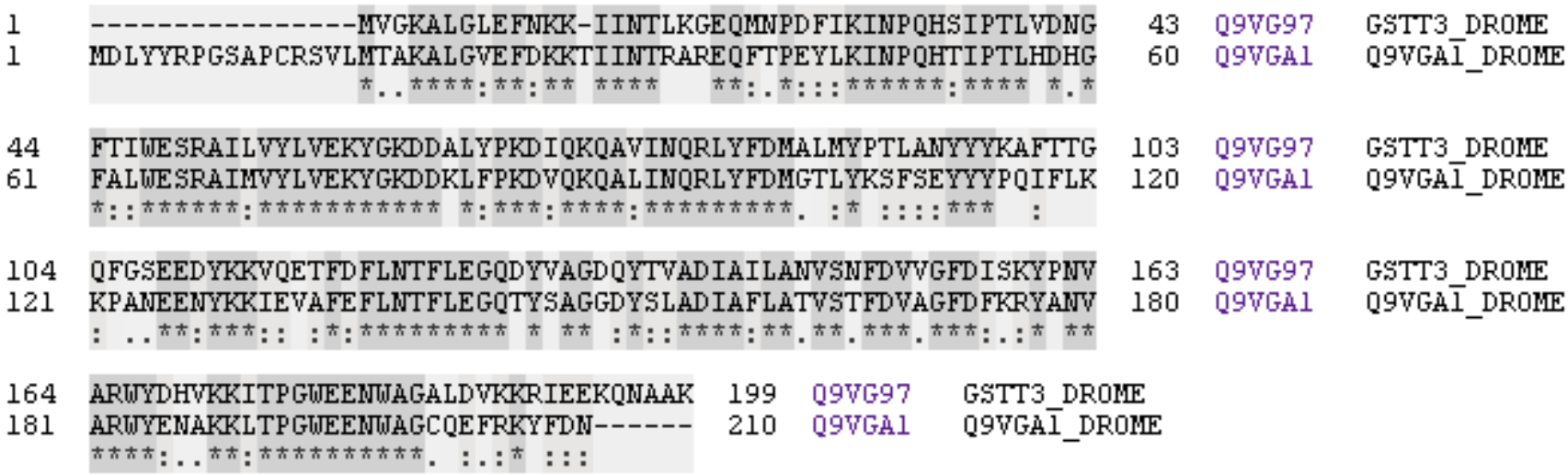

Figure 4.6: Pairwise sequence alignment between GSTD3 (Q9VG97) and GSTD10 (Q9VGA1). Sequences highlighted in grey color and * represent similarity. Note that, in GST isoform 10, there is a gap introduced.

4.2.2 Analysis of template structure, 3EIN.

3EIN is the PDB ID for glutathione S-transferase enzyme from Delta class isoform 1 which originated from *Drosophila melanogaster* and appears as homodimeric structure (Figure 4.7). It consists of 209 amino acid which involves in the reactions of glutathione transferase activity and DDT-dehydrochlorinase activity. Apart from that, GSTD1 also plays a role in protein binding, transferase activity, as well as lyase activity. The domains include in DmGSTD1 are N-terminal domain and C-terminal domain. The N-terminal domain is at position 2-75 whereas the C-terminal domain starts at position 89-205. Usually, the glutathione substrate binds with the residues in the vicinity of N and C-terminal domains, as the residues that important in catalytic activity are resides in the N-terminal domain. Concerning about the GSH binding site and substrate binding pocket, they appeared at positions 10 – 66 and 102 – 204, respectively (Figure 4.8).

In terms of secondary structure of glutathione S-transferase Delta class isoform 1, from the STRIDE (Protein Secondary Structure Assignment from Atomic Coordinates), it predicted that 3EIN comprises of 10 alpha-helices, 4 beta-strands, and 7 turns (Figure 4.9). Also, regarding this secondary structure of 3EIN, 61% of amino acid is α-helices, while 7% of amino acid is β-sheets (Wongsantichon et al. 2012). The remaining 32% of amino acid constitutes turns and loops. In addition, in respect to the amino acid that responsible in catalytic activity of 3EIN they scattered around N-terminal domain, which consistent with what had been confirmed before. Furthermore, the residues that have contacts with the ligand in ligand binding to N-terminal domain are glutamic acid (position 65), arginine (position 67), serine (position 66 and 10), histidine (position 51), isoleucine (position 53), threonine (position 52), proline (position 12 and 54), methionine (position 102) and histidine (position 39)(Figure 4.10). The first 6 residues form hydrogen bond with the

ligand, whereas the rest five residues form van der Waals bond with the ligand. Figure 4.11 shows the conformation of N-terminal, C-terminal, and GSH binding site in 3D form.

Ligand usually binds onto N-terminal domain of GST1-1 is glutathione. Glutathione is a tripeptide protein that has plays many roles in maintaining the cells. Normally, it conjugates the hydrophobic toxic compounds into more water soluble and hydrophilic compounds which later can be easily excreted from the cell. Furthermore, glutathione may also act as a cofactor for some enzyme glutathione peroxidase in which they assist in biochemical transformation by transferring the electron group from lipid hydroperoxide (Figure 4.12).

Glutathione S-transferase Delta class isoform 1 has two three dimensional structures being deposited into the RCSB PDB database which have been experimentally determined (Wongsantichon et al. 2012). The PDB identification codes for GST1-1 are 3EIN and 3MAK. The differences between those two are crystallographic resolution and binding of ligand. According to the crystallographic resolution, 3EIN has a better resolution which is 1.13 Å compared to the 3MAK which is 1.8 Å. In addition, 3EIN is an apo-protein of GST1-1 while 3MAK is the crystal structure of GST1-1 in complex with glutathione. Due to these different properties mentioned above, 3EIN was chosen as template for homology modeling of GST Delta isoform 3.

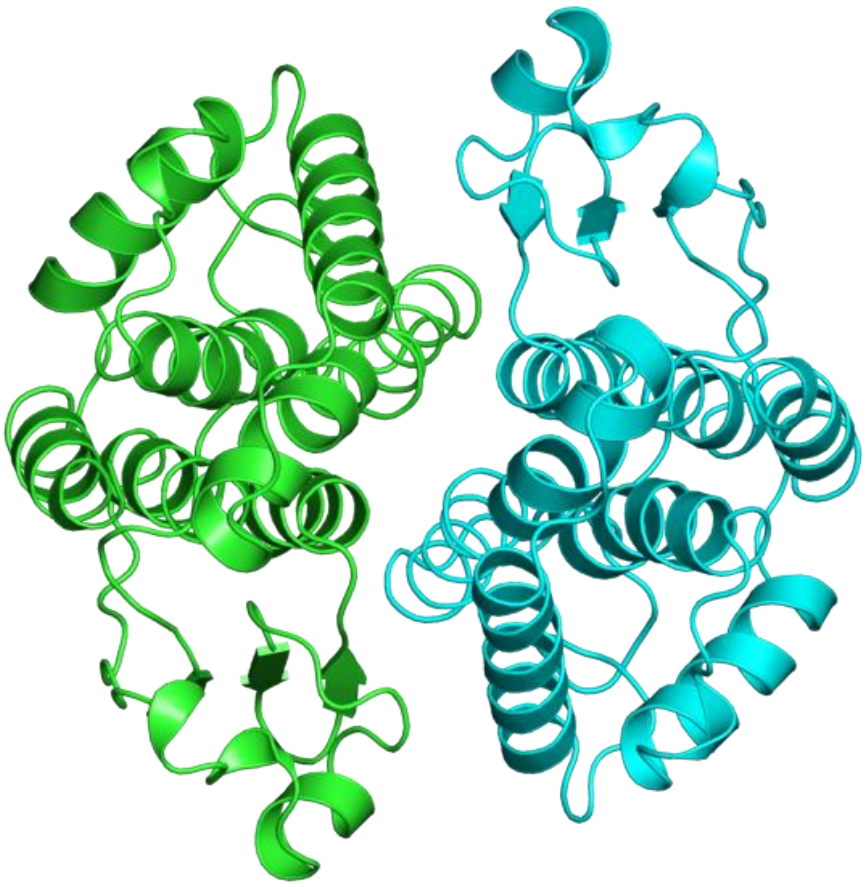

Figure 4.7: Quaternary structure of 3EIN. They appeared as homodimer molecule.

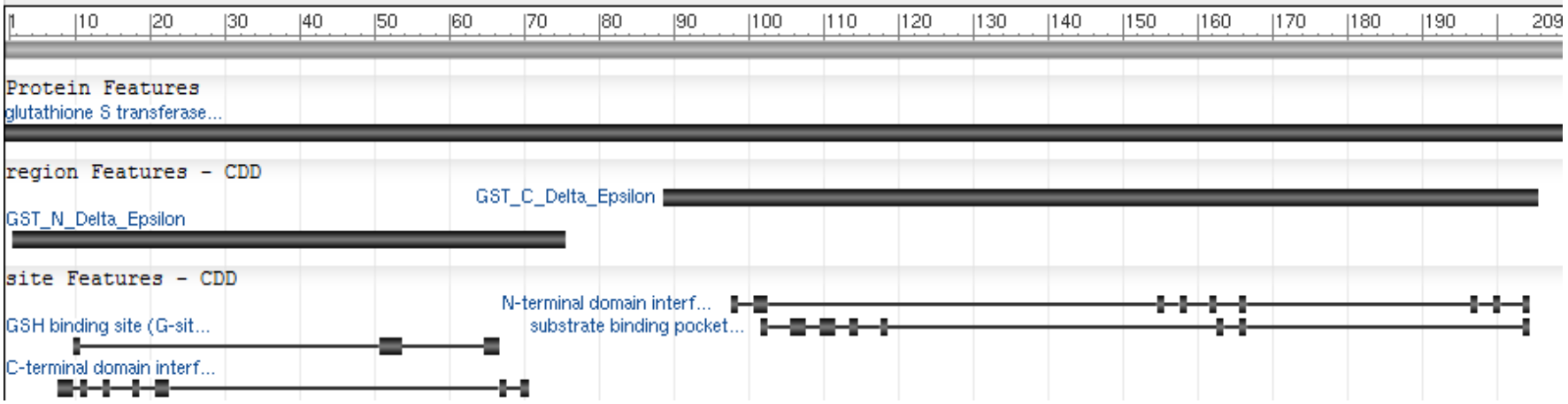

Figure 4.8: Position of GSH binding site, substrate binding pocket, C-terminal domain, and N-terminal domain in GSTD1 (NCBI Protein).

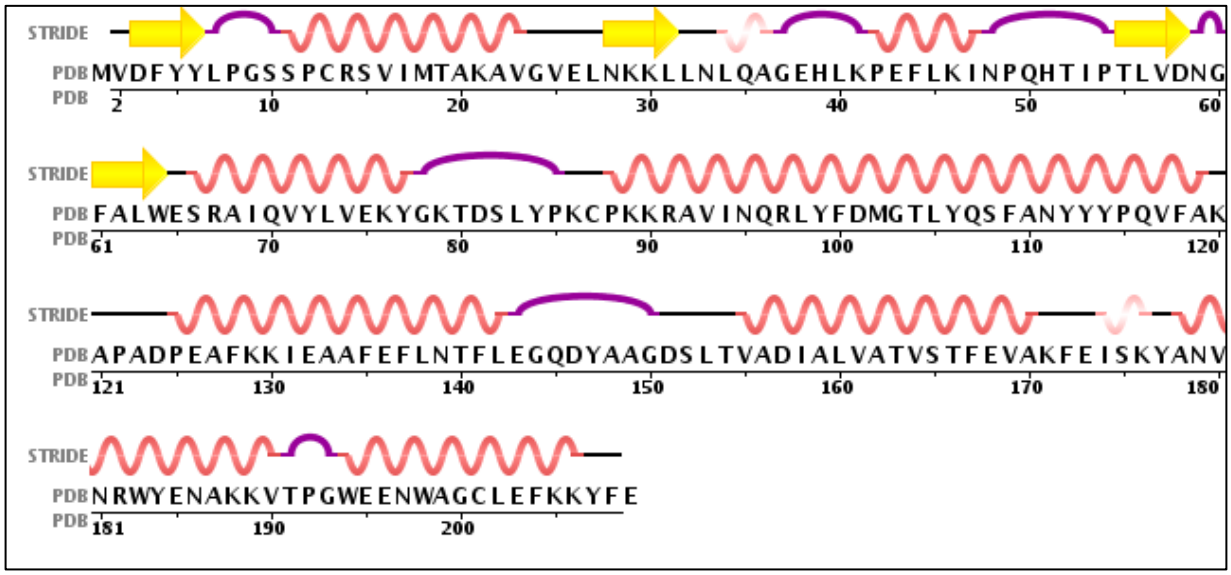

Figure 4.9: Secondary structure of GST isoform 1. Yellow arrow represents beta sheet, purple line represents turn, red curved line represents alpha helix, pink-shiny curved lined represents 3/10 helix (STRIDE)

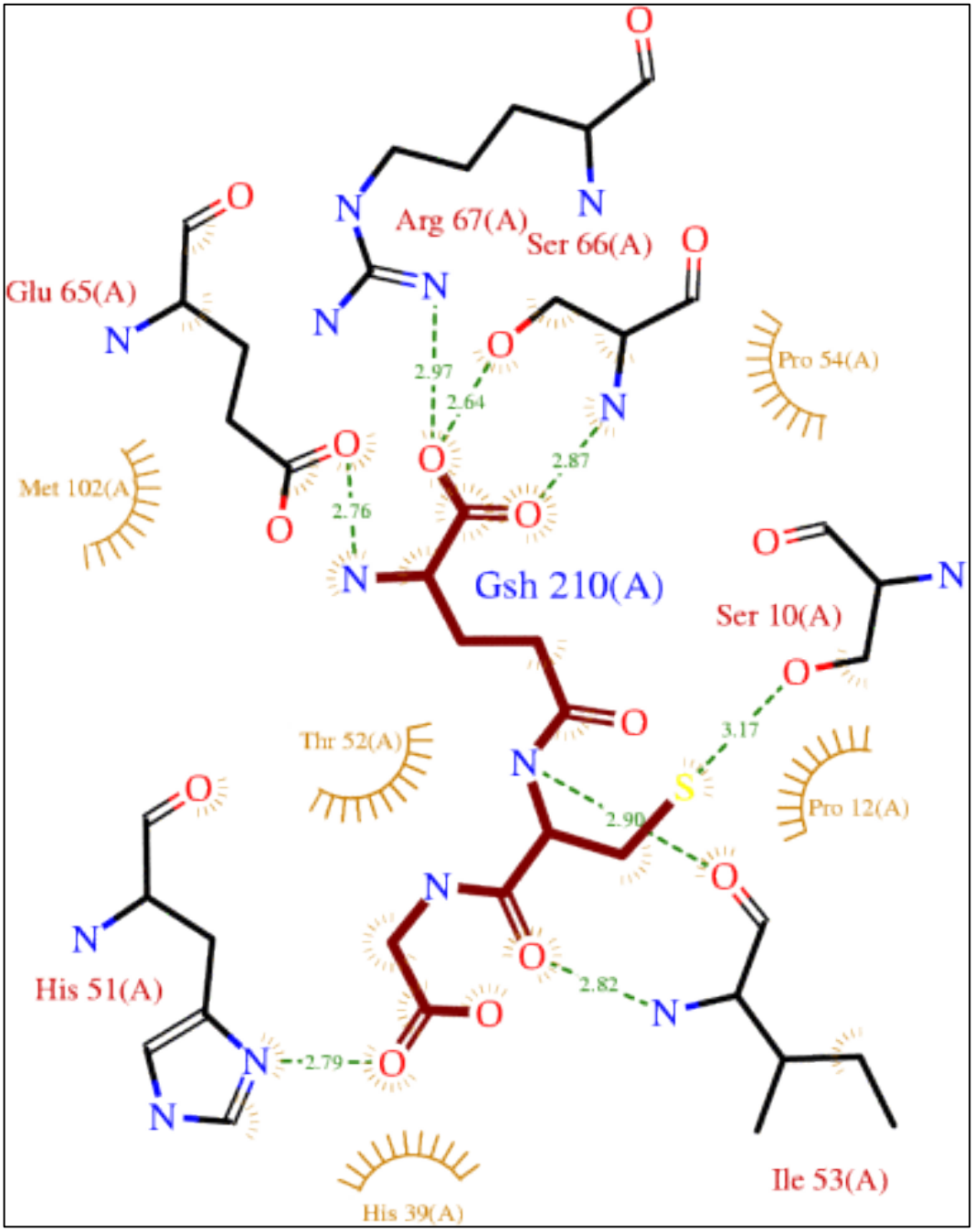

Figure 4.10: Structure of ligand binding and residues involved (LigPlot).

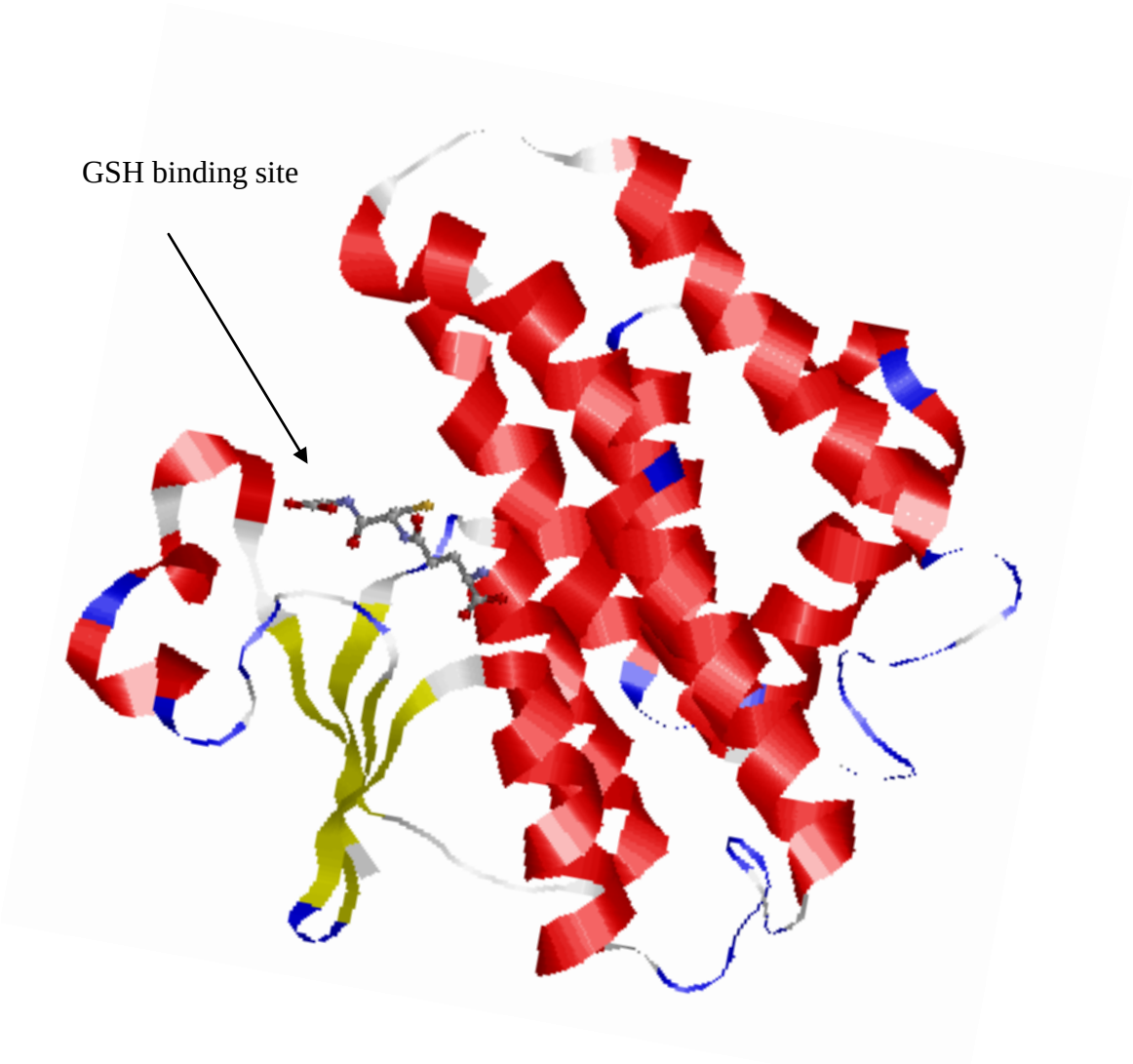

Figure 4.11: 3D structure of 3EIN. Alpha helices were represented in red color, beta sheets represented in yellow color, turns represented in blue color. GSH, which is glutathione, is represented in wireframe and colored by CPK (RasWin)

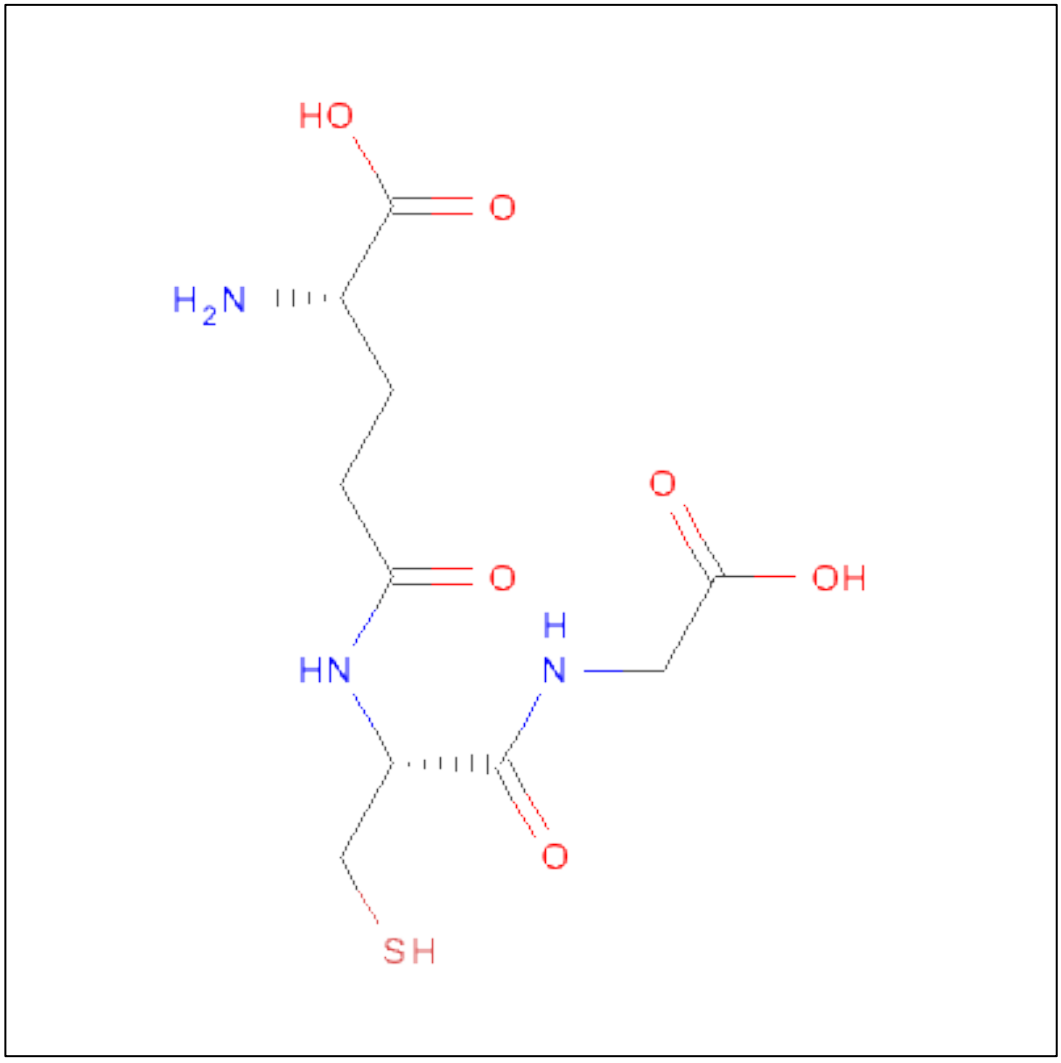

Figure 4.12: Structure of glutathione, $C_{10}H_{17}N_3O_6S$

4.3 Modeling the query sequence

After template sequence has been determined and identified, the modeling of query sequence was carried out by Modeller program. In modeling the GST Delta isoform 3, the alignment between template and query was carried out using Modeller as well. Due to high sequence identity between query and template sequence which is 63%, there is no gaps introduced in the alignment. Consequently, the 3D model for query is build using this target-template alignment file as the reference. A set of 100 structures were generated and the best structure was choose by determining the low energy structure of DOPE potential score and highest GA341 score. The model obtained (which is in pdb file format) from Modeller program was evaluated by using PDBsum.

From the Python script which is buildmodel.py, a set of 100 structures were generated. The best fit model was determined and identified by determining the DOPE score and GA341 score. DOPE (Discrete Optimized Protein Energy) score is used in assessing the good model of target protein. Usually, the lowest DOPE score depicted the good model from bad model. Meanwhile, the GA341 method score uses percentage of sequence identity between target and template alignment. The ranges of GA341 in differentiating good model from bad model is good model having score of 1 whereas the bad model having score of 0. However, GA341 method is not excellent as DOPE score. From the log file of buildmodel.py, the best model has been identified as target.B9990076.pdb whereby its DOPE score and GA341 score were -22787.36328 and 1, respectively.

4.4 Model Evaluation

Model that was generated using buildmodel.py was evaluated using PROCHECK program. PROCHECK program provides analysis of stereochemical properties and quality of all residues in the model file.

4.4.1 Ramachandran plot

The Ramachandran plot of target pdb file showed 94.4% of the residues fall in most favored regions which are region A, B, and L. These regions correspond to core alpha, core beta, and core-left handed alpha. Several amino acids that fall into these regions were alanine, arginine, asparagines and others. 3.9% of 199 amino acid falls into allowed regions which are *a, b, l,* and *p,* whereas only 1 residue falls into generously allowed region. The rest of amino acid, which is 2 amino acid fall into disallowed region. Since 94.4% of total residues were in these allowed regions, therefore, this model has a better stereochemical quality (Figure 4.13).

4.4.2 G-Factors

In distinguished normal stereochemical properties than abnormal stereochemical properties, G-factors were calculated and measured. The measurements include dihedral angles and main chain covalent bonds. In each parameter, they were assessed on phi-psi distribution, chi1-chi2 distribution, chi1 distribution, chi3 and chi 4 distribution, and omega distribution, as well as bond lengths and bond angles, respectively. According to Ramachandran plot, many residues were falling to the allowed regions, as a result, the G-factor for this model have a high score, which is 0.09 (Figure 4.14).

4.5 GSTD3 Secondary Structure Prediction

The pdb file of model obtained from Modeller was used in STRIDE program to predict its secondary structure. In secondary structure prediction, it shows that the query sequence, DmGSTD3 contains two beta sheets in which the positions are threonine at position 38 to aspartic acid at position 41 and phenylalanine (position 44) to tryptophan (position 47). This structure prediction is somewhat different from I template sequence, in which the template sequence contains four beta sheets. Out of 199 amino acids, 127 of them are alpha helix, 8 of them are beta sheets, and the rest are turn or loop.

4.5.1 Comparisons of query and target

The predicted secondary structure of query was compared to secondary structure of 3EIN, the template sequence. From the comparison, it shows that the query structure lacked of 2 beta sheets compared to 3EIN which has 4 beta sheets. In addition, the positions of beta sheets were also different. Moreover, the amount of alpha helix in the query structure is 11 whereby 3EIN only have 13 alpha helices (Figure 4.15).

Besides that, as mentioned above, the amino acid residues in 3EIN that responsible in catalytic activity and interacts with ligand during ligand binding is different from query sequence. In 3EIN, the catalytic residue is tyrosine at positions 5 and 6. However, this catalytic residue was not present in query sequence due to truncation of 15 amino acid at the N-terminal domain. Meanwhile, residues that responsible in ligand binding in 3EIN which are glutamic acid (65), serine (66), arginine (67), serine (10), isoleucine (53), histidine (51 and 39), threonine (52), proline (12 and 54) and methionine (102), in which some of them are present in query sequence as well. Based on the pairwise sequence alignment between target sequence and template sequence, I identified that the residues

having contact with ligand in query sequence were histidine (51), isoleucine (53), proline (54), glutamic acid (65), serine (66), arginine (67), and methionine (102). However, residues at position 52, 39, 10, and 12 were not the same as 3EIN (Figure 4.16). At these positions, they were being replaced by serine, and glutamine respectively, but, at two latter positions, there were no residues. Nevertheless, the residue that replaced basic histidine at position 39 is a polar residue. On the other hand, the residue at position 52 in query is a polar residue, which the same as template.

Although the residues that crucial for ligand binding were highly similar between query and target, however, their secondary structures were a bit different even though there were some similarities. The secondary structures for those importance residues in 3EIN were turn/loop, alpha-helix, turn/loop, turn/loop, turn/loop, turn/loop, turn/loop, turn/loop, alpha-helix, alpha-helix, and alpha-helix according to their position respectively. In the meantime, the secondary structures for query sequence were alpha-helix, turn/loop, turn/loop, turn/loop, turn/loop, turn/loop, alpha-helix, alpha-helix, and alpha-helix (Table 4.4).

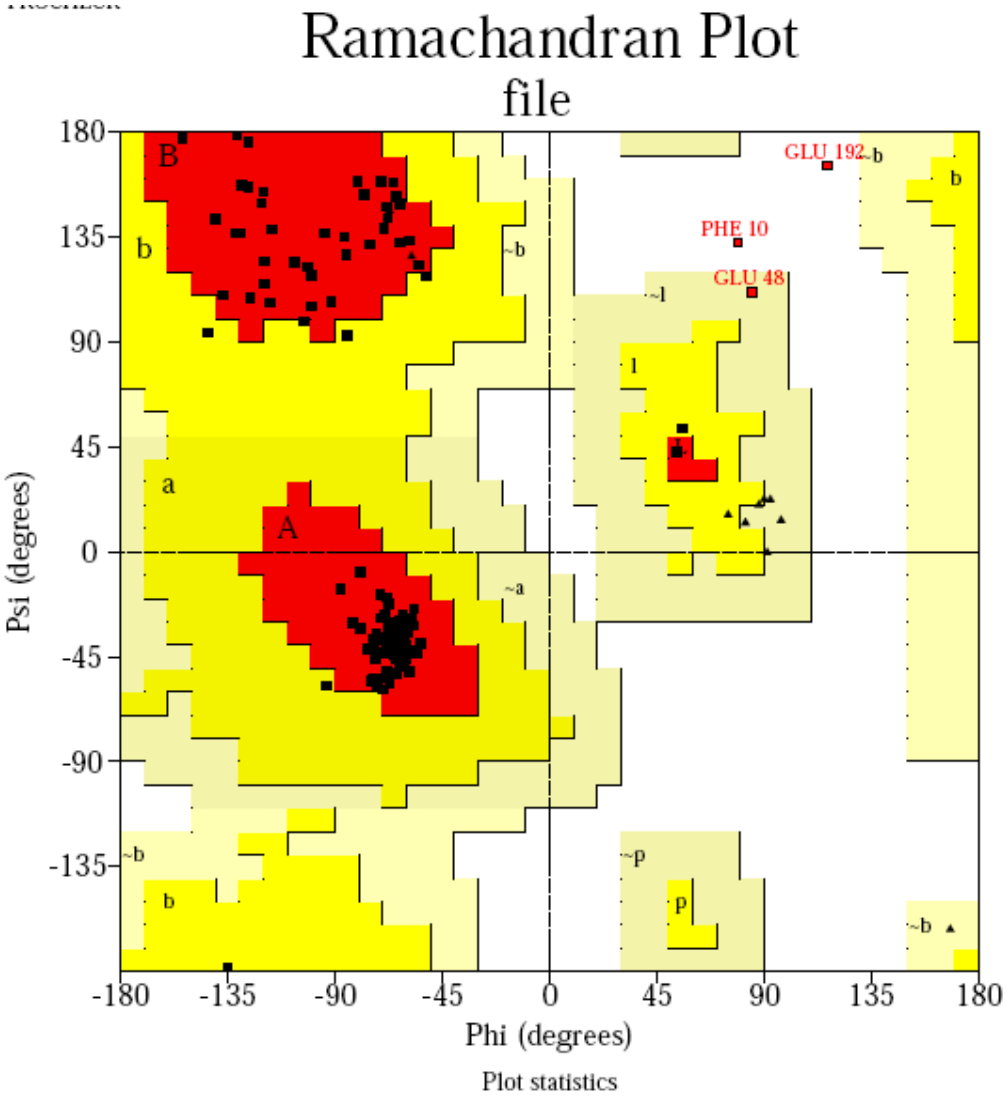

Figure 4.13: Ramachandran plot of target structure. Regions in red color represent the allowed region, yellow regions represent the additional allowed region, and the rest of the regions represent generally allowed and disallowed regions. Residues in disallowed regions are highlighted in red color with its three code name and position.

2. G-Factors

Parameter	Score	Average Score
Dihedral angles:-		
Phi-psi distribution	0.46	
Chi1-chi2 distribution	0.03	
Chi1 only	-0.07	
Chi3 & chi4	0.64	
Omega	-0.07	
		0.20
Main-chain covalent forces:-		
Main-chain bond lengths	-0.05	
Main-chain bond angles	-0.14	
		-0.10
OVERALL AVERAGE		0.09

Figure 4.14: Results of G-factors. It is the log-odd score of distribution of stereochemical properties.

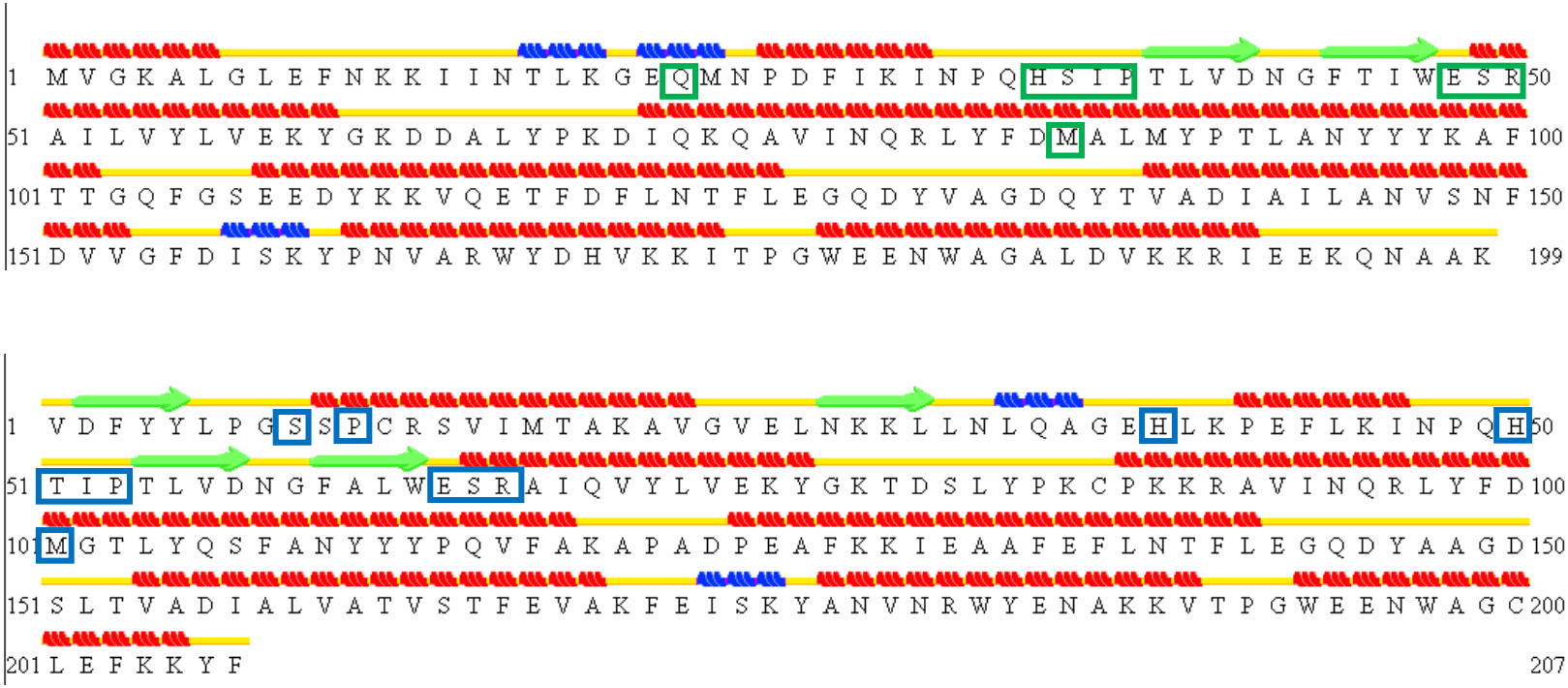

Figure 4.15: Residues involved in ligand binding. Red and blue represent alpha helix, whereas green represents beta sheets. Yellow represent turn or coil. This secondary structure is designed by STRIDE program.

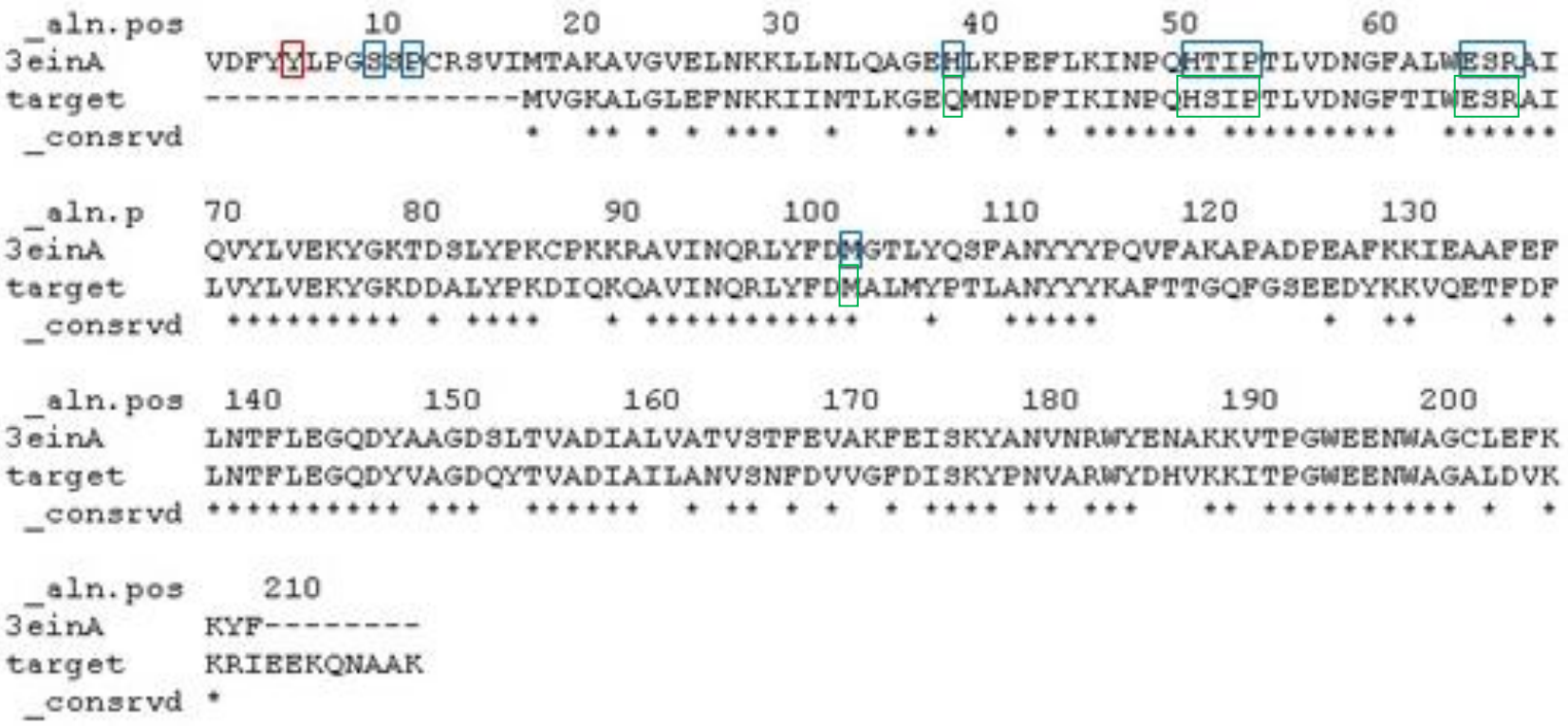

Figure 4.16: Comparison of query and template sequences. Residue in red box is catalytic residue, and residues in blue box represent residues responsible in ligand binding. Meanwhile, residues in the green box represent residues that complement the residues in blue box.

Table 4.4: Comparison of secondary structure between query and 3EIN based on residues important in ligand binding.

	10	12	39	51	52	53	54	65	66	67	102
3EIN	Turn/ loop	α-helix	Turn/ loop	Turn/ loop	Turn/ loop	Turn/ loop	Turn/ loop	Turn/ loop	α-helix	α-helix	α-helix
QUERY	-	-	α-helix	Turn/ loop	Turn/ loop	Turn/ loop	Turn/ loop	Turn/ loop	α-helix	α-helix	α-helix

4.6 3D structure of target sequence

The topology of query sequence is shown in Figure 4.17. According to the 3D structure of the target sequence, they are consisting of 11 alpha helices, 2 beta sheets, and turn or loops (Figure 4.18). The RMSD value calculated from the target sequence, by taking the sequences of 3EIN as reference is 0.22 Å. Regarding the catalytic residue which is tyrosine at positions 5 and 6 in 3EIN, I hypothesized that there were other tyrosine residues from different positions which are 89 and 97 that have substitute the tyrosine of 5 and 6. This is because; from the structural alignment, the query sequence that aligned to the template sequence was starting at position 18 which codes for methionine (Figure 4.19). Moreover, when the 3D structure of query was superimposed with template 3D structure, there is extra of amino acid at N-terminal and C-terminal, for target and template, respectively. Furthermore, further analyzing by looking at the 3D structure of D3 and its sequence, it shows that tyrosine residue of adjacent positions to GSH binding site such as 82, 89, 95, 96, 97, and others were present. The side chains of tyrosine from 89 and 97 were projecting inside into the GSH binding site (Figure 4.20). Regarding to the rest of tyrosine residues in position 82, 95, and 96, however, their side chains were projecting outside the GSH binding site and far from the GSH binding site (Figure 4.21).

Supplementary analyzing the 3D structure of model, the query sequences were modified at position 89 and 97 with alanine and glycine at both positions. However, since the side chain of glycine consist of hydrogen, thus, it is not appropriate to become the catalytic residue in D3, same goes to glycine. Since there are two tyrosine at position 89 and 97 predicted to be the catalytic residue, auxiliary analysis is carried out in which the force field energy of target sequences was calculated. The force field energy was carried out by GROMOS 96 implemented in Swiss-PDBViewer in which it calculates the bond and

angles score (Table 4.5). The total force field energy for tyrosine at position 89 is 34.889 whereas the total force field energy for tyrosine 97 is 34.247. Even though there is no major difference of total force field energy between tyrosine 89 and tyrosine 97, however, tyrosine 97 is more stable compared to tyrosine 89 since it has the lowest energy.

From the topology as well, its topology is different from topology of template structure. The topology of query is in αβα form for its N-terminal domain in domain I, since it only contained 2 β-sheets that were flanked with bundle of α-helices whereas, the C-terminal domain in domain II contains bundle of α-helices.

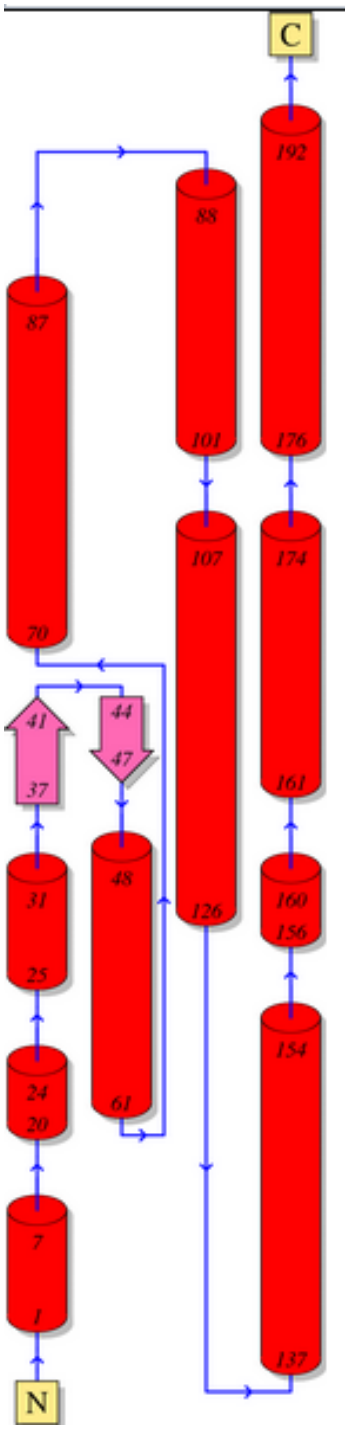

Figure 4.17: Topology of query sequence

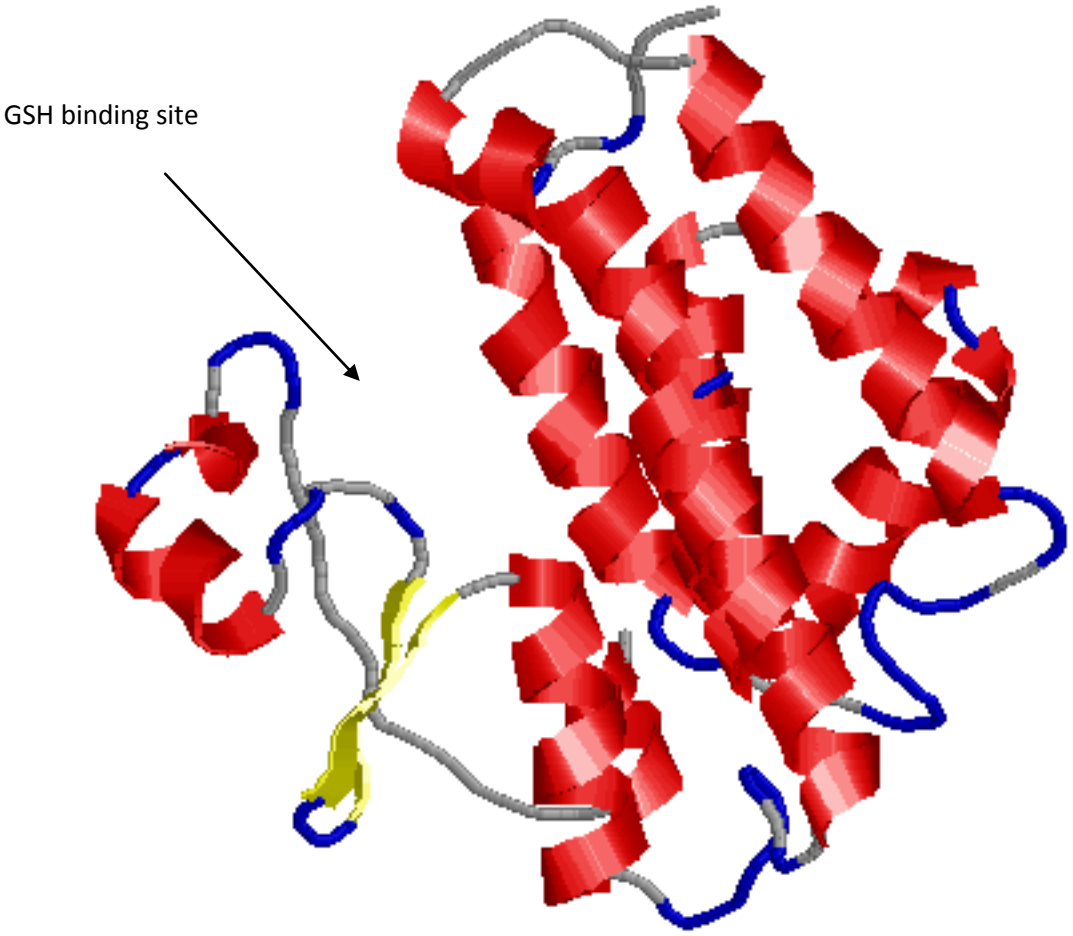

Figure 4.18: 3D structure of query sequence. Alpha helices are representing as red, beta sheets as yellow, and turns/loops as blue.

```
 _aln.pos          10         20         30         40         50         60
3einA     VDFYYLPGSSPCRSVIMTAKAVGVELNKKLLNLQAGEHLKPEFLKINPQHTIPTLVDNGFALWESRAI
target    -----------------MVGKALGLEFNKKIINTLKGEQMNPDFIKINPQHSIPTLVDNGFTIWESRAI
 _consrvd                  *  ** * * ***  *   **   * * ****** *********  ******

 _aln.p   70         80         90        100        110        120        130
3einA     QVYLVEKYGKTDSLYPKCPKKRAVINQRLYFDMGTLYQSFANYYYPQVFAKAPADPEAFKKIEAAFEF
target    LVYLVEKYGKDDALYPKDIQKQAVINQRLYFDMALMYPTLANYYYKAFTTGQFGSEEDYKKVQETFDF
 _consrvd  ********* * ****   * ***********   *   *****           *  **    * *

 _aln.pos  140        150        160        170        180        190        200
3einA     LNTFLEGQDYAAGDSLTVADIALVATVSTFEVAKFEISKYANVNRWYENAKKVTPGWEENWAGCLEFK
target    LNTFLEGQDYVAGDQYTVADIAILANVSNFDVVGFDISKYPNVARWYDHVKKITPGWEENWAGALDVK
 _consrvd ********** ***  ******  * ** * *  * **** ** ***   ** ********** *  *

 _aln.pos    210
3einA     KYF--------
target    KRIEEKQNAAK
 _consrvd *
```

Figure 4.19: Sequence alignment between query and template (Modeller).

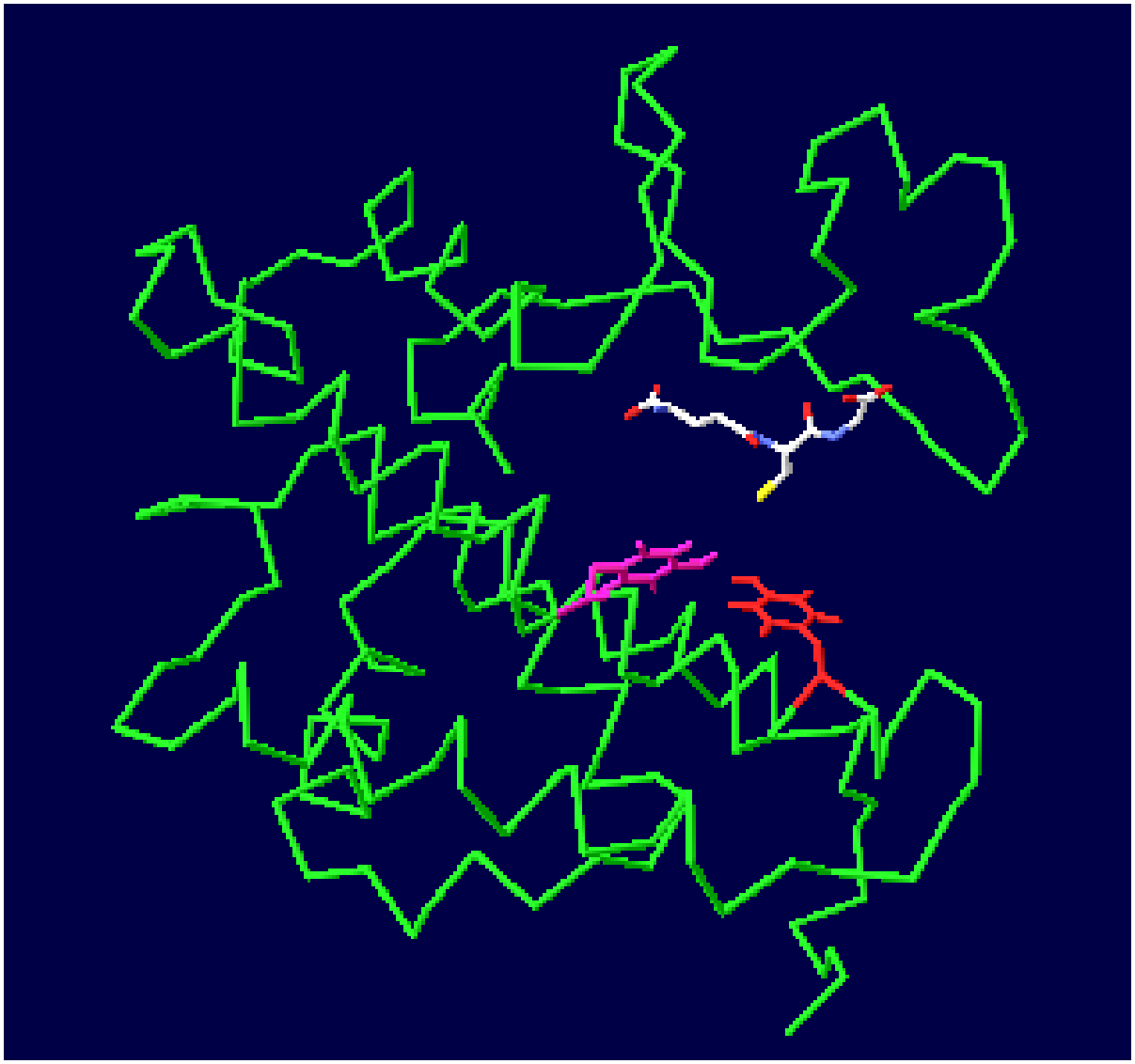

Figure 4.20: 3D structure of target sequence with sidechain of tyrosine 89 and 97. Glutathione substrate was represent in cpk color, tyrosine 89 in pink color and tyrosine 97 in red color.

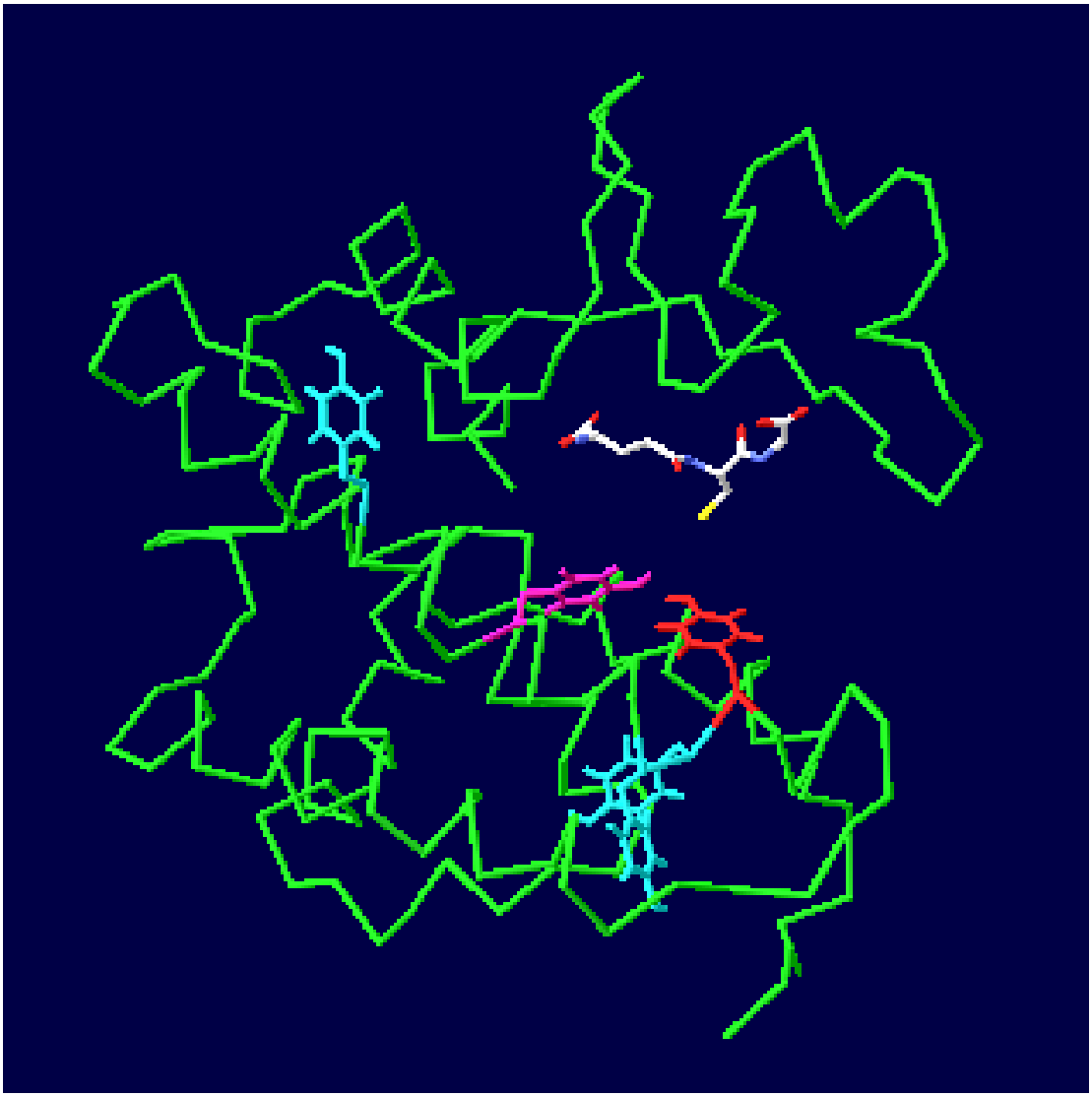

Figure 4.21: 3D structure of target sequence with sidechain of tyrosine 82, 89, 95, 96 and 97. As presented above, other tyrosine residues (colored in blue) that are nearby the GHS binding site, their side chains were protruding outside of the binding site.

Table 4.5: Force Field Energy for Tyrosine position 89 and 97.

Residue	Position	Bond	Angles	Total Energy
TYR	89	13.044	21.845	34.889
TYR	97	17.291	16.955	34.247

4.7 Control Experiment

In control experiment, since there are several truncations of amino acid at N-terminal domain of GSTD3, I obtained several amino acids at N-terminal from the template sequence, which is 3EIN, and connects them together with the target sequence (labeled as query1). The sequence then was submitted to the Modeller program to select the suitable model being generated. From the result, it shows that the sequence alignment between query1 and template sequences begins at position 1 for both of the sequences (Figure 4.22).

From the 3D structure of query1, it shows that this new query sequence has 4 beta sheets. However, regarding the conserved tyrosine residue at positions 5 and 6, their sidechain were not projecting inside the GSH binding sites (Figure 4.23). Furthermore, the speculated tyrosine of positions 97 and 89 were also not projecting inside the GSH binding sites. This is because, the sequence alignment, had caused the 3D structure of query1 model to be different from the 3D structure of original query sequence. Also, the tyrosine residues of position 89 and 97 (query1) aligned with the tyrosine residues of positions 106 and 114. Therefore, I conclude that, tyrosine 89 and 97 are the only tyrosine residues that substitute the missing tyrosine of positions 5 and 6.

```
query               1 MVDFYYLPGSSPCRSVIMVGKALGLEFNKKIINTLKGEQMNPDFIKINPQ     50
                      ||||||||||||||||..||:|:|.|||::|...||.:.|:|:|||||
template            1 MVDFYYLPGSSPCRSVIMTAKAVGVELNKKLLNLQAGEHLKPEFLKINPQ     50

query              51 HSIPTLVDNGFTIWESRAILVYLVEKYGKDDALYPKDIQKQAVINQRLYF    100
                      |:|||||||||.:||||||.|||||||||.|:||||..:|:|||||||||
template           51 HTIPTLVDNGFALWESRAIQVYLVEKYGKTDSLYPKCPKKRAVINQRLYF    100

query             101 DMALMYPTLANYYYKAFTTGQFGSEEDYKKVQETFDFLNTFLEGQDYVAG    150
                      ||..:|.:.|||||...........|.:||::..|:||||||||||.||
template          101 DMGTLYQSFANYYYPQVFAKAPADPEAFKKIEAAFEFLNTFLEGQDYAAG    150

query             151 DQYTVADIAILANVSNFDVVGFDISKYPNVARWYDHVKKITPGWEENWAG    200
                      |..||||||::|.||.|:|..|:||||.||.|||::.||:||||||||||
template          151 DSLTVADIALVATVSTFEVAKFEISKYANVNRWYENAKKVTPGWEENWAG    200

query             201 ALDVKKRIEEKQNAAK    216
                      .|:.||..|
template          201 CLEFKKYFE-------    209
```

Figure 4.22: Sequence alignment between query1 and template sequence. Matching denoted as horizontal line, "." represents semi-conserved substitutions, and ":" represent conserved substitutions. (EMBOSS Needle-Alignment)

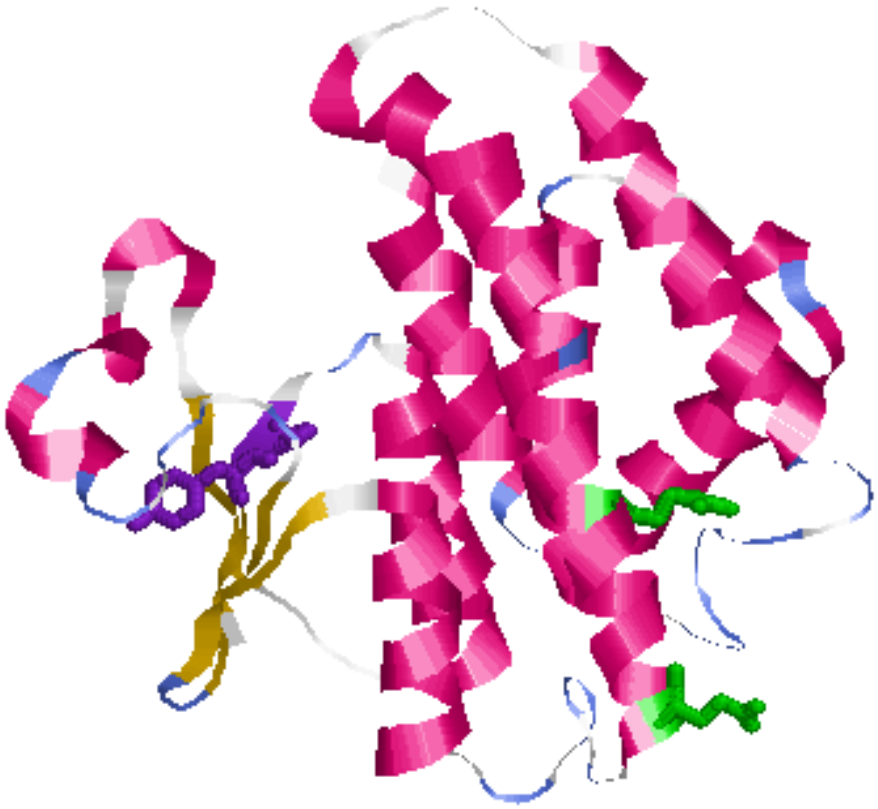

Figure 4.23: 3D structure of query1. Purple represents sidechain of tyrosine residues of position 5 and 6, whereas green represents sidechain of tyrosine residues at position 89 and 97 (RasMol).

CHAPTER 5

CONCLUSION

Glutathione S-transferase is a detoxifying enzyme that responsible in catalyzing the conjugation of electrophilic substrate with thiol group of glutathione into a hydrophilic so that it can be excreted easily from the cells via ion transport channel. In *Drosophila melanogaster*, delta and epsilon classes of GSTs are organism specific in which they are unique and having specific functions as well as specific expression pattern. The members of delta class consist of 10 isoforms, whereby each isoform has their own properties and characterizations. Along conducting this project, I discovered that two residues had substitute the missing tyrosine which is tyrosine residues of position 89 and 97. Since the residue at position 97 is more suitable, by referring to its lowest force field energy value, therefore, I concluded that tyrosine from position 97 had replaced the tyrosine at positions 5 and 6. Nevertheless, given that the force field energy between tyrosine 89 and 97 were not really too differ from each other, thus, there is possibility that both of them are involved in catalytic activity of GST Delta isoform 3. Although the residues have been identified, however, docking analysis need to be carried out in order to confirmed that tyrosine 97 involve in catalytic activity. Furthermore, the docking experiment should be carried out with different substrates such as GSH, DDT, paraquat, and others, in order to observe the catalytic activity and expression of GSTD3 towards these substrates. In addition, to ensure the reliability of 3D structure of target sequence, another approach of determining the 3D structure should be carried out such as crystallographic experiment.