JOURNAL OF TROPICAL LIFE SCIENCE
2022, Vol. 12, No. 2, 241 – 252
http://dx.doi: 10.11594/jtls.12.02.10

Research Article

In Silico Molecular Characterization of a Putative Haloacid Dehalogenase Type II from Genomic of *Mesorhizobium loti* Strain TONO

Sefatullah Zakary [1,4], Hamida Mashal [1], Abdul Rahman Osmani [2]*, Habeebat Adekilekun Oyewusi [3], Fahrul Huyop [4], Muzhgan Mohammad Nasim [5]

[1] Department of Botany, Faculty of Biology, Kabul University, 1006 Dehbori, Kabul, Afghanistan
[2] Department of Zoology, Faculty of Biology, Kabul University, 1006 Dehbori, Kabul, Afghanistan
[3] Department of Biochemistry, School of Science and Computer Studies, Federal Polytechnic Ado Ekiti, Ado Ekiti PMB 5351, Ekiti State, Nigeria
[4] Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia
[5] Department of Chemical and Environmental Engineering, Faculty of Malaysia Japan International Institute of Technology, University Technology Malaysia, 50100 UTM, Kuala Lumpur, Malaysia

Article history:
Submission November 2021
Revised November 2021
Accepted January 2022

**Corresponding author:*
E-mail: 'aosmani@ku.edu.af

ABSTRACT

Halogenated organic compounds are found as waste in the biosphere and can cause numerous dilemmas because of their toxicity and persistence in the environment. They play a major role in the quality of life of both, human beings and other living organisms. Degradation of these compounds by microorganisms is significant to reduce recalcitrant and cost. Thus, in the current study, an in-silico approach was used for homology modelling and docking assessment of a newly identified DehLt4, type II dehalogenase to predict its ability to degrade selected haloalkanoic acids and haloacetates. The study aimed to establish the catalytic tendencies of the enzyme to optimally degrade the selected halogenated haloacids. The refined modelled structure of DehLt4 using GROMACS 5.1.2 software revealed satisfactory scores of ERRAT (94.73%), Verify3D (90.83%) and PROCHECK (99.05 %) assessments. Active site prediction by blind docking and multiple sequence alignment indicated the catalytic triads for DehLt4 were Asp9-Lys149-Asn175. Both L-2-chloropropionic acid (L-2-CP) and trichloroacetate (TCA) docked with DehLt4 exhibited binding energy of -3.9 kcal/mol. However, the binding energy for D-2-chloropropionic acid (D-2-CP) and monochloroacetate (MCA) was -3.8 kcal/mol and -3.1 kcal/mol, respectively. Thus, the findings of the study successfully identified the catalytic important residues of DehLt4 for possible pollutant degradation. The in-silico study as such has a good potential for characterization of newly identified dehalogenases based on basic molecular structure and functions analysis.

Keywords: Dehalogenase, Haloacid dehalogenase, Mesorhizobium loti strain TONO, Protein structure

Introduction

The enormous utilization of chemical herbicides and pesticides by the agricultural sector for weed and pest management has led to serious concerns owing to their harmful effect on the ecosystem. In agricultural fields, there are more than 500 various types of pesticide formulations are used, of which most of them are non-biodegradable, recalcitrant and highly toxic [1]. The concerns are compounded by the global growth of large-scale industries, contributing more to the halogenated compounds that are naturally present. These compounds cause numerous dilemmas in the environment. The number of these compounds is severely rising, from fewer than 50 naturally produced

How to cite:
Zakary S, Mashal H, Osmani AR, et al., (2022) In Silico Molecular Characterization of a Putative Haloacid Dehalogenase Type II from Genomic of *Mesorhizobium loti* Strain TONO. Journal of Tropical Life Science 12 (2): 241 – 252. doi: 10.11594/jtls.12.02.10.

compounds in the year 1968 to more than 5000 in 2015 and still on the increase [2-4]. Bacterial dehalogenases are the key elements in the decontamination of organ halogen compounds. This process of decontamination is called dehalogenation [5]. According to the substrate specificities, the haloacid dehalogenases have been classified into different groups [5]. For instance: D-2-haloacid dehalogenases react with D-2-haloalkanoic acids to produce L-2-hydroxyalkanoic acids, L-2-haloacid dehalogenases act on L- 2-haloalkanoic acids to yield the corresponding D-2-hydroxyalkanoic acids, DL-2-haloacid dehalogenases (retention type) catalyze the dehalogenation of both D- and L-2-haloalkanoic acids to the corresponding D- and L-2- hydroxyalkanoic acids, DL-2-haloacid dehalogenases (inversion type) dehalogenate 2-haloalkanoic acids and yield products as L- and D-2- hydroxyalkanoic acids respectively.

Enzymes catalyze the conversion of various chemical compounds in a very specific and effective manner, making them applicable in the chemical industry, pharmaceutics, and bioremediation. Dehalogenases are microbial enzymes that catalyse the critical step in the degradation of priority halogenated organic pollutants cleavage of the carbon halogen bond [6]. Among these, α-haloacid (αHA) dehalogenases have been widely investigated and are placed into groups, I and II [7]. Both groups dehalogenate low molecular weight organohalogens and act on Cα [8]. In contrast to Cα there was dehalogenase from *Arthrobacter* sp. S1 that can degrade both Cα and Cβ of haloalkanoic acids isolated from the Philippines soil contaminated area [9]. Common group II dehalogenase was stereoselective dehalogenases acting only on L-2-haloacids while the group I dehalogenases were non-stereospecific dehalogenases and specific for D and L-2-haloacids. The group I members can react with L and D forms of substrates producing D- and L-hydroxy acids, respectively, or retain the chemical's isomeric pattern [7].

It was a common feature that many soil bacteria produce more than one dehalogenase, as previously reported [10-12]. The existence of more than one dehalogenase in a single microorganism is far from clear. Previously, we analysed the genome of *Mesorhizobium loti* strain TONO for dehalogenase enzyme to predict its function as a bioremediating agent [13]. The results revealed four haloacid dehalogenase type II and one haloalkane dehalogenase [14]. Among all dehalogenases type II (namely, DehLt1, DehLt2, DehLt3 and DehLt4) possess high sequence identity from 48.18% to 42.73%, with the well studied DehIVa and L-DEX, respectively. Therefore, DehLt4 was chosen for further characterization in the current study. The 3D structure of DehLt4 from *M. loti* strain TONO and its catalytically important residues have not been characterized. Understanding the structure of DehLt4 in *M. loti* strain TONO is essential for understanding of its function and mechanism. Thus, we constructed a homology model of the DehLt4 using DehIVa from *Burkholderia cepacia* MBA4 as a template. The DehLt4 model provided an insight of folding and also identification of the substrate-binding location, which helps in the recognition of the catalytic residues. Therefore, the catalytically important amino acids function in the catalytic mechanism of DehLt4 is essential to understand the dehalogenation process in *M. loti* strain TONO.

Material and Methods

Amino Acid Sequences Retrieval and Analysis

The amino acid sequences of DehLt4 (*M. loti* strain TONO; WP_096456195.1), L-DEX (*Pseudomonas* sp. Strain YL; S74078.1) [15] and DehIVa (*B. cepacia* MBA4; X66249.1) [16] were retrieved from the National Centre for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) GenBank. The amino acid sequence of DehLt4 was then uploaded to Protparam tools for characterization of the physicochemical properties on the ExPASy server (https://web.expasy.org/protparam/) [17]. Multiple sequence alignment of DehLt4 with crystalized DehIVa and L-DEX was performed using MultAlin version 5.4.1 [18].

Secondary and 3D Structure Prediction

Secondary structure was predicted with integrated Webware Network Protein Sequence Analysis (NPSA) in Pole Bioinformatics Lyonnais-Gerland [19] by GOR4 server [20]. The NCBI blast searched for the high similar homology model in the Protein Data Bank (PDB) with the highest percentages of similarity with DehLt4. The SWISS- MODEL Automatic Protein Modelling Server website (https://swissmodel.expasy.org) [21] was used to construct DehLt4 model, based on the homology DehIVa of *B. cepacia* MBA4. The built model was visualized by PyMOL software [22].

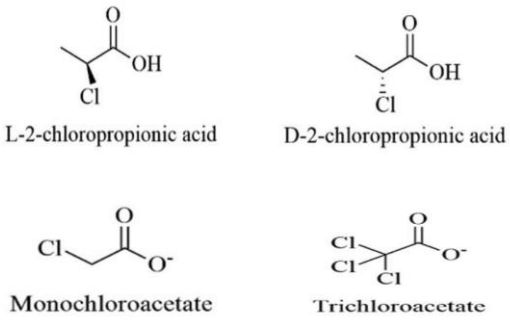

Figure 1. Structures of halogenated compounds used in this study

3D model refinement and validation

A molecular dynamic simulation was run on GROMACS 5.1.2 software and 54a7 force field was applied [23, 24]. The protein was placed in an appropriately sized cubic simulation box and solvated in 15214 simple point-charges (SPC) water molecules. Three Na^+ ions were added to the system to neutralize the charge. The energy of the system converged at 355 steps of steepest descent. All simulations were carried out at constant pressure and temperature. 50 ns molecular dynamics simulation was run at a pressure of 1 atm and temperature of 300 K.

For electrostatic interactions, the particle mesh Ewald method was used, and the linear constraint solver (LINCS) was utilized to constrain the bond length. The integration time step was 2 femto-second (fs), and the neighbor list was updated every tenth step using the grid option and a cut-off distance of 1.4 Å. A periodic boundary condition was employed with a constant number of particles in the systems, pressure, and temperature simulation criteria (NPT). Structural frames were generated after every 2 ps to update the trajectory. The simulation results were analyzed using Xmgrace graphs [25]. The Lemkul published protocol was used for step-by-step configurations [26]. Snapshots of DehLt4 extracted from the MD trajectory was visualised in PyMOL software [22]. The root mean square deviation (RMSD) was used to analyse the model's dynamic behaviour and structural stability and changes. The quality of the refined structure was then validated by using ERRAT [27], PROCHECK [28] and Verify3D [29]. The validated structure was viewed using Chimera (UCSF) software [30].

Preparation of ligands 3D structure

The 3D structures of the ligands, haloacids (D-2-chloropropionic acid; D-2-CP), (L-2-chloropropionic acid; L-2-CP) and haloacetate (trichloroacetate; TCA), (monochloroacetate; MCA) were downloaded from PubChem database (Figure 1). The ligands were converted to PDB file via PyMOL [22].

Molecular docking analysis

In the current study, the Autodock Tools 1.5.6 and Autodock version 4.2.6 software were employed for the docking of DehLt4-substrate complex. After removing the water molecules from the DehLt4 model, polar and non-polar hydrogens were added. Then the Kollman and Gasteigher charges were assigned. The same process was done to ligands to ensure correct adoption of torsions for rotation during docking. The grid box in Autogrid tool for DehLt4 was fixed at ± 1.000 Å from the 40.619 Å, 40.354 Å and 39.061 Å coordinates with sizes 48, 56, and 50 (x, y, and z positions respectively) to cover all amino acid residues. The AutoDock Vina was used for docking analysis, where the best result for each substrate was selected as the largest conformation cluster revealing the lowest binding energy (kcal/mol). Finally, the "pdbqt" file for each DehLt4 substrate-complex was converted to the "pdb" file format and viewed by PyMOL [22]. The hydrophobic interactions and hydrogen bonds for all DehLt4 substrate complexes were analysed using LigPlot.

Results and Discussion

Physicochemical property of DehLt4

Physiochemical analysis is important to understand the nature of a protein. The DehLt4 from *M. loti* strain TONO consists of 220 amino acid residues. Analysis by the ProtParam tools [17] showed that DehLt4 has a theoretical molecular weight of 24645.06 Daltons and a pI of 5.46. Notably, a pI value of less than 7 shows acidic nature [17]. DehLt4 had GRAVY value (grand average of hydropathicity) of ─ 0.112. The instability index was computed 41.93 and a high aliphatic index of 89. A protein instability index near 40 shows stability and aliphatic index above 40 indicates thermal stability of protein. [12, 17]. Furthermore, molecular formula of DehLt4 was $C_{1122}H_{1715}N_{297}O_{320}S_5$ with 3459 total number of atoms. Total numbers of positively and negatively charged residues were 23 and 26, respectively (Ta-

ble 1).

Table 1. Physicochemical summary of DehLt4

Details	DehLt4
Amino acid residues	220
Theoretical pI	5.46
Molecular weight (Da)	24,645.06
Positively charged residues	23
Negatively charged residues	26
Total number of atoms	3459
Molecular formula	$C_{1122}H_{1715}N_{297}O_{320}S_5$
Instability index (%)	41.93
Aliphatic index (%)	89
GRAVY	-0.112

Table 2. The important amino acid residues in the crystallized haloacid dehalogenases and their predicted conserved residues in DehLt4

L-DEX	DehIVa	DehLt4
10	D11	D9
T14	T15	T13
R41	R42	R40
S118	S119	S116
K151	K152	K149
Y157	Y158	Y155
S175	S176	S173
N177	N178	N175
D180	D181	D178

Multiple Sequence Alignment of DehLt4, DehIVa and L-DEX

The multiple sequence alignment of DehLt4 of *Mesorhizobium loti* strain TONO with DehIVa (*Burkholderia cepacia* MBA4) [16] and L-DEX (*Pseudomonas* sp. Strain YL) [15] was performed (Figure 2). The D11, T15, R42, S119, K152, Y158, S176, N178 and D181 were reported to be catalytically important residues in DehIVa [11, 16, 31]. These residues were conserved with L-DEX (*Pseudomonas* sp. Strain YL) [15, 32] (Table 2). Surprisingly, of these conserved amino acid residues, all of these residues were conserved with DehLt4 (Table 2). The overall amino acid sequence identity of DehLt4 with DehIVa and L-DEX were 48.18% and 42.73%, respectively. The amino acid residues of DehLt4 were compared to DehIVa, and L-DEX to check the amino acid profiles, as shown in Figure 3. It revealed that the amino acid composition of DehLt4 was approximately similar to DehIVa, having of nine conserved catalytically important residues with DehIVa and L-DEX (Table 2). The findings therefore allowed the prediction of the amino acid residues essential for DehLt4 catalysis.

Secondary Structure Analysis

The secondary structure of a protein is formed by hydrogen bonds between the carboxyl and amide groups of amino acid residues in polypeptide backbone of enzymes. The secondary structure prediction of DehLt4 from *M. loti* strain TONO by GOR4 server [20] is shown in Figure 4. The percentage of strands (e) was 14.09%, helices (h) was 47.73% and coil (c) was 38.18%. The total α-helices of DehLt4 were nine which the second was the longest and the eighth and ninth helix were the shortest having the same number of residues.

Homology Modelling of DehLt4

Homology modelling plays an important role in building protein 3D structure. It aims to construct three-dimensional protein structure models using experimentally crystalized structures of related family members as templates. Protein structure prediction based on target-template alignment was notably constructive and convenient [33] as most of the newly sequenced proteins probably possess similar structures with one that had already been experimentally determined. Hence in the present study, the DehLt4 protein structure was built using Swiss-Model webserver [34] (Figure 5). The DehIVa (2NO5) from *Burkholderia cepacia* MBA4 (having 48.62 similarity in Swiss-Model) was selected as a template for homology modelling of DehLt4. The GMQE value revealed 0.75 while QMEAN showed -2.05. The QMEAN value was based on four criteria, they were Cβ (-2.29), All atom (0.39), Solvation (0.24) and Torsion (-1.86). From this template, DehLt4 model (residues 1-218) was generated with 0.99% coverage. The DehIVa structure is formed of two domains that are known as the core and cap domains (Figure 5).

The cap domain of DehIVa consists of a four-helix bundle, formed by residues 17 to 97 (15-95 in DehLt4). The core domain contains central six-stranded parallel open twisted β-sheet (strand order is β6-β5-β4-β1-β2-β3), flanked on both sides, by a total of five α-helices (1– 16 and 96–218 residues in DehLt4). The core domain of DehIVa is structurally similar to the other HAD superfamily structures. However, differences occur along this superfamily in the cap domain, which seems to

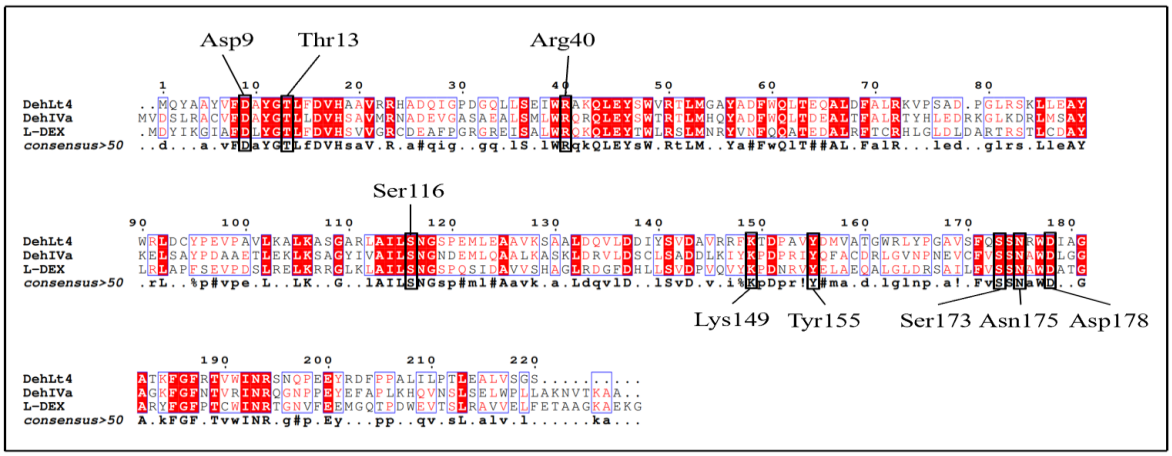

Figure 2. Multiple sequence alignment of DehLt4 with DehIVa, and L-DEX. The conserved catalytic important residues are shown in rectangles. Highly conserved amino acids are highlighted in red; low conserved values are blue, and neutral are highlighted in black

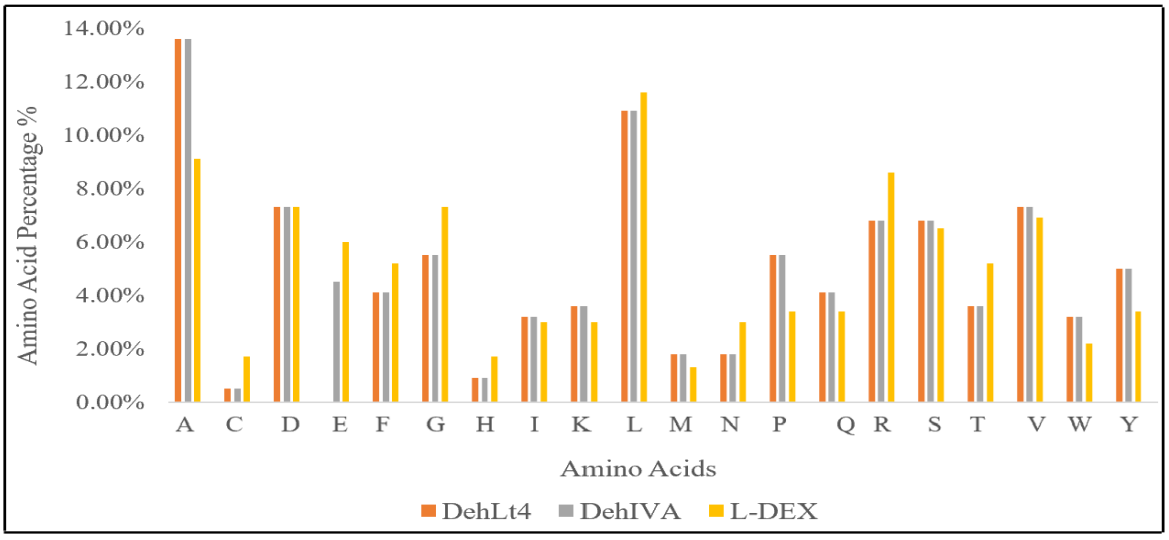

Figure 3. The bar graph comparing the percentage of each amino acid of DehLt4 with DehIVa and L-DEX

```
         10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MQYAAYVFDAYGTLFDVHAAVRRHADQIGPDGQLLSEIWRAKQLEYSWVRTLMGAYADFWQLTEQALDFA
ccccceeeccccchhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhhhcccchhhhhhhhhhhhhh
LRKVPSADPGLRSKLLEAYWRLDCYPEVPAVLKALKASGARLAILSNGSPEMLEAAVKSAALDQVLDDIY
hhccccccchhhhhhhhhceeccccchhhhhhhhhhhhheeeeccchhhhhhhhhhhhhhhhhchh
SVDAVRRFKTDPAVYDMVATGWRLYPGAVSFQSSNRWDIAGATKFGFRTVWINRSNQPEEYRDFPPALIL
hhhhhhhccccceeeeeeccccccccceeeeccchhhhccccccceeeeeeccccccccccccccc
PTLEALVSGS
hhhheeeeec
```

Figure 4. Secondary structure prediction of DehLt4 which consists of coil (c), strand (e) and helix (h)

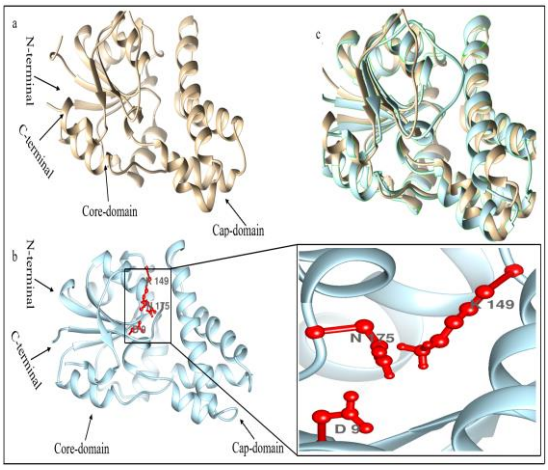

Figure 5. The 3D structure of DehIVa (template) and DehLt4 (model) after refinement. (a) The ribbon representation of DehIVa (Tan colour). (b) The ribbon representation of DehLt4 (Light Blue colour), the catalytic triad is represented in ball and stick. (c) The superimpose representation of DehIVa (Tan colour) and DehLt4 model (light blue colour). The illustration was prepared in UCSF Chimera [30].

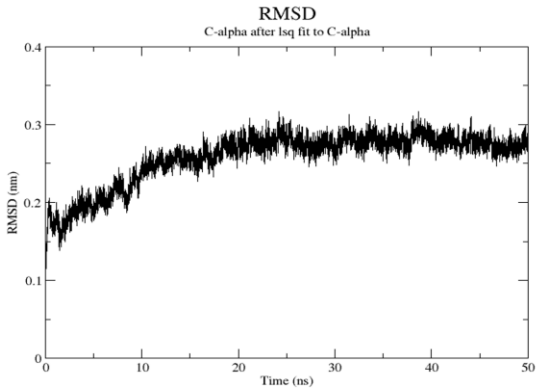

Figure 6. RMSD plot of DehLt4 backbone atoms, throughout MD simulation

play role in the diversity observed in HAD superfamilys' reaction mechanisms and substrate specificity. The cap domain of DehIVa is structurally similar to L-DEX YL [31]. The superimpose DehIVa and DehLt4 show very similar folding in both core and Cap domains. Depiction of the DehLt4 (Light Blue) superimposed with DehIVa (Tan) and also their N-terminal and C-terminal tails are shown in Figure 5.

DehLt4 Model Refinement

Molecular dynamics simulation for 50 ns was run to assess and refine the stability of the DehLt4

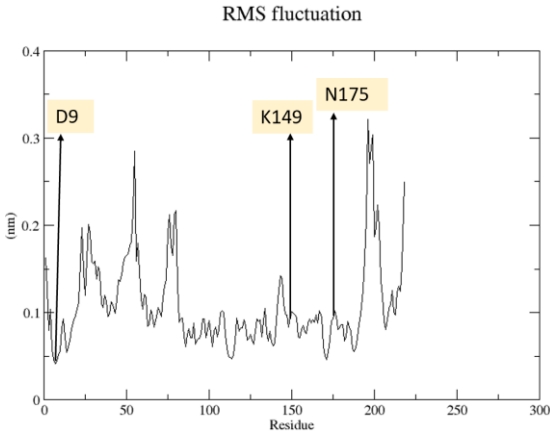

Figure 7. RMSF plot of DehLt4 backbone atoms, and its labelled catalytic triad throughout MD simulation

structure. The stability of DehLt4 model was characterized based on the root mean square deviation (RMSD) and the root mean square fluctuations (RMSF) of the protein backbone. It should be noted that the RMSD of DehLt4 model structures did not change remarkably (within the range 0.2-0.3 nm) during the simulations as is shown in Figure 6. Notably, a low RMSD value (RMSD 0.2-0.3 nm) shows high stability of the structures [12]. These RMSD values revealed that the employed simulation time was long enough to achieve an equilibrium structure of DehLt4. In addition, the root mean square fluctuation (RMSF) was significant for determining amino acid fluctuations along the protein chain during simulation. On this plot, peaks indicated amino acids that fluctuated the most during the simulation. The C-terminal and N-terminal tails were realized fluctuated much more than any other part of the protein (Figure 7). It has been reported in DehIVa and L-DEX, first 10 residues of C-terminal play no essential role in the structure or the catalytic function of these proteins [35]. It is essential to highlight here, the highest value of RMSF shows a higher degree of movement, while a low RMSF value represents a more stable residues due to less fluctuations during simulation. RMSF of > 0.05nm (0.5 Å) is the threshold value where a significant change in residue specific flexibility occurs.

In addition, for 50 ns simulation an average temperature of 300 K for the simulated system was equal to 300 ± 0.5 K. Thus, the obtained DehLt4 modelled structure at 300 K for DehLt4 was extracted under stable temperature conditions (Figure 8).

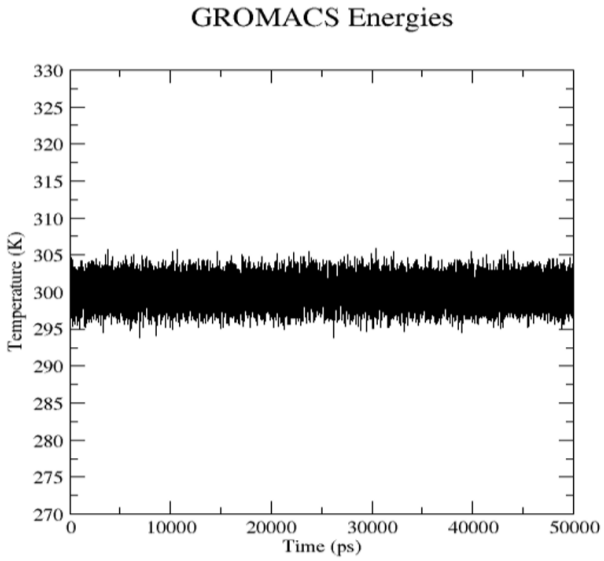

Figure 8. Variation of the temperature stability during 50 ns molecular simulation

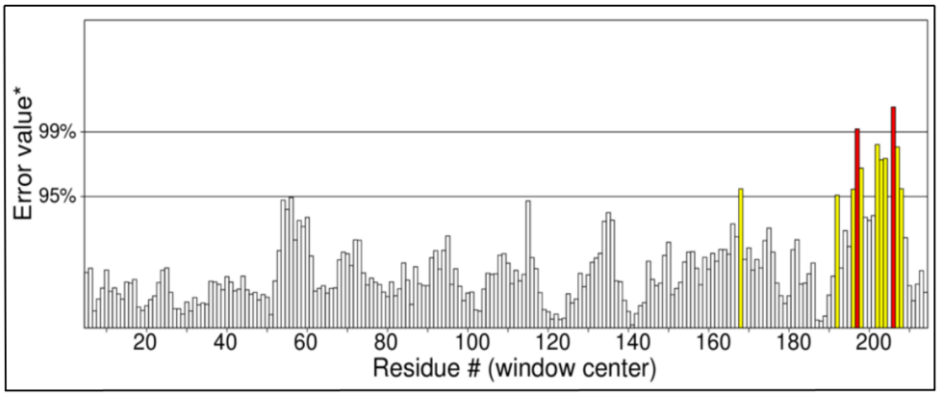

Figure 9. ERRAT plot of DehLt4 model

The validation of DehLt4 structure

The DehLt4 model was validated after 50 ns molecular dynamics simulation. The Verify3D result showed that 90.83% of the residues had an averaged 3D-1D score over 0.2, and the remaining residues did not gain this score. Residues with a score above 0.2 is considered reliable. According to the structure validation analysis using Verify3D, the environment profile of the model was good. ERRAT tools computed the nonbonded atomic interactions by comparing the statistics of highly refined structures. DehLt4 analysis revealed an ERRAT score of 94.73% that was acceptable in the normal range (Figure 9). The ERRAT score for a good model was over 50%, whereas a higher score indicated a better quality. The DehLt4 was analyzed by the PROCHECK tools which calculated psi/phi angles (ψ, ϕ), and therefore generating Ramachandran plot. The Ramachandran plot was a two-dimensional (2D) plot of the phi (ϕ) - psi (ψ) torsion angles of the protein backbone that provides a simple view of the conformation of a protein. Based on the results, the stereochemical evaluation of backbone psi and phi dihedral angles of the DehLt4 showed that 90.2%, 8.3%, 1.0%, and 0.5% of residues were within the most favoured regions, additionally allowed regions, generously allowed regions, and disallowed regions, respectively. In total, 99.05 % of DehLt4 amino acid residues located in allowable regions. A score close to 100% indicated the good stereochemical quality of the structure (Figure 10). This signified that the model was acceptable.

Active site residue prediction

The current investigation predicted the catalytic triad of DehLt4 of *M. loti* strain TONO using multiple sequence alignment (Figure 2) and dock-

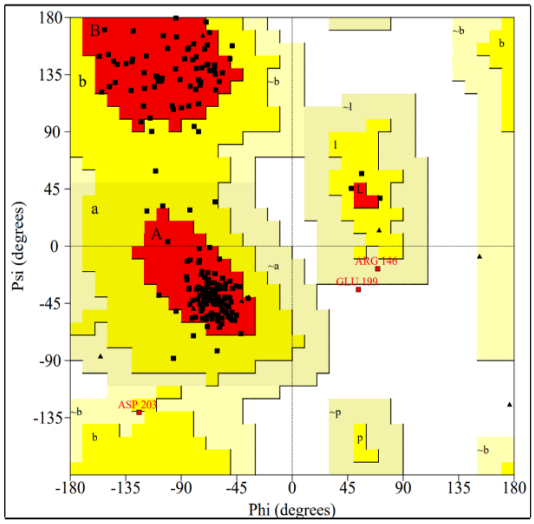

Figure 10. Ramachandran plot of DehLt4 model. Red (most favoured regions), yellow (additionally allowed regions) and pale-yellow (generously allowed regions), white colour (disallowed region)

ing. Blind docking of substrates haloacetate (TCA and MCA) and haloalkanoic acids (D-2-CP and L-2-CP), were run over the whole surface of the DehLt4 to identify the possible active sites, because there is no information about DehLt4 active site [36]. Site specific docking was also performed to doublecheck the blind docking results in order to ensure the substrates had bonded properly to the active site of the target [12]. The residues K149 and N175 were observed forming hydrogen bonds with all substrates (Figure 11).

Molecular Docking Analysis

The catalytic triad plays an important role in the bioremediation of organohalides [37]. Thus, it is crucial that the active site amino acid is well comprehend, in order to fully understand the catalytic capability of an enzyme. The literature review described that location of catalytic essential residues are usually relative and point in the same direction of the active site. Notably, catalytically important amino acids are highly conserved in dehalogenase type II [11]. The study found that there are nine catalytically conserved amino acid residues in the sequence of DehLt4 aligned with β-dehalogenases, DehIVa and L-DEX (Table 2). Research using molecular docking could explain the binding between a ligand and an enzyme and resultantly can ease understating the degradative mechanism of the enzyme [38]. The AutoDock Vina software utilizing the AutoGrid tools was run for the docking of DehLt4 model with ligand. This study observes for the lowest binding energy (kcal/mol) which shows a strong affinity of enzyme towards ligands [39]. Results showed that the K149 and N175 form hydrogen bond with the carbonyl group of the substrate (Figure 11). These amino acids were hypothesised to be important in the successful binding of the DehLt4 active site. The four different ligands were haloacetate (TCA and MCA), haloalkanoic acid (D-2CP and L-2CP), used in this study. It should be mentioned here that the number of predicted amino acids responsible for the binding with each assessed substrate in this study, were high enough to warrant successful protein-ligand interactions. In our case, the three amino acids namely D9 (D9), K149 (K149), and N175 (N175) that closely matched the residues predicted (i.e., residues predicted by multiple sequence alignment and docking) to be the active site of DehLt4.

The enlists of the docking result of the DehLt4-ligand complexes with respective hydrogen bond distances, where up to two hydrogen bonds were formed. The LigPlot analysis which shows non-ligand residues involved in the hydrophobic interaction and those formed hydrogen bond with its equivalent distance is shown in Figure 11. The lowest binding energy shows the formation of a stronger DehLt4-ligand complex interaction. The DehLt4-L-2CP complex yielded the binding energy of -3.9 kcal/mol. This corresponded to two hydrogen bond distances of 3.02 Å and 3.14 Å that formed with residues K149 and N175, respectively. The docking of DehLt4 with D-2CP also produced hydrogen bond distances of 2.96 Å and 3.16 with K149 and N175 with comparable binding energy of -3.8 kcal/mol. The binding energy for TCA was -3.9 kcal/mol forming the two hydrogen bonds of 3.25 Å and 2.80 Å with K149 and N175, respectively. MCA forms hydrogen bond of 3.01 Å and 3.03 Å with K149 and N175 with binding energy of -3.1 kcal/mol. It is essential to know that a shorter distance between two atoms increases the likelihood of a hydrogen bond being formed, since one atom acts as an electron acceptor while the other is an electron donor.

It should be acknowledged here that the hydrogen bond distances established in DehLt4-ligand complexes are all within the reasonable distance for the formation of hydrogen bonds (< 3.5 Å) [40]. While a longer hydrogen bond distance (> 3.5 Å) shows a lower affinity of an enzyme to

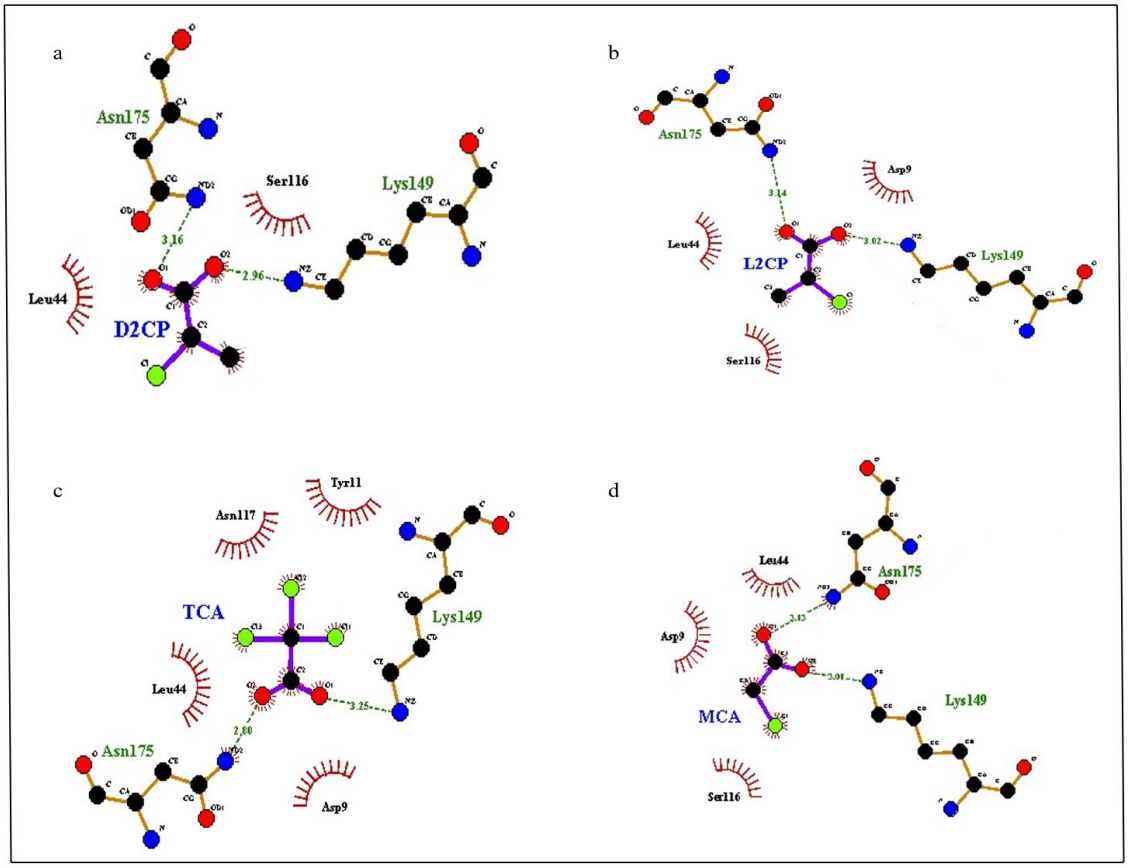

Figure 11. LigPlot analysis for DehLt4-ligand interaction showing residues involved in the hydrophobic interaction and hydrogen bonds with its equivalent distance. (a) D-2CP (b) L-2CP (c) TCA and (d) MCA

wards a substrate and is less likely to catalyze the compound. It should be noted that although some substrate may bind with low binding energy, they are not degraded by the respected enzyme. This issue is related to stereospecificity of dehalogenases. For example, DehIVa cannot degrade D2CP even tough having a low binding energy [31]. The stereospecificity is not in scope of this study. Therefore, the current study believes that in the dehalogenation by DehLt4, D9 acts as nucleophile by attacking C_2 atom of the substrate. This attack will happen when substrate is bond to the binding residues to remove the chloride ion and form substrate intermediate. This phenomenon of dehalogenation has been previously reported [3, 11, 31, 32, 35, 41, 42]. It is worth mentioning that mutation of D11 and N178 residues in DehIVa which are conserved with the predicted active site residue (D9 and N175 of DehLt4) turns the protein to inactive [35]. Thus, this study believes that the absence of any one of the D9, K149 and N175 residue may have affect to the catalytic activity of DehLt4, therefore, likely affecting the degradation of halogenated compounds.

Conclusion

This is the first reported study of the analysis of dehalogenase type II and homology modelling of the DehLt4 from the genomic study of *M. loti* strain TONO. The overall perspective comprehends the basic structure and functions of DehLt4 for degradation of halogenated compounds. It was revealed that DehLt4 significantly interacts with D-2CP, L-2CP, TCA and MCA as shown by the formation of a stable enzyme-substrate complex. Thus, this study finds the catalytic triad of DehLt4 which possibly plays a vital role in bioremediation. The DehLt4 structure showed similar folding as DehIVa and L-DEX consisting of two domains known as Core-domain and Cap-domain. The findings of this study have highlighted the need for crystallographic and in-depth studies of the detailed structure of DehLt4 complexes with its substrate binding mechanism and the substrate specificities. MD simulation is extensively used to study protein–ligand interactions. Therefore, future work will involve MD simulation and free binding energy to calculate the favourability of the ligand at the binding site of DehLt4 and substrates

used. Moreover, to rationally engineer DehLt4, full molecular details of its catalytic mechanism and understanding of the potential amino acid residues will also be examine using mutagenesis that would influence its catalytic efficiency and enantioselectivity.

Acknowledgement

The authors would like to express their gratitude to UTM-GUP Grant no. Q.J130000.2414.08G59 for financial assistance. SZ thanks the Afghanistan Higher Education Ministry for MSc. studentship.

References

1. Douglas SH, Dixon B, Griffin D (2018) Assessing the abilities of intrinsic and specific vulnerability models to indicate groundwater vulnerability to groups of similar pesticides: a comparative study. Physical Geography 39(6): p. 487-505. doi: 10.1080/02723646.2017.1406300.
2. Gribble GW (2015) A recent survey of naturally occurring organohalogen compounds. Environmental Chemistry 12(4): p. 396-405. doi: 10.1071/EN15002.
3. Adamu A, Wahab RA, Aliyu F et al. (2020) Haloacid dehalogenases of *Rhizobium* sp. and related enzymes: catalytic properties and mechanistic analysis. Process Biochemistry 92: 437 – 446. doi: 10.1016/j.procbio.2020.02.002.
4. Zakary s, Oyewusi HA, Huyop F (2021) Dehalogenases for pollutant degradation: A mini review. Jornal of Tropical Life Science 11(1): p. 17-24. doi: 10.11594/jtls.11.01.03.
5. Slater JH, Bull AT, Hardman DJ (1995) Microbial dehalogenation. Biodegradation 6 (3): 181-189. doi:10.1007/BF00700456.
6. Weightman AJ, Topping AW, Hill KE et al. (2002) Transposition of DEH, a broad-host-range transposon flanked by ISPpu12, in *Pseudomonas putida* is associated with genomic rearrangements and dehalogenase gene silencing. Journal of Bacteriology 184(23): p. 6581-6591. doi: 10.1128/JB.184.23.6581-6591.2002.
7. Hill KE, Marchesi JR, Weightman AJ (1999) Investigation of Two Evolutionarily Unrelated Halocarboxylic Acid Dehalogenase Gene Families. Journal of Bacteriology 181 (8): 2535-2547. doi: 10.1128/JB.181.8.2535-2547.1999.
8. Janssen DB, Oppentocht JE, Poelarends GJ (2001) Microbial dehalogenation. Current Opinion in Biotechnology 12(3): p. 254-258.doi: 10.1016/S0958-1669(00)00208-1.
9. Bagherbaigi S, Gicana RG, Lamis RJ et al. (2013) Characterisation of *Arthrobacter* sp. S1 that can degrade α and β-haloalkanoic acids isolated from contaminated soil. Annals of Microbiology 63(4): p. 1363-1369. doi: 10.1007/s13213-012-0595-4.
10. Hardman DJ, Sater JH (1981) Dehalogenases in soil bacteria. Microbiology 123(1): p. 117-128. doi: 10.1099/00221287-123-1-117.
11. Adamu A, Wahab RA, Huyop F (2016) L-2-Haloacid dehalogenase (DehL) from *Rhizobium* sp. RC1. Springer Plus 5(1): 1-17. doi: 10.1186/s40064-016-2328-9.
12. Oyewusi HA, Huyop F, Wahab RA (2020) Molecular docking and molecular dynamics simulation of *Bacillus thuringiensis* dehalogenase against haloacids, haloacetates and chlorpyrifos. Journal of Biomolecular Structure and Dynamics 1-16. doi: 10.1080/07391102.2020.1835727.
13. Shimoda Y, Hirakawa H, Sato S et al. (2016) Wholegenome sequence of the nitrogen-fixing symbiotic *Rhizobium Mesorhizobium loti* strain TONO. Genome Announcements 4 (5). doi: 10.1128/genomeA.01016-16.
14. Zakary S, Oyewusi HA, Huyop F (2021) Genomic analysis of *Mesorhizobium loti* strains reveals dehalogenases for bioremediation. Jornal of Tropical Life Science 11(1) 67 – 77. http://dx.doi: 10.11594/jtls.11.01.09.
15. Nardi-Dei V, Kurihara T, Okamura T et al. (1994) Comparative studies of genes encoding thermostable L-2-halo acid dehalogenase from *Pseudomonas* sp. strain YL, other dehalogenases, and two related hypothetical proteins from Escherichia coli. Appl. Environ. Microbiol 60(9): p. 3375-3380. doi: 10.1128/aem.60.9.3375-3380.1994.
16. Murdiyatmo U, Asmara W, Tsang JS et al. (1992) Molecular biology of the 2-haloacid halidohydrolase IVa from *Pseudomonas cepacia* MBA4. Biochemical journal, 284(1): p. 87-93. doi: 10.1042/bj2840087.
17. Gasteiger E, Hoogland C, Gattiker A et al. (2005) Protein identification and analysis tools on the ExPASy server, in The proteomics protocols handbook. Springer p. 571-607. DOI: 10.1385/1-59259-890-0:571.
18. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 16 (22): 10881-10890. doi: 10.1093/nar/16.22.10881.
19. Combet C, Blanchet C, Geourjon C et al. (2000) NPS@: network protein sequence analysis. Trends in Biochemical Sciences 25(3): p. 147-150. doi: 10.1016/S0968-0004(99)01540-6.
20. Garnier J, Gibrat JF, Robson B (1996) [32] GOR method for predicting protein secondary structure from amino acid sequence, in Methods in Enzymology. Elsevier p. 540-553. doi: 10.1016/S0076-6879(96)66034-0.
21. Biasini MS, Bienert A. Waterhouse K, Arnold et al. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research 42(W1): p. W252-W258. doi: 10.1093/nar/gku340.
22. DeLano WL (2002) Pymol: An open-source molecular graphics tool. CCP4 Newsletter on Protein Crystallography 40(1): p. 82-92.
23. Schmid N, Eichenberger AP, Choutko A et al. (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. European Biophysics Journal 40(7): p. 843-856.
24. Van Der Spoel D, Lindahl E, Hess B et al. (2005) GROMACS: fast, flexible, and free. Journal of Computational Chemistry 26(16): p. 1701-1718.
25. Lindahl E, Hess B, Van Der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. Molecular Modeling Annual 7(8): p. 306-317. doi: 10.1007/s008940100045.

26. Lemkul J, (2018) From proteins to perturbed Hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package [article v1. 0]. Living Journal of Computational Molecular Science 1(1): p. 5068. doi: 10.33011/livecoms.1.1.5068.
27. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Science 2(9): p. 1511-1519.
28. Laskowski RA, MacArthur MW, Moss DS et al. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26(2): p. 283-291. doi: 10.1107/S0021889892009944.
29. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364): p. 83-85. doi: 10.1038/356083a0.
30. Pettersen EF, Goddard TD, Huang CC et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry 25(13): p. 1605-1612. doi: 10.1002/jcc.20084.
31. Schmidberger JW, Wilce JA, Tsang JS et al. (2007) Crystal structures of the substrate free-enzyme, and reaction intermediate of the HAD superfamily member, haloacid dehalogenase DehIVa from *Burkholderia cepacia* MBA4. Journal of Molecular Biology 368(3): p. 706-717. doi: 10.1016/j.jmb.2007.02.015.
32. Kurihara T, Liu JQ, Nardi-Dei V et al. (1995) Comprehensive site-directed mutagenesis of L-2-halo acid dehalogenase to probe catalytic amino acid residues. The Journal of Biochemistry 117(6): p. 1317-1322. doi: 10.1093/oxfordjournals.jbchem.a124861
33. McGuffin LJ (2007) Benchmarking consensus model quality assessment for protein fold recognition. BMC Bioinformatics 8(1): p. 345. doi: 10.1186/1471-2105-8-345.
34. Waterhouse A, Bertoni M, Bienert S et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research 46(W1): p. W296-W303. doi: 10.1093/nar/gky427.
35. Pang BC, Tsang JS (2001) Mutagenic analysis of the conserved residues in dehalogenase IVa of *Burkholderia cepacia* MBA4. FEMS Microbiology Letters 204(1): p. 135-140. doi: 10.1111/j.1574-6968.2001.tb10876.x.
36. Yan C, Xu X, Zou X (2016) Fully blind docking at the atomic level for protein-peptide complex structure prediction. Structure 24(10): p. 1842-1853. doi: 10.1016/j.str.2016.07.021.
37. Yu J, Shi J, Zhang Y, Yu Z et al. (2020) Molecular Docking and Site-Directed Mutagenesis of Dichloromethane Dehalogenase to Improve Enzyme Activity for Dichloromethane Degradation. Applied Biochemistry and Biotechnology 190(2): p. 487-505. doi: 10.1007/s12010-019-03106-x.
38. Lemmon G, Meiler J (2013) Towards ligand docking including explicit interface water molecules. PloS one 8(6): p. e67536. doi: 10.1371/journal.pone.0067536.
39. Mishra R, Mazumder A, Mazumder R et al. (2019) Docking Study and Result Conclusion of Heterocyclic Derivatives Having Urea and Acyl Moiety. Asian J Biomed Pharmaceut Science 9(67): p. 13.
40. Fu Y, Zhao J, Chen Z (2018) Insights into the molecular mechanisms of protein-ligand interactions by molecular docking and molecular dynamics simulation: a case of oligopeptide binding protein. Computational and Mathematical Methods In Medicine 2018.
41. Muslem WH, Edbeib MF, Aksoy HM et al. (2019) Biodegradation of 3-chloropropionic acid (3-CP) by *Bacillus cereus* WH2 and its in silico enzyme-substrate docking analysis. Journal of Biomolecular Structure and Dynamics 38(11), pp.3432-3441. doi: 10.1080/07391102.2019.1655482.
42. Ridder IS, Rozeboom HJ, Dijkstra BW et al. (1995) Crystallization and preliminary X-ray analysis of L-2-haloacid dehalogenase from *xanthobacter autotrophicus* GJ10. Protein Science 4(12): p. 2619-2620. doi: 10.1002/pro.5560041220.

This page is intentionally left blank.