IN SILICO DETERMINATION AND ANALYSIS OF PUTATIVE HALOALKANOIC ACID TRANSPORT PROTEIN OF *RHIZOBIUM SP*. RC1

MUHAMMAD ADAM IZZUDDIN BIN MOHD NASIR

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Faculty of Science
Universiti Teknologi Malaysia

JULY 2019

DEDICATION

This work is dedicated to my family, for without their never-ending support and guidance this work will never would have come into fruition. I dedicate this to my close friends, despite their busy schedules, helped me in completing this work. And to my Professors, who provided me with the knowledge and skills needed to complete this work.

ACKNOWLEDGEMENT

I would like to express my very special appreciation to my supervisor, Professor Fahrul Zaman Huyop for his guidance and support throughout the duration of the completion of this thesis. Without him, all of this would not be possible.

A very special thanks to my co-supervisors, Assoc. Prof. Dr. Roswanira Abdul Wahab and Dr. Zarita Zakaria that they have been very insightful and supportive in the writing of this thesis and their inputs have been invaluable for the completion of the thesis.

I am very grateful to my family who have given me all the care in the world in order to give me the opportunity to even have the chance to continue my studies and have been trusting in my capabilities all this time.

A very special mention to friends that I made along the way that not only made this journey better, but also have opened my mind to new possibilities. I hope we meet again at the top.

ABSTRACT

There is a growing concern regarding the lack of an efficient solution to solve halogenated compound pollution in the environment. A Gram-negative bacterium, *Rhizobium sp.* RC1, which uses 2,2-DCP as one of its primary sources of carbon was previously isolated. However, the process of transporting haloacids into *Rhizobium sp.* RC1 has yet to be confirmed. A putative haloacid transport gene, *dehrP*, inside *Rhizobium sp.* RC1 is speculated to be responsible for this process. The aim of this research was to elucidate the function of this gene for the transport of haloacids into the cell. To achieve this, *dehrP* was initially analysed using several BLAST tools and then aligned using T-Coffee against other known transport proteins. The subsequent protein of this gene, DehrP, was concluded to belong in the Major Facilitator Superfamily (MFS) and Metabolite:H^+ symporter (MHS) family of proteins. DehrP was determined to have nine transmembrane helices with MFS unique motifs between helices two and three, and helices eight and nine. Evolutionary analysis of DehrP was determined to have close relations to MHS family haloacid transporters, DehP, Deh4p and Dehp2 in *Burkholderia caribensis* MBA4. DehrP was modelled using Phyre^2 with the transport protein XylE from *Escherichia coli* as the reference model. DehrP was compared with XylE in order to determine the proton and haloacid binding sites. The proton binding site of DehrP is made up of two residues, Asp36 and Arg130 whereas the assumed haloacid binding site residues are (Glu33, Trp34, Phe37, Arg75, Tyr271 and Ser402). To verify the assumption for the haloacid binding site, the binding site residues were replaced with alanine and the new sequence was named DehrPa. The 3D structures of DehrP and DehrPa were refined using 3Drefine in order to prepare them for docking simulations using AutoDock Vina. Docking simulations were done with four haloacids (2,2-DCP, MCA, D-2DCP and L-2DCP). The assumed substrate binding residues of DehrP was validated due to the lower binding affinity and lower binding accuracy of DehrPa. Unexpectedly, it was also found that 2,2-DCP was still able to bind to three other residues that was not mutated inside DehrPa. This study confirms haloacid binding site for DehrP of previous work with additional discovery of alternative binding residues specifically for 2,2-DCP.

ABSTRAK

Kurangnya penyelesaian yang berkesan untuk pencemaran alam sekitar oleh sebatian halogen adalah semakin membimbangkan. Bakteria Gram-negatif, *Rhizobium sp.* RC1 yang menggunakan 2,2-DCP sebagai salah satu sumber utama karbon telah berjaya dipencilkan. Bagaimanapun, proses pengangkutan 2,2-DCP ke dalam sel *Rhizobium sp.* RC1 masih belum dikenal pasti. Satu gen putatif pengangkutan haloasid, *dehrP*, di dalam *Rhizobium sp.* RC1 dianggap bertanggungjawab bagi proses ini. Kajian in bertujuan untuk menjelaskan fungsi gen ini bagi pengangkutan haloasid ke dalam sel. Untuk mencapai matlamat ini, gen *dehrP* dianalisa menggunakan beberapa alat BLAST dan kemudian disejajar menggunakan perisian T-Coffee dibandingkan dengan protin pengangkutan lain. Protin daripada gen ini, DehrP, telah disimpulkan tergolong di dalam kumpulan protin *Major Facilitator Superfamily* (MFS) dan simpot Metabolit:H^+ (MHS). DehrP didapati mempunyai sembilan heliks transmembrane dengan motif unik MFS di antara heliks kedua dan ketiga, dan heliks kelapan dan kesembilan. DehrP telah didapati berkait rapat dengan protin pengangkut haloasid lain daripada kumpulan MHS, DehP, Deh4p dan Dehp2 dalam *Burkholderia caribensis* MBA4. Model DehrP telah dibina menggunakan Phyre^2 dengan protin pengangkutan XylE dari *Escherichia coli* sebagai model rujukan. DehrP telah dibandingkan dengan XylE untuk menentukan lokasi tapak pengikat proton dan haloasid. Lokasi tapak pengikat proton DehrP terdiri daripada dua residu, Asp36 dan Arg130 manakala andaian tapak pengikat haloasid adalah Glu33, Trp34, Phe37, Arg75, Tyr271 dan Ser402. Bagi mengesahkan andaian tapak pengikat bagi haloasid, residu-residu tersebut telah ditukar kepada alanin dan jujukan protin baru ini diberi nama DehrPa. Model 3D DehrP dan DehrPa telah dikemaskan dengan menggunakan 3Drefine sebagai persediaan simulasi dok menggunakan AutoDock Vina. Simulasi mengedok dilakukan menggunakan empat haloasid (2,2-DCP, MCA, D-2DCP dan L-2DCP). Kesimpulannya, andaian residu tapak pengikat dalam DehrP telah berjaya disahkan melalui keafinan ikatan dan kejituan yang rendah dalam DehrPa. Tanpa diduga, 2,2-DCP didapati masih dapat mengikat kepada tiga residu lain yang tidak dimutasikan dalam DehrPa. Kajian ini mengesahkan tapak pengikat haloasid bagi DehrP oleh kajian terdahulu dengan penemuaan terbaru residu pengikat alternatif khusus bagi 2,2-DCP.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2,2-DCP	-	2,2-dichloropropionic
3D	-	Three dimensional
ABC	-	ATP-Binding Cassette
BLAST	-	Basic Local Alignment Search Tool
bp	-	Base pairs
D-2CP	-	D-2-chloropropanoic acid
Da	-	Dalton
DDM	-	n-dodecyl-β-D-maltoside
ESPript	-	Easy Sequencing in PostScript
FA	-	Fluoroacetate
GLUT	-	Glucose transporter
GRAVY	-	Grand Average of Hydropathy
L-2CP	-	L-2-chloropropanoic acid
MAST	-	Motif Alignment and Search Tool
MBA	-	Monobromoacetate
MCA	-	Monochloroacetic acid
MEGA	-	Molecular Evolutionary Genetics Analysis
MEME	-	Multiple EM for Motif Elicitation
MFS	-	Major Facilitator Superfamily
MHS	-	Metabolite:H^+ Symporter
MSA	-	Multiple Sequence Alignment
NADV	-	Nucleotide Amino acid Derived Visualization
NCBI	-	National Centre for Bioinformatics Information
NMR	-	Nuclear magnetic resonance
OG	-	n-octyl-β-D-glucoside

PCR	-	Polymerase chain reaction
PDB	-	Protein Data Bank
Pfam	-	Protein families
Pgp	-	P-glycoprotein
PSI-BLAST	-	Position-Specific Iterative Basic Local Alignment Search Tool
PSIPRED	-	PSI-BLAST based secondary structure prediction
RMSD	-	Root Mean Square Deviation
sp.	-	Species
TCA	-	Trichloroacetic acid
TCDB	-	Transmembrane Classification Database
THM	-	Trihalomethanes
TM	-	Transmembrane
UCSF	-	University of California, San Francisco
UDS	-	Ultracentrifugation dispersity sedmentation
UniprotKB	-	Universal Protein Resource Knowledgebase

LIST OF SYMBOLS

#	-	Number
α	-	Alpha
β	-	Beta
H^+	-	Hydrogen ion
k	-	Kilo
μ	-	Micro
pH	-	Potential of Hydrogen
pI	-	Isoelectric point

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Halogenated compounds have been used extensively in both industrial and personal use cases (herbicides, cleaning agents, etc.). Halogenated compounds are both made artificially and naturally where the latter is the result of geothermal processes such as volcanic eruptions and forest fires (Gribble, 2003). As halogenated compounds are also a type of carbon source, some microorganisms have taken this to advantage which resulted in using halogenated compounds as a source of energy through biodegradation. Types of halogenated compounds would be haloacids or haloacetate which can be found in everyday uses such as 2,2-dichloropropionic acid (2,2-DCP) or better known as Dalapon®, is widely used as a herbicide and while haloacetate can be found in household cleaners. The amount of pollution that is caused by haloacids and haloacetate is concerning due to their alarming toxic effects towards living organisms that have little to no means of disposing these compounds from their systems (Plewa, Kargalioglu, Vankerk, Minear, & Wagner, 2002).

Bioinformatics tools are one of the main methods of protein analysis in general. By using these tools, information such as protein family and 3D structures can be obtained without having to physically analyse the protein itself. Usually bioinformatics tools go in tandem with preliminary laboratory work to work out the properties of a protein. This is especially useful when analysing a protein that is hard to analyse such as membrane proteins.

The bacterium used for the research is *Rhizobium* sp. RC1 which is a Gram-negative bacterium that was isolated from the soil using 2,2-dichloropropionic acid (2,2-DCP). These bacteria break down haloacids and use it as a source of carbon and is capable of doing this due the dehalogenase enzymes that are present inside the cell (Huyop & Nemati, 2010). It has been predicted that this organism has a specific transport protein that is responsible of transporting haloacids into the cell, but it has not been extensively researched (Tsang & Pang, 2001). The putative transport gene of *Rhizobium* sp. RC1, *dehrP*, is used as the starting point of the research.

Analysis of proteins such as DehrP have brought upon the formation of families that groups these proteins together in a family. The Major Facilitator Superfamily (MFS) was established so that proteins that are responsible for transport inside bacteria are grouped together and is used as reference to log other proteins similar to the members of the family (Pao, Paulsen, & Saier, Jr., 1998). Sub-families were then established under MFS in order to group the proteins based on the distinctive traits of each proteins such as transport system and substrate transported. One f these families is the Metabolite:H^+ Symporter (MHS) family which groups proteins transport metabolites with protons simultaneously into the cell.

Members of the MHS family include Deh4p and Dehp2 in *Burkholderia cepacia* MBA4, XylE in *E. coli* and GlcPse in *Staphylococcus epidermidis*. All of these proteins exhibit the unique characteristics that categorise them as members of the MFS and MHS family of proteins. All of these proteins should possess a unique MFS protein between helices 2 and 3, and helices 8 and 9 and have a unique cytoplasmic loop (Pao, Paulsen, & Saier, Jr., 1998). It is inferred that DehrP inside *Rhizobium* sp. RC1 relates to these proteins based on prior research and it is highly probable that DehrP would also exhibit the same traits.

1.2 Problem Statement

It is well established that the consumption of food or nutrients that are contaminated with haloalkanoic acids (haloacids) could have toxic and carcinogenic effects towards the organism that consumed it. This is also a problem since not all organisms can degrade haloacids in their system. Microorganisms are gradually being extensively used in order to breakdown haloacids in places such as in soil due to their ability to transport haloacids into the cytoplasm. However, only specific microorganisms have ability to transport these compounds into the cell. In order to be able to transport these substrates, the microorganism should possess a certain transport gene to be expressed to produce a specific haloacids transport protein due to haloacids not being a natural substrate for these microorganisms. By expressing these genes, the microorganism is able to produce an alternative transport pathway to accommodate for the haloacids.

The cell used for the research, *Rhizobium sp.* RC1, is known to be able to transport haloacids such as 2,2-DCP into the cell as a carbon source. However, only the process of transporting haloacids into the cell is confirmed. The full mechanism during the process is not yet well documented and extensively researched. By researching the mechanism of the transport process, it can lead to the better understanding of the transport system of the cell itself and can be used to efficiently study haloacid transport in *Rhizobium sp.* RC1 in real life situations.

1.3 Significance of Research

By the means of this research, it is hoped that the structure and functions of the proteins involved in the transport mechanism of haloacids into *Rhizobium sp.* RC1 can be clarified. Research regarding the putative transport protein of *Rhizobium sp.* RC1 have been minimal and the latest research has inferred the structure and proton binding

abilities. In solidifying these hypotheses, we could start to build a better picture of the process of transporting haloacids into *Rhizobium sp.* RC1 operates. By gaining more understanding of the structure and the haloacid transport process of haloacid transporters of *Rhizobium sp.* RC1, it would be useful in gaining better efficiency of using the microorganism to uptake haloacids in order to better degrade haloacids pollutants in designated areas.

1.4 Research Objectives

i) To analyse the amino acid sequence of haloacid transporter (DehrP) from *Rhizobium sp.* RC1 and subsequently determine the classification and family of DehrP.

ii) To determine the 3D structure and the topological arrangements of DehrP in *Rhizobium sp.* RC1 from its primary sequence.

iii) To determine and analyse the substrate and proton binding sites of DehrP.

1.5 Scope of the Research

Analysis of *Rhizobium sp.* RC1 putative transport gene, dehrP, is purely computational and software based. This is because the laboratory work to determine the nucleotide sequence of the putative transport gene in *Rhizobium sp.* RC1 is already documented and stored in an online database. This require downloading the determined nucleotide sequence that was submitted into the GenBank Database and translate it into the appropriate amino acid sequence. Using the information given, the nucleotide sequence would then be used to do further computational analysis using bioinformatics software packages during the research. No further laboratory work will

be done since it has been determined that doing extensive laboratory and physical analysis of the putative transport protein would take a long time and would require specialised tools and materials beforehand. Doing computational analysis is a good alternative that will still yield the results that we have predicted prior to the research.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 403-410.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, L., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, W344-W350.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Reseearch*, W529-W533.

Bailey, T. L., & Elkan, C. (1994). *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* California: AAAI Press.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (pp. 28-36). Menlo Park, California: AAAI Press.

Bailey, T. L., & Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, 48-54.

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.

Benkert, P., Blasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 343-350.

Berry, E. K., Allison, N., & Skinner, A. (1979). Degradation of the Selective Herbicide 2,2=Dichloropropionate (Dalapon) by a Soil Bacterium . *Journal of General Microbiology* , 39-45.

Bhattacharya, D., Nowotny, J., Cao, R., & Cheng, J. (2016). 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Research*, W406-W409.

Biovia, D. S. (2017). *Discovery Studio Modeling Environment.* San Diego: Dassault Systemes.

Blaney, J. (2012). A very short history of structure-based design: How did we get here and where do we need to go? *Journal of Computer-Aided Molecular Design*, 13-14.

Buchan, D., Minneci, F., Nugent, T. C., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, W340-W348.

Bull, R. J., Reckhow, D. A., Rotello, V., Bull, O. M., & Kim, J. (2006). *Use of Toxicological and Chemical Models to Prioritize DBP Research.* Washington: IWA Publishing.

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

Caspari, T., Stadler, R., Sauer, N., & Tanner, W. (1994). Structure-function relationship of the Chlorella glucose/H+ symporter. *The Journal of Biological Chemistry*, 3498-3502.

Chufan, E. E., Kapoor, K., Sim, H.-M., Singh, S., Talele, T. T., Durell, S. R., & Ambudkar, S. V. (2013). Multiple Transport-Active Binding Sites Are Available for a Single Substrate on Human P-Glycoprotein (ABCB1). *PLOS One*, e82463.

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/

Coldiron, B. (1992). Thinning of the ozone layer: Facts and consenquences. *American Academy of Dermatology*.

Colovos, C., & Yeates, T. G. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society*, 147-150.

Combet, C., Blanchet, C., Geourjon, C., & Deleage, G. (2000). NPS@: Network Protein Sequence Analysis. *Trends in Biochemical Sciences*, 147-150.

Consortium, U. (2015). Uniprot: a hub for protein information. *Nucleic Acids Research*, D204-D212.

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, 1188-1190.

Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., . . . Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature*, 734-738.

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, W389-W394.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 783-791.

Fendt, S.-M., Bell, E. L., Keibler, M. A., Olenchock, B. A., Mayers, J. R., Wasylenko, T. M., . . . Vander Heiden, M. G. (2013). Reductive glutamine metabolism is a function of the α-ketoglutarate to citrate ratio in cells. *Nature Communications*, 1-21.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., . . . Mitchell, A. L. (2017). InterPro in 2017 — beyond protein family and domain annotations. *Nucleic Acids Research*, D190-D199.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., . . . Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, D279-D285.

Foloppe, N., & Hubbard, R. (2006). Towards predictive ligand design with free-energy based computational methods? . *Current Medicinal Chemistry*, 3583-3608.

Fonseca, A., Spencer-Martins, I., & van Uden, N. (1991). Transport of lactic acid in Kluyveromyces marxianus: evidence for a monocarboxylate uniport. *Yeast*, 775-780.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyl, 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.

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. In *The Proteomics Protocols Handbook* (pp. 571-607). New Jersey: Humana Press.

Gribble, W. G. (2003). The diversity of naturally produced organohalogens. *Chemosphere*, 289-297.

Gricorescu, A. S., Lapara, T. M., & Hozalski, R. M. (2010, October). BIodegradation of Haloacetic Acids and Potential Applicability to Drinking Water Treatment. *Romanian JOurnal of Biochemistry, 47*(2), 165-177.

Griifth, J. K., Baker, M., Rouch, D. A., Page, M. G., Skurray, R. A., Paulsen, I. T., . . . Henderson, P. J. (1992). Membrane transport proteins: implications of sequence comparisons. *Current Opinion in Cell Biology*, 684-695.

Gururprasad, K., Reddy, B. V., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection,* 155-161.

Gutmann, D. A., Mizohata, E., Newstead, S., Ferrandon, S., Henderson, P. J., van Veen, H. W., & Byrne, B. (2007). A high-throughput method for membrane protein solubility screening: The ultracentrifugation dispersity sedimentation assay. *Protein Science*, 1422-1428.

Haggar, R. J., & Elliott, J. G. (1978). The effects of dalapon and stocking rate on the species composition and animal productivity of a sown sward. *Grass and Forage Science*, 23-33.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 95-98.

Hashimoto, S., Azuma, T., & Otsuki, A. (1998). Distribution, sources, and stability of haloacetic acids in Tokyo Bay, Japan. *Enviromental Toxicology and Chemistry*, 798-805.

Henderson, P. J., & Baldwin, S. A. (2013). This is about the in and the out. *Nature Structural & Molecular Biology*, 654-655.

Henderson, P. J., & Maiden, M. C. (1990). Homologous sugar transport proteins in Escherichia coli and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 391-410.

Hill, K. E., Marchesi, J. R., & Weightman, A. J. (1999, April). Investigation of Two Evolutionarily Unrelated Halocarboxylic Acid Dehalogenase Gene Families. *Journal of Bacteriology, 181*(8), 2535-2547.

Hirai, T., Heymann, J. A., Maloney, P. C., & Subramaniam, S. (2003). Structural Model for 120Helix Transporters Belonging to the Major Facilitator Superfamily. *Journal of Bacteriology*, 1712-1718.

Huyop, F., & Cooper, A. R. (2003). Potential use of dehalogenase D (DehD) from *Rhizobium sp.* for industrial process. *Jurnal Teknologi*, 1-8.

Huyop, F., & Nemati, M. (2010). Properties of dehalogenase from *Rhizobium sp.* RC1. *African Journal of Microbiology* , 2836-2847.

Huyop, F., & Nemati, M. (2010). Properties of dehalogenase from *Rhizobium sp.* RC1. *African Journal of Microbiology*, 2836-2847.

Iancu, C. V., Zamoon, J., Woo, S. B., Aleshin, A., & Choe, J.-y. (2013). Crystal structure of a glucose/H+ symporter and its mechanism of action. *Proceedings of the National Academy of Sciences of the United States of America*, 17862-17867.

Iancu, C. V., Zamoon, J., Woo, S. B., Aleshin, A., & Choe, J.-y. (2013). Crystal structure of a glucose/H+ symporter and its mechanism of action. *Proceedings of the National Academy of Sciences of the United States of America*, 17862-17867.

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, 1895-1898.

Jensen-Marshall, A. E., Paul, N. J., & Brooker, R. J. (1995). The conserved motif, GXXX(D/E)(R/K)XG[X](R/K)(R/K), in hydrophilic loop 2/3 of the lactose permease. *Journal of Molecular Biology*, 16251-16257.

Jing, N. H., Wahab, R. A., Hamdan, S., & Huyop, F. (2010). Cloning and DNA sequence analysis of the haloalkanoic permease uptake gene from *Rhizobium sp.* RC1. *Biotechnology*, 319-325.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 195-202.

Juuti, S., Yrjo, N., Helle, T., & Ruuskanen, J. (1996). Trichloroacetic acid in conifer needles and arboreal lichens in forest environments. *Science of the Total Environment*, 117-124.

Karimian, M., & Ornston, L. N. (1981). Participation of the β-ketoadipate transport system in chemotaxis. *Journal of General Microbiology*, 93-96.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 845-858.

Kumar, S., Stecher, G., & Tamura, K. (2016). Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 1870-1874.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 1547-1549.

Kumar, S., Tsai, C.-J., & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein Engineering, Design and Selection*, 179-191.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 105-132.

Lam, V. M., Daruwalla, K. R., Henderson, P. J., & Jones-Mortimer, M. C. (1980). Proton-Linked D-Xylose Transport in Escherichia coli. *Journal of Bacteriology*, 396-402.

Landau, E. M., & Rosenbusch, J. P. (1996). Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 14532-14535.

Leigh, J. A., Skinner, A. J., & Cooper, R. A. (1986). Isolation and partial characterisation of dehalogenase-deficient mutants of a *Rhizobium sp. FEMS MIcrobiology*, 163-166.

Lewis, J. A., Horswill, A. R., Schwem, B. E., & Escalanter-Semerena, J. C. (2004). The tricarballylate utilization (tcuRABC) genes of Salmonella enterica serovar Typhimurium. *Journal of Bateriology*, 1629-1637.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Scott, M. P., Bretscher, A., & Ploegh, H. (2008). Proteins Interact with Membranes in Three Different Ways. In *Molecular Cell Biology Sixth Edition* (pp. 421-422). Basingstoke: Freeman.

Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., . . . Richardson, D. C. (2003). Structure validation by Clapha

geometry: phi, psi and Cbeta deviation. *Proteins: Structure, Function, and Genetics*, 437-450.

Lundgren, B. R., Villegas-Penaranda, L. R., Harris, J. R., Mottern, A. M., Dunn, D. M., Boddy, C. N., & Nomura, C. T. (2014). Genetic Analysis of the Assimilation of C5-Dicarboxylic Acids in Pseudomonas aeruginosa PAO1. *Journal of Bacteriology*, 2543-2551.

Ly, A., Henderson, J., Lu, A., Culham, D. E., & Wood, J. M. (2004). Osmoregulatory Systems of Escherichia coli: Identification of Betaine-Carnitine-Choline Transporter Family Member BetU and Distributions of betU and trkG among Pathogenic and Nonpathogenic Isolates. *Journal of Bacteriology*, 296-306.

MacMillan, S. V., Alexander, D. A., Culham, D. E., Kunte, H. J., Marshall, E. V., Rochon, D., & Wood, J. M. (1999). The ion coupling and organic substrate specificities of osmoregulatory transporter ProP in Escherichia coli. *Biochimica et Biophysica Acta*, 30-44.

Marcoline, F. V., Bethel, N., Guerriero, C. J., Brodsky, J. L., & Grabe, M. (2015). Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations. *Structure*, 1526-1537.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., & Goodsell, D. S. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 2785-2791.

Musa, M. A., Abdul Wahab, R., & Huyop, F. (2018). Homology modelling and in silico substrate-binding analysis of a *Rhizobium sp.* RC1 haloakanoic acid permease. *Biotechnology & Biotechnological Equipment*, 339-349.

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.

Omasits, U., Ahrens, C. H., Muller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, 884-886.

Omasits, U., Ahrens, C. H., Muller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, 884-886.

Omasits, U., Ahrens, C. H., Muller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visulaization and integration with experimental proteomic data. *Bioinformatics*, 884-886.

Pao, S. S., Paulsen, I. T., & Saier, Jr., M. H. (1998). Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews*, 1-34.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 785-786.

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

Plewa, M. J., Kargalioglu, Y., Vankerk, D., Minear, R. A., & Wagner, E. D. (2002). Mammalian cell cytotoxicity and genotoxicity analysis of drinking water disinfection by-products. *Enviromental and Molecular Mutagenesis* , 134-142.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, 276-277.

Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R., & DeMarini, D. M. (2007). Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: A review and roadmap for research. *Mutation Research/Reviews in Mutation Research*, 178-242.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server . *Nucleic Acids Research*, W320–W324.

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., & Gelfand, M. S. (2002). Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. *Journal of Biological Chemistry*, 48949-48959.

Rost, B. (2001). Protein secondary structure prediction continues to rise. *Journal of Structural Biology*, 204-218.

Saier Jr, M. H. (2003). Tracing pathways of transport protein evolution. *Molecular Micrbiology*, 1145-1156.

Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., & Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Research*, D372-D379.

Saier, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., & Moreno-Hagesieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Research*, D372-D379.

Saint, C. P., & Romas, P. (1996). 4-Methylphthalate catabolism in Burkholderia (Pseudomonas) cepacia Pc701: a gene encoding a phthalate-specific permease forms part of a novel gene cluster. *Microbiology*, 2407-2418.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 406-425.

Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. L., . . . Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein databsae searches with composition0based statistics and other refinements. *Nucleic Acids Research*, 2994-3005.

Seol, W., & Shatkin, A. J. (1992). Site-directed mutants of Escherichia coli α-ketoglutarate permease (KgtP). *Biochemistry*, 3550-3554.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 1-6.

Sievers, F., Wilm, A., Dineen, D., GIbson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 1-6.

Sjodin, A., Hagmar, L., Kasson-Wehler, E., Bjork, J., & Bergman, A. (2000). Influence of the consumption of fatty Baltic Sea fish on plasma levels of halogenated environmental contaminants in Latvian and Swedish men. *Environmental Health Perspectives*, 1035-1041.

Stringfellow, J. M., Cairns, S. S., Cornish, A., & Cooper, R. A. (1997). Haloalkanoate dehalogense II (DehE) of a *Rhizobium sp. European Journal of Biochemistry*, 789-793.

Su, X., Deng, L., Kong, K. F., & Tsang, J. S. (2013). Enhanced Degradation of Haloacid by Heterologous Expression in Related Burkholderia Species. *Biotechnology and Bioengineering*, 2687 - 2697.

Su, X., Kong, K.-F., & Tsang, J. S. (2012, November). Transports of acetate and haloacetate in Burkholderia species MBA4 are operated by distinct systems. *BMC Microbiology, 12*(267).

Su, X., Li, R., Kong, K. F., & Tsang, J. (2016). Transport of haloacids across biological membranes. *Biochimica et Biophysica Acta*, 3061-3070.

Su, X., Li, R., Kong, K. F., & Tsang, J. S. (2016). Transport of haloacids across biological membranes. *Biochimica et Biophysica Acta*, 3061-3070.

Sugihara, J., Smirnova, I., Kasho, V., & Kaback, H. (2011). Sugar recognition by CscB and LacY. *Biochemistry*, 11009-11014.

Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y., & Yan, N. (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, 361-366.

Tamura, K., Nei, M., & Kumar, S. (1987). Prespects for inferring very large phylogenies by using the neighbor-joining method. . *Proceedings of the National Academy of Sciences* , 11030-11035.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 4673-4680.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 455-461.

Tsang, J. H., Sallis, P. J., & Bull, A. T. (1988, September). A monobromoacetate dehalogenease form Pseudomonos cepacia MBA4. *Archieves of Microbiology*, *150*(5), 441-446.

Tsang, J. S., & Pang, B. C. (2001). Mutagenic analsis of conserved residues in dehalogenase IVa of *Burkholderia cepacia* MBA4. *FEMS Microbiology Reviews*, 135-140.

Tsang, J. S., Hardman, D. J., & Bull, A. T. (1988). *Cloning and expression of 2-haloalkanoic acid dehalogenase of Pseudomonoas cepacia MBA4 in Escherichia coli and in Pseudomonas pulida.* (S. T. Chang, K. Y. Chan, & N. Y. Woo, Eds.) Hong Kong: Chinese University Press.

Tsirigos, K. D., Peters, C., Shu, N., Kall, L., & Elofsson, A. (2015). The TOPCONS web server for combined membrane protein topology and signal peptide prediction. *Nucleic Acids Research*, W401-W407.

Ubarretxena-Belandia, I., & Stokes, D. L. (2010). Present and future of membrane protein structure determination by electron crystallography. *Advances in Protein Chemistry and Structural Biology*, 33-60.

van der Rest, M. E., Schwarz, E., Oesterhelt, D., & Konings, W. N. (1990). DNA sequence of a citrate carrier of Klebsiella pneumoniae. *European Journal of Biochemistry*, 401-407.

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.

Whipp, M. J., Camakaris, H., & Pittard, A. J. (1998). Cloning and analysis of the shiA gene, which encodes the shikimate transport system of escherichia coli K-12. *Gene*, 185-192.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, W407-W410.

Wisedchaisri, G., Park, M.-s., Iadanza, M. G., Zheng, H., & Gonen, T. (2014). Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nature Communications*, 1-10.

Wu, C., & Schaum, J. (2001). *Sources, Emission and Exposure for Trichloroethylene (TCE) and Related Chemicals.* Washington: DIANE Publishing.

Yu, M., Yun-WIng, F., Chung, W. Y., & Tsang, J. S. (2007, AUgust). Isolation and Characterization of a Novel Haloacid Permease from *Burkholderia cepacia* MBA4. *Applied and Environmental Micorbiology, 73*(15), 4874-4880.

Zheng, H., Wisedchaisri, G., & Gonen, T. (2013). Crystal structure of a nitrate/nitirite exchanger. *Nature*, 361-366.