UNIVERSITI PUTRA MALAYSIA

A MOLECULAR MODELLING APPROACH FOR DESIGNING A NOVEL SEMISYNTHETIC METALLOENZYME BASED ON THERMOLYSIN

AHMAD HANIFF JAAFAR

FS 2006 22

A MOLECULAR MODELLING APPROACH FOR DESIGNING A NOVEL SEMISYNTHETIC METALLOENZYME BASED ON THERMOLYSIN

AHMAD HANIFF JAAFAR

MASTER OF SCIENCE
UNIVERSITI PUTRA MALAYSIA

2006

Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the Degree of Master of Science.

A Molecular Modelling Approach for Designing a Novel Semisynthetic Metalloenzyme Based on Thermolysin

By

AHMAD HANIFF JAAFAR

July 2006

Chairman : Assoc. Prof. Dr. Mohd Basyaruddin Abdul Rahman

Faculty : Sciences

Current computational chemistry tools were used to solve the problem of screening for the best conformation of potential protein-ligand-metal complex in designing a novel semisynthetic metalloenzyme. The computational tools used were Computational Atlas Topography of Protein (CASTp), a sophisticated molecular modeling environment InsightII, a conventional drug-docking algorithm Autodock 3.05 and a schematic diagram for protein-ligand interactions for a given PDB file LIGPLOT. Overall 48 protein pockets on the thermolysin structure were measured using CASTp and the four biggest pockets based on their number of residues and surface area were identified to be suitables site for the modification. Ten different sizes and multifunctional groups of chemical ligands were studied for their thermodynamic valuation using the AutoDock 3.05 program.

For further modification, phosphoethanolamaine (PSE), phenylalanine (PHE), phenylacetic acid (PAC) and phenanthroline (PHN) were chosen as they possessed the lowest docking energy of -8.49, -8.34, -7.33 and -7.06 kcal/mol, respectively. Non-covalent interactions included hydrogen bonding and hydrophobic interaction between the ligands and the thermolysin were determined using CASTp. The result showed that larger ligands with multifunctional groups such as PSE and PHE showed higher number interactions compared to the smaller ligands. In terms of specific pockets for the modification, different protein-ligand complexes showed different suitable pockets; complex of thermolysin and PSE ligand at pocket 45, complex of thermolysin and PAC ligand at pocket 48 and both complexes of thermolysin with PHE and PHN ligands at pocket 45, respectively. To verify the final metal ion orientation, three procedures were conducted to narrow down the number of possible conformations for the modification. From four tested metal ions (Ca^{2+}, Mg^{2+}, Fe^{2+} and Zn^{2+}), Ca^{2+} was identified to be the most favorable metal ion for the modification. It had orientated within an allowed geometry in all tested protein ligand complexes. Meanwhile, both Mg^{2+} and Fe^{2+} were identified as favorable metal ions in KEI-PSE and KEI-PAC complexes, respectively. Zn^{2+} however, showed non favorable docking in all tested complexes due to improper parameterized file for zinc ion in AutoDock.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains.

Pendekatan Molukular Modeling ke Atas Thermolysin Untuk Merekabentuk Semisintetik Metalloenzim Yang Baru

Oleh

AHMAD HANIFF BIN JAAFAR

Julai 2006

Pengerusi : Prof. Madya Dr. Mohd Basyaruddin Abdul Rahman

Fakulti : Sains

Perisian kimia pengkomputeran terkini digunakan untuk menyelesaikan masalah saringan bagi menentukan potensi kedudukan protin-ligand yang terbaik untuk mencipta semisintetik metaloenzim yang baru. Kajian ini melibatkan penggunaan beberapa perisian komputer termasuklah 'Computational Atlas Topography of Protein' atau CASTP, Insight II, AutoDock 3.05 dan LIGPLOT. Satu siri protin dari struktur thermolysin telah dikenalpasti melalui CASTp dan empat poket yang terbesar dan amino acid yang terlibat dipilih sebagai poket yang sesuai di dalam kajian ini. 10 ligand yang berbeza dari segi saiz dan kumpulan berfungsi telah di kaji melalui kajian termodinamik menggunakan program AutoDock 3.05.

Untuk modifikasi selanjutnya, phosphoethanolamine (PSE), phenylalanine (PHE), phenylacetic acid (PAC) dan phenanthroline (PHN) telah dipilih berdasarkan kepada nilai E_{docked} yang terendah yang dicatatkan iaitu -8.49, -8.34, -7.33 dan -7.06 kcal/mol.. Interaksi non-covalen seperti ikatan hidrogen dan interaksi hidrophobik yang dinilai menggunakan program LIGPLOT menunjukan ligand yang lebih besar dan lebih fleksibel seperti PSE dan PHE menghasilkan lebih banyak interaksi sekaligus menyumbang kepada kestabilan percantuman. Kajian dari segi poket protein yang sesuai menunjukan hasil yang berlainan untuk setiap kompleks seperti berikut; kompleks protin dan ligand PSE di poket 45, komplex protin dan ligand PAC di poket 48 manakala kedua-dua kompleks protein-ligand PHN dan PSE di poket 47. Untuk mengesahkan keputusan ion logam yang sesuai, tiga protokol pemilihan telah dijalankan untuk menyaring konformasi yang terbaik untuk modifikasi ini. Daripada empat ion logam yang dianalisis (Ca^{2+}, Mg^{2+}, Fe^{2+} and Zn^{2+}), Ca^{2+} telah dikenalpasti sebagai ion logam yang paling sesuai. Ianya telah menunjukan orientasi yang sesuai terhadap kesemua protein kompleks yang diuji. Sementara itu, Mg^{2+} dan Fe^{2+} menunjukan orientasi yang sesuai di kompleks KEI-PSE dan KEI-PAC sahaja. Manakala Zn^{2+}, tidak menunjukan orientasi yang sesuai untuk semua protein kompleks yang diuji dan ini berkemungkinan disebabkan parameter untuk ion logam ini tidak tepat di dalam AutoDock.

ACKNOWLEDGEMENTS

APPROVAL SHEET 1

I certify that an Examination Committee met on 7th of November 2006 to conduct the final examination of Ahmad Haniff Jaafar on his Master of Science thesis entitled "*In Silico* Protein Engineering: A Molecular Modelling Approach for Designing a Novel Semisynthetic Metalloenzyme Based on Thermolysin" in accordance with Universiti Pertanian Malaysia (Higher degree) Act 1980 and Universiti Pertanian Malaysia (Higher degree) Regulations 1981. The committee recommends that the candidate be awarded the relevant degree.

Members of the Examination Committee are as follows:

Abdul Halim Abdullah, PhD
Associate Professor,
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Karen Badri , PhD
Professor,
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Intan Safinar Ismail, PhD
Lecturer,
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Sharifuddin M. Zin, PhD
Associate Professor,
Faculty of Science
Universiti Malaya
(External Examiner)

HASANAH MOHD GHAZALI , PhD.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

APPROVAL SHEET 2

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

MOHD BASYARUDDIN ABDUL RAHMAN, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

ABU BAKAR SALLEH, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

MAHIRAN BASRI, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

_____
AINI IDERIS, PhD
Professor/Dean
Schoolof Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION SHEET

I hereby declare that the thesis is based on my original work except for quotations and citations, which have been duly acknowledge. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

_____

AHMAD HANIFF JAAFAR

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ΔGbinding	Free energy of binding
AutoDock	Automated docking of flexible ligands to receptors
CASTp	Computer atlas topography of protein
LIGPLOT	Schematic diagram for protein-ligand interactions
PDB	Protein data bank
E_{docked}	Docked energy
kcal/mol	kilocalorie per mol
RMSd	root mean square deviation (s)
PSE	Phosphoethanolamine
PHN	Phenanthroline
PHE	Phenylacetic acid
PAC	Phenylalanine
PBZ	P-aminobenzamidine
BEN	Benzamidine
ETA	Ethanolamine
EOH	Ethanol
PDO	Propandiol
POL	Propanol
ALA (A)	Alanine
ARG (R)	Arginine
ASN (N)	Asparagine

ASP (D)	Aspartic Acid
CYS (C)	Cysteine
GLU (Q)	Glutamic acid
GLY (G)	Glysine
HIS (H)	Histidine
ILE (I)	Isoleucine
LEU (L)	Leusine
LYS (K)	Lysine
MET (M)	Methionine
PHE (F)	Phenylalanine
PRO (P)	Proline
SER (S)	Serine
THR (T)	Threonine
TRP (W)	Tryptophan
TYR (Y)	Tyrosine
VAL (V)	Valine

CHAPTER I

INTRODUCTION

Enzymes are mainly biomolecular proteins which are able to catalyse chemical reactions. For many years, researchers have been planning to utilise the diverse chemical reactions driven by enzymes in biotechnological industries. One such industrial application known as bioprocessing, aims to exploit enzymes rather than chemicals as the catalysts that are part of many industrial processes. The development of such enzymatic tools, however, requires a detailed structural and chemical understanding of the enzyme (Haki and Rakshit, 2004).

Enzymes perform chemical reactions with high specificity and rate enhancement in aqueous media at ambient temperature and neutral pH. These features have made these biocatalysts attractive for a variety of purposes in pharmaceuticals, fine chemicals, cosmetics and bio-related industries. However, the usage of natural enzymes are restricted by their inherent specificity. To circumvent this limitation, the development of artificial enzymes has received considerable attention. One approach for the design of new enzymes is to modify a known enzyme at a defined site with a cofactor or new functional group to create a new generation of catalyst enzymes (Davies and Distefano, 1997).

Metalloenzymes are proteins that function as enzymes and contain metals that are tightly attached and always isolated with the protein (Davies *et al.*, 1999). The metal main function is to serve in electron transfer as electrophiles and nucleophiles. The electrostatic environment in the active sites is the major factor that guides the substrate to the binding site in the correct position. Metal ions can contribute to a positive result in this process, often binding groups in a stereochemically rigid manner, thereby helping to control and enhance the activity of the enzyme. The importance of metallobiomolecules in biological systems to the environmental, medical, pharmaceutical, agricultural and biotechnological industries is widespread and still rapidly growing especially over the past decade.

Understanding the structural and functional significance of these metal sites requires a specialised array of sophisticated instrumentation and techniques, as well as the expertise to use them. It is only through a detailed understanding of structure and function that enzymes can be selected or redesigned to perform industrially relevant catalysis (Kazlauskas, 2000). By determining the biomolecular structure of these enzymes at atomic resolution, we can try to understand the fundamental basis of the protein's enzymatic activity and its stability under various solution conditions *e.g.* high temperature, high salt concentration, extreme pH and with organic solvents.

Our research is to develop a new type of enzymatic catalyst that is based on the binding of ligand and metal on the enzyme to produce a semisynthetic metalloenzyme. We are endeavouring to understand the nature of specific protein-ligand interactions through a structural prediction by combinatorial computational chemistry, and molecular modelling. Consequently we are exploring protein-ligand-metal interactions at an atomic level and endeavour to design novel protein-ligand-metal complexes in order to improve their novel properties for be use in bio-based and biotechnology related industries.

Objectives

1. To develop a new method for designing a semisynthetic metalloezyme with novel characteristic by using current computational modeling tools.

2. To screen for favorable pockets in thermolysin and intermediate ligands for modification.

3. To study the non-covalent interactions within the semisynthetic metalloenzyme complex through computer-aided molecular modeling.

4. To identify the most favorable metal ions for the modification.

CHAPTER II

LITERATURE REVIEW

Enzymes

Enzymes are one example of natural polymers functioning as biological catalysts. They speed up chemical reactions, often very dramatically. In nature, enzymes evolved to be extremely specific catalysts. They have special catalytic sites which are specifically designed for single reactant. They also operate within defined temperature and pH ranges and are often easily destroyed by chemical reactions and heat. The reactant substances upon which an enzyme acts are termed the substrates (Figure 1). The substances produced as a result of the reaction are the products. Enzyme-controlled reactions are mostly reversible and involve the formation of an intermediate enzyme-substrate complex. (http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/E/Enzymes.html /Enzymes)

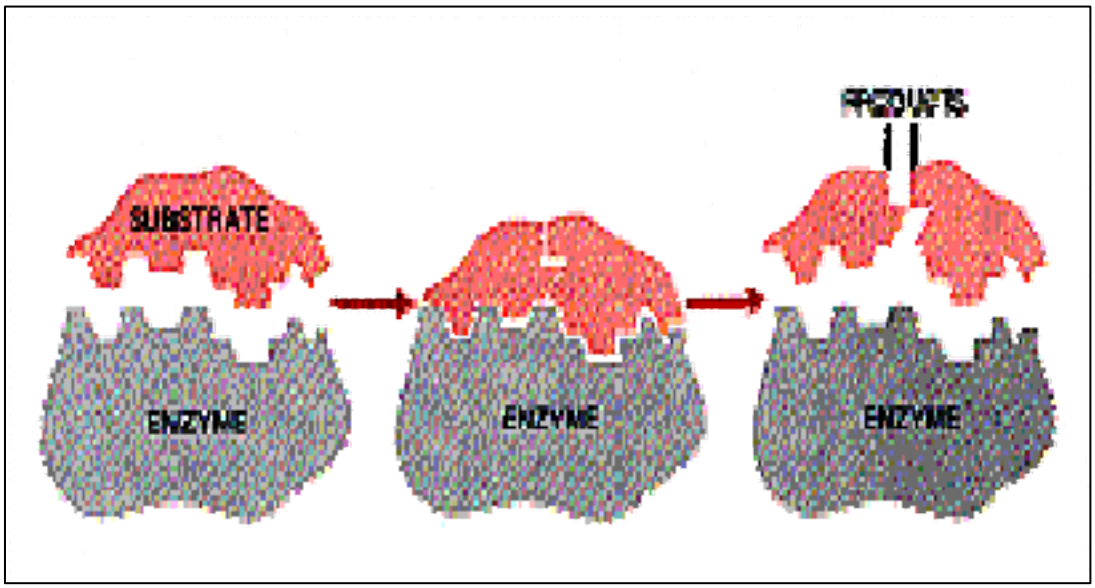

Figure 1: Visualization of enzymatic catalysis reaction (http://users.rcn.com).

Enzyme Classifications

Traditionally, enzymes were simply assigned names by the investigator who discovered the enzyme. As knowledge expanded, systems of enzyme classification became more comprehensive and complex. As summarized in Table 1 enzymes are grouped into six functional classes by the International Union of Biochemists (I.U.B.). (Medical Biochemistry Page: mking@medicine.indstate.edu*)*

Table 1**:** Enzyme classification and biochemical properties.

Numbers	**Classification**	**Biochemical Properties**
1	Oxidoreductases	Act on many chemical groupings to add or remove hydrogen atoms. e.g.- Glucose oxides.
2	Transferases	Transfer functional groups between donor and acceptor molecules. Kinases are specialized transferases that regulate metabolism by transferring phosphate from ATP to other molecules. e.g.- Glucokinase.
3	Hydrolases	Add water across a bond, hydrolyzing it. e.g.- Alpha-amylase.
4	Lyases	Add water, ammonia or carbon dioxide across double bonds, or remove these elements to produce double bonds. e.g.- Pectate lyase.
5	Isomerases	Carry out many kinds of isomerization: L to D isomerizations, mutase reactions (shifts of chemical groups) and others. e.g.- Glucose (xylose) isomerase.
6	Ligases	Catalyze reactions in which two chemical groups are joined (or ligated) with the use of energy from ATP. e.g.- DNA ligase.

Proteases

Proteases refer to a group of enzymes whose catalytic function is to hydrolyze (break down) peptide bonds of proteins. They are also called proteolytic enzymes or proteinases. They differ in their ability to hydrolyze various peptide bonds. Each type of protease break specific kind of peptide bond. Based on the functional group present at the active site, proteases are further classified into four prominent groups, e.g., serine proteases, aspartic proteases, cysteine proteases, and metalloproteases (Hartley, 1960).

Serine proteases

Serine proteases are characterized by the presence of a serine group in their active sites. They are numerous and widespread among viruses and bacteria. Serine proteases are generally active at neutral and alkaline pH, with optimal pH between 7 and 11. They have broad substrate specificity including esterolytic and amidase activity. Their molecular masses range between 18-35 kDa. The isoelectric points for serine protease are generally between pH 4 and 6. (Rao *et al., 1998*)

Cysteine Proteases

Cysteine proteases are widely distributed in all living organisms. Cysteine proteases are small proteins with molecular weight range from 20 000-35 000 KDa and most of them have neutral pH optima. They occur in both prokaryotes and eukaryotes such as bacteria, parasites, plants (papain is one of the well-characterized cysteine proteases from the latex of *carica papaya*), invertebrates, and vertebrates (Berti and Store, 1995). In mammals, the major cysteine proteinases are the lysosomal cathepsins. They are involved in many physiological processes such as protein degradation (Kirschke *et al.*,