

EVOLUTIONARY TRACE ANALYSIS OF  
A ZINC METALLOENZYME – ALCOHOL DEHYDROGENASE

SAKINAH JAMIL

FACULTY OF SCIENCE  
UNIVERSITY OF MALAYA  
KUALA LUMPUR

2012

EVOLUTIONARY TRACE ANALYSIS OF  
A ZINC METALLOENZYME – ALCOHOL DEHYDROGENASE

SAKINAH JAMIL  
(SGJ080011)

SUBMITTED TO  
INSTITUTE OF BIOLOGICAL SCIENCES  
FACULTY OF SCIENCE  
UNIVERSITY OF MALAYA

IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR  
THE DEGREE OF MASTER OF BIOINFORMATICS

2012

## **ACKNOWLEDGEMENTS**

Although this is a personal project paper, the success of this research has been contributed by the pool of energy and ideas from various organizations. I wish to dedicate this research especially to my parents Mr Abdul Jamil Md Ali and Madam Juriah Ismail who have been supportive in giving me the motivation and drive to complete this research.

I convey my deepest appreciation to my supervisor Dr. Saharuddin Mohamad who never gave up on me and was always there to assist me in my research. In addition to that, I personally thank all my lecturers for coaching me throughout my learning escapade in the Bionformatics and Biocomputing Division, Institute of Biological Sciences, University of Malaya.

A heartfelt thank you to the laboratory staff who have given their helping hands in this research. I convey my deepest gratitude to my course mates for their out of the world guidance and knowledge sharing sessions we had. Finally, I wish to thank everyone else including family and friends for their modest assistance.

## ABSTRACT

Molecular evolution is the study of the evolutionary patterns and processes of nucleic acids and proteins to understand the phylogenetics, population genetics, biogeography, and other areas of research at the DNA, RNA and protein levels of the biological organization. Evolutionary trace method is a technique to identify evolutionary important residues and their positions in a group of protein alignment. In this study, the zinc metalloenzyme alcohol dehydrogenases (ADH) of the eukaryote *Mus musculus* (mouse) was used as the sequence of interest against 48 other bacteria species. BLAST was used to identify the 48 bacteria species with the closest ADH sequences to the *Mus musculus* ADH. Multiple sequences alignment was done in order to construct the phylogenetic tree to view the evolutionary relationship of the sequences. Partitions on the Cladogram were created to trace back to where the organisms branch off pointing to a common ancestor. Each partition was then split into smaller groups to identify similar species in the family. 29.8% of the residues in the first partition were fully conserved whilst another 47.8% of the active site residues were fully conserved.

## ABSTRAK

Evolusi molekular ialah kajian corak-corak bersifat evolusi dan pemahaman proses-proses asid nukleik dan protein melalui kajian filogenetik, genetik populasi, biogeografi dan lain-lain penyelidikan di tahap DNA, RNA dan protein organisasi biologi. Kaedah jejak evolusi (*Evolutionary Trace Method*) ialah satu teknik mengenal pasti jejak-jejak penting bersifat evolusi dan kedudukan amino-amino asid dalam jajaran protein. Dalam kajian ini, zink enzim-metal alcohol dehydrogenase (ADH) eukariot *Mus musculus* (tikus) digunakan sebagai jujukan rujukan kepada 48 spesis-spesis bakteria. Program BLAST digunakan bagi mengenal pasti 48 spesis-spesis bakteria yang mempunyai jujukan-jujukan ADH terhampir dengan *Mus musculus*. Jujukan berbilang penjajaran (*Multiple sequences alignment*) dilakukan bagi membina pokok filogenetik untuk melihat hubungan evolusi antara jujukan-jujukan. Segmentasi melalui Kladogram diwujudkan bagi mengesan nenek moyang asal yang sama organisma-organisma sebelum berpecah kepada ranting. Setiap segmen dibahagikan kepada kumpulan-kumpulan yang lebih kecil bagi mengenal pasti spesis-spesis yang serupa dalam keluarga itu. 29.8% residu dalam segmen pertama adalah sepenuhnya terpelihara (*fully conserved*) manakala satu lagi 47.8% residu tapak aktif sepenuhnya terpelihara (*fully conserved*).

**UNIVERSITI MALAYA**  
**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: SAKINAH BINTI ABD JAMIL (850123-14-5180)

Registration/Matric No: SGJ080011

Name of Degree: MASTER OF BIOINFORMATICS

TITLE: “EVOLUTIONARY TRACE ANALYSIS OF A ZINC  
METALLOENZYME – ALCOHOL DEHYDROGENASE”

Field of Study: BIOINFORMATICS

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature  
Subscribed and solemnly declared before,

Date:

Witness’s Signature  
Name:  
Designation:

Date:

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	II
ABSTRACT .....	III
ABSTRAK .....	IV
ABSTRAK .....	IV
LIST OF TABLES .....	IX
LIST OF FIGURES .....	X
<b>1 CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
1.1 PURPOSE OF STUDY .....	1
1.2 MOLECULAR EVOLUTION.....	1
1.3 METALLOENZYME .....	2
1.4 CONSERVED SITE .....	2
1.5 OBJECTIVES OF STUDY .....	2
<b>2 CHAPTER 2: LITERATURE REVIEW.....</b>	<b>3</b>
2.1 INTRODUCTION TO ALCOHOL DEHYDROGENASE.....	3
2.2 MECHANISM OF ALCOHOL DEHYDROGENASE.....	4
2.3 THE ALCOHOL DEHYDROGENESE ACTIVE SITE.....	6
<b>3 CHAPTER 3: MATERIALS AND METHODS .....</b>	<b>9</b>

3.1	MATERIALS.....	9
3.1.1	Hardware.....	9
3.1.2	Software.....	10
3.2	METHODS.....	12
3.3	TASK 1: DETERMINING THE PHYLOGENETIC RELATIONSHIP OF ALCOHOL DEHYDROGENASE.....	15
3.3.1	Identification of Alcohol Dehydrogenase Containing Protein.....	15
3.3.2	Verifying 1E3I FASTA Sequence from PDB with UniProt.....	15
3.3.3	BLAST Search.....	15
3.3.4	Collecting FASTA Format Sequences of Bacteria.....	16
3.3.5	Multiple Sequence Alignment.....	16
3.3.6	Construction of Phylogenetic Tree.....	17
3.4	TASK 2: IDENTIFYING CONSENSUS SEQUENCES AND TRACE ANALYSIS.....	19
3.4.1	Partitioning the Cladogram and Species Grouping.....	19
3.4.2	Single Group Multiple Sequences Alignment.....	19
3.4.1	Identification of Consensus Sequences and Evolutionary Trace.....	21
3.5	TASK 3: PROTEIN MAPPING.....	21
3.5.1	Secondary Structure of 1E3I.....	21
3.5.2	The Zinc Atom, NAD Ligand and CXF Ligand with Surrounding Residues	23
3.5.3	Comparison of Trace Analysis, Rasmol and Active Sites Identified from PDB SUM.....	24
3.5.4	Mapping of Conserved Residues on the Protein Structure.....	24
<b>4</b>	<b>CHAPTER 4: RESULTS.....</b>	<b>25</b>
4.1	TASK 1: DETERMINING THE PHYLOGENETIC RELATIONSHIP OF ALCOHOL DEHYDROGENASE.....	25
4.1.1	Identification of Alcohol Dehydrogenase Containing Protein.....	25
4.1.2	Verifying 1E3I FASTA Sequence from PDB with UniProt.....	25
4.1.3	BLAST Search.....	28
4.1.4	Collecting FASTA Format Sequences of Bacteria.....	28
4.1.5	Multiple Sequence Alignment.....	32
4.1.6	Construction of Phylogenetic Tree.....	32
4.2	TASK 2: IDENTIFICATION OF CONSENSUS SEQUENCES AND EVOLUTIONARY TRACE.....	35
4.2.1	Partitioning the Cladogram and Species Grouping.....	35
4.2.2	Single Group Multiple Sequences Alignment.....	39

4.2.3	Identification of Consensus Sequences and Evolutionary Trace .....	39
4.3	TASK 3: PROTEIN MAPPING .....	42
4.3.1	Secondary Structure of 1E3I .....	42
4.3.2	The Zinc Atom, NAD Ligand and CXF Ligand with Surrounding Residues 42	
4.3.3	Comparison of Trace Analysis, Rasmol and Active Sites Identified from PDB Sum.....	45
4.3.4	Mapping of Conserved Residues on the Protein Structure .....	48
<b>5</b>	<b>CHAPTER 5: DISCUSSION .....</b>	<b>50</b>
5.1	TASK 1: DETERMINING THE PHYLOGENETIC RELATIONSHIP OF ALCOHOL DEHYDROGENASE.....	50
5.2	TASK 2: IDENTIFICATION OF CONSENSUS SEQUENCES AND EVOLUTIONARY TRACE .....	54
5.3	TASK 3: PROTEIN MAPPING .....	56
<b>6</b>	<b>CHAPTER 6: SUMMARY.....</b>	<b>59</b>
	<b>REFERENCES.....</b>	<b>61</b>
	<b>APPENDICES .....</b>	<b>1</b>

## LIST OF TABLES

Table 2.1: The Human Alcohol Dehydrogenase (ADH) Genes and Proteins.....	5
Table 3.1: The Software being Used.....	10
Table 4.1: BLAST Results of Selected Similar Bacterial Sequences. The Mus musculus Query Sequence is Highlighted in Purple at the Top of the Table. ....	29
Table 4.2: Comparison of P0 Trace Analysis Conserved Residues, Rasmol Analysis and PDB SUM Active Site Residues. Green Shows Amino Acid Residues that Exist in Both Rasmol and are Evolutionary Trace Conserved residues, Blue shows the Amino Acid are PDB SUM Active Site Residues and Evolutionary Trace Conserved Residues, and Yellow Shows the Amino Acid Residues are Found in Rasmol, are PDB SUM Active Residues and Evolutionary Trace Conserved Residues.....	46
Table 4.3: Comparison of P4 Trace Analysis Conserved Residues, Rasmol Analysis and PDB SUM Active Site Residues. Green Shows Amino Acid Residues that Exist in Both Rasmol and are Evolutionary Trace Conserved residues, Blue shows the Amino Acid are PDB SUM Active Site Residues and Evolutionary Trace Conserved Residues, and Yellow Shows the Amino Acid Residues are Found in Rasmol, are PDB SUM Active Residues and Evolutionary Trace Conserved Residues.....	47

## LIST OF FIGURES

Figure 3.1: The Clustal Omega Multiple Sequence Alignment Web Based Program on EMBL-EBI.....	11
Figure 3.2: The Flowchart of the Methodology.....	14
Figure 3.3: Standard Protein BLAST Suite.....	18
Figure 3.4: The ClustalW2 Multiple Sequence Alignment Program on EMBL-EBI Server. ....	20
Figure 3.5: The Rasmol Molecular Visualization Tool. ....	22
Figure 4.1: Secondary Structure of 1E3I Viewed using Jmol. Pink represents the Alpha Helix and Yellow represents the Beta Sheets.....	26
Figure 4.2: FASTA Sequence of 1E3I. ....	27
Figure 4.3: Sample FASTA of Some of the 48 Selected Similar Bacterial Sequences. ....	31
Figure 4.4: Pre Alignment (top) and Post Multiple Sequences Alignment (bottom) using ClustalX.....	33
Figure 4.5: Rectangular Cladogram using PhyloDraw. The <i>Mus musculus</i> Query Sequence is Highlighted in Green.....	34
Figure 4.6: Rectangular Cladogram Partitioning.The <i>Mus musculus</i> Query Sequence is Highlighted in Green. ....	37
Figure 4.7: Rectangular Cladogram Grouping from the Partitions. The Stars show the Groupings from the Partitions.....	38
Figure 4.8: Sample Multiple Sequences Alignment of Partition P1 by Group.....	40
Figure 4.9: Consensus Sequences for Each of the Partitions.....	41

Figure 4.10: Secondary Structure of 1E3I with Zinc Atoms, NAD Ligand and CXF Ligand with Surrounding Residues on Consecutive 90° Rotations about the X-axis.	43
Figure 4.11: Zinc Atoms, NAD Ligand and CXF Ligand with Surrounding Residues. ....	44
Figure 4.12: Conserved Residues Mapped on the Protein Surface on Consecutive 90° Rotations About the X-axis. ....	49
Figure 5.1: Phylogram as a Result of Multiple Sequence Alignment in ClustalW. The Mus musculus Query Sequence is Highlighted in Green. ....	52
Figure 5.2: Cladogram as a Result of Multiple Sequence Alignment in ClustalW. The Mus musculus Query Sequence is Highlighted in Green. ....	53
Figure 5.3: Sample Section of Conserved Amino Acid Residues Clusters. ....	57
Figure 5.4: Amino Acid Conserved Residues Form Clusters on the Protein Surface. ....	58