Malaysian Journal of Microbiology
Published by Malaysian Society for Microbiology
(In SCOPUS since 2011)

In silico molecular analysis of novel L-specific dehalogenase from *Rhizobium* sp. RC1

Azza Hanif Harisna[1], Mohamed Faraj Edbeib[2], Aliyu Adamu[2], Azzmer Azzar Abdul Hamid[3], Roswanira Abdul Wahab[4]*, Widodo[1] and Fahrul Huyop[2]*

[1]Department of Biology, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Jl Veteran, Malang, 65145, East Java, Indonesia.
[2]Department of Biotechnology and Medical Engineering, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia.
[3]Department of Biotechnology, Faculty of Science, IIUM, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia.
[4]Department of Chemistry, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia.
Email: fahrul@utm.my

Received 4 July 2016; Received in revised form 12 August 2016; Accepted 6 September 2016

ABSTRACT

Aims: This study presents the first structural model and proposed the identity of four important key amino acid residues, Asp13, Arg51, Ser131 and Asp207 for the stereospecific haloalkanoic acid dehalogenase from *Rhizobium* sp. RC1.
Methodology and results: The enzyme was built using a homology modeling technique; the structure of crystallized L-DEX YL from *Pseudomonas* sp. strain YL as a template. Model validation was performed using PROCHECK to generate the Ramachandran plot. The results showed 80.4% of its residues were located in the most favoured regions suggested that the model is acceptable. Molecular dynamics simulation of the model protein was performed in water for 10 nanoseconds in which Na^+ was added to neutralize the negative charge and achieved energy minimization. The energy value and RMSD fluctuation of Cα backbone of the model were computed and confirmed the stability of the model protein.
Conclusion, significance and impact of study: *In silico* or computationally based function prediction is important to complement with future empirical approaches. L-haloacid dehalogenase (DehL), previously isolated from *Rhizobium* sp. RC1 was known to degrade halogenated environmental pollutants. However, its structure and functions are still unknown. This structural information of DehL provides insights for future work in the rational design of stereospecific haloalkanoic acid dehalogenases.

Keywords: DehL, *Rhizobium* sp. RC1, protein structure, protein functions, dehalogenase

INTRODUCTION

Many halo-organic compounds have caused numerous cases of environmental pollution due to accidental spillage and improper disposal. In addition, the industrial product is highly toxic like dioxins and polychlorinated biphenyls (PCBs) and therefore, was banned from the market. Bacterial dehalogenases are the key elements in bioremediation and the process is known as dehalogenation (Slater *et al.*, 1995).

The dehalogenation reactions have been classified into different types according to their substrate specificities (Slater *et al.*, 1995). For instance on the basis of the specificity toward the configurations of chiral carbons as follows; L-2-haloacid dehalogenases act on L-2-haloalkanoic acids to yield D-2-hydroxyalkanoic acids, D-2-haloacid dehalogenases catalyze the dehalogenation of D-2-haloalkanoic acids to the corresponding L-2-hydroxyalkanoic acids, DL-2-haloacid dehalogenases (inversion type) act on both isomers of 2-haloalkanoic acids and yield products with inversion of the C_2-configurations of the substrate, DL-2-haloacid dehalogenases (retention type) dehalogenate both D- and L-2-haloalkanoic acids to the corresponding D- and L-2-hydroxyalkanoic acids. L-2-haloacid dehalogenases have been isolated from several strains and most extensively studied so far. L-DEXs are similar to one another in their primary structure (Nardi-Dei *et al.*, 1994) and also to haloacetate dehalogenase H-II from *Moraxella* sp. (Kawasaki *et al.*, 1981). Recently, the introduction of the mutation S188V into DehE from *Rhizobium* sp. RC1 has revealed the improvement in substrate specificity towards β-halogenated compounds, for instance, 3-chloropropionic acid (Hamid *et al.*, 2015).

CORRECTED PROOF

A systematic approach has been described to amplify two different families of α-halocarboxylic acid (α-HA) dehalogenase genes of group I and group II based on the knowledge of conserved residues among different dehalogenase (Hill *et al*., 1999). Group I dehalogenases are non-stereospecific, whereas group II showing stereospecificity dechlorinating only L- but not D-2-chloropropionic acid. There was no sequence homology in two separate classes though high similarities in the amino acid sequence have been observed among the dehalogenases (Kurihara *et al*., 1995; Hill *et al*., 1999). Adoption of the molecular method is useful in identifying cryptic or silent dehalogenase genes presence in the bacteria (Hill *et al*., 1999). The hydrogen-bonded triad consisting of nucleophilic amino acid, histidine, and acidic amino acid residues plays an essential role in the catalysis of at least four groups of hydrolases: serine proteases such as trypsin and chymotrypsin, cysteine proteases such as papain, subtilisins, and a/b hydrolase fold enzymes such as haloalkane dehalogenase (Hisano *et al*., 1996).

The dehalogenation of L-2-haloacid catalyzes by L-DEX YL from *Pseudomonas* sp. strain YL probably proceeds through an S_N2 mechanism; the carboxyl group of Asp-10 is assumed to approach the C_2 atom of the substrate from the opposite side to the halogen atom. The groups other than the leaving group that are attached to the C_2 atom become planar with the atom during the transition state of the reaction. The nucleophile $O^{\delta 2}$ of Asp-10 attacks the C_2 atom from a route perpendicular to the plane while the leaving halide anion is abstracted concertedly by a residue placed on the opposite side of the plane (Li *et al*., 1998). The catalytic mechanism of L-DEX YL has been analyzed by comprehensive site-directed mutagenesis (Kurihara *et al*., 1995), ^{18}O incorporation (Liu *et al*., 1995), and chemical modification with hydroxylamine (Liu *et al*., 1997). These studies revealed that Asp-10, Thr-14, Arg-41, Ser-118, Lys-151, Tyr-157, Ser-175, Asn-177, and Asp-180 are important for the enzymatic activity and that Asp-10 acts as a nucleophile in the enzymatic reaction (Hisano *et al*., 1996). In contrast, no similar studies were done on DehL from *Rhizobium* sp. RC1. *Rhizobium* sp. RC1 produced three haloalkanoic acid dehalogenases (Cairns *et al*., 1996). The presence of more than one dehalogenases in one microorganism is far from clear.

This paper focused on DehL since the three-dimensional structure of DehL from *Rhizobium* sp. RC1 and its active mechanism have not been described. Understanding the structure of L-haloacid dehalogenase in *Rhizobium* sp. RC1 is indispensable as a basic understanding of its mechanism and function. However, the three-dimensional structure of protein solved. Therefore, we predict a homology model of the L-haloacid dehalogenase using L-DEX YL from *Pseudomonas* sp. YL as a template. The structural model of DehL will provide details of a new fold and also recognition of the substrate-binding location, which allows the identification the key catalytic amino acids. Therefore, the key amino acids function in the catalytic mechanism of L-haloacid dehalogenase in *Rhizobium* sp. RC1 is essential so that it will help to predict the mechanism of L-haloacid dehalogenase enzyme action in *Rhizobium* sp. RC1. The proposed model of the enzyme is also structurally analyzed and simulated using molecular dynamics simulation.

MATERIALS AND METHODS

L-specific dehalogenase amino acid sequence

The amino acid sequences of L-specific dehalogenase from group II were obtained from the National Centre for Biotechnology Information (NCBI) database (GenBank accession number as listed in Table 1). The amino acid sequence of *Rhizobial* DehL (Accession number CAA63794) was subjected to the ProtParam web program via the Expasy (Expert Protein Analysis System) Proteomic Server for characterization (Gasteiger *et al*., 2005). The phylogenetic tree was constructed with the neighbour-joining method by MEGA6 software package (Tamura *et al*., 2013) to establish a molecular phylogenetic classification.

Amino acid sequence alignment was carried out in the form of multiple and pairwise. The multiple sequence alignment was conducted using MultAlin version 5.4.1 (Corpet, 1988). Pairwise sequence alignment was carried out using BioEdit version 7.2.5 (Hall, 1999).

Table 1: List of dehalogenase group II and its sources.

Source of dehalogenase	Dehalogenase	Accession No.	References
Rhizobium sp. RC1	DehL	CAA63794	(Cairns *et al*., 1996)
Pseudomonas putida strain AJ1	HadL	AAA25832	(Jones *et al*., 1992)
Pseudomonas sp. strain CBS3	DehCI	AAA63640	(Schneider *et al*., 1991)
Pseudomonas sp. strain CBS3	DehCII	AAA25833	(Schneider *et al* ., 1991)
Xanthobacter autotrophicus strain GJ10	DhlB	AAA27590	(Van Der Ploeg *et al*., 1991)
Pseudomonas putida strain 109	DehH109	BAA04474	(Kawasaki *et al*., 1994)
Pseudomonas cepacia strain MBA4	Hdl IVa	CAA46976	(Murdiyatmo *et al*., 1992)
Moraxella sp. strain B	DehH2	D90423	(Kawasaki *et al* ., 1992)
Pseudomonas sp. strain YL	L-DEX YL	AAB32245	(Nardi-Dei *et al* ., 1994)

CORRECTED PROOF

Hydrophobicity profile

The hydrophobicity profiles were calculated according to Kyte and Doolittle hydrophobicity mean profile using BioEdit version 7.2.5 (Hall, 1999). On the other hand, the analysis also been conducted using ProtParam (Gasteiger *et al*., 2005) for the grand average of hydropathicity.

Secondary structure prediction and building 3D model

Secondary structure prediction was performed with Integrated Webware (NPSA) in Pole Bioinformatic Lyonnais-Gerland by GOR4 (Garnier *et al*., 1996). The amino acid sequence of DehL was submitted to the SWISS-MODEL comparative protein modeling server for three-dimensional structure building (Biasini *et al*., 2014) based on the expected homology L-DEX of *Pseudomonas* sp. strain YL enzyme.

Model refinement

The constructed model was refined by molecular dynamics (MD) simulation using GROMACS 4.6.3 software package (Van Der Spoel *et al*., 2005) in the Gromos96 54a7 force field. The protein was put into a suitable sized simulation cubic box and immersed with simple point-charges (SPC) water model. In addition, Na^+ ions were added to neutralise the charge of the system. The energy of the entire system converged at 458 step of steepest descent. All simulations were performed at a constant temperature and pressure with a non-bonded cut-off of 1.4 Å. The molecular dynamics (MD) simulation was carried out for 10 ns at 300 K. The linear constraint solver (LINCS) was used to constrain the bond length, and the particle mesh Ewald method was employed for the electrostatic interactions. The integration time step was 2 femto-second (fs), and the neighbour list was updated every fifth step using the grid option and a cut-off distance of 1.4 Å. A periodic boundary condition was used with a constant number of particles in the systems, pressure, and temperature simulation criteria (NPT). During the simulation, every 1.0 ps of the actual frame was stored. The stabilised structure was taken from the trajectory system for the quality determination of the protein geometry and the structure folding reliability. Subsequently, the dynamic behaviour and the structural changes of the protein were analysed by the calculation of the root mean square deviation (RMSD). Monitoring the RMSD of the protein can give insights into its structural conformation throughout the simulation.

Structural assessment of the DehL model

A model quality estimation was evaluated using Qualitative Model Energy Analysis (QMEAN) server implemented by Swiss Institute of Bioinformatics (SIB) which able to handle oligomeric structures and absolute quality measures. The quality of the structural model was validated by ERRAT (Colovos *et al*., 1993), Verify3D (Eisenberg *et al*., 1997) and PROCHECK (Laskowski *et al*., 1996). In PROCHECK, a Ramachandran plot was generated. The final structure of DehL was visualised using UCSF-Chimera version 1.10.2 (Pettersen *et al*., 2004).

RESULTS AND DISCUSSION

Evolutionary relationships of L-specific dehalogenases

L-specific dehalogenases belong to group II. Figure 1 shows the evolutionary relationships of all L-specific dehalogenases. Studies on L-DEX YL and DhlB were well established. Both enzymes shared common ancestral lineage to DehL that provide the basis of further analysis.

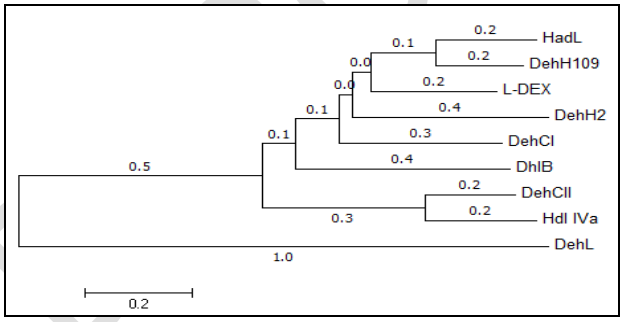

Figure 1: The evolutionary history was inferred using the Neighbour-Joining method by MEGA6 software package. The optimal tree with the sum of branch length = 4.11 is shown. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 9 amino acid sequences. All positions containing gaps and missing data were eliminated.

Analysis of DehL protein by ProtParam

The DehL from *Rhizobium* sp. RC1 consists of 279 amino acid residues. Analysis by the program ProtParam indicated that DehL has a molecular weight of 30,888.2 Da and a theoretical pI of 8.62. DehL was found to have a negative GRAVY value (grand average of hydropathicity) of 0.272, which is generally hydrophobic. There was a total of 31 negatively charged and 34 positively charged residues in the amino acid sequence. Furthermore, the total number of atoms was 4,319 with a molecular formula of $C_{1353}H_{2151}N_{399}O_{402}S_{14}$ and an Aliphatic Index of 81.58.

Multiple sequence alignment of group II dehalogenase

The multiple sequence alignment of DehL (*Rhizobium* sp. RC1) (Cairns *et al*., 1996), HadL (*P. putida* strain AJ1)

CORRECTED PROOF

(Jones *et al*., 1992), DehCI (*Pseudomonas* sp. strain CBS3) (Schneider *et al*., 1991), DehCII (*Pseudomonas* sp. strain CBS3) (Schneider *et al*., 1991), DhlB (*Xanthobacter autotrophicus* strain GJ10) (Van Der Ploeg *et al*., 1991), DehH109 (*P. putida* strain 109) (Kawasaki *et al*., 1994), HdlVa (*P. cepacia* strain MBA4) (Murdiyatmo *et al*., 1992) and DehH2 (*Moraxella* sp. strain B) (Kawasaki *et al*., 1992) were carried out (Figure 2).This alignment has a minimum sequence length of 224, maximum sequence length of 279 and the average sequence length of 236. As shown in Figure 2, the conserved regions are of F (Phe12), D (Asp13), G (Gly16), T (Thr17), W (Trp46), L (Leu59), (Leu29), (Leu88) and (Leu 118).

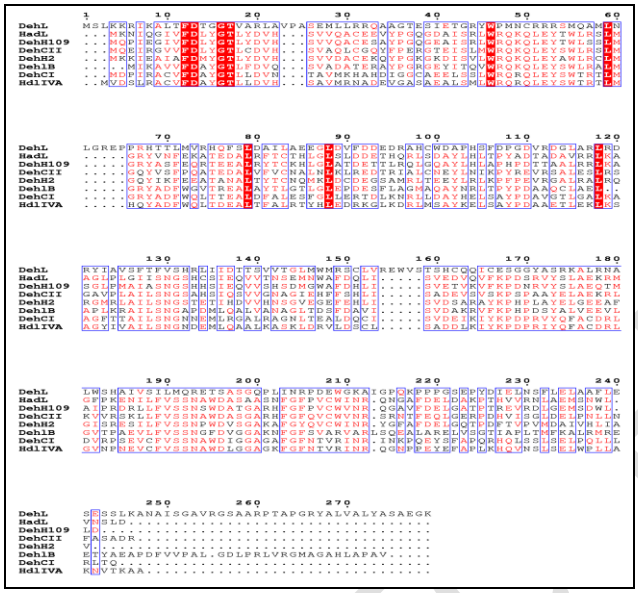

Figure 2: An overview of multiple sequence alignment of nine related dehalogenase amino acids from Group II. Conserved amino acids residues are F (Phe), D (Asp), G (Gly), T (Thr), W (Trp) and L (Leu); Amino acids with high consensus value are highlighted in red; those low consensus value and neutral are highlighted in black.

Pairwise sequence alignment of DehL and L-DEX YL

DehL amino acid sequence was compared to the well-studied L-DEX YL (Hisano *et al*., 1996; Li *et al*., 1998), and identified three key amino acids D(Asp), R(Arg) and S(Ser) which play important role in the L-haloacid dehalogenase mechanism. Those amino acids in L-DEX YL are best aligned at position D10, R41, S118 and D180 which are equivalent to D13, R51, S131, and D207 in *Rhizobium* sp. RC1, respectively (Figure 3). The pairwise results of both L-haloacid dehalogenase from *Rhizobium* sp. RC1 and *Pseudomonas* sp. YL gave only 18% sequence identity (29% similarity). However, the key amino acids were identical to each other and were hypothesized to have similar functions.

Hydrophobicity profile

Hydrophobicity analysis by BioEdit showed a comparison chart of hydrophobicity of *Rhizobium* sp. RC1 with *Pseudomonas* sp. strain YL. In Figure 4 the hydrophobicity in *Pseudomonas* sp. strain YL is higher than *Rhizobium* sp. RC1. Analysis using ProtParam showed that DehL was found to have a negative GRAVY value (grand average of hydropathicity) of 0.272, which is generally hydrophobic. On the other hand, GRAVY value of *Pseudomonas* sp. strain YL is -0.248, which is less hydrophobic.

Secondary structure of DehL

The secondary structure in a protein is described by the patterns of hydrogen bonds between the amide and carboxyl groups in the invariant parts of the amino acids in the polypeptide backbone or main chain. The secondary structure predictions for DehL from *Rhizobium* sp. RC1 using GOR4 server (Garnier *et al*., 1996) showed in Figure 5. The predicted quantity of helices (h) was 40.14%, strands (e) 16.13% and coil (c) 43.73% respectively.

Homology modeling of DehL

We design one monomer of DehL (residues 6-239) with coverage 0.84. In the present study, the three-dimensional model of DehL was built based on the target-template alignment, L-DEX YL using Swiss-Model, a protein modeling server. For structure based L-DEX YL template showed the GMQE value is 0.96 while QMEAN is -10.44 as shown in Figure 6A. The QMEAN value is based on four criteria, Cß (-7.54), all atom (-4.69), solvation (-6.07) and torsion (-6.69). Accordingly, Figure 6B showed the form of N-terminal of DehL based on L-DEX YL template form as a strand and the C-terminal form as a helix. The final total energy of the predicted model was -6822.463 kJ/mol suggesting the 3D structure generated was adequate.

Refinement of the initial DehL model

A 10 nano-second (ns) molecular dynamics calculation was conducted to refine and asses the stability of the initial Swiss-Model structure. The global behaviour of the model was analyzed based on the RMSD and the root mean square fluctuations (RMSF) of the protein backbone. It is noteworthy that the RMSD of the resulting structures did not change significantly (within the range 0.2-0.4) throughout the simulations as is illustrated in Figure 7A.

The RMSF plot of the proteins as a function of residue number in Figure 7B were found to be within average and no discrepancies found between the RMSF graph and the 3D structure in terms of loops distribution. Since these are tolerable fluctuations in the backbone, the protein model stability is verified.

CORRECTED PROOF

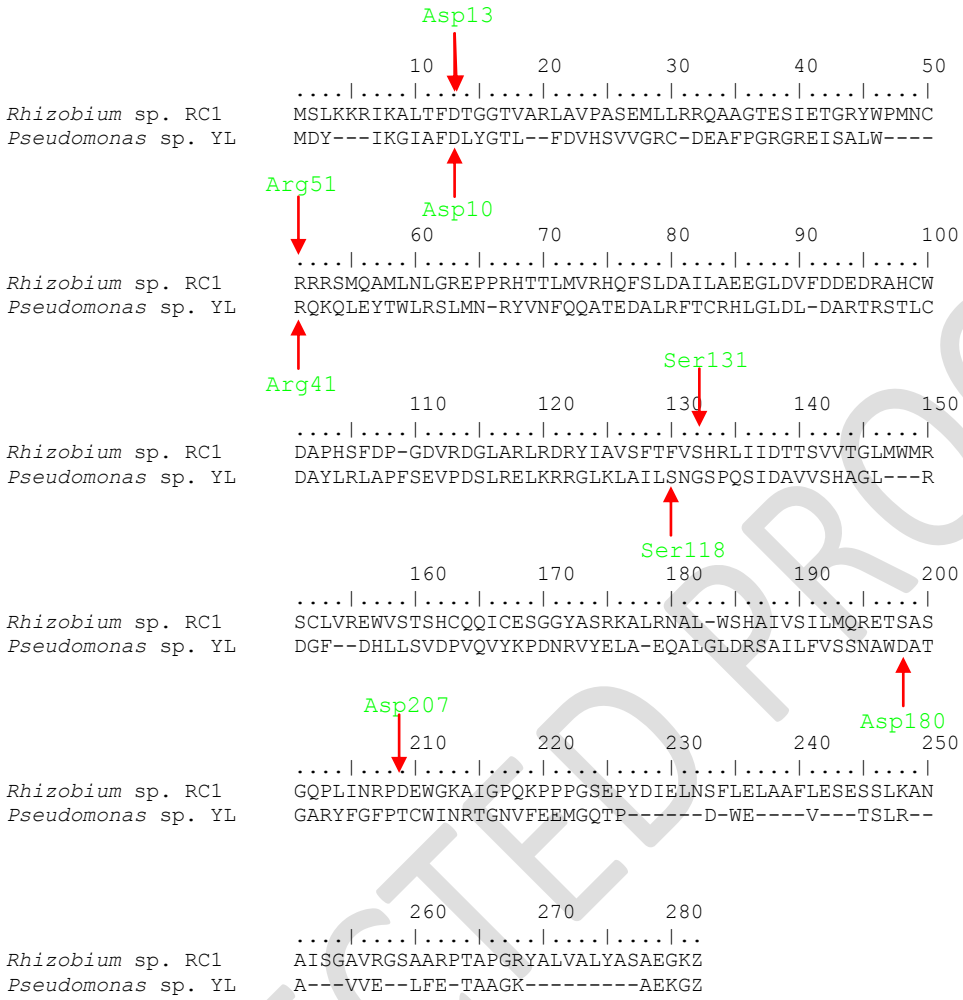

Figure 3: The pairwise sequence alignment results within DehL from *Rhizobium* sp. RC1 (Cairns *et al*., 1996) and L-DEX YL from *Pseudomonas* sp. strain YL (Nardi-Dei *et al*., 1994). Z indicates stop codon.

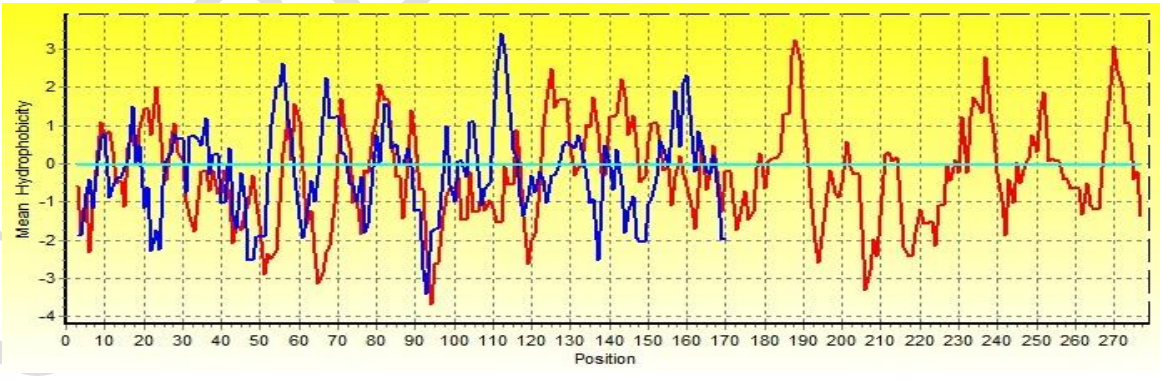

Figure 4: Mean hydrophobicity profile of DehL (red) and L-DEX YL (blue).

CORRECTED PROOF

```
1         10         20         30         40         50         60         70
|         |          |          |          |          |          |          |
MSLKKRIKALTFDTGGTVARLAVPASEMLLRRQAAGTESIETGRYWPMNCRRRSMQAMLNLGREPPRHTT
ccccceeeeeeccccccchhhhhhhhhhhhhhhhhhcccceeeeccccccceeeechhhhhhccccccccch

LMVRHQFSLDAILAEEGLDVFDDEDRAHCWDAPHSFDPGDVRDGLARLRDRYIAVSFTFVSHRLIIDTTS
hhhhhhhhhhhhhhhhhccccccccccccccccccccccccchhhhhhhhhhheeeccccccceeeeccee

VVTGLMWMRSCLVREWVSTSHCQQICESGGYASRKALRNALWSHAIVSILMQRETSASGQPLINRPDEWG
eeeceeeeeeeeeeeeeccccceeccccccchhhhhhhhhhhhhhhhhhhhhcccccccceeccccccc

KAIGPQKPPPGSEPYDIELNSFLELAAFLESESSLKANAISGAVRGSAARPTAPGRYALVALYASAEGK
ccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhhhhhceec
```

Figure 5: Prediction secondary structure of DehL which contain helix (h), strand (e) and coil (c).

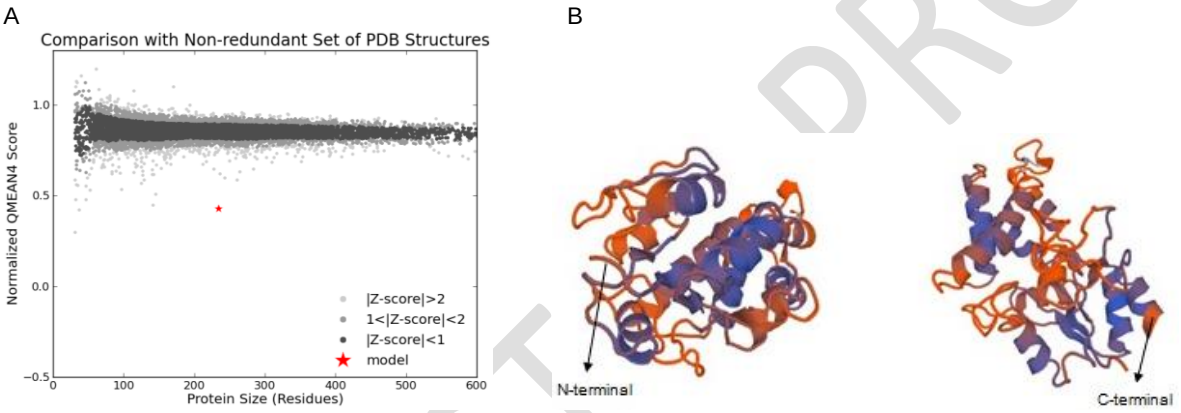

Figure 6: Comparison with a non-redundant set of structure (A). The structure of DehL which built according to L-DEX YL as a template showing possible N- and C- terminal regions (B).

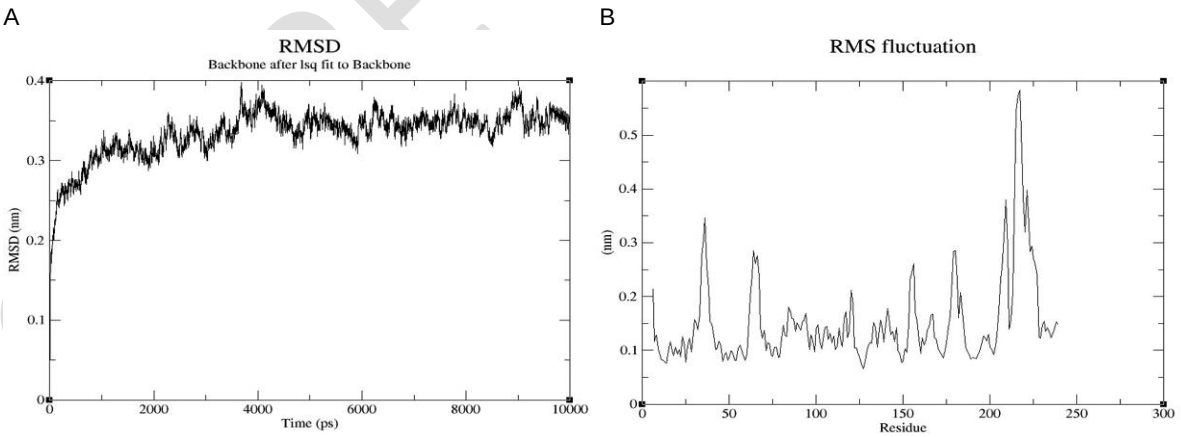

Figure 7: Molecular dynamics simulation of DehL at 10 ns. Backbone root mean square deviation (RMSD) of the model during the simulation (A). C-α backbone root mean square fluctuations (RMSF) of the model during the simulation (B).

.

In addition, an average temperature of 10 ns simulation at 300 K for the studied system was equal to 300 ± 0.5 K (Figure 8). Therefore, the extracted equilibrium structure at 300 K for DehL was achieved under stable temperature conditions.

The validation of DehL structure

The QMEAN score for a three-dimensional model of DehL was 0.469. The DehL model was validated after a molecular dynamics simulation using several validation web servers. First, the results of analysis by Verify3D revealed that 73.93% of the residues had an averaged 3D-1D score above 0.2, and the remaining residues did not attain this score. Residues with a score over 0.2 in Verify3D should be considered reliable. The three-dimensional structure of DehL was analysed using the PROCHECK tools which calculates phi/psi angles (ϕ, ψ), and thus generating Ramachandran plot (Figure 9). Based on the results, the stereochemical evaluation of backbone phi and psi dihedral angles of the DehL revealed that 80.4, 16.6 and 2.9% of residues were located within the most favoured regions, allowed regions and disallowed regions, respectively. The overall quality of the model was also assessed by the ERRAT program. The ERRAT score was 51.77%. A score close to 100% implies the good stereochemical quality of the model and therefore, the current model is acceptable (Figure 10).

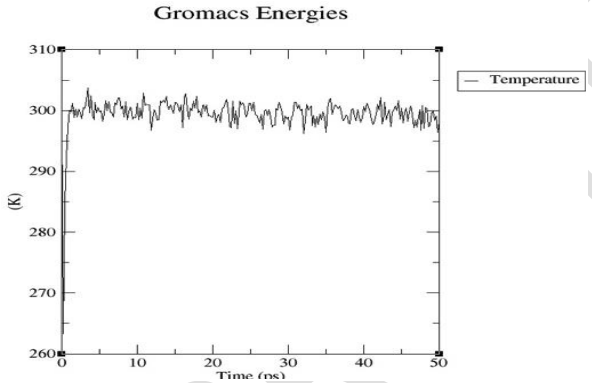

Figure 8: Variation of the temperature stability during molecular simulation.

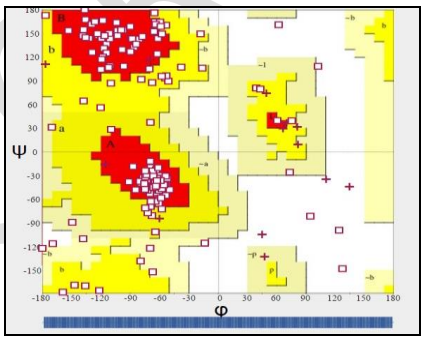

Figure 9: Ramachandran plot of the three-dimensional model of DehL shows the distribution of φ - ψ values for all the residues in the structure. The shading indicates the favourable and unfavourable regions of the plot, the darker the shading the more favourable the region.

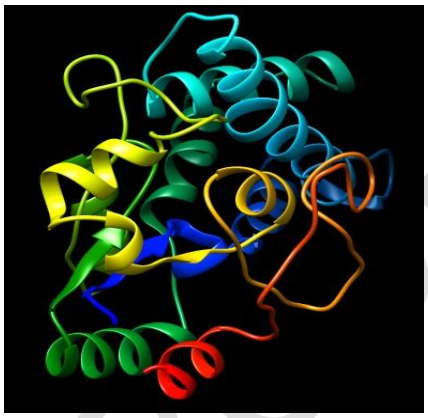

Figure 10: The three-dimensional structure of DehL coloured as a spectrum from N-terminus (blue) to C-terminus (red).

DISCUSSION

A dehalogenase is a microbial enzyme that catalyses the breakdown of high priority halogenated organic pollutants by cleaving the carbon–halogen bond. *Rhizobium* sp. RC1 produces more than one dehalogenases, dehalogenase D, E and L (Cairns *et al*., 1996). Current investigation focuses on further characterisation of DehL being stereospecific for L-2-chloropropionic acid (L-2CP). The DehD and DehL from *Rhizobium* sp. RC1 are completely different by comparing their amino acids sequence. The amino acids sequence of DehE show little similarity to DehL (16%) and DehD (14%) (Personal communication). Furthermore, it was proposed that *dehE* gene might have been evolved from *dehD* and *dehL* genes and gain new ability to react with both D- and L-2-CP. A comprehensive molecular phylogenetic classification was established and identified two different evolutionary families: group I and group II *deh* genes. Group I *deh* genes do not share any obvious feature with group II *deh* genes in terms of DNA or deduced amino acid sequences, suggesting that they are not evolutionarily related. They also seem to be functionally distinct in that all of group I *deh* genes tested encoded dehalogenases that dechlorinated D-2-chloropropionic acid, whereas all group II dehalogenases tested lacked this activity (Hill *et al*., 1999). Almost all of the α-haloalkanoic acid (αHA) dehalogenases encoded by genes now assigned to either the group I or group II *deh* genes. Only two exceptions were identified: the DehH1 gene from a *Moraxella* sp. (Kawasaki *et al*., 1992) and DehL from a *Rhizobium* sp. (Cairns *et al*., 1996).

The DehL from a *Rhizobium* sp. RC1 shows the same stereospecificity as group II, (i.e., dechlorination of L- but not D-2-chloropropionic acid). The results from phylogenetic tree (Figure 1) indicates that the *dehL* gene

was found to be closely related to the group II *deh* genes from the HAD superfamily. Identifying the origin of novel enzyme activities and the adoption of bacteria to degrade xenobiotic compounds are importance as a result of α-haloalkanoic acid (αHA) dehalogenase evolution. The role of an unknown protein can be inferred from its sequence and structural homology to other known proteins. Therefore, it is crucial to determine the true evolutionary relationships between dehalogenase genes and to develop methods by which adaptive processes involving *deh* genes can be studied in the natural environment. Recent study has reported that *Ancylobacter aquaticus* strain UV5 also produced L-2-haloacid dehalogenase (L-2-DhlB) (Kumar *et al*., 2016). The substrate specificity of L-2-DhlB suggested that it belongs to L-2-dehalogenase. The enzyme converted L-2-chloropropionic acid almost 100% but showed no activity towards D-2-chloropropionic acid.

Multiple sequence alignment showed an overview of the relationships between nine sequences L-specific dehalogenase from group II (Figure 2). There are four key amino acids were proposed to be important for the formation of enzyme-substrate based on the mechanism of L-haloacid dehalogenase from *Pseudomonas* sp. strain YL (Figure 3). These amino acids are predicted to have similar function with its superimposed DehL from *Rhizobium* sp. RC1 and/or possibly in each of bacterial produces L-haloacid dehalogenases in the same group. The key amino acids like D10 (D13) which has shown to act as a nucleophile α-carbon of the substrate to form an ester intermediate in the L-haloacid dehalogenase mechanism. R41 (R51) which acted as substrate uptake, whereas, S118 (S131) acts to stabilize the substrate carboxyl moiety and D180 (D207) which activated the hydrolysis of ester intermediate (the equivalent residue numbers for DehL are enclosed in parentheses). Similar observation was proposed as reported by Hamid *et al*. (2015), D189 is homologous to D194 in D, L-DEX 113 located at the active site location of DehI. The D189 side chain is likely to be directly involved in catalysis.

The amino acid sequence alignment patterns provided information on the conservation of core residues, such as the hydrophobic core and regions that are important for protein function. The amino acid sequence of DehL was observed to be similar to the previously crystallized structure to L-specific dehalogenase from *Pseudomonas* sp. strain YL. The similarity of DehL to other proteins was used to assess the statistical significance of the global alignment to generate sequence pairs of the appropriate length and composition by chance. Analysis from BioEdit indicated that DehL was only 18% identical to the conserved sequences with L-DEX YL and a sequence similarity of 29%. This observation suggested that this protein may or may not have the same functional residues responsible for the dehalogenation process. Hydrophobic interactions are a major force in protein folding and numerous hydropathy scales have been developed to quantify the relative hydrophobicity of the amino acids. Hydropathy profiles can be used to examine the surface features of proteins in order to generate hypotheses that can be confirmed experimentally.

The hydrophobic effect is responsible for the separation of a mixture of oil and water into its two components. It also responsible for the stability of cell membranes, drives protein folding as well as the insertion of membrane proteins into the non-polar lipid environment and finally stabilizes protein-small molecule interactions.

In this study, the hydrophobicity between DehL from *Rhizobium* sp. RC1 and *Pseudomonas* sp. strain YL was compared. The hydrophobicity curve represents the average of a residue-specific hydrophobicity index calculated over the chosen window. The hydrophobicity in *Pseudomonas* sp. strain YL, with GRAVY value -0.248 is higher than *Rhizobium* sp. RC1 with GRAVY value -0.272. The value under 0 indicates less hydrophobicity. The hydrophobicity has responsible for the stability of protein structure. *Rhizobium* sp. RC1 showed low instability index (II), 48.23, which classified the protein as unstable. In contrast, *Pseudomonas* sp. strain YL showed higher instability index (II), 39.19, suggesting that the protein is stable. It gives information that the L-haloacid dehalogenase encoded by DehL and L-DEX YL might give different protein characteristic. It also affects the protein-folding and gives different structure for both.

The longest α-helices for DehL were the sixth helix, and the shortest α-helices were the second helix and the seventh helix. For the secondary structure predictions, DehL illustrates similar local contacts to its closely related enzyme, the crystallised structure of L-DEX YL from *Pseudomonas* sp. strain YL. Both enzymes contain mostly α-helices.

The fundamental of comparative or homology model building is that similar sequences implement similar protein structures. Protein structure prediction based on target-template alignment is considerably constructive and convenient, as most of the newly sequenced proteins are probably sharing similar structures with one that has already been experimentally determined and crystallized. Hence, in the present study, *Pseudomonas* sp. strain YL was chosen as a template for homology modeling of DehL. A regularization technique has been used to predict coordinates in the positions where the key amino acids occurred in the alignment.

The GMQE value showed 0.96 while QMEAN showed -10.44. The QMEAN value is based on four criteria, they are; Cß (-7.54), All atom (-4.69), Solvation (-6.07) and Torsion (-6.69). This might be the cause of the low value in identity and similarity of the sequence. Using L-Dex YL as template, one monomer of DehL (residues 6-239) was generated with coverage 0.84. There is no ligand which match with the model because of the binding site is not conserved. Energy minimization was performed for the model and resulted in a model with reduced energy. The global minimum level of total energy as protein structures is preferred to be more stable at low energy levels by utilising the energy minimization step. The native conformation of proteins corresponds to global minima of their free energy as indicated by thermodynamic

CORRECTED PROOF

hypothesis (Frisch *et al*., 1997). Proteins carry out their function efficiently only in their native state.

A molecular dynamics simulation was conducted for 10 nanoseconds to improve and verify the stability of the initial DehL structure. The three-dimensional structure of DehL was analyzed based on RMSD and RMSF. The RMSD is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. The black line in Figure 7A represents the RMSD of the protein backbone during the simulation, which remains within an acceptable range. These RMSD values denote that the employed simulation time was long enough to achieve an equilibrium structure of DehL. Meanwhile, the RMSF is useful for characterizing local changes along the protein chain. On this plot, peaks indicate areas of the protein that fluctuated the most during the simulation. The C-terminal tail was observed fluctuated much more than any other part of the protein, and it showed in Figure 10 that the α-helix is the most stable part of the protein. Thus, the applied molecular dynamics was required to specify the geometry of DehL.

A model quality estimation is an essential component of protein structure prediction since ultimately the accuracy of a model determines its usefulness for specific applications. According to the QMEAN scoring function (Benkert *et al*., 2008), the stable protein structure has a range number 0-1 and QMEAN score for DehL was 0.469. The average 3D-1D profile score for each residue in a 21-residue sliding window showed in the vertical axis of Verify3D from -1 (bad score) to 1+ (good score). The score for Verify3D is between -0.1 and 0.7 suggesting the DehL profile model was good. Analysis using ERRAT tools was computed for non-bonded atomic interactions by comparing the statistics of highly refined structures. DehL model scored 51.77%. Any value above 50% or higher, indicates a better quality model. The stereochemical quality of the final model was analysed with PROCHECK to generate the Ramachandran plot. The Ramachandran plot is a two-dimensional plot of the Phi (ϕ) - Psi (ψ) torsion angles of the protein backbone that provides a simple view of the conformation of a protein. The plot indicated that 80.4% of residues located in the most favoured regions, thus verified model is acceptable.

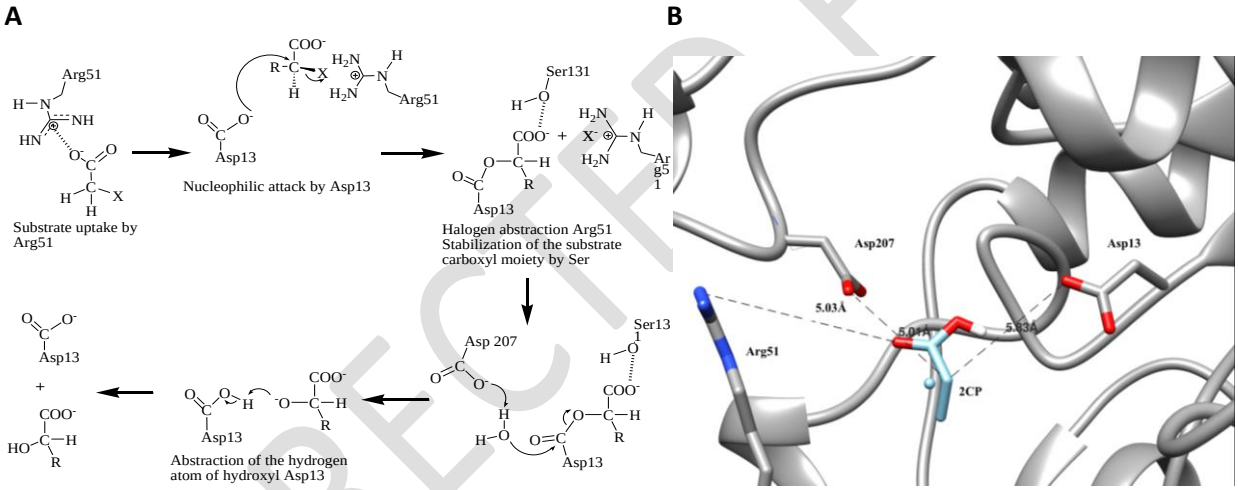

Figure 11: A, A proposed mechanism of the reaction of DehL. There are four key amino acid residues involved in the dehalogenation reaction Arg51, Asp13, Ser131 and Asp207. B, A predicted active site contacts for DehL protein from *Rhizobium* sp. RC1 deduced from UCSF Chimera (v1.10.2) for substrate L-2CP (or 2CP) and its important amino acids. The distance are measured as shown in (B). Asp13 was predicted to be a nucleophilic attack α-carbon of the substrate to form an ester intermediate. Arg51 acts as substrate uptake, while Ser131 (behind Asp13) stabilize the substrate carboxyl moiety. Asp207 was predicted to activate the hydrolysis of the ester intermediate.

CONCLUSION

In this study, reported the first *in silico* model of the DehL from *Rhizobium* sp. RC1. The overall perspective demonstrated the basic structure and functions of DehL for potential bioremediation agents, which contributes to future rational design studies. A proposal of basic mechanism of DehL was deduced based on important amino acids (Figure 11 A). A 3D model of DehL complete with its important amino acids to prove a possible amino acids interaction with L-CP (2CP) for basic mechanism of enzyme action was in Figure 11 B. Therefore, should be any further research about the electrostatic properties of the enzyme active site, including the quantum-chemistry calculations of the inter-atomic interaction and docking experiments should lead to a better understanding of this protein. Findings from this study have emphasized the need to perform crystallographic and in-depth studies of the details structure of DehL complexes with its substrate binding mechanisms. The conserved region and polar

CORRECTED PROOF

residues should be subjected to site-directed mutagenesis to confirm the catalytic residues.

ACKNOWLEDGEMENT

We gratefully thank Ministry of Education Malaysia & Universiti Teknologi Malaysia for Research University Grant Scheme (RUG) **Vot Q.J130000.2545.12H91, 09H95, 14H56** and **11H94** for financial supports. AHH thanks University of Brawijaya for international studentship award.

REFERENCES

Benkert, P., Tosatto, S. C. E. and Schomburg, D. (2008). Qmean: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics* **71, 261-277.**

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. and Schwede, T. (2014). Swiss-model: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42, W252-258.**

Cairns, S. S., Cornish, A. and Cooper, R. A. (1996). Cloning, sequencing and expression in *Escherichia coli* of two *Rhizobium* sp. genes encoding haloalkanoate dehalogenases of opposite stereospecificity. *European Journal of Biochemistry* **235, 744-749.**

Colovos, C. and Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science* **2, 1511-1519.**

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16, 10881-10890.**

Eisenberg, D., Lüthy, R. and Bowie, J. U. (1997). VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology* **277, 396-404**.

Frisch, C., Schreiber, G., Johnson, C. M. and Fersht, A. R. (1997). Thermodynamics of the interaction of barnase and barstar: Changes in free energy versus changes in enthalpy on mutation1. *Journal of Molecular Biology* **267, 696-706.**

Garnier, J., Gibrat, J. F. and Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods in Enzymology* **266, 540-553**.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein identification and analysis tools on the expasy server. In: The proteomics protocols handbook. J. M. Walker (ed.). Humana Press, Totowa, NJ. **pp. 571-607.**

Hamid, A.A.A., Tengku Abdul Hamid, T. H., Abdul Wahab, R., Omar, M. S. S., and Huyop, F. (2015). An S188V mutation alters substrate specificity of non-stereospecific α-haloalkanoic acid dehalogenase E (DehE). *PloS one* **10, e0121687.**

Hall, T. A. (1999). Bioedit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/nt. In: Nucleic acids symposium series. Oxford University Press, **pp. 95-98**

Hill, K. E., Marchesi, J. R. and Weightman, A. J. (1999). Investigation of two evolutionarily unrelated halocarboxylic acid dehalogenase gene families. *Journal of Bacteriology* **181, 2535-2547.**

Hisano, T., Hata, Y., Fujii, T., Liu, J.-Q., Kurihara, T., Esaki, N. and Soda, K. (1996). Crystal structure of l-2-haloacid dehalogenase from *Pseudomonas* sp. Yl an α/β hydrolase structure that is different from the α/β hydrolase fold. *Journal of Biological Chemistry* **271, 20322-20330.**

Jones, D. H., Barth, P. T., Byrom, D. and Thomas, C. M. (1992). Nucleotide sequence of the structural gene encoding a 2-haloalkanoic acid dehalogenase of *Pseudomonas putida* strain AJ1 and purification of the encoded protein. *Journal of General Microbiology* **138, 675-683.**

Kawasaki, H., Tone, N. and Tonomura, K. (1981). Purification and properties of haloacetate halidohydrolase specified by plasmid from *Moraxella* sp. Strain B. *Agricultural and Biological Chemistry* **45, 35-42.**

Kawasaki, H., Toyama, T., Maeda, T., Nishino, H. and Tonomura, K. (1994). Cloning and sequence analysis of a plasmid-encoded 2-haloacid dehalogenase gene from *pseudomonas putida* No 109. *Bioscience, Biotechnology and Biochemistry* **58, 160-163.**

Kawasaki, H., Tsuda, K., Matsushita, I. and Tonomura, K. (1992). Lack of homology between two haloacetate dehalogenase genes encoded on a plasmid from *Moraxella* sp. strain B. *Journal of General Microbiology* **138, 1317-1323.**

Kumar, A., Pillay, B. and Olaniran, A. O. (2016). L-2-haloacid dehalogenase from *ancylobacter aquaticus* UV5: Sequence determination and structure prediction. *International Journal of Biological Macromolecules* **83, 216-225.**

Kurihara, T., Liu, J.-Q., Nardi-Dei, V., Koshikawa, H., Esaki, N. and Soda, K. (1995). Comprehensive site-directed mutagenesis of l-2-halo acid dehalogenase to probe catalytic amino acid residues. *Journal of Biochemistry* **117, 1317-1322.**

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996). Aqua and procheck-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8, 477-486.**

Li, Y.-F., Hata, Y., Fujii, T., Hisano, T., Nishihara, M., Kurihara, T. and Esaki, N. (1998). Crystal structures of reaction intermediates of L-2-haloacid dehalogenase and implications for the reaction mechanism. *Journal of Biological Chemistry* **273, 15035-15044.**

Liu, J.-Q., Kurihara, T., Miyagi, M., Esaki, N. and Soda, K. (1995). Reaction mechanism of L-2-haloacid dehalogenase of *Pseudomonas* sp. YL identification of asp as the active site nucleophile by 18° incorporation

CORRECTED PROOF

experiments. *Journal of Biological Chemistry* **270, 18309-18312.**

Liu, J.-Q., Kurihara, T., Miyagi, M., Tsunasawa, S., Nishihara, M., Esaki, N. and Soda, K. (1997). Paracatalytic inactivation of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL by hydroxylamine evidence for the formation of an ester intermediate. *Journal of Biological Chemistry* **272, 3363-3368.**

Murdiyatmo, U., Asmara, W., Tsang, J. S. H., Baines, A. J., Bull, A. T. and Hardman, D. J. (1992). Molecular biology of the 2-haloacid halidohydrolase IVa from *Pseudomonas cepacia* MBA4. *Biochemical Journal* **284, 87-93.**

Nardi-Dei, V., Kurihara, T., Okamura, T., Liu, J.-Q., Koshikawa, H., Ozaki, H., Terashima, Y., Esaki, N. and Soda, K. (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*. *Applied and Environmental Microbiology* **60, 3375-3380.**

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004). UCSF chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25, 1605-1612.**

Schneider, B., Müller, R., Frank, R. and Lingens, F. (1991). Complete nucleotide sequences and comparison of the structural genes of two 2-haloalkanoic acid dehalogenases from *Pseudomonas* sp. strain CBS3. *Journal of Bacteriology* **173, 1530-1535.**

Slater, J. H., Bull, A. T. and Hardman, D. J. (1995). Microbial dehalogenation. *Biodegradation* **6, 181-189.**

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30, 2725-2729.**

Van Der Ploeg, J., Van Hall, G. and Janssen, D. B. (1991). Characterization of the haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 and sequencing of the *dhlb* gene. *Journal of Bacteriology* **173, 7925-7933.**

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. and Berendsen, H. J. C. (2005). Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry* **26, 1701-1718.**

CORRECTED PROOF