

Putative Dehalogenase uptake gene from *Rhizobium* sp. RC1

Adam Izzuddin Nasir¹, Aliyu Adamu², Yilmaz Kaya³, Mohamed Faraj Edbeib⁴,
Fahrul Huyop⁵

^{1,5}Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor Bahru Johor, Malaysia

²Department of Microbiology, Kaduna State University, Kaduna, Nigeria

^{3,4}Department of Agricultural Biotechnology, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

⁴Department of Animal Production, Faculty of Agriculture, Baniwalid University, Libya

¹Corresponding email: adamsizzuddinnasir@gmail.com

Abstract

Halogenated compounds that are polluting the environment is becoming a huge concern that demands an efficient solution. The rising priority in the use of microorganisms that have the potential to be the most environmental-friendly way to degrade and dispose of halogenated compounds is crucial in solving this problem. *Rhizobium* sp. RC1 is a Gram-negative bacterium that was isolated from soil using 2,2-dichloropropionic acid (Dalapon) and uses it as one of its primary sources of carbon. It is known that *Rhizobium* sp. RC1 is able to biodegrade halogenated compounds but the transport process of halogenated compounds into the cell is mostly unknown. There is a putative haloacid transport gene that is present in *Rhizobium* sp. RC1, *dehrP*, and the properties of the protein from the gene, *DehrP*, is ambiguous. It was concluded from this study that *DehrP* is a protein that belongs in the Major Facilitator Superfamily (MFS) and Metabolite:H symporter (MHS) classification of proteins by analysing the sequence and identifying the motifs that classify *DehrP* as a part of this classification of proteins. It was also found that the domains between *DehrP* and MFS/MHS proteins are conserved which provides a solid relation between them.

Keywords: *dehrP*, *DehrP*, 2,2-dichloropropionic acid (Dalapon), Major Facilitator Superfamily (MFS), Metabolite:H symporter (MHS)

1. Introduction

Halogenated compounds are a carbon source that are made artificially or naturally in which the latter is the result of events such as volcanic eruptions or other geothermal processes (Gribble, 2003). These compounds are used extensively in industrial and personal use cases and are also used as an energy source for microorganisms through biodegradation. Halogenated compounds are common in everyday applications. Dalapon or 2,2-dichloropropionic acid and haloacetates are usually used in herbicides and household cleaners respectively. Due to the heavy reliance of halogenated compounds in the modern world, the pollution that has amounted from this is a cause for concern due to their proven toxic effects towards living organisms and lack the ability to properly dispose of these compounds from their systems (Plewa, Kargalioglu, Vankerk, Minear, & Wagner, 2002). *Rhizobium* sp. RC1 is a Gram-negative bacterium was found to have the ability to biodegrade halogenated compounds and use it along with glucose as a primary source of carbon (Huyop & Nemati, 2010). Since *Rhizobium* sp. RC1 has the ability to transport haloacids into the cytoplasm, it is assumed that a specific transport protein is responsible for

this process. It is found that *dehrP*, a type of putative haloacid transport gene, was sequenced and the protein, *DehrP*, was classified as a member of the Major Facilitator Superfamily (MFS) class of protein that possess a distinct sugar motif in its domain.

This aim for this research is to establish that *DehrP* does belong to Major Facilitator Superfamily (MFS) class of proteins by deducing the motifs and full classification of the protein in contrast to established MFS proteins by using a variety of Bioinformatics tools.

2. Methods

Retrieval and Analysis of Protein Sequence

The sequence for transport protein gene of *Rhizobium* sp. RC1, *dehrP*, was retrieved from GenBank Database using the accession number AM260971 (Clark, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2016) and the nucleotide sequence was then translated using Expasy translate tool (Gasteiger, et al., 2003) to determine its amino acid notation. The protein sequence was then downloaded from Uniprot Knowledgebase Database with accession number Q1M2W6 (Consortium, 2015) and was run through various protein

BLAST tools in order to compare the sequence with other known proteins. The physical and chemical properties of DehrP was established with the use of ExPASy's ProtParam tool (Gasteiger, et al., 2005). InterPro 70.0 was used to establish conserved domains within DehrP to make an initial hypothesis of the class of protein that DehrP is a part of (Finn, et al., 2016).

MEME and MAST (MEME Suite 5.0.1) are then used to analyse DehrP by looking at the intersection of Metabolite:H symporter (MHS) specific motifs between DehrP and the other 12 MHS proteins (Bailey, et al., 2009).

Multiple Sequence and Pairwise Alignments of Proteins

Multiple Sequence Alignments (MSA) between DehrP and other 12 was done with the use of T-Coffee (Notredame, Higgins, & Heringa, 2000) and is visualised with the use of ESPrpt 3.0 (Robert & Gouet, 2014) and the same steps were then carried out with the same class of MFS and MHS family of proteins. Using Transitive Consistency Score (TCS) web server as further measure, we determined that the results are indeed reliable (Chang, Di Tommasso, Lefort, Gascuel, & Notredame, 2015).

3. Results and Discussion

```
1      M _ T _ T _ T _ L _ V _ A _ R _ T _ S _ S _ A _ G _ R _ M _ T _ R _ E _ E _ R _  
1      ATGACTACGACTCTAGTCGCCCCGCTACTTCATCAGCCGGTCGCATGACACGCGAGGAGCGC  
21     K _ V _ I _ F _ A _ S _ S _ L _ G _ T _ V _ F _ E _ W _ Y _ D _ F _ F _ L _ Y _  
61     AAAGTGATCTTCGCCCTCCTCGCTCGGTACTGTCTTCGAATGGTACGATTTCTTTCTATAT  
41     G _ S _ L _ A _ A _ I _ I _ G _ A _ T _ F _ F _ K _ D _ F _ P _ A _ T _ Q _  
121    GGCTCACTCGCCGCTATCATCGGCGCGACCTTTTTCGAAGGACTTTCCGCCAGCCACACAA  
61     A _ I _ F _ A _ L _ L _ A _ F _ A _ A _ G _ S _ L _ V _ R _ T _ F _ G _ A _ L _  
181    GCCATATTCGCGCTCCTTGCTTTTCGCGGCTGGCTCGCTTGTTCGGACTTTTCGGCGCACTC  
81     I _ F _ G _ R _ L _ G _ D _ M _ I _ G _ R _ K _ Y _ T _ F _ L _ V _ T _ I _ L _  
241    ATTTTTGGCCGTCTCGCGCATATGATTGGGCGCAAATATACCTTCCTCGTAACCATCCTG  
101    I _ M _ G _ L _ S _ T _ F _ V _ V _ G _ L _ L _ P _ G _ S _ D _ T _ I _ G _ L _  
301    ATCATGGGTCTGTTCGACGTTCTGGTTCGGCCTTCTACCGGGTTTCGGACACCATTTGGACTT  
121    A _ A _ P _ T _ I _ L _ I _ L _ L _ R _ L _ L _ Q _ G _ L _ A _ L _ G _ G _ E _  
361    GCGGCCCTACGATCCTGATTCTCCTCCGCTGCTTCAGGGCCTCGCGCTCGCGCGTGAG  
141    Y _ G _ G _ A _ A _ V _ Y _ E _ A _ E _ H _ A _ P _ P _ G _ R _ R _ G _ F _ Y _  
421    TACGGGGGCGCCGCTGTTTACGAGGCGGAGCATGCTCCGCCCGGACGCCGCGCTTTTAT  
161    T _ S _ W _ I _ Q _ T _ T _ A _ T _ G _ G _ L _ F _ L _ S _ L _ L _ V _ I _ L _  
481    ACAAGTTGGATTCAAACAACGCGACGGGTGGCCTCTCTCTATCGCTCCTTGTCTATTCTC  
181    G _ T _ R _ S _ L _ L _ G _ E _ E _ S _ F _ T _ S _ W _ G _ W _ R _ V _ P _ F _  
541    GGAACCTCGCTCATTTGCTGGGCGAGGAATCCTTCACGCTTGGGGCTGGCGCTGCCGTTTC  
201    L _ L _ S _ V _ V _ L _ L _ G _ I _ S _ V _ W _ I _ R _ M _ Q _ L _ N _ E _ S _  
601    TTGTTGTCCGTGGTACTGCTCGGCATCTCGGTATGGATCCGCATGCAGCTCAACGAGTCG  
221    P _ V _ F _ Q _ R _ M _ K _ A _ E _ G _ K _ A _ S _ K _ A _ P _ L _ R _ E _ A _  
661    CCCGTGTTCCAGCGCATGAAGGCGAGGGCAAGGCGTCCAAGGCTCCCTTGAGGGAGGCA  
241    F _ A _ H _ W _ P _ N _ A _ R _ L _ A _ L _ V _ A _ L _ F _ G _ M _ V _ A _ G _  
721    TTTGCGCACTGGCCTAACGCAAGGCTCGCTCTGTTGCCCTTTTCGGCATGGTCGCCGGC  
261    Q _ A _ V _ V _ W _ Y _ T _ G _ Q _ F _ Y _ V _ L _ F _ F _ L _ Q _ S _ I _ L _  
781    CAAGCCGTGGTCTGGTACACGGGCCAATTCTATGTGCTGTTCTTCTCCTGCAGAGCATCTC  
281    K _ V _ D _ G _ F _ T _ T _ L _ L _ I _ C _ W _ S _ L _ L _ L _ G _ S _ G _  
841    AAGGTAGACGGCTTCACGACGACGCTACTTATCTGTTGGTCTTTGCTGCTGGGATCTGGC  
301    F _ F _ V _ F _ F _ G _ W _ L _ S _ D _ R _ I _ G _ R _ K _ P _ I _ M _ I _ A _  
901    TTCTTCGTTTCTTTGGCTGGTTATCCGATCGCATCGGACGCAAGCCGATTATGATCGCC  
321    G _ C _ L _ L _ A _ V _ V _ T _ Y _ F _ P _ I _ F _ E _ A _ I _ T _ E _ R _ A _  
961    GGCTGCCTCCTTGCTGTTGTGACTTACTTTCCGATCTTCGAGGCAATCACCGACGAGCG  
341    N _ P _ T _ L _ A _ K _ A _ I _ S _ E _ V _ K _ V _ T _ R _ G _ L _ P _ T _ R _  
1021   AATCCACTTTAGCCAAGGCGATTTCGGAGGTGAAGGTACCCGTTGGTCTCCCGACCCGG  
361    S _ N _ A _ G _ I _ Y _ S _ I _ R _ S _ E _ P _ A _ C _ S _ H _ R _ L _ V _ T _  
1081   TCGAATGCGGAATCTATTCAATCCGGTCGGAACCGCGTGTTCACATCGTCTTGTGACG  
381    L _ P _ E _ P _ I _ L _ L _ R _ V _ R _ C _ D _ M _ I _ V _ R _ Q _ V _ L _ Q _  
1141   TTGCCCGAGCCTATCTTGCTCAGAGTTTCGGTGGCATATGATCGTCAGGCAGGTGCTGCAG  
401    E _ S _ Q _ L _ A _ C _ S _ S _ M _ A _ P _ R _ * _  
1201   GAGAGCCAACTCGCGTGCTCGTCAATGGCACCGAGGTAA
```

Figure 1. The nucleotide sequence and their corresponding amino acid notations of DehrP from *Rhizobium* sp. RC1 that was retrieved from GenBank Database (accession number: AM26097) (Clark, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2016) and translated using ExPASy (Gasteiger, et al., 2003). Visualisation were done with use of Nucleotide Amino acid Derived Visualization (NADV) (Cleveland, 2018).

Using the exact protein sequence from Figure 1, the physical and chemical traits of DehrP was established which is valuable in further understanding the protein in question. One of the more notable property is the hypothesised isoelectric point (pI) of the protein which is determined by the 24 negatively charged residues and 37 positively charged residues that are made up by Asp + Glu and Arg + Lys respectively. This is important in determining the exact pH in future downstream processing of the protein that would be carried out to purify the protein. By running the protein sequence of DehrP

through various protein BLAST tools, it is quickly hypothesised that DehrP indeed belongs in the MFS class of proteins due to the high sequence identity with the highest sequence identity of them all amounted to 98%. This is further confirmed using InterPro 70.0 which shows that the domains of MFS proteins are indeed present in DehrP (Finn, et al., 2016). By using MEME Suite 5.0.2, the locations of these motifs and further hypothesising the relation to MHS class of proteins are determined and visualised (Bailey, et al., 2009).

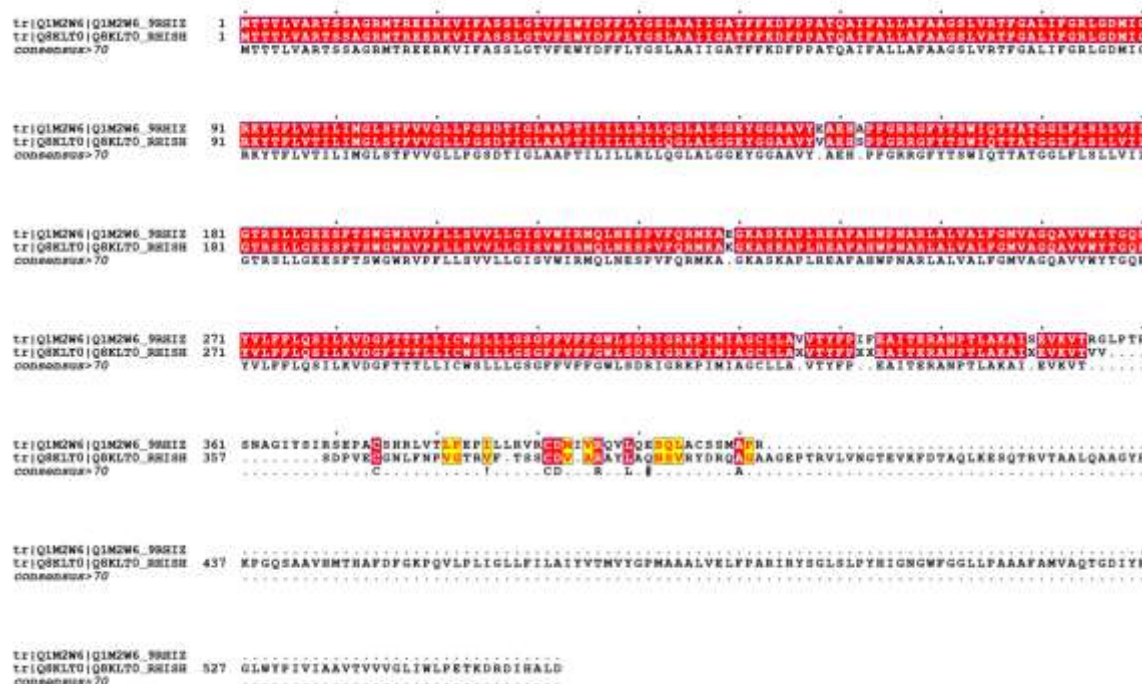


Figure 2. Pairwise sequence alignment between proteins DehrP (Uniprot accession number: Q1M2W6) and DehP (Uniprot accession number: Q8KLT0) which are from *Rhizobium* sp. RC1 and *Agrobacterium* sp. NHG3 respectively (Consortium, 2015). The alignments were generated by T-Coffee (Notredame, Higgins, & Heringa, 2000) and visualised by ESPript v.3.0 (Robert & Gouet, 2014).

By using T-Coffee, the specific amino acids that make the motifs are properly visualised and compared between DehrP and 12 members of the MHS class of proteins (Notredame, Higgins, & Heringa, 2000). From the MSA that was done, it was observed that the motifs between DehrP and the 12 MHS proteins are conserved, especially near the N terminus and towards the middle of the sequences. The same step was repeated with DehP, a known haloacid transporter protein from *Agrobacterium* sp. NHG3 that had the highest sequence identity at 98% which indicates that DehrP is indeed a transport

protein. The high sequence identity can be attributed to the fact that *Rhizobium* sp. and *Agrobacterium* sp. are only different due to mobile DNA elements such as plasmids and they are interchangeable by modifying these genes (Velazquez, et al., 2005). From Figure 2, it can be seen that DehrP has a high sequence identity to DehP and the sequences are highly conserved from the N terminus and towards the middle part of the sequence. Due to the fact that DehP is a known haloacid transporter within *Agrobacterium* NHG3, it can be hypothesised that DehrP does possess the same function. A thing to note is that the protein

sequence of DehP is longer at 558 amino acids compared to DehrP which is 412 amino acids long.

4. Conclusion

Bioinformatics tools have been key in providing a good understanding and providing general predictions to unknown proteins. This is evident in classifying DehrP as a part of the MFS and MHS class of proteins due to previous works and research done since the overall prediction is based on similarity between already known properties of other proteins. This conclusion is devised by analysing closely the motifs and amino acids that are conserved between DehrP and other proteins from the MFS and MHS family. It can be concluded that DehrP has a high possibility to be classed in the same protein family, but it is becoming apparent that this is not sufficient, and further theorisation and research is needed to raise confidence in our hypothesis. By further completing the research of DehrP, it could also prove to be useful in achieving better understanding transport proteins in general by contributing to the Bioinformatics platform.

5. References

- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, W202-W208.
- Chang, J. M., Di Tommaso, P., Lefort, V., Gascuel, O., & Notredame, C. (2015). TCS: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction. *Nucleic Acids Research*, W3-W6.
- Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids Research*, 67-72.
- Cleveland, S. B. (2018). *NADV - Nucleotide Amino acid Derived Visualization*. Retrieved September 20, 2018, from NADV: <http://nadv.herokuapp.com/>
- Consortium, U. (2015). Uniprot: a hub for protein information. *Nucleic Acids Research*, D204-D212.
- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., . . . Mitchell, A. L. (2016). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research*, D190-D199.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Balroch, A. (2003). ExPASyL the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 3784-3788.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. New Jersey: Humana Press.
- Gribble, W. G. (2003). The diversity of naturally produced organohalogens. *Chemosphere*, 289-297.
- Huyop, F., & Nemati, M. (2010). Properties of dehalogenase from *Rhizobium* sp. RC1. *African Journal of Microbiology*, 2836-2847.
- Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 205-517.
- Plewa, M. J., Kargalioglu, Y., Vanker, D., Minear, R. A., & Wagner, E. D. (2002). Mammalian cell cytotoxicity and genotoxicity analysis of drinking water disinfection by-products. *Environmental and Molecular Mutagenesi*, 134-142.
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, W320-W324.
- Velazquez, E., Peix, A., Zurdo-Pineiro, J. L., Palomo, J. L., Mateos, P. F., Rivas, R., . . . Martinez-Molina, E. (2005). The Coexistence of Symbiosis and Pathogenicity-Determining Genes in *Rhizobium rhizogenes* Strains Enables Them to Induce Nodules and Tumors or Hairy Roots in Plants. *Molecular Plant-Microbe Interactions*, 1325-1332.