THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Biochemistry)

DEGREE

Biochemistry	Biochemistry
FIELD	**DEPARTMENT**

TITLE: Expression of a Recombinant Mulberry Leaf Lectin in Bacterial Host and Computational Structure Analysis

NAME: Miss Tuanta Getsang

THIS THESIS HAS BEEN ACCEPTED BY

..... **THESIS ADVISOR**

(Assistant Professor Amornrat Promboon, Ph.D.)

..... **COMMITTEE MEMBER**

(Associate Professor Sunanta Ratanapo, Ph.D.)

..... **COMMITTEE MEMBER**

(Mr. Kiattawee Choowongkomon, Ph.D.)

..... **COMMITTEE MEMBER**

(Assistant Professor Lertluk Ngernsiri, Ph.D.)

..... **DEPARTMENT HEAD**

(Associate Professor Sunanta Ratanapo, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON .....

..... **DEAN**

(Associate Professor Vinai Artkongharn, M.A.)

EXPRESSION OF A RECOMBINANT MULBERRY LEAF LECTIN IN BACTERIAL HOST AND COMPUTATIONAL STRUCTURE ANALYSIS

TUANTA GETSANG

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Biochemistry)
Graduate School, Kasetsart University
2007

Tuanta Getsang 2007: Expression of a Recombinant Mulberry Leaf Lectin in Bacterial Host and Computational Structure Analysis. Master of Science (Biochemistry), Major Field: Biochemistry, Department of Biochemistry. Thesis Advisor: Assistant Professor Amornrat Promboon, Ph.D. 92 pages.

N-glycolylneuraminic acid (Neu5Gc) specific lectins, named MLL1 were previously purified, characterized and full-length cDNA were successfully cloned using RACEs strategy. In this thesis, the coding sequence of the mature MLL1 were subcloned, sequenced and expressed in *Escherichia coli* BL21 (DE3). Moreover, this study is focused on homology modeling and the interaction between MLL1 and Neu5Gc was studied by molecular docking simulation using AutoDock 3.0.5 program. The resulting, in order to obtain the high amount of soluble recombinant protein, the recombinant cells were induced with 0.5 mM IPTG and 0.4% glucose at 25 °C. The recombinant MLL showed an apparent molecular mass of about 20 kDa in Tricine-SDS-PAGE analysis, while calculated mass by computing program is 25 kDa. Purified MLL was obtained from DEAE-ion exchange and Ni^{2+}-NTA affinity column chromatography. This protein did not show any hemagglutination activity. The result from computational analysis showed that MLL preferred to bind with Neu5Gc a sialic acid derivative better than GalNac and Gal, respectively. The result is in good agreement with carbohydrate-binding specificity of the mulberry leaf lectins. Serine and Isoleucine in binding site played an important role in sugar binding specific binding. In molecular docking, Gly-6 forms strong H-bond with glycolyl group of Neu5Gc. The Ile- 84 and Trp-128 contributed hydrophobic interactions to the sugar. Serine in binding site of MLL may be may be played an important role in the recognition of the sialic acid carboxylate. Additionally, the electrostatic distribution on molecular surface of MLL is neutral nearly positive charge. These reasons may be the important recognition processes for the strong affinity and specificity of MLL for acidic sugar as Neu5Gc a sialic acid derivative.

_____ _____ ___ / ___ / ___

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

My thanks are extended to the Department of Biochemistry and the Department of Genetics, Faculty of Science, Kasetsart University for providing laboratories for the experiments.

I am very grateful to express my deeply indebted and special appreciation to my major subject committee Assistant Professor Amornrat Promboon for her great help, valuable suggestion, comment, expert guidance, intensive supervision and solving problems in laboratory.

I would like to express my special sincere gratitude and appreciation to my dissertation advisor Assistant Professor Associate Professor Sunanta Ratanapo and Assistant Professor Lertluk Ngernsiri for their kindness, valuable guidance, attentive interest, understanding and support throughout the course of this research which enable me to complete this dissertation and this thesis could not have been completed without his advice and the supply of computation analysis on my research, Dr. Kiattawee Choowongkomon

My appreciation is also expressed to staff, undergraduate and graduate students in the Department of Biochemistry, Faculty of Science, Kasetsart University, especially all member in 7214 and 7216 laboratories for their great helps and wonderful friendship.

Most of all, I would like to express my extremely grateful feeling to member of my family for their love, understanding, constant encouragement and care which have enable me to carry out this study successfully.

Tuanta Getsang

April, 2007

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

G-1-P	Glucose-1-phosphate
G-6-P	Glucose-6-phosphate
H_2O_2	Hydrogen peroxide
PCR	Polymerase chain Reaction
bp	base pair
ºC	degree Celsius
Da	Dalton
dATP	deoxyadenosine-5′-triphosphate
dCTP	deoxycytosine-5′-triphosphate
dGTP	deoxyguanosine-5′-triphosphate
dTTP	deoxythymidine-5′-triphosphate
dNTP	deoxynucleotide-5′-triphosphate
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetic acid
EtBr	ethidium bromide
hr	hour
IPTG	isopropyl thiogalactoside
kb	kilobase pair
LB	Luria-Bertani medium
µl	microlitre
µg	microgram
µM	micromolar
mg	milligram
min	minute
ml	millilitre
mM	millimolar
MW	molecular weight
OD	optical density

LIST OF ABBREVIATIONS (Continued)

rpm	revolution per minute
SDS-PAGE	Sodium dodesyl sulfate Polyacylamide Electrophoresis
TAE	Tris boric acid EDTA electrophoresis buffer
TE	Tris EDTA
Tm	melting temperature
Tris	Tris (hydroxymethyl) aminomethane
U	unit
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

EXPRESSION OF A RECOMBINANT MULBERRY LEAF LECTIN IN BACTERIAL HOST AND COMPUTATIONAL STRUCTURE ANALYSIS

INTRODUCTION

Lectins are multivalent carbohydrate-binding proteins or glycoproteins of non-immune origin. They are widely found in plants, animals, and microorganisms. Ratanapo and coworkers (1998) were successfully purified two new kinds of mulberry leaf lectins specific to N-glycolylneuramic acid (Neu5Gc) namely, MLL1 and MLL2. Neu5Gc-specific lectins are very rare among sialic acid-specific lectins and mostly absence in plant but are a component of animal glycoprotein. Neu5Gc is formed in the glycoconjugates of many human tumors. Moreover, Neu5Gc-containing gangliosides were discovered in various human tumors. The lectin MLL1 can induce the agglutination of a specific phytopathogenic bacterium, *Pseudomonas syringae pv mori* while the biological functional role of the lectin MLL2 is not known yet.

To study the exact functional roles of mulberry leaf lectins, cloning and sequencing of the lectin genes are necessary. However, the extraction and the purification of lectins are time consuming, labor-intensive as well as very low amount obtained. These restrict to the study of lectins at molecular levels such as nucleotide and amino acid sequences. Full-length MLL cDNAs of *Morus rotunbiloba* (Mon-Noi) was successfully cloned using RACEs. One of these, MLLc20, provided the sequence homology to *Morus nigra* galactose binding-lectin, *Maclura pomifera* agglutinin and jacalin with 95, 78 and 76 %, respectively (Kankamol, 2003).

This thesis, the coding sequence of the mature MLL1 were subcloned, sequenced and expressed in *E. coli* (BL21 (DE3)). Moreover, this work is focused on the properties of this lectin to indicate direct interactions between specific sugars Neu5Gc and residues of this lectin by using the homology modeling and computational analysis. Since no crystal structure is available for a given

protein/saccharide complex, it is possible to model the structure if enough 3D information is known from database. The aim of this work is then to predict the best possible interaction between the protein and ligand while taking into account the flexibility of the saccharide and the amino acid side chains inflexibility. (Bush and Martin-Pastor, 1999)

MLL1 gene will further be cloned into suitable eukaryotic expression vector in order to produce a large amount of the lectins to study of these genes at molecular levels the present work has been designed in the following objectives.

The objectives of this study are:

1. To express and characterize MLL in *E. coli* system
2. To analysis computational structure and binding of MLL

LITERATURE REVIEW

1. Lectins

The occurrence in nature of erythrocyte-agglutinating proteins has been known since the 1960s, such proteins were referred to as hemagglutinins, or phytoagglutinins, because they were originally found in extracts of plants. It is generally believed that the earliest description of such a hemagglutinin was isolated by Stillmark (presented in 1888) from seeds of the castor tree (*Ricinus communis*) and was named ricin. Subsequently, H. Hellin demonstrated the presence of a toxic galactose-specific hemagglutinin, abrin, in extracts of the jequirity bean (*Abrus precatorius*). The ability of plant agglutinins to distinguish between erythrocytes of different blood types led Boyd and Shapleigh to propose for them the name lectins, from the Latin legere, to pick out or choose. (Sharon and Lis, 2004)

Lectins are carbohydrate-binding proteins of non-immune origin that bind glycans of glycoproteins, glycolipids, or polysaccharides with high affinity (Goldstein and Hayes, 1978), which agglutinate cells or precipitate glycoconjugates. This definition was too restrictive as it excluded some poorly agglutinating toxins (such as ricin, abrin, modeccin, etc.) which were known to contain lectin subunits, and so the definition was extended to include these toxins. Also, some lectins contain a second type of binding site that interact with non-carbohydrate ligands and lectins were redefined as carbohydrate-binding proteins other than antibodies or enzymes because of their binding specificity, they have the capability to serve as recognition molecules within a cell, between cells or between organisms. It is assumed that lectins play fundamental biological roles in plants because they are found in many different species and in many different organs and tissues. (Chrispeels, 1991)

2. Plant lectins: a special class of plant proteins

Many plant lectins have been extensively characterized. Most are secretary proteins, meaning that they enter the secretary system and subsequently accumulate either in vacuoles or in the cell wall and intercellular spaces. For example the well-known such as lectins phytohemagglutinin, concanavalin A, soybean agglutinin, pea lectin, and favin are all presented at relatively high levels and accumulate in vacuoles in the cotyledons (8% to l0% of total protein) and at lower levels in the embryonic axes of the seeds. These lectins are synthesized during seed development together with the more abundant seed storage proteins. During germination and seedling growth, both storage proteins and lectins are broken down to provide amino acids for the growing seedling. Lectins are also often quite abundant in vegetative plant organs such as roots, leaves, rhizomes, and stems. Some of these are vacuolar while others such as the chitin-binding Datura seed lectin are extracellular. Vacuolar lectins also occur in cereal seeds, but are much less abundant (1 pg/dry grain) and occur only in specific cell layers of the embryo (e.g.,wheat germ agglutinin in the coleorhiza and rootcap of the wheat embryo) (Etzler, 1986)

Plant lectins have been the subject of broad interdisciplinary research since their discovery. Initially, the interest in these particular proteins focused on their specific carbohydrate-binding properties and biological activities. More recently, in order to understand and exploit the unique properties of plant lectins, other aspects such as their biochemistry, cell biology, physiology, genetics, chemotaxonomy, etc. Recent advances in the molecular cloning of lectins and lectin-related proteins argue for an update of the definition of lectins.

3. Lectins on the basis of structural/functional criteria

Firstly, it has been demonstrated that some plant enzymes are fusion proteins of a carbohydrate-binding domain tandemly arrayed with a catalytic domain. Class I chitinases, for instance, are built up of an N-terminal chitin-binding domain linked through a hinge region to a catalytic domain. Similarly, the so-called Type 2

ribosome-inactivating proteins (RIP, e.g. ricin and abrin) are fusion proteins of an N-terminal toxic A chain, which has the N-glycosidase activity characteristic of all RIP, and a C-terminal carbohydrate binding B chain.

Secondly, several carbohydrate-binding proteins possess only one binding site and are therefore unable to precipitate glycoconjugates or agglutinate cells. For instance, the non-agglutinating mannose-binding proteins from orchids are very similar to the dimeric mannose-specific lectins from the same species, except that they occur as monomers and are hence incapable of agglutinating cells.

Thirdly, the molecular cloning of lectin genes led to the discovery of so-called lectin-related proteins. Some legumes contain proteins that are evolutionary and structurally related to the lectins but are devoid of carbohydrate-binding activity, because of alterations in the structure of the biding sites. Well-known examples of this group of proteins are the *Phaseolus vulgaris* arcelins and α-amylase inhibitor.

The structure and molecular evolution of lectins and lectin genes confirm the need for a definition that is based on the presence of functionally active carbohydrate-binding domains. As a consequence, the presence of at least one non-catalytic domain, which binds reversibly to a specific carbohydrate, should be the only criterion for a protein to be considered as lectins. Accordingly, plant lectins can be defined as plant proteins possessing at least one non-catalytic domain, which binds reversibly to a specific mono- or oligosaccharide. This new definition is much less restrictive than the previously proposed definitions since it includes a broad range of proteins, with different agglutination and/ or glycoconjugate precipitation properties. On the basis of the overall structure of the lectins subunits (or more precisely the primary translation product of the lectin genes), four major types of lectins are distinguished, namely 1. merolectins, 2. hololectins, 3. chimerolectins and 4. superlectins.

1. Merolectins are small proteins consisting exclusively of a single carbohydrate binding domain. Due to their monovalent nature, merolectins cannot

precipitate glycoconjugates or agglutinate cells. At present, only a few merolectins have been described. Well-known examples are the chitin-binding hevein from the latex of the rubber tree *(Hevea brasiliensis)* and the monomeric mannose-binding proteins from orchids.

2. Hololectins also are exclusively composed of carbohydrate-binding domains. However, in contrast to merolectins, they contain two or more carbohydrate binding domains which are identical or very homologous and bind either the same or structurally similar sugars. Since hololectins have multiple binding sites they fully capable of agglutinating cells and/ or precipitating glycoconjugates. Most plant lectins belong to the subgroup of hololectins.

3. Chimerolectins are basically fusion proteins composed of a carbohydrate binding domain tandemly arrayed with an unrelated domain. The latter domain may have a well-defined catalytic (or another biological) activity but acts independently of the carbohydrate-binding domain. Chimerolectins with multiple carbohydrate-binding sites behave as hololectins, e.g. Type 2 RIP, which possess two carbohydrate-binding sites on each B chain and agglutinate cells. On the contrary, chimerolectins with a single carbohydrate-binding site behave as merolectins. Class I plant chitinases, for instance, which have only one chitin binding domain per molecule, cannot precipitate glycoconjugates or agglutinate cells.

4. Superlectins are special type of chimerolectins. They are fusion proteins built up of two tandemly arrayed carbohydrate-binding domains which are structurally different and recognize structurally unrelated sugars. The primary translation product of the gene encoding this lectin is composed of an N-terminal mannose-binding domain tandemly arrayed with an unrelated GalNAc-binding domain.

4. The carbohydrate-binding specificity of plant lectins

Since the applications of lectins in biological and biomedical research, as well as their potential use in biotechnology, are determined primarily by their carbohydrate

- binding specificity, this issue needs special attention. Usually the specificity of plant lectins is expressed in terms of their ability to bind one or more specific monosaccharides. On the basis of this criterion, lectins have classically been subdivided into glucose/ mannose-specific lectin, Gal/ GalNAc-specific lectin, Glc/GlcNAc-specific lectin, fucose-specific lectin and sialic acid-specific lectin.

Classification of lectins in specificity groups on the basis of their binding to monosaccharide suffers another drawback. The fact that some lectins bind simple sugars does not imply that this monosaccharide is their true receptor molecules since even these lectins usually have a much higher affinity for oligosaccharides containing two or more monosaccharide units. Moreover, some lectins bind exclusively to carbohydrates that are not common or are totally absent in plants. For example, Chitin-binding lectins recognize a typical polysaccharide from the cell wall of fungi and the exoskeleton of invertebrates. Similarly, the sialic acid-biding lectins from *Maackia amurensis* and elderberry (*Sambucus* sp.) recognize a sugar which absent in plants, but which is a predominant carbohydrate component of animal glycoprotein. The same holds true for all lectins that bind exclusively to the complex (modified) oligosaccharide side chains of typical animal glycoproteins. This obvious preference of plant lectins for carbohydrates from animal or microbial origin has two important consequences. First, it indicates that most lectins are destined to bind foreign glycoconjugates rather than endogenous receptors. Second, it makes plant lectins important and useful tools for the isolation and analysis of human and animal glycoconjugates (including clinical applications).

5. Structure and evolution of plant lectins and their genes

Due to the important progress in the biochemistry and molecular biology of plant lectins, the structural and evolutionary relationships between the different members of this apparently very heterogeneous group of proteins have become increasingly evident. An overview of the sequence data obtained through molecular cloning and protein sequencing has led to the rather unexpected conclusion that the majority of all currently known plant lectins can be classified into four groups of evolutionary related proteins.

5.1 Legume lectins

Legume lectins are a large family of homologous proteins which are confined to species of the plant family Leguminoseae. Legume lectins have been purified most from mature seeds. Legume lectins strongly differ from each other with respect to their carbohydrate-binding specificity. Several so-called specificity groups are distinguished according to their preferential binding to monosaccharide (s). The majority of legume lectins belong to the mannose/glucose-binding lectins (e.g. *Canavalia ensiformis* lectin) and the galactose/ N-acetylgalactosamine specific lectins (e.g. *Arachis hypogaea* lectin). Others recognize specifically chitobiose (e.g. *Cytisus sessilifolius* lectin), L-fucose (e.g. *Ulex europaeus* lectin UEA-I) or Neu5Ac α(2, 3) Gal/GalNAc. Finally, there are also examples of legume lectins, which do not bind to any simple sugar but interact exclusively with oligosacharides (e.g. *Phaseolus vulgaris lectin).*

Based on the structure of their subunits, legume lectins are often (artificially) subdivided into one-chain and two-chain lectins. One-chain legume lectins are built up of two or four (either identical or highly homologous) subunits with a molecular mass of 25-30 kDa. Two-chain lectins are composed of two or four heavy subunits (β chains) and an equal number of light subunits (α chains). The light and heavy subunits of the two-chain lectins are proteolytic cleavage products from precursor molecules of about 30 kDa, which are very similar to the subunits of the one-chain lectins. All legume lectins contain divalent cations at specific metal-binding sites. Each subunit possesses both Mn^{2+} and Ca^{2+} ions, which are held in place by interactions with specific amino acid residues and are essential for the carbohydrate binding activity of the lectins. The importance of the metal-binding sites is also reflected by the fact that the amino acid residues, which are involved in the binding of the Mn^{2+} and Ca^{2+} ions, are highly conserved in all legume lectins.

The results of molecular cloning of the genes and studies of the biosynthesis, processing and topogenesis of the proteins indicate that all legume lectins are synthesized on the endoplasmic reticulum (ER) as preproteins or

preproproteins, which after post-translational modifications are converted into the mature lectins. Depending on the presence and accessibility of glycosylation sites, oligosaccharide chains can be N-linked to the precursor molecules whilst still present in the ER. Further processing by proteolytic cleavage of N-terminal, C-terminal or internal peptides and/ or deglycosylation or modifications of the glycan chains takes place after the precursors leave the ER and are transported via the Golgi system to their final destination, which in the case of most legume lectins is the vacuole or a vacuole-derived organelle.

Judging from the amino acid sequences available, it is evident that the legume lectins are a group of homologous proteins. To unravel their three-dimensional structure several legume lectins have been studied by X-ray crystallography. Furthermore, predictive methods such as the hydrophobic cluster analysis method allowed the construction of three-dimensional models using the amino acid sequences of different legume lectins. The most important conclusion is that all legume lectins, irrespective of their sugar-binding specificity, are composed of monomers with a strikingly similar three-dimensional structure, which is typified by the occurrence of two anti-parallel pleated sheets (β structures). Importantly, the majority of the residues that are not involved in the β structures are located in turns and loops interconnecting the stands of the pleated sheets. The different legume lectins not only have a very similar overall structure but also exhibit markedly conserved Mn^{2+} and Ca^{2+} binding sites. On the contrary, the amino acids involved in the carbohydrate-binding site are less conserved (though often very similar).Most probably, the amino acid substitutions in the carbohydrate-binding specificity. Finally, the striking sequence homology between the legume lectins, even when derived from taxonomically distant species, not only confirms the evolutionary conservatism within this group of proteins but also offers the opportunity to trace the taxonomic relationships between the different members of this family using a set of well-defined protein or DNA sequences.

5.2 Chitin-binding lectins containing hevein domains

Chitin-binding lectins which are composed of one or more so-called hevein domains. The name of the latter domain is derived from hevein, a small chitin-binding protein from the latex of the rubber tree *Hevea brasiliensis.* The hevein sequence contains only 43 amino acids and is very rich in cysteine and glycine. In spite of its small size, hevein is fully capable of binding chitin. Molecular cloning of hevein revealed that the mature protein is derived from a large precursor, which contains the information for a signal peptide, the hevein sequence and an additional 144 amino acids. Since most chitin binding proteins contain this 43 amino acid motif, the hevein sequence is generally referred to as the hevein domain or the chitin-binding domain.

Chitin-binding lectins with hevein domains have been isolated from several taxonomically unrelated plant families such as gramineae, Urticaceae, Solanaceae, Papaveraceae, Euphorbiaceae, Phytolaccaceae and Viscaceae. Although they are usually referred to as chitin-binding lectins, most of these lectins also react with GlcNAc, GlcNAc-oligomers and Neu5Ac. As well as the true agglutinins (hololectins), the chitin-binding lectins also comprise a few merolectins (which because of their monovalency are unable to agglutinate cells) and an extended group of chimerolectins, namely the Class I chitinases which are composed of a hevein domain tandemly arrayed with a catalytic domain with chitinase activity.

5.3 Monocot mannose-binding lectins

The super-family of so-called monocot mannose-binding lectins is a relatively new group of portins. Intensive studies of numerous monocot mannose-binding lectins, and molecular cloning of their corresponding genes, have shown that they all belong to a single super-family of evolutionary related proteins, which not only have a marked sequence homology but also exhibit an exclusive specificity towards mannose.

Although all monocot mannose-binding lectins are composed of lectin polypeptides of approximately 12 kDa, there are important differences in the processing a post-translational modifications of the primary translation products of their genes. At present, different molecular forms can be distinguished within the class of monocot mannose-binding lectins. Most of the lectins are homomers composed of two or four identical subunits of about 12 kDa, which are synthesized as separate polypeptides. However, these homomeric lectins are synthesized as preproproteins which are converted into the mature lectin polypeptides by the co-translational cleavage of a signal peptide and the post-translational removal of a C-terminal peptide, which in most instances in non-glycosylated. There are also several types of heteromeric forms. First, there are the heterodimers built up of two different (but highly homologous) subunits of about 12 kDa, each derived from separate pre-proproteins that undergo a processing similar to that of precursors of the homomeric lectins (e.g. *Allium ursinum* lectin I).A second type are the heterodimers or heterotetramers which are composed of two different types of subunits that are derived from a single precursor with two distinct lectin domains. The heterodimeric garlic (*A. sativum)* lectin ASA-I is synthesized as a large precursor, which contains two tandemly arrayed, highly homologous (84% at the amino acid level) mannose-binding domains. Formation of the mature lectin requires the removal of a glycosylated linker between the two domains as well as the removal of a non-glycosylated C-terminal peptide. The *Arum maculatum* lectin is also synthesized as a large precursor, which contains two tandemly arrayed mannose-binding domains. However, in this case, the homology between the two domains is only 40% and the linker sequence is not glycosylated. Finally, the tulip lectin TxLC-I represents a unique protein. This lectin is a tetramer of four identical subunits of 28 kDa, which contains a mannose-binding domain tandemly arrayed with an unrelated GalNAc-binding domain. Since most of the 28 kDa polypeptides are cleaved into two smaller subunits, the native lectin behaves as a hetero-octamers.

Unlike all other plant lectins studied thus far, the monocot mannose-binding lectins are encoded by large families of closely related genes. Sequence analyses of different cDNA clones encoding the lectins from *Galanthus, Narcissus*

and *Hippeastrum* have shown that the expression of different lectin genes in these species can explain the occurrence of multiple lectin isoforms at the protein level.

Molecular cloning of the orchid lectin genes has also led to the discovery of lectin-related mannose-binding proteins, which occur as monomers and are therefore unable to agglutinate red blood cells, or cross-link glycoconjugates or cells. In contrast to the non-carbohydrate-binding lectin-related proteins found in some legumes, the non-agglutinating monomeric mannose-binding proteins from orchids clearly possess carbohydrate-binding activity.

5.4 Type 2 ribosome-inactivating proteins

A fourth group of structurally related lectins are the Type 2 ribosome inactivating proteins (RIP). They possess a highly specific rRNA N-glycosidase activity and are capable of catalytically inactivating eukaryotic (and in some instances also prokaryotic) ribosomes. Type 1 and Type 2 RIP are different (in terms of subunit composition) in their molecular structure. Type 1 RIP is monomeric proteins consisting of a single catalytically active subunit (of about 30 kDa); they occur in many species and represent an extended family of evolutionary related proteins. Type 2 RIP has a more complex structure. They are composed of one, two or four identical units. Each of these units consists of two structurally and functionally different polypeptides, called the A and B chains, which are covalently linked through a disulphide bridge. The A chain possesses N-glycosidase activity and exhibits sequence homology to the Type 1 RIP, whereas the B chain is catalytically inactive but exhibits a carbohydrate binding activity comparable to that of lectins. Hence only Type 2 RIP can be regarded as lectins. Recently, conclusive evidence has been obtained that RIP not only de-adenylate ribosomal RNA but are also capable of removing adenine residues from DNA and several other polynucleotide substrates. It has been proposed, therefore, to rename RIP as polynucleotide: adenosine: glycosidases (PAG).

The A and B chains composing Type 2 RIP are synthesized as large precursors containing two functional domains separated from each other by a so-called linker sequence. Processing of these precursors requires the excision of a peptide linking the C-terminus of the first domain and the N-terminus of the second domain; however, even after removal of the linker sequence, the A and B chains are still held together by a disulphide bridge. Type 2 RIP different from each other in the number of A-B subunits. For instance, ricin, SNA-V and SNA-I are composed of one, two and four A-B units, respectively.

Ricin is undoubtedly the best-characterized Type 2 RIP. Its three-dimensional structure has been resolved by X-ray crystallography which showed that the ricin A chain is a globular protein with extensive secondary structures and a well-defined cleft, which presumably corresponds to the enzymatially active site. The ricin B chain folds into two globular domains, each of which can bind lactose in a shallow cleft. Since both domains are similar, it is assumed that they have evolved through gene duplication. Similar to WGA, the ricin B chain is stabilized through multiple disulphide bonds.

5.5 Jacalin and related lectins

A highly unusual polypeptide structure has been found for the seed lectins from the Moraceae species *Artocarpus integrifolia (*Jackfruit) and *Maclura pomifera.* Both lectins are tetramers composed of α chains (133 residues) which are non-covalently associated to β chain (20 residues) subnunits. Molecular cloning of the lectin gene from *Artocarpus integrifolia* revealed an unusual primary structure for the jacalin precursor, and meant that the problem of the molecular structure of jacalin could be solved. A comparison of the deduced amino acid sequence of the prolectin and α and β chains of the mature lectins demonstrated that both mature polypeptides originate from a single precursor through a complex post-translational processing, which involves the removal of an N-terminal peptide, an internal peptide and a C-terminal peptide. Apparently the different isolectins in jacalin are encoded by different genes, indicating that the Jackfruit lectin is encoded by a family of genes. The expression of a lectin gene family can also explain the high degree of micro

heterogeneity in the lectin preparations. However, in addition to genetic variation, the presence of different isoforms is also due to differences in post-translational modifications. X-ray diffraction analysis of jacalin indicated that each promoter has a threefold symmetric β-prism fold made up of three four-stranded β sheets.

Recently, a mannose/ maltose-binding lectin was isolated from the rhizomes of the *Convolvulaceae* species *Calystegia sepium.* Molecular cloning and sequence analysis of the lectins revealed a striking sequence similarity with jacalin, suggesting that both lectins might belong to the same group of lectins. The apparent presence of evolutionary related lectins in *Moraceae* and *Convolvulaceae* family indicates that lectins belonging to the jacalin group may occur in different taxonomic groups.

5.6 Cucubitaceae phloem lectins

A dimeric chitin-binding lectin has been characterized from the phloem sap of *Cucubita* sp. (Cucurbitaceae). Molecular cloning and sequence analysis revealed that, despite its carbohydrate-binding specificity towards chitin, the pumpkin lectin does not show significant sequence homology to the chitin-binding domains from chitinases or Gramineae lectins. Southern blot analysis and studies at the genomic level have shown that the *Cucurbita* lectins are encoded by a small gene family.

5.7 Amaranthin group

Several *Amaranthus* species contain closely related GalNAc-specific seed lectins composed of lectin polypeptides of 33-36 kDa. N-terminal sequencing of the *A. caudatus* lectin revealed that it corresponds to a previously cloned seed-specific protein from *A. leucocarpus.* Since the deduced sequence of the latter protein has no similarity with any other plant lectins, it is obvious that the *Amaranthus* lectins represent a distinct group of plant lectins. X-ray diffraction studies of the *A. caudatus*

lectin further confirmed that its three-dimensional structure definitely differs from that of the other known lectins structures.

6. Plant lectin function

The physiological role of plant lectins has made an important contribution to the development of lectinology. As soon as it was found that plant lectins recognize specific carbohydrates, numerous hypotheses were put forward suggesting different functions on the basis of this particular biological activity. The proposed roles were based on either endogenous or exogenous lectin-carbohydrate interactions. Interactions between lectins and carbohydrates from the same plant, for instance, were believed to be essential for

The transport of sugars
The accumulation of storage compounds
Cell-cell interactions, regulation of cell division
Defence against bacterial, fungal, and viral pathogens or animal predators
Rhizobium-legume symbisis
Packing or mobilization of storage material during seed maturation or germination
Maintenance of seed dormancy
Carbohydrate-specific enzymes
Mitogenic stimulators of plant embryonic cells
etc. (Etzler, 1992)

Similarly, interactions between lectins and carbohydrates of exogenous origin were considered as the determining factor in the establishment of plant-microbe interactions and in defense mechanisms against various organisms. Unfortunately, most of the hypotheses that were put forward were founded on hypothetical considerations rather than experimental evidence. Moreover, in some instances, the evidence in favor of a particular hypothesis was obtained from experiments with

impure lectins preparations. As a result, the biological activity of a contaminating protein (e.g. enzyme) was attributed erroneously to otherwise inactive lectins.

A breakthrough in the search for the physiological role of plant lectins was achieved when it was realized that most plant lectins not only play a role in the plant itself (e.g. as a store of nitrogen or as a specific recognition factor) but are also capable of interfering with the functioning of foreign organisms through an interaction with glycoconjugates on the surface or in the digestive tract of these organisms. Accordingly, most plant lectins are probably involved in the plant's defense. Whereas direct interference with viruses and microorganisms is rather exceptional, the deleterious effects of plant lectins on both predatory invertebrates and higher animals are well documented. To explain the abundance of lectins in storage organs and their storage protein-like behavior, it has been proposed that plants accumulate part of their nitrogen reserve in the form of carbohydrate binding proteins, which can be used as passive defense proteins.

Although there is not doubt about the defense-related function of (some) plant lectins, it is unlikely that they all play a defensive role. Lectins occurring at low concentrations might well be involved in specific recognition processes either within or outside the plant. Legume root lectins, for instance, could be involved in the recognition and/or binding of *Rhizobium* and *Bradyrhizobium* species and hence be a determining factor in establishing symbioses.

Some plant lectins are considered as efficient resistance factors against insects, nematodes or mammalian pests. However, most lectins with promising potential in crop protection are only found in wild species which cannot be crossed with our common crop plants. Therefore, lectin genes can be used as resistance factors only when transferred to and expressed in the desired crop plant. For instance, genetically modified potato plants which express the snowdrop lectin gene at a sufficiently high level are protected against sucking insects and nematodes. (Peumans and Van Damme, 1995)

7. **Mulberry lectins**

7.1 Mulberry leaf lectin

Ratanapo *et al*. (1998) purified and characterized two new *N*-Glycolylneuraminic acid (Neu5Gc) binding lectins, named MLL1 and MLL2, from mulberry (*Morus rotunbiloba*) leaves. Both MLL1 and MLL2 activities were most effectively inhibited by NeuGc at 0.78 and 1.56 mM, respectively. GalNac and dD-Gal at higher concentrations also inhibited the lectin activities. They are glycoproteins, with neutral sugar contents of 8.8 and 40%, respectively, having a native molecular weight at 51 kDa. Each consisted of subunit with molecular weight of 16.5 kDa. Both MLLs require no metal ion for their lectin activities. Each MLL is probably a tetramer.

Ratanapo *et al*. (2001) studied two new *N*-Glycolylneuraminic acid (NeuGc) specific lectins, MLL1 and MLL2, for anti-bacterial activity againt *P.syringae pv mori,* aspecific pathogen bacterium of mulberry leaves. MLL1 but not MLL2 was found to induce the agglutination only at the exponentail phase of bacterial growth in a liquid medium and the agglutination was specifically inhibited by Neu5Gc at 12.5 mM and bovine submaxillary murin at 0.05 μg/ml.

Kankamol *et al*. (2003) were first and successfully cloned the full-length Mll1 cDNAs of *Morus rotunbiloba* (Mon-Noi) using RACE, yielding 300 clones. Four of 216 residuces each deduced amino acid sequences shared 94-98% homology, indicating 4 isoforms of the lectin genes. One of these clons, MLLc20, provided the sequence homology to *Morus nigra* galactose binding-lectin, *Maclura pomifera* agglutinin and jacalin with 95, 78 and 76 %, respectively, Moreover, Kankamol *et al* studied and characterized the expression of lectin genes using the full-length cDNA as a probe. Several hybridizing bands with various intensities were detected in each digested DNA sample. The data from Southern blot and nucleotide sequence analysis supported the existence of at least 4 isoforms or MLL genes in the leaf of *M. rotunbiloba* family. Northern blot analysis indicated that the expression level was the

highest in shoot, lesser in root and in flower, but not found in bark and in mature leaf. The result is in good agreement with a strong hemeagglutination activity in young leaves but less activity in mature leaves (Ratanapo, unpublished data). Northern blot analysis of all lectin genes showed an identical size of their transcripts

Physiological function of the MLL in mulberry plant is not yet defined. By Northern blot analysis, mRNA transcripts of MLL were found in all those varieties and it suggested that this lectin has important biological functions. Some selected mulberry varieties had significant difference in the expression level of MLL1 genes. Khun-Pai, Buri Ram 60, and Buri Ram 4/2 varieties showed the highest expression of MLL1 genes. It has been reported that, Khun-Pai and Phai-Ubon have strong resistance against Root-Rot disease; Buri Ram 60 has strong resistance against Powder mildew disease and Mosaic disease (Agricultural Extension department, Annual report) whereas Mon-Noi (*M. rotunbiloba*) is susceptible to Root-Rot disease. The data suggested that lectins in natural environment mulberry may possibly have function as a defense molecule against its phytopathogenic bacteria. The high expression of mRNA in the mulberry shoot supported the antibacterial role of MLL in the plant. The data is in good agreement with previous report that that MLL can agglutinate a bacterium, *P. syringae pv mori*, a pathogen of bacterial blight disease.

Five pairs of primers were designed from MLL cDNA to clarify genomic DNA sequences and developed some specific primers. The nucleotide sequences suggested that MLL genomic DNA contain at least two introns at the position 1-132 and over 534 of cDNA sequence. A pairs of specific primer was designed at the position of 334-693 of MLL cDNA. Three bands of 800, 700 and 600 bp were found in 13 diverse mulberry varieties with different PCR pattern among local varieties and the others. The PCR products were larger than the expected size. This primer is specific for Morus species since it could not amplified in related Moraceae (*Ficus altissima Blume* and *Broussonetia papyrifera*) plant species. Nucleotide sequencing of the PCR products were compared with over 90% identity. This MLL gene will be useful as a specific primer for identify mulberry varieties in combination with other primers.

Ratanapo *et al*. (2005) MLL1 and MLL2 have inhibitory effects to purified trypsin-like alkaline proteases from digestive fluid of the silkworm. The inhibition of the proteases by both mulberry leaf lectin was not recovered by the specific lectin binding sugars, N-glycolylneuramic acid and N-acetylgalactosamine.

7.2 Mulberry lectins in other parts of plants

Yeasmin *et al*. (2001) discovered three galactose-specific lectins, MSL-1, MSL-2 and MSL-3, from mulberry seeds. The molecular masses were 175000, 120000 and 89500 and contained 5.7, 5.4 and 4.5% neutral sugars, respectively. MSL-1 is dimmer in nature, with the two monomers held together by disulfide bond(s), while MSL-2 and MSL-3 contain four nonidentical subunits that are held together by nonionic hydrophobic interactions. The lectins exhibited strong cytotoxic effect in brine shrimp lethality bioassay.

Van Damme *et al*. (2001) demonstrated that the bark of the black mulberry tree (*Morus nigra*) accumulates large quantities of a galactose-specific (MornigaG) and a mannose (Man)-specific (MornigaM) jacalin-related lectin. MornigaG resembles jacalin with respect to its molecular structure, specificity, and co- and posttranslational processing indicating that it follows the secretory pathway and eventually accumulates in the vacuolar compartment. In contrast, MornigaM represents a novel type of highly active Man-specific jacalin-related lectin that is synthesized without signal peptide or other vacuolar targeting sequences, and accordingly, accumulates in the cytoplasm.

Wu *et al*. (2003) isolated Morniga M, a jacalin-related and mannose-specific lectin, from the bark of the mulberry *(Morus nigra)*. The di-, tri-, and oligomannosyl structural units of N-glycans such as those of the bovine α_1-acid glycoprotein and lactoferrin were the most active glycoproteins, but not the O-glycans or polysaccharides including mannan from yeast.

Razzaque *et al*. (2005) describes the purification, characterization and N-terminal sequences of a mannose/glucose specific lectin from the seeds of mulberry. Its MW was estimated to be 22,000. The lectin is glycoprotein with neutral sugar content of 28.57%. The lectin also exhibited cytotoxic effect in brine shrimp lethality bioassay. The N-terminal sequences of the lectin upto 45-residues except the positions of 21, 39, 42 and 44 were identified.

Singh *et al*. (2007) was studied the subfamily of galactose-specific jacalin-related lectin isolated from the bark of black mulberry (*Morus nigra*) (Morniga G) in detail by enzyme-linked lectinosorbent and inhibition assays using panels of monomeric saccharides, mammalian polyvalent glycotopes and polysaccharides. Morniga G reacted best with glycoproteins presenting a high density of tumor-associated carbohydrate antigens. It also reacted well with many multi-antennary *N*-glycans with II (Galβ1-4GlcNAc) termini, ABH histo-blood group antigens and their precursors containing high densities of I/II and T/Tn glycotopes, and sialylated T/Tn. Among the mono-, di- and oligosaccharides tested, Thomsen-Friedenreich (T) disaccharide with aromatic aglycon [Galβ1-3GalNAcα1-benzyl (Tα1-benzyl)] and Tn glycopeptides were the best inhibitors. Molecular modeling and docking studies indicated the occurrence of a primary GalNAcα1- and Galβ1-3GalNAc glycotope-binding site in Morniga G.

8. **Sialic acid-specific lectins**

Sialic acids (Sia) play important roles as ligands in various cells. More than 30 kinds of sialic acid are known: *N*-glycolylneuraminic acid (Neu5Ac), *N* -Glycolylneuraminic acid ,1 deaminoneuraminic acid (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid) (Neu5Gc)and their analogs, which are modified by *O*-acetylation, *O*-lactylation, *O*sulfation, and lactonization (Figure.1), that are typically located at the terminal positions of a variety of glycoconjugates. Naturally occurring sialic acid can also mediate a variety of specific recognition processes. For instance, as the terminal residues on many glycoconjugates, can mask underlying structures, as observed for erythrocytes and other blood cells, as well as serum glycoproteins show

an immense diversity of structure, and this reflects their involvement in a variety of biologically important processes (Lehmann, 2006). Neu5Ac and Neu5Gc are the most common sialic acid in most mammals, but normal human tissues do not contain Neu5Gc. Neu5Gc is formed in the glycoconjugates of many human tumors, where it may be responsible for the formation of Hanganutziu-Deicher antibodies. Neu5Gc-containing gangliosides were discovered in various human tumors including colon cancer, retinoblastoma, melanoma, breast cancer, and yolk sac tumor. Lectins that recognize sialic acid are known in vertebrates, arthropods, mollusks, protozoa, plants, bacteria, virus and some of them have been used as biological and medical probes for sugar or sugar chain detection. Neu5Gc-specific lectins are very rare among sialic acid-specific lectins. Although some Neu5Ac-specific lectins have been studied and are commercially available, Neu5Gc-specific lectins are difficult to obtain commercially. A few Neu5Gc-specific lectins have been isolated from several living sources. (Kobayashi, 2004)

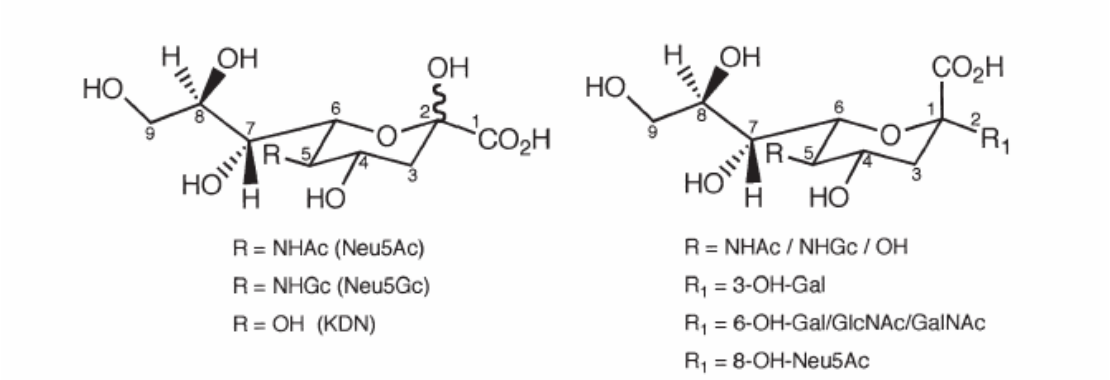


Figure 1 The structural diversity of sialic acid. Sialic acid is generated by a combination of variations at C-5 with modifications of any of the hydroxyl groups at C-4, C-7, C-8 and C-9. sialic acid is predominantly found glycosidical linked via *a*2, 3-, *a*2, 6- or *a*2, 8- linkages to underlying sugars as shown.

Source: Lehmann *et al*. (2006)

9. Sialic acid-specific plants lectins

Even though only a handful of sialic-specific lectins have been identified and isolated from plants (Table 1). A popular theory used to account for the presence of sialic acid-specific lectins in plants concerns their involvement in plant defense. Some arguments in favor of this role include the fact that these lectins specifically bind sialic acid, a carbohydrate that plants themselves do not express. This may provide plants with a meaning of recognizing and combating sialylated pathogens. Further, the digestive tracts of animals capable of feeding on plants are covered with highly sialylated mucins, providing numerous ligands for sialic-specific lectins. Presumably, it is this binding of sialic-specific lectins from elderberry (*Sambucus nigra*) bark and wheat germ agglutinin that initiates the severe toxicity symptoms observed upon ingestion of plant lectins in higher organisms. Moreover, sialic acid-specific plant lectins like other plant lectins are predominantly localized in regions of the plant that are most susceptible to attack. Thus, they require an adequate protection strategy. For instance, the lectin from elderberry and the leguminous plant *Maackia amurensis* are found in the bark and seed, respectively. Peumans and Van Damme have suggested that this aspect of plant physiology has a direct influence on viability, arguing that a growing plant that is half eaten may (still) survive and even produce viable offspring. All of the above hypotheses are based on the assertion that a family of plant lectins actually exists that specifically binds sialic acid. However, this view is not universally shared. The presence of sialic acid is thought by some, to only provide an acidic group that enhances the interaction. That is, the interaction with sialic acid-containing glycoconjugates is believed to be a purely coincidental one. Evidence supporting this assertion includes the observation that free sialic acid does not interact with putative sialic acid-specific plant lectins, with Galactose or lactose being the real binding partner. The crystal structure of *M. amurensis* lectin complexed with sialoglycoconjugates shows that a Galactose residue occupies the primary binding site. A sulfate group at C3 of Galactose instead of sialic acid was found to bind *M. amurensis* lectin, indicating that only a charged group is required rather than a complete sialic acid molecule. However, this would mean that the presence of a sialic acid molecule, regardless of linkage, would elicit the same effect. Finally, sialic acid-

specific lectins appear not to be as widespread in plants as would be expected given their proposed importance in plant defense. It is therefore reasonable to suggest that due to evolutionary pressure placed on these plants by sialylated pathogens and/or predators, they have developed extremely specific defense mechanisms (Lehmann *et al*., 2006)

Table 1 Sialic acid-specific plants lectins

Species	Lectin[1]	Specificity/ligand
Maackia amurensis	MAL	Neu5Acα2,3Galß1,4GlcNAc
Maackia amurensis	MAH	Neu5Acα2,3Galß1,3[Neu5Acα2,6]GalNAc
Sambucus nigra	SNA	Neu5Aca2,6Gal
Sambucus canadensis	SCA	Neu5Acα2,6Gal
Sambucus sieboldiana	SSA	Neu5Acα2,6Gal
Trichosanthes japonica	TJA	Neu5Acα2,6Galβ1,4GlcNAc
Viscum album	ML-I	Neu5Acα2,6Galβ1,4GlcNAc
Saraca indica	saracin	Neu5Acα2,6/3Galβ1,4GlcNAc
Artocarpus integrifolia	jacalin	Gal and Man > Neu5Ac
Triticum vulgaris	WGA	internal GlcNAc > Neu5Ac
Morus alba	MLL	Neu5Gc
Lactuca scariole	PLA	Sia

[1] HA-A, haemagglutinin activity observed.

Source: Lehmann *et al*. (2006)

10. Protein structure prediction

Ever since the first protein structure was determined, computational biologists and computational chemists have attempted to develop software that could predict the three-dimensional structure of proteins, using only their sequence as input. Indeed, many of the first bioinformatics programs ever written were directed at trying to solve the protein folding problem. Even though the field is more than 40 years old, protein structure prediction continues to be one of the most active areas in all of bioinformatics research, with hundreds of papers being published on the subject each year. Encouragingly, some progress has been made and it is now possible to predict or model the three-dimensional structure of proteins using at least three different

methods: homology (or comparative) modeling, threading, or ab initio methods. All three methods are fundamentally predictive, meaning that the structures generated are models and are not based on raw experimental data derived from X-ray diffraction or NMR experiments. Rather, each of these predictive approaches attempts to build on prior knowledge about protein structure and to extrapolate these principles toward the generation of new structures.

10.1 Homology modeling

Of the three predictive methods that are currently available, the most powerful and accurate approach is homology modeling. Homology (or comparative) modeling is a robust technique for predicting or generating detailed three-dimensional structures of proteins based on the coordinates of known homolog found in PDB. In homology modeling, the quality of the model strongly depends on the degree of similarity between the query sequence and the matching database sequence, with proteins sharing the highest degree of similarity being modeled best. As a general rule, the average coordinate agreement between the modeled structure and the actual structure drops by approximately 0.3 °A for each 10% reduction in sequence identity. Further more, homology modeling generally cannot be used for predicting structures of protein having less than 30% sequence identity to a target protein already in PDB. However in curtain rare cases, homology modeling can be used to generate a reliable three dimensional structural model of a protein with less than 20% sequence identity. (Baxevanis and Ouellette, 2005)

10.2 Molecular docking study

Molecular docking can be defined as the prediction of the structure of receptor-ligand complexes, where the receptor is usually a protein or a protein oligomer and the ligand is either a small molecule or another protein. Different simplifications are used to make molecular docking tractable in different applications. Initially, molecular docking was used to predict and reproduce protein-ligand complexes. There are two key parts to any docking program, namely a search of the

configurationally and conformational degrees of freedom and the scoring or evaluation function. The search algorithm must search the potential energy landscape in enough detail to find the global energy minimum. In rigid docking this means that the search algorithm explores different positions for the ligand in the receptor active site using the translational and rotational degrees of freedom. Flexible ligand docking adds exploration of torsion degrees of freedom of the ligand to this process. The scoring function has to be realistic enough to assign the most favorable scores to the experimentally determined complex. Usually, the scoring function assesses both steric complementarities between the ligand and the receptor as well as their chemical complementarity. (Brooijmans and Kuntz, 2003)

10.2.1 AutoDock

The program AutoDock was originally written in FORTRAN-77 in 1990 by David S. Goodsell here in Arthur J. Olson's laboratory. It was designed to perform automated docking of a flexible ligand to a binding site of a rigid protein, given the region of the protein containing the binding site and the substrate. For maximum predictive accuracy as well as reasonable computational time, it employs the AMBER force field in conjunction with free energy scoring functions and a large set of protein-ligand complexes with known protein-ligand constants. AutoDock consists of two main programs: AutoGrid pre-calculates a three-dimensional grid of interaction energies based on the macromolecular target using the AMBER force field. Since the structure of the receptor protein is rigid and known, interaction energies between the probe and surrounding amino acids can be calculated at each point in the grid and stored in a table. Additional tables are made for each atom type in the ligand, taking into account dispersion/repulsion and hydrogen bonding energies. A second grid is made to allow for electrostatic effects, using a probe with a single positive charge. After the grid has been completed, AutoDock can begin the simulation. First, the ligand randomly moves in any one of six degrees of freedom (either translation or rotation) and the energy of the new ligand state is calculated. If the energy of the new state is higher than the old state, the new one is automatically accepted as the next step in docking.

A weakness with this method is that the initial freedom of movement of the ligand makes it likely that the most energetically favored binding conformation will not be observed. This problem is addressed by performing separate simulations starting from random initial states to find consistent conformations. However, finding unknown binding sites in a protein is not AutoDock's strength. Four different docking algorithms are currently available in AutoDock: SA, the original *Monte Carlo* simulated annealing; GA, a traditional Darwinian genetic algorithm; LS, local search; and GA-LS, which is a hybrid genetic algorithm with local search. The GA-LS is also known as a Larmarckian genetic algorithm, or LGA, because children are allowed to inherit the local search adaptations of their parents. (Huey and Morris, 1996)

MATERIALS AND METHODS

Materials

1. Equipments for analysis

Autoclave: Model HA-300M
Autopipette: Pipetteman, Gilson, France
Balance: Satorious
Centrifuge, refrigerated centrifuge: Model BR 4i (JOUAN SA)
Centrifuge, microcentrifuge: Model Spectrafuge 16M
Dialysis bag
Electrophoresis unit: Submerged Agarose Gel Electrophoresis System
Electrophoresis unit: Vertical SDS-PAGE system
Fraction collector (Pharmacia Biotech)
Freezer -20 °C (Brandt)
Freezer -80 °C
Gel documentation & analysis (Pharmacia Biotech)
Horizontal Electrophoresis Chamber
Hot air oven: Model 838F (Fisher Scientific)
Hot plate (Fisher Scientific)
Hybrization oven: Model 1012 (Shel Lab)
Incubator: Model IPR 150.XX2.C
Incubator shaker: Model IOC400.XX2.C (GALLENKAMP PLC)
Incubator water bath: Model INNOVA 3100 (New Brunswick Scientific)
Larminar flow: BH-120 (GelmanScience)
Magnetic stirrer
Microwave (Turbora)
pH meter (Denver Instrument)
Power supply (600 V, 500 mA)
Shaker water bath (Grant)
Sonicator

Thermal cycle: Model 480 (PerkinElmer)

Thermal cycle: Model 9700 (GeneAmp PCR system)

UV/ Visible spectrophotometer (Beckman)

Vortex mixer (Fisher scientific)

2. Chemicals

Acrylamide (Pharmacia Biotech, Sweden)

Agarose (Sekem)

Ammonium persulfate (Amersham)

Ampicillin (Sigma, USA)

Bacto-Agar (Merck, Germany)

Bovine Serum Albumin (BSA) (Promega,USA)

5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal)
(Sigma, USA)

Chloroform (Merck, Germany)

Calcium Chloride (Merck, Germany)

Coomassie® brilliant G 250

DEAE-cellulose resin (Sigma, USA)

Deoxyribonucleic acid (dATP, dCTP, dGTP, and dTTP) (Promega,USA)

DNA marker: DNA, 1 Kb plus DNA ladder and 100 bp ladder
(Gibco,USA)

Ethanol, absolute (Merck, Germany)

Ethidium bromide (EtBr) (Sigma, USA)

Ethylene diamine tetraacetic acid, disodium salt dihydrate (EDTA)
(Merck, Germany)

Ethyl alcohol absolute (Merck, Germany)

Extract II kit (MACHEREY – NAGEL, Germany)

Glacial acetic acid (Carlo Erba)

Glycerol (Carlo Erba)

Glycine (Amresco)

Hydrochloric acid (Carlo Erba)

Imidazole (Acro organics, USA)

Isopropyl-β-D-thiogalactoside (IPTG) (Sigma, USA)

Isoamyl alcohol (Merck, Germany)

Kanamycin (Maiji Phamaceutical, Japan)

Metanol (Carlo Erba)

β-Mercaptoethanol (Merck, Germany)

Molecular weight protein marker (Amersham, UK)

Ni^{2+}-NTA resin (Pierce)

N,N' –methyl-bis-acrylamide (Pharmacia Biotech, Sweden)

NNN'N'- tetramethyl-1,2-diaminoethane (TEMED) (Pharmacia Biotech, Sweden)

Phenol (Sigma, USA)

Phenol Red (Sigma, USA)

Sodium chloride (Ajax chemicals)

Sodium dihydrogen orthophosphate (Ajax chemicals)

Sodium hydrogen diphosphate (Ajax chemicals)

Sodium Dodecyl Sulfate (SDS) (Bio Basic Inc.)

Tricine (Acros Organics, USA)

Tris-(hydroxy methyl)-aminomethane (Sigma, USA)

Tryptone (Merck, Germany)

Urea (Amersham)

Yeast extracts (Merck, Germany)

All chemicals and reagents used throughout this study were molecular biology grade or analytical grade.

3. Enzymes

Taq DNA polymerase (Promega)

Restriction endonuclease *Nhe*I (Biolabs)

Restriction endonuclease *Xho*I (Biolabs)

RNase A (Sigma)

T_4 DNA ligase (Biolabs)

4. Plasmid vectors

4.1 pGEM-T easy vector: vector for MLL gene cloning

The pGEM-T Easy Vector System is convenient systems for the cloning of PCRproducts. The vectors are prepared by cutting Promega's pGEM-5Zf (+) (b) and pGEM-T Easy Vectors with EcoR V and adding a 3´ terminal thymidine to both ends. These single 3´-T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products generated by certain thermostable polymerases (Promega, 1998)

4.2 pET33b (+) vector: vector for MLL overexpression

The pET-33b(+) vector is derived from pET-28b(+) and carries a 15 bp sequence encoding the protein kinase A (PKA) site RRASV, located between the thrombin cleavage and *Nde* I sites. Proteins expressed in pET-33b (+) can be easily purified by metal chelation chromatography (*via* either N- or C-terminal His•Tag sequences) and efficiently labeled with 32_P- or 33_P-gATP and the catalytic subunit of cAMP-dependent protein kinase from heart muscle. Labeled proteins can be used as direct probes in protein-protein interaction studies. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circle map. The f1 origin is oriented so that infection with helper phage can produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer. (Novagen, 1998)

5. Oligonucleotide primers

Primers: F 5′*Nhe*I and R5′*Xho*I were synthesized by BSU

F: 5′-GGAAGCTAGCATGCTGATGGCTTCGTC-3′

R: 5′-CCTTACTCGAGGCTGCCGCGGCCCACCAGAAGTGCTA AG TGAAACCCAATG-3′

6. Bacterial strains

Escherichia coli JM 109

Escherichia coli Bl21(DE3)

7. Computer and program

Computer

GENO3D homology modeling program

SPDBV program

ViewerPro program

AutoDock 3.0.5 program

Methods

1. Preparation of MLL

1.1 Amplification of MLL gene by polymerase chain reaction (PCR) technique

1.1.1 Primer design

The MLL coding region was amplified by appropriate primers base on MLL1 sequences (Kankamol *et al*., 2003). The forward primer was designed from 5′ nucleotide sequence of MLL1 and modified by adding *Nhe*I restriction site before the full length gene. The reverse primer was designed from 3′end of MLL1 nucleotide sequence and added *Xho*I restriction site following thrombin site. Primers are short nucleotide sequences that are usually 15-25 nucleotides long and have ~ 50% G + C content. Special care must be taken to assure that the primer sequences do not from duplex structures with each other or hairpin loops within themselves.

1.1.2 PCR amplication

The polymerase chain reaction (PCR) was carried out in a Thermal Cycler (Px2Thermal cycle: Thermo hybrid). The PCR reaction was performed in a 100 μl containing 200 ng of MLL20 plasmid template, 1.25 mM of dNTPs (dATP, dTTP, dCTP and dGTP), 25 mM $MgCl_2$, 20 pmol of each forword and reverse primer, 10x PCR reaction buffer and 10 Unit of Tag Polymerase. The condition was consisted of preheating the reaction mixtures at 94 °C for 4 min, the cycle was started with denaturation at 94 °C for 30 seconds, annealing at 50 °C for 1 minute and extension at 72°C for 1 minute. After completion of 30 cycles, the extension phase was prolonged to10 minute at 72 °C. The PCR product size was analyzed by comparing with the relative eletrophoretic mobility of 100 bp DNA ladder on 2.0 % agarose gel

1.1.3 Analysis of DNA fragment by agarose gel electrophoresis

Electrophoresis through agarose is the standard method used to separate, identify, and purify DNA fragments. Agarose gel has range of DNAs separation from 200 bp to approximately 50 kb in length. The sufficient electrophoresis buffer 1x TAE was prepared. The agarose powder was added the correct amount of agarose. (2.0-0.7 % (w/v) agarose gel has the efficient separation range of linear DNA molecules from (0.1-2) to (0.8-10) kb). Agarose with electrophoresis buffer was heated until the agarose dissolved. The solution was cooled to 60 °C, and if desired, ethidium bromide was added to final concentration of 0.5µg/ml and mixed thoroughly then poured the solution into the mold. The gel should be between 3 mm to 5 mm thick. The comb was positioned 0.5-1.0 mm above the plate and no air bubbles were observed in the gel. After the gel was completely set (30-45 minutes at room temperature), the comb was carefully removed and put the gel in the electrophoresis tank. Enough electrophoresis buffer was added just to cover the gel to a depth of about 1mM. The samples of DNA were mixed with gel-loading buffer. The mixture was slowly loaded into the slots of the submerged gel. The lid of the gel tank was closed and the electrical leads were attached. The gel was run at 100 volts (10 V/cm) until the bromophenol blue migrated the appropriate distance through the gel. The electric current was turned off. If ethidium bromide was present in the gel and electrophoresis buffer, the gel was examined by ultraviolet light.

1.1.4 Purification of PCR product

NucleoSpin Extract II gel extraction kit was used for direct purification of PCR products performed according to protocol. Briefly, two volumes of NT buffer were mixed with one volume of PCR product. The sample was loaded into the column and centrifuged for 1 min at 11,000 rpm. At this step, the DNA was bound to the membrane. The flow-through was discarded. The DNA was washed by adding 600 µl NT3 buffer and centrifuged for 1 min at 11,000 rpm. The flow-through was discarded and the membrane was dried by centrifuged for 2 min at 11,000 rpm. Finally, 15-50 µl of elution NE buffer was added to the center of the membrane. The

column with elution buffer was incubated to increase yield of DNA at room temperature for 1 min. The incubated column was centrifuged for 1 min at 11,000 rpm. The purified PCR product was collected.

2. Cloning of MLL gene into pGEM-T easy vector

2.1 Construction of recombinant pGEM-MLL

The MLL and pGEM-T Easy vector (Figure 2) were ligated in 20 µl reaction volume containing 50ng of pGEM-T Easy vector, 250 ng of MLL gene fragment, 1x ligation buffer and T_4 DNA ligase. The reaction mixture was incubated at 16°C for 16 h. The ligated products were used for transformation.

2.2 Transformation of ligation product into *E.coli* (JM 109)

2.2.1 Preparation of competent cells

The cell culture containing desired *E. coli* cells was growth overnight at 37° C in LB (no antibiotics) to stationary phase. The overnight culture was diluted 1:50 in fresh LB and grown at 37° C until OD_{600} reached 0.6. The cells were harvested by centrifugation, resuspended in 1/2 volume of cold 100 mM $MgCl_2$ and incubated on ice for 20 minutes. Pellet cells were resuspended in 1/10 volume of cold 100 mM $CaCl_2$ and incubated on ice for 1 hour. The competent cells were immediately used for heat shock transformation, or sterile glycerol was added to a final concentration of 15%, aliquot and stored at -80° C.

2.3 Transformation by heat shock method

At least two LB-Amp plates containing ampicillin at 50-100 µg/ml were prepared: one was for the transformation and the other for a negative control. After, the competent cells were stored at 2-8 °C, to equilibrate at room temperature. One microlitres of plasmid was added to a sterile 1.5 ml tube on ice. Another tube on ice was not added plasmid. Fifty microlitres of competent cells was added. The tubes

were gently flicked to mix, and placed on ice for 30 minutes. The tubes were incubated for 45 seconds to 2 minutes in a water bath at exactly 42 °C and immediately put back on ice for 2-10 minutes. Nine hundred and fifty microlitres of room temperature SOC media was added to the tubes and incubated 1-1.5 h at 37 °C with 250 rpm shaking. Before plating the transformant 30 minutes, 0.4 % X-Gal and 100 mM IPTG was plated onto LB-Amp agar plate. One hundred microlitres of each transformation culture were plated onto antibiotic plates. The plates were incubated overnight (16-24 h) at 37 °C**.**

2.4 Selection of MLL positive clone

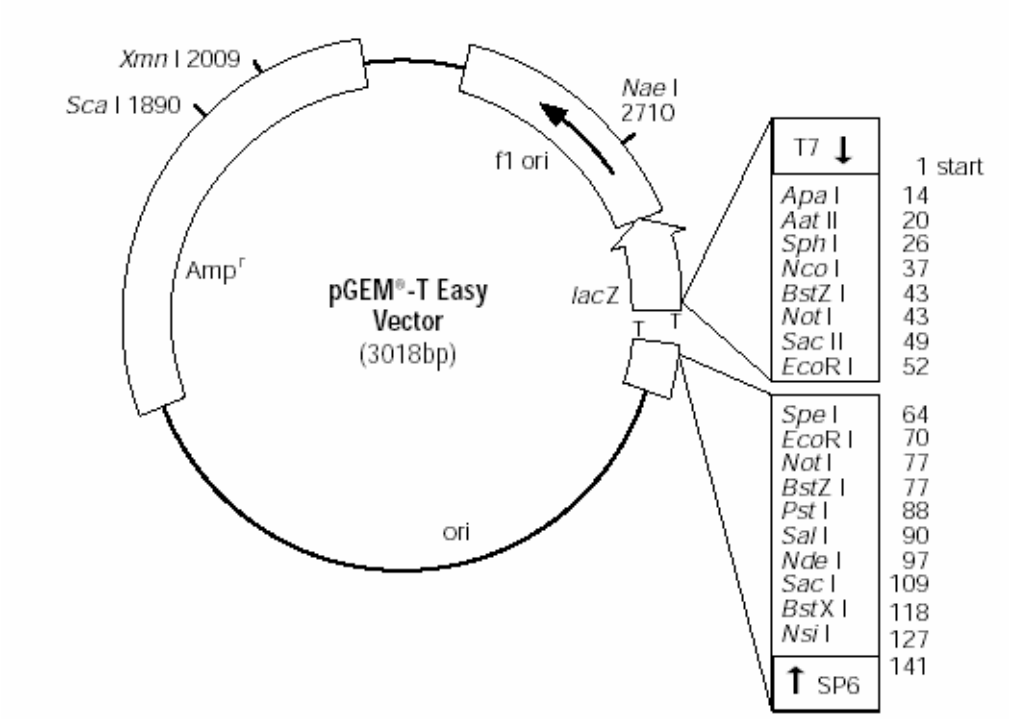
2.4.1 Blue/white colony selection

Direct selection of the recombinant was provided of screening of colonies. Cells containing pGEM-T Easy plasmid that has amplicinlin resistant gene can grow on amplicilin selection plate. The β-galactosidase gene of pGEM-T Easy plasmid can be induced by IPTG to produce β-galactosidase that can cleave the chromogenic substrate X-gal, resulting in the appearance of blue colonies. Upon the insertion of DNA fragment into the β-galactosidase gene locus, the β-galactosidase in recombinant plasmid becomes inactivated and the cells lose an ability to cleave X-gal which is colorless. Therefore the cell containing the recombinant plasmid was grown as white colony. Thus, white colony clones were selected to proceed for another step.

2.4.2 Colony PCR

Transformed white colonies were randomly picked into PCR reaction mixture tube. The reaction mixture was carried out in a final volume of 10 μl. The same condition was used as the step 1.1.2 for 15 cycles. After completion, the PCR product was analyzed by agarose gel electrophoresis.

A



B

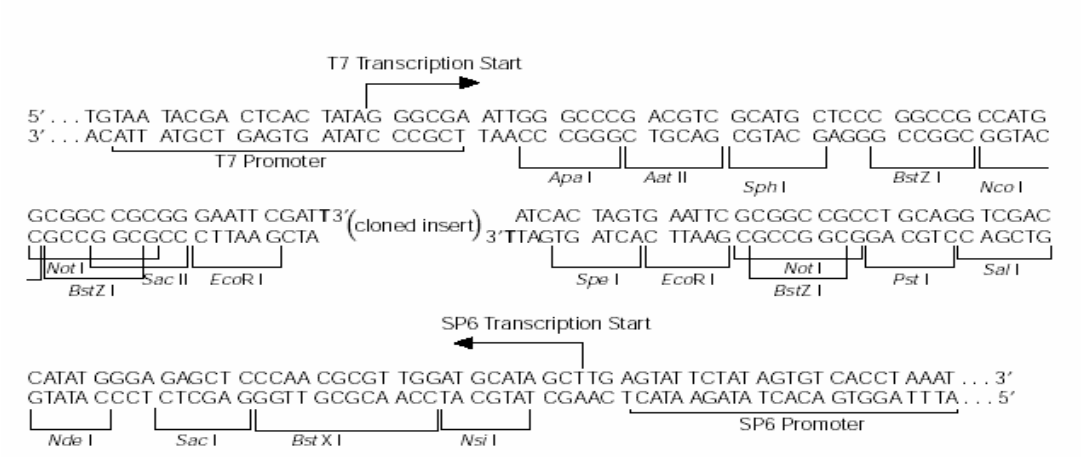


Figure 2 Map of pGEM-T Easy vector. pGEM-T Easy vector circle map(A). Promoter and multiple cloning sequence of the pGEM-T Easy vector (B). The top strand of the sequence corresponded to the RNA synthesized by T7 RNA polymerase. The bottom strand corresponded to the RNA synthesized by SP6 RNA.

Source: Promega (1998)

2.4.3 Plasmid extraction and restriction pattern analysis

Plasmid was extracted by the alkaline lysis method of Sambrook and Russell, (2001). The transformed white colony which could be amplify by PCR was picked to culture in a test tube containing 1.5 ml amplicinlin-LB media at 37°C for overnight and cells were harvested, transferred to a new 1.5 microcentrifuge tube and centrifuged at 4°C, 5,000 rpm for 3 minutes. The pack cells were resuspended in 100 µl of solution I, mixed well by vortex and kept on ice for 15 minutes. Two hundred microlitres of freshly prepared solution II was added, mixed by inversion and kept on ice for 5 minutes and centrifuged at 4°C, 12,000 rpm for 5 minutes, the supernatant was transferred to a new 1.5 ml microcentrifuge tube and extracted by equal volume of phenol: chloroform: isoamyl alcohol (25:24:1), was well mixed by vortex. The two phases were separated by centrifuged at 4 °C, 12,000 rpm for 10 minutes and the upper phase (DNA fraction) was transferred to a new microcentrifuge tube. Two volumes of 95 % ethanol were then added and the tube was kept at -20 °C for 20 minutes. The plasmid was pelleted by centrifugation at 4 °C, 12,000 rpm for 10 minutes. The pellet was washed with 70% ethanol, centrifuged at 4°C 12,000 rpm for 10 minutes and air dried. The plasmid was finally dissolved in TE buffer and stored at 4 °C until used.

The analysis of DNA fragments was generated by restriction endonuclease treatment. After the digestion of plasmid DNA, the fragments were determined by 1.5 % agarose gel electrophoresis to evaluate the size of the recombinant plasmid, the original vectors and the insert fragment was indicated. Finally analysis, The nucleotide sequence was confirmed.

3. High expression of MLL using pET vector

3.1 pET33b and pGEM-MLL plasmid extraction

NucleoSpin plasmid kit was used for pET33b (Figure. 3) and pGEM-MLL plasmid extraction. The overnight *E.coli* cells culture which contained pET33b and

pGEM-MLL plasmid were centrifuged at 12,000 rpm for 30 second. The media was discarded and 250 µl of A1 buffer was added. After cell pellet was re-suspended by vigorous vortex, 250 µl of A2 buffer was added into the tube. The tube was mixed gently by inverting 6-8 times and incubated at room temperature for 5 minutes. Then, 300 µl of A3 buffer was added and the tube was mixed gently by inverting 6 - 8 times. The supernatant was collected after the tube was centrifuged at 12,000 rpm for 5-10 minutes at room temperature and loaded into the NucleoSpin plasmid column. The DNA was bound to the membrane by centrifugation the tube at 11,000 rpm for 1 min. The plasmid was washed by adding 600 µl of A4 buffer containing ethanol and the column was centrifuged for 1 min at 11,000 rpm. The membrane was dried by centrifuged for 2 min at 11,000 rpm. Finally, 15-50 µl of elution AE buffer was added into the center of membrane. The column with elution buffer was incubated to increase yield of DNA at room temperature for 1 min. The incubated column was centrifuged for 1 min at 11,000 rpm. The plasmid was collected and stored at 4 °C.

3.2 Preparation of pET33b and MLL gene fragment

*Nhe*I- *Xho*I double digests were performed both for pGEM-MLL and pET33b plasmid. One hundred microlitre of reaction mixture containing 3 µg of plasmid, 10x restriction enzyme buffer, BSA and 10-20 U of *Nhe*I-*Xho*I enzymes was incubated at 37 °C for 2-4 h. Enzymes were heat to inactivate at 65 °C for 15 minutes. DNA fragment was analyzed by agarose gel electrophoresis.

3.3 Extraction of DNA fragment from agarose gels

NucleoSpin Extract II gel extraction kit was used for extraction of DNA fragment from agarose gel. The DNA fragment from agarose gel was carefully excised to minimize the gel volume. The weight of the gel slice was determined and the gel was transferred to a clean tube. Two hundred microlitres of NT buffer was added for each 100 mg of agarose gel. The tube was incubated for 5-10 minutes at 50 °C until the gel slice was completely dissolved. The solution was loaded into the

column. The DNA was bound to the membrane by centrifugation the tube at 11,000 rpm for 1 min. The membrane was washed by adding 600 µl of NT3 buffer and the column was centrifuged for 1 min at 11,000 rpm. The membrane was completely dried by centrifuged for 2 min at 11,000 rpm. Finally, 15-50 µl of elution NE buffer was added. The column with elution buffer was incubated to increase yield of DNA at room temperature for 1 min. The incubated column was centrifuged for 1 min at 11,000 rpm. The DNA was collected. After gel purification, the DNA was re-suspended in the appropriate volume of TE buffer that in order to adjust the plasmid concentration to be the same as the insert DNA concentration (~0.1µg/µl).

3.4 Construction of recombinant pET-MLL

For ligation, the ideal insert DNA-to-vector ratio of DNA is variable; however, a reasonable starting point is 2:1(insert DNA-to-vector molar ratio). The control and experimental 10 µl of ligated reactions were prepared by adding the components: 10x ligase buffer, 100 mM DTT, 10 mM ATP, 50 ng/µl of pET33b vector, 50 ng/µl of insert DNA and T_4 DNA ligase into sterile 1.5 ml microcentrifuge tubes. The reactions were incubated for overnight at 16°C.

3.5 Transformation of ligated product into *E.coli* BL21(DE3)

Two microlitres of pET-MLL ligated product and 50 µl of *E.coli* (JM 109) competent cells were added to a sterile 1.5 ml tube on ice. The tube was gently flicked to mix, and placed on ice for 30 minutes. Then, the tube was incubated for 45 seconds to 2 minutes in a water bath at exactly 42 °C and immediately put back on ice for 2-10 minutes. Nine hundred and fifty microlitres of room temperature SOC media was added to the tubes and incubated 1-1.5 h at 37 °C with 250 rpm shaking. One hundred microlitres of transformation culture were plated onto 30 µg/ml kanamycin plates. The plate was incubated overnight (16-24 h) at 37 °C.

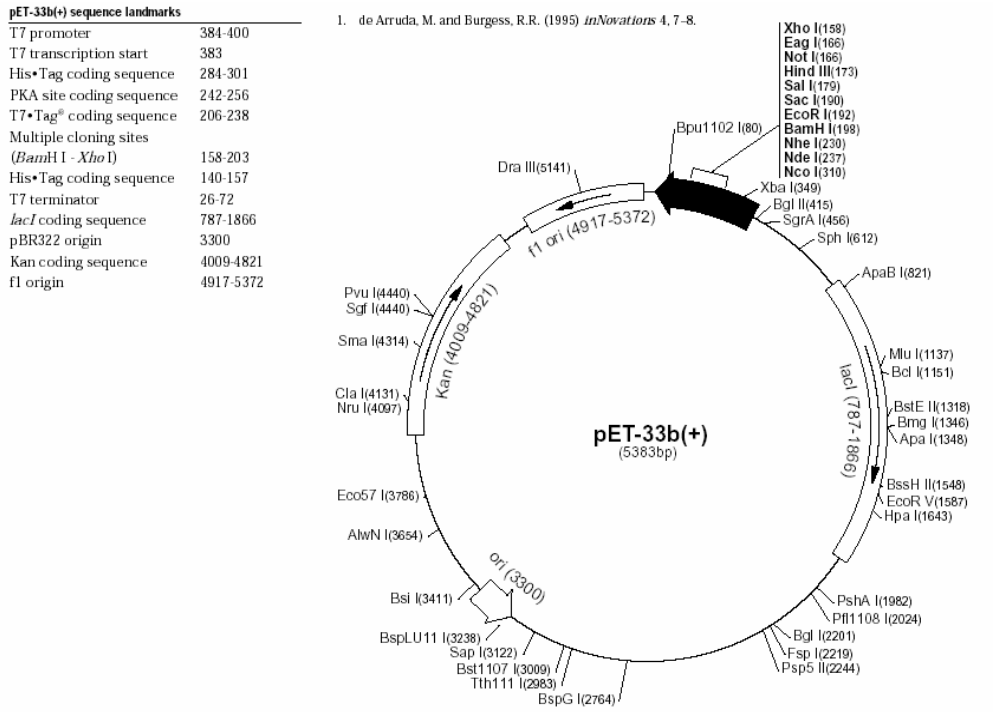
3.6 Characterization of positive clone

For characterization of pET-MLL, bacteria cells were first checked directly by colony PCR. The amplified colony was cultured and the plasmid was extracted as described in section 2.4.3. The extracted plasmid was then digested by both *Nhe*I-*Xho*I restriction endonuclease. After the digestion of plasmid DNA, the fragment was determined by 1.5 % agarose gel electrophoresis, compared the size with 100 bp ladder marker. Then, DNA sequencing was performed to make sure that there is no mutation.

3.7 Cultivation of recombinant clone

MLL was engineered in the form of 6xHis fusion from the prokaryotic pET expression vector system. This vector has a strong T7 promoter, and is designed to work with *E.coli* BL21 (DE3) strain. Thus, this *E.coli* strain was used as an expression host cells for express MLL protein. The preparation of competent cells and transformation method followed 2.2. The entire transformation reaction was plated onto 30 μg/ml kanamycin agar plate. The plates were incubated overnight at 37 °C. After selection the putative MLL recombinant clones, the single colony was pre-cultured overnight at 37 °C in 20 ml LB media containing 30 mg/ml of kanamycin and added to 200 ml of LB media at 37 °C. When the OD_{600} nm reached about 0.6, the cells were induced by adding IPTG to 1 mM final concentration. The culture was harvested by centrifugation at 5,000 rpm for 30 minutes. The cell pellet was suspended in 20 ml of resuspension buffer (20mM phosphate buffer, pH 8.0). The cells were disrupted by sonicator and centrifuged at 14,000 rpm for 10 minutes. The protein concentration of cell lysate was measured by Lowry's method. The inducible protein was analyzed by SDS PAGE and a sample from non-induced cell was used as control.

A



B

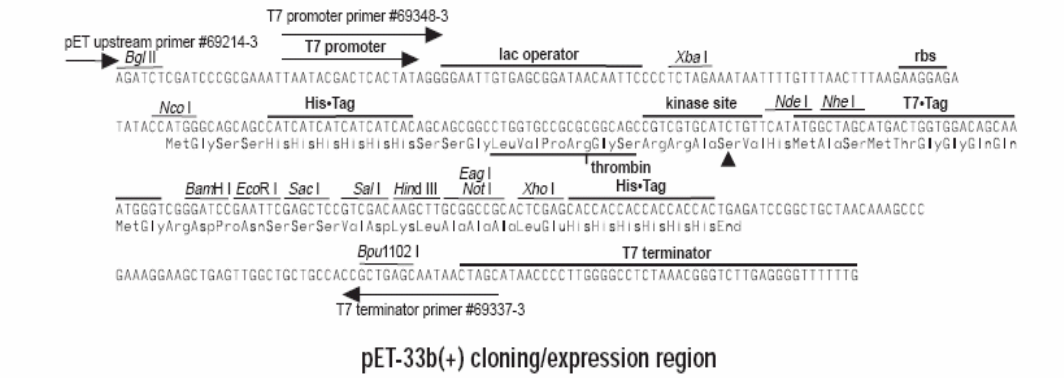


Figure 3 Map of pET33b vectors. (A) pET33b vector circle map (B) Promotor and multiple cloning sequence of the pET33b vector. The top and middle strands of the sequence corresponded to T7 promoter, lac operator, multiple cloning sites with restriction sites indicated, 6xhis tag sequence and kinase site. The bottom strand corresponded to the RNA synthesized by T7 terminater.

Source: Novagen (1998)

3.8 Optimization of gene expression

3.8.1 Inducing time

Cells containing the recombinant plasmid were cultured until OD_{600} reached about 0.6. Subsequently the cells were induced with 1 mM IPTG at 37 °C. The inducing time was varied, the culture was harvested at 1, 2, 3, and 4 h respectively after induced with IPTG. Cells were lysed and screened for MLL.

3.8.2 IPTG concentration

Cells containing the recombinant plasmid were cultured and induced with IPTG as described previously. In this experiment, the concentration of IPTG was varied to a final concentration of 0.2, 0.5, and 1.0 mM respectively**.** The culture was harvested after 4 h and cells were lysed and screened for MLL.

3.8.3 Temperature for expression

The recombinant cells were cultured were cultured and induced with IPTG as described previously but the inducing temperature was varied. After induced with 1mM IPTG, cells were cultured at 25 and 37 °C respectively. The culture was harvested after 4 h and cells were lysed and screened for MLL.

3.8.4 OD of cell culture in initial induction

Cells containing the recombinant plasmid were cultured at 37 °C. In this experiment, the absorbance of OD_{600} was varied. The cultures were induced with 1mM IPTG when OD_{600} reached about 0.3, 0.6, 0.9 and 1.1. The culture was harvested after 4 h and cells were lysed and screened for MLL.

3.8.5 The basal expression level

Adding high amount of glucose into the growth medium will decrease the basal expression level associated with the upstream *lac* promoter but will not affect basal level expression from the *tac* promoter. The presence of glucose should not significantly affect overall expression following the induction with IPTG. In this experiment, 0.4% glucose and 0.5 mM IPTG was added into the culture which was cultured at 37 °C until OD_{600} reached about 0.6. The culture was harvested after cultured at 25 °C for overnight. Cells were lysed and screened for MLL

4. Determination of recombinant protein

4.1 Protein measurement

Protein concentration was determined by Lowry's method (Lowry, 1951) which modified by Pierce Chemical Co. Technical Bulletin. To make a standard curve, a serie of tubes containing standard BSA solution was prepared in 1.2 ml total volume. Tubes containing 1.2 ml of appropriately diluted protein solution were prepared. Six hundred microlitres of Alkaline Copper Reagent was freshly prepared within 1 h of using, was immediately mixed well and waited for 10 min. Two millilitres of Folin-Ciocalteu Reagent were added and immediately mixed thoroughly. After 30 min, the absorption at 750 nm was measured. The standard curve was plotted and the concentration of protein was determined.

4.2 Tricine–SDS-PAGE

Tricine–SDS-PAGE is commonly used to separate proteins in the mass range of 1–100 kDa. It is the preferred electrophoretic system for the optimal separation of proteins <30 kDa (Schagger, 2006).

4.2.1 Preparation of separating gel

The amounts of reagents required for 12% acrylamide separating gel containing 1.8 ml gel monomer, 1.5 ml gel buffer, 0.72 ml distilled water. Freshly prepared ammonium persulfate (APS) solution and TEMED should be added last, immediately before pouring to the gel cassettes. The poured gel was overlaid with several drops of water. The gel was left for about 30 min to polymerize.

4.2.2 Preparation of stacking gel

After the separating gel was polymerized, the layer of water was removed. The polymerized separating gel was overlay directly with a 4% sample (stacking) gel, prepared by mixing 0.25 ml gel monomer, 0.5 ml gel buffer and 1.25 ml distilled water. Freshly prepared ammonium persulfate (APS) solution and TEMED should be added last. The comb was carefully inserted into gel sandwich. After the stacking gel was polymerized, the comb was carefully removed.

4.2.3 Sample preparation and protein loading

Protein concentration was adjusted so that a suitable amount of protein could be loaded onto the gel. The sample was preferentially concentrated by techniques that do not increase the salt concentration such as roughly 0.2–1 ng of protein for each protein band was sufficient for Coomassie staining. Accordingly, the desired protein concentration in the sample was 0.1 mg/ml for each protein band. The samples was mixed with SDS-containing 2x sample buffer and heated at 95 °C for 5 min. The gel was mounted in the vertical electrophoresis apparatus and anode buffer was added as the lower electrode buffer and cathode buffer as the upper electrode buffer. The samples were loaded under the cathode buffer.

4.2.4 Electrophoresis conditions

Gel was started electrophoresis with an initial current may be as high as 80 mA and maintained at this voltage until the sample has completely entered the stacking gel. The next appropriate voltage step can then be applied (30mA).Gels may warm up, but the temperature should not exceed 35–40 °C.

4.2.5 Protein visualization

Protein can be visualized directly in the gel by Coomassie staining. Before staining, the gel was incubated in fixing solution for 15 min. The gel was stained with staining solution for 30 min. The gel was then destained twice in destaining solution. Incubation should be least 15–60 min. The gel was transferred to water.

5. MLL purification

5.1 Preparation of crude protein

Escherichia coli BL21 (DE3) cells harboring the desired plasmid were cultured at 37 °C in 200 ml LB media containing 30 μg/ml kanamycin. When the OD_{600} reached about 0.4–0.7, IPTG and glucose was added into the culture to a final concentration of 0.5 mM and 0.4%, respectively. The culture was harvested, the cell pellet was suspended in 10 ml of resuspension buffer (phosphate buffer pH 8.0, 1mM EDTA and 1mM of 2-mercaptoethanol). The cells were disrupted by sonicator and centrifuged at 10,000 rpm for 30 min. The supernatant (crude extract) were collected and stored at 4 °C for further analysis and purification. The pellet (inclusion bodies) was washed twice with resuspended buffer and solubilized by adding 10 ml of resuspended buffer containing 8 M urea. Insoluble materials were removed by centrifugation at 10,000 rpm for 30 min. The supernatant was subjected to further purification.

5.2 Column chromatography purification

Both soluble and inclusion bodies fraction was subjected to ion-exchange and affinity column chromatography.

5.2.1 Anion exchange column chromatography

The crude protein was applied to DEAE-cellulose ion exchange column which was equilibrated by resuspended buffer (50 mM phosphate buffer, pH7.4, 1 mM EDTA, 1 mM mercaptoethanol). The contaminating proteins were removed with resuspended buffer. To purify inclusion bodies, initial refolding were carried out using a modification of the procedures described by Kassab*, et al.* (2004). Refolding was performed on the column by exchange to non-denaturing buffer conditions with a linear 8–0 M urea gradient before elution of the bound protein. The recombinant MLL was eluted with the resuspended buffer containing 0 - 0.5 M NaCl. The elution profile was monitored for protein by measuring the absorbance at 280 nm. The fractions were collected and submitted to 12 % tricine-SDS–PAGE.

5.2.2 Affinity column chromatography

Ni^{2+}-NTA affinity column chromatography (2 ml bed volume) performed with His-Bind resin modified to the manufacturer's instructions (IMAC).The fractions containing MLL protein were pooled and dialyzed against the binding buffer(50 mM phosphate buffer pH 7.4, 0.15M NaCl, 10 mM imidazole), before applied to a nickel affinity, the column was equilibrated with the same buffer. Unbound protein was removed by the washed buffer (50 mM phosphate buffer pH 7.4, 0.15M NaCl, 30 mM imidazole). The bound protein was eluted with the PBS buffer containing 50-350 mM imidazole and different pH buffer, (pH 6.0 and pH 4.0) stepwise. The fractions were collected and submitted to 12 % tricine-SDS–PAGE and analyzed for hemagglutination activity.

6. Characteration of MLL

6.1 Hemaglutination activity

The hemagglutinating assayed for recombinant MLL. Twenty-five microlitres of recombinant protein was serially twofold diluted on a 96-well microtitre U-shaped plate. Twenty-five microlitres of 2% suspension of rabbit erythrocytes was added .The agglutination was visually monitored after 45 min at room temperature. The activity was expressed as the minimum concentration of protein that promoted visible hemagglutination. (Ratanapo *et al*. 1998)

7. Molecular docking

The binding energies between the flexible ligands and the large acceptor molecule were calculated using AUTODOCK 3.0.5, an atom-based docking simulation program. To accomplish a quick energy evaluation, AUTODOCK precalculates grid-based molecular affinity potentials. For each atom type in the substrate molecule, a different three-dimensional grid map is generated by assigning the energy of interaction of a single probe atom to the grid point. Similarly, a grid for the electrostatic potential is calculated. The ligand is defined by a rigid root, from which rotatable bonds sprout. The program features different search methods (simulated annealing, genetic search algorithms, and local search), from which the Lamarckian genetic algorithm was selected. This algorithm starts the docking process with a random population of a limited number of individuals. These individuals represent molecules with uniformly distributed random values for torsion angles, quaternions, and translation vectors. The values for torsion angles, quaternion, and translation vectors represent the genes of an individual and can be inherited by the upcoming population (Neuman *et al*., 2002)

7.1 Preparing the ligand files

Several known ligands were chose for docking experiments. These ligand found to inhibit red blood cell agglutination induced by MLL have been reported in the literature, namely a *N*-glycolyl- neuraminic acid (Neu5Gc), *N*-acetyl- neuraminic acid (NeuNAc), N-acetyl-b-D-galactosamine (GalNAc), glucose (Glu), Galactose (Gal), Mannose (Man). All ligand structures were built using a SYBYL Program (Tripos Inc.). These structures were optimized by quantum-mechanical calculations of increasing accuracy and calculated electrostatic potential derived charges. Then all hydrogen atoms were computed by a Gasteiger charge method. Finally, the final ligand files were saved as *.out.pdbq for input into AutoDock Program.

7.2 Preparing the macromolecule file

The three-dimension structure of MLL was taken from the GENO3D. The artificial intelligence based server at http://geno3d-pbil.ibcp.fr/cgi bin/geno3d_automat.pl?page=/ GENO3D/geno3d_home.html was also used to create homology models. The homology file was checked before used it as an input macromolecule to make sure that all missing residues were fixed by DeepView Only polar hydrogen atoms and solvation parameters were added to macromolecule by Kollman United Atom charge method. The macromolecule file was written in a *.pdbqs file for used by AutoGrid.

7.3 Docking

In order to model hydrogen bonds, polar and nonpolar hydrogens were used for both ligand and receptor resulting in two different hydrogen maps. Parameters for pairwise atomic interaction energies were taken from the AUTODOCK 3.0.5 User Guide according to Morris *et al*, (2002). During the docking process, the torsion angles for ring systems were kept fixed, while all other groups were freely rotatable to allow the formation of as many hydrogen bonds as possible. To assess the suitability of for determining the binding site for carbohydrates, all

ligands were subjected to 100 docking runs using large grid maps of 40 x 50 x 50. Each ligand conformation of the best interaction energy was saved as a result of each run. This grid map was large enough to allow the ligand to rotate freely even when the ligand was in its most fully extended conformation. The grid map was at position which was assumed as a binding site. The grid parameter file was written in a *.gpf file and launched by AutoGrid, a subset of AutoDock Program. The appropriately genetic algorithm specific parameters and Docking Run Parameters were set. The name of the *.dpf file was written. This file would contain docking parameters and instructions for a Genetic Algorithm-Local Search (GA-LS) docking, known as the Lamarckian Genetic Algorithm (LGA). The AutoDock job was started and the result file had the extension *. dlg.

7.5 Analyzing AutoDock Results

Reading a docking log or a set of docking logs was the first step in analyzing the results of docking experiments by conventing dlg files to be pdb files. The best docking result out of 100 runs was considered from the final docked energies which is equal to the summary of the Final Intermolecular Energy and the Final Internal Energy of the ligand.

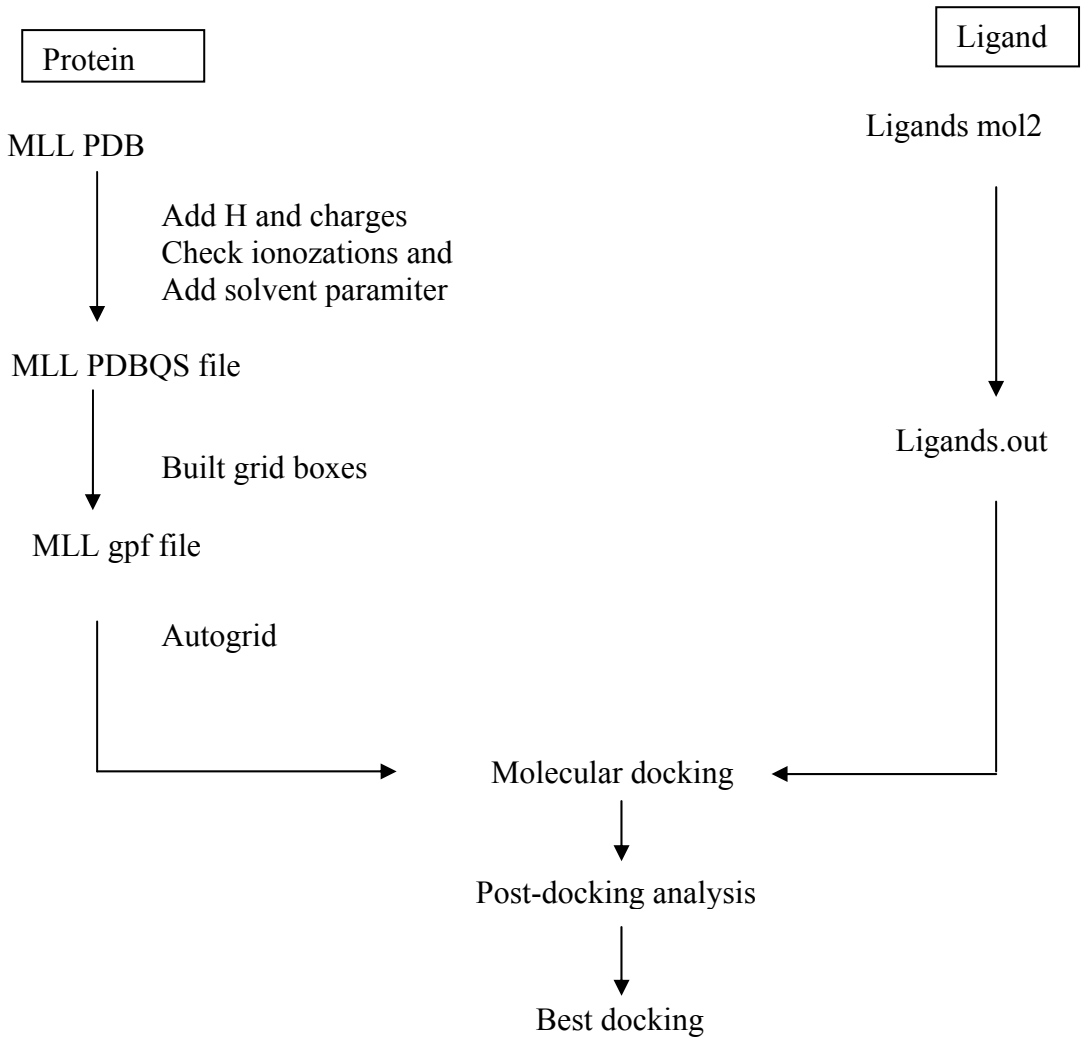


Figure 4 Molecular docking steps.

RESULTS

1. Preparation of MLL

1.1 Amplification, purification and characterization of MLL gene

The MLL coding region was amplified by PCR using MLL20 clone (Kankamol *et al*., 2003) as a template using primers which were designed containing *Nhe*I and *Xho*I restriction sites; F5´*Nhe*I (5´GGAAGCTAGCATGCTGATGGCTT CGTC 3´) and F5´ *Xho*I (5´CCTTACTCGAGGCTGCCGCGGCCCACCAGAAG TGCTAAGTGAAACCCAATG 3´).The PCR optimization was done. The optimal annealing temperature was 50°C. The amplification gave PCR product about 670 bp was purified by using condition as described in section 1.1.2 and analyzed by 2 % agarose gels electrophoresis (Figure 5). The purified PCR product was kept for further experiments.

2. Cloning MLL using pGEM-T Easy vector

The purified PCR product was ligated with pGEM-T Easy vector as described in section 2.1. Two microlitres of ligated products were used for transformation into *E. coli* JM 109 competent cells by heat shock method. Subsequently, ten colonies from total ninety colonies were randomly selected and analyzed.

2.1 Screening and characterization of positive clone

For screening recombinant colonies, ten white colonies on selected plate containing amplicilin, X-gal and IPTG were assumed to contain putative recombinant plasmids were randomly picked for direct colony PCR. The result in Figure 6 shows strong bands of about 670 bp fragments accept lane 7 but lane 1, 2, 3, 4 and 5 have unexpected bands. Therefore, the colony which showing a single band at the expected size band (lane 10) was picked to culture in amplicinlin-LB at 37 °C for overnight

and harvested for plasmid extraction. The selected plasmid was linearized digestion by *Nhe*I had higher molecular weight than pGEM-T plasmid. Moreover, linear pGEM-T and about 670 bp fragment were obtained by *Nhe*I and *Xho*I double digested. The result is shown in the Figure 7.

For further verification, the recombinant plasmid, was called pGEM-MLL, was sequenced using SP6- forward and T7- reverse primer system. The sequence had 100% identity compared with MLL1 cDNA (Kankamol *et al*., 2003). The fragment contained 672 nucleotides with 5′*Nhe*I restriction site at start ATG codon and 3′*Xho*I restriction site at stop codon. (The sequence data not show).

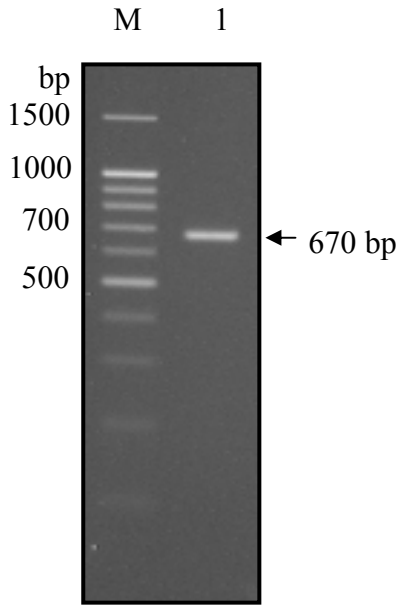


Figure 5 PCR product from MLL20 clone. PCR amplification started with predenaturing at 94°C for 4 min, amplification was 30 cycles with denaturation at 94°C for 30 seconds, annealing at 50°C for 1 min and extension at 72 °C for 1 min. PCR product about 670 bp (lane 1) and DNA ladder 100 bp + 1.5kb marker (lane M) was subjected to 2% agarose gel electrophoresis.

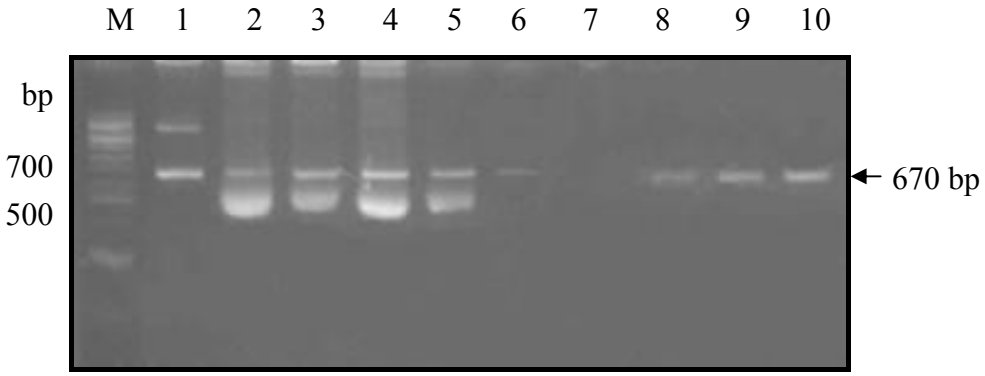


Figure 6 Randomly colonies PCR. Picked white colonies on selected plate containing amplicilin, X-gal and IPTG were assumed to contain putative recombinant plasmids for amplification by PCR. Randomly white colonies PCR (lane1-10) were subjected to 2% agarose gel electrophoresis compared with DNA ladder 100 bp + 1.5kb (lane M)

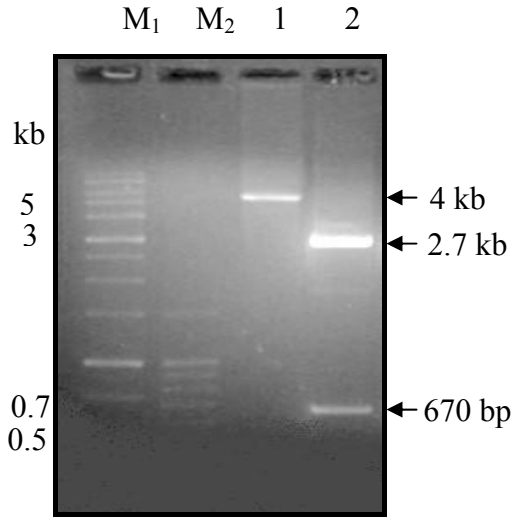


Figure 7 Plasmid extraction and digestion. The colony which shown strong expected size band was picked for plasmid extraction, linearized with *Nhe*I (lane 1) and double digested with *Nhe*I and *Xho*I (lane 2) were subjected to 1.5% agarose gel electrophoresis compared with DNA ladder 100 bp + 1.5kb (lane M_1) and 10 kb marker (lane M_2).

2. Cloning and expression of MLL using pET vector

2.1 Cloning of MLL gene using pET33b

pET33b plasmid and the correctly selected clone plasmid which contained pGEM-MLL were digested by *Nhe*I- *Xho*I double digestion. The digested fragments were purified for ligation. The ligation products were used for transformation into *E. coli* JM 109 competent cells by heat shock method. The colonies were randomly selected, analyzed and sequenced. The correct recombinant plasmid was called pET-MLL and then transformed into *E.coli* BL21 (DE3) strain for protein expression.

2.2 Characterization of positive clone

Direct selection of recombinant provides screening of colonies. Cells containing pET33b plasmid can grow on kanamycin selection plate due to its kanamycin resistant gene. The bacterial colonies were randomly picked for direct colony PCR. The result is shown in the Figure 8, the colony 2, 8 and 10 show strong expect size PCR band For further experiment, colony 2 and 8 were picked and cultured in kanamycin-LB broth at 37 °C for overnight. Then, the plasmid was extracted and linearized by *Nhe*I digested and digested fragment had higher molecular weight than wild type pET33b plasmid (Figure 9 A). Moreover, it could be *Nhe*I and *Xho*I double digested to yield linear pET33b and about 670 pb fragments. (Figure 9 B)

For further expression, the recombinant plasmid pET-MLL was sequenced using T7 promoter - forward and T7 terminator - reverse primer system to confirm the correction of the sequence of gene encoding for MLL. The amplified fragment contained 670 nucleotide composing of 5´*Nhe*I restriction site, start ATG codon, thrombin, His tag and 3´ *Xho*I (The sequence data not show). The gene encoded for 250 amino acids included of thrombin and C and N-His tag terminus fragments. The sequence was aligned with cDNA of MLL20 by using ClustalW program. The molecular mass of deduced amino acids was predicted by using

program computed pI/MW on website http://br.expasy.org/tools/pi_tool.htm. The result is shown in the Figure 10. The expected theoretical pI and MW were 8.70 and 28 kDa respectively.

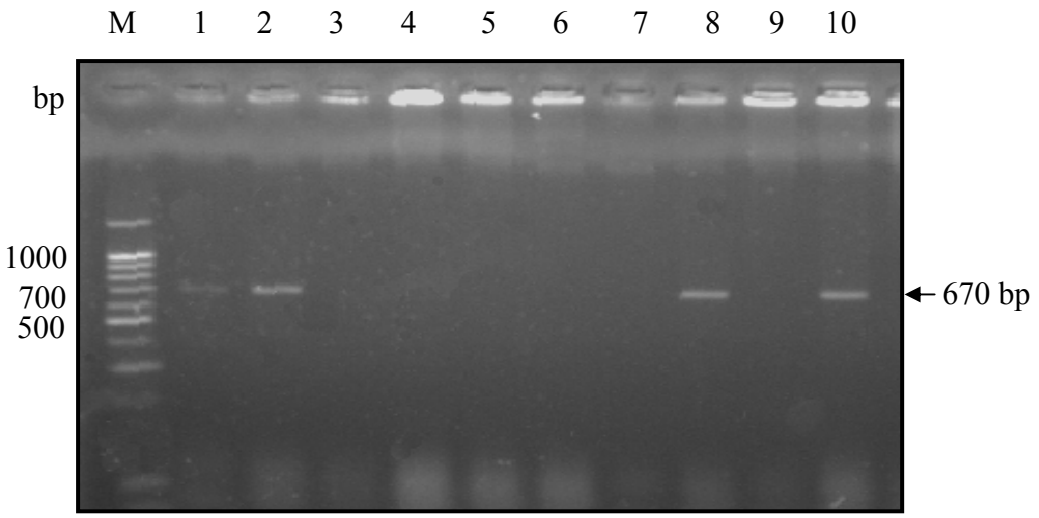


Figure 8 Randomly colonies PCR. Colonies which grown on kanamycin selected plate was randomly picked, assumed to contain putative recombinant plasmids for amplification by PCR (lane1-10) were subjected to 2% agarose gel electrophoresis compared with DNA ladder 100 bp + 1.5kb (lane M).

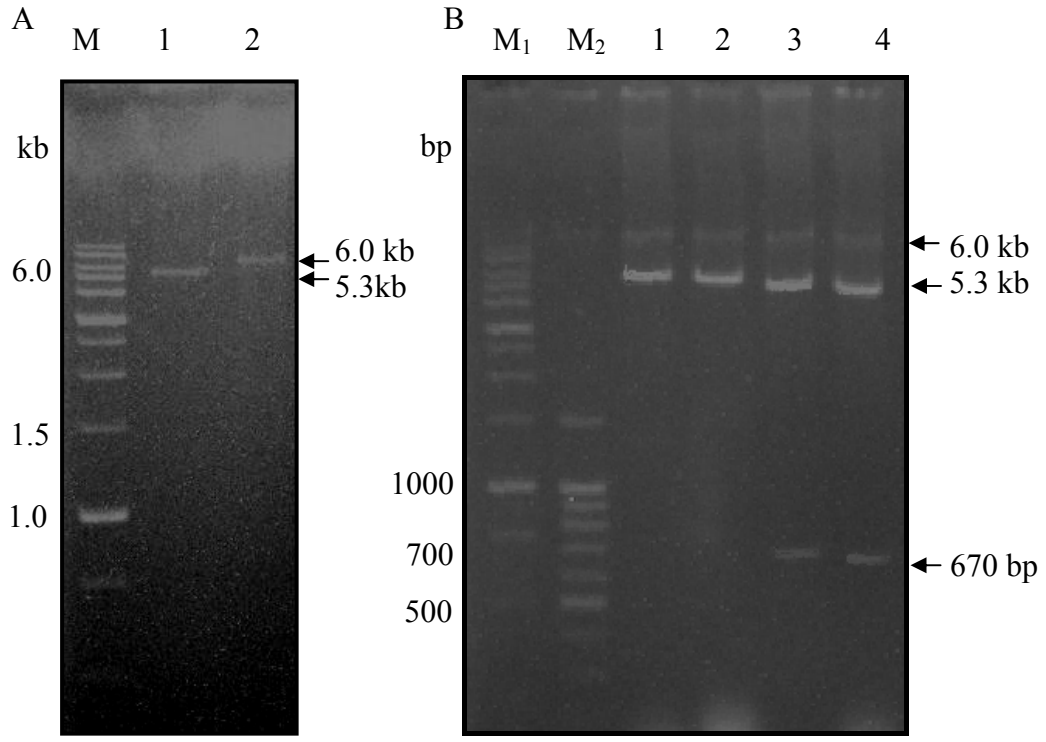


Figure 9 Plasmid extraction and digestion.

(A) pET33b and recombinant MLL plasmid, digested with *Nhe*I plasmid without transformation. (lane 1), pET-MLL plasmid digested by *Nhe*I (lane 2). Digested DNA were compared with DNA ladder (10 kb) (lane M).

(B) Recombinant MLL plasmid (colony 2 and 8) were digested by *Nhe*I (lane 1, 2) and double digested by *Nhe*I and *Xho*I (lane 3, 4). DNA ladder (10 kb) (lane M_1) and DNA ladder 100 bp + 1.5kb (lane M_2).

Compute pI/Mw

Theoretical pI/Mw (average) for the user-entered sequence:

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSR RASVHHAMLM ASSSFLSLSF LVLLFSISSA NTRKWSLSNV

        70         80         90        100        110        120
LDQKPISIIE AAIGVSEDLL NLNGMEAKNN QQSGKSQTIV VGTWGAQATS SNGVAFDDGS

       130        140        150        160        170        180
YTGIREINFE YNNETAIGSI QVTYDVNGTP FEAKKHASFI TGFTPVKISL DFPSEYIVEV

       190        200        210        220        230        240
SGYTGKVSGY IVVRSLTFKT NKETYGPYGV TSGTHFKLPI QNGLIVGFKG SVGYWLDYIG

    250
FHLALLVPRG SHHHHHH
```

Theoretical pI/Mw: 8.70 / 28002.52

Figure 10 The deduced amino acids sequences of MLL gene clone from cDNA of MLL20. The expected theoretical pI was 8.70 and MW was 28 kDa. Computed by http://br.expasy.org/tools/pi_tool.htm.

2.3 Induction of pET-MLL in *E.coli*

The pET33b construct carries the T7 promoter. After a target plasmid was established in a λDE3 lysogen (BL21DE3). The expression of the target DNA could be induced by adding IPTG (Novagen, 1999). The inducible protein was analyzed by SDS PAGE and visualized by Coomassie Brilliant Blue dye. The result in Figure 10 is shown that, the recombinant protein fragment was observed at about 20 kDa in induced recombinant cell but not in cell containing pET33b plasmid and non-induce recombinant cell. The recombinant protein was expressed in 2 forms, soluble and insoluble form. However, most of the protein was in insoluble form. The refolded recombinant protein partially aggregates accordingly expressed in inclusion bodies. The optimization of induced condition to gain high amount of recombinant was done by varied induction time, IPTG concentration, temperature for expression and OD of cell culture in initial induction .

2.3.1 Inducing time

The recombinant cells were cultured as previous described. The inducing time was varied and the culture was harvested at 1, 2, 3, and 4 h, respectively after induced by IPTG .The expressed protein was submitted to 12% SDS–PAGE. The Figure 12 shows that the recombinant MLL expression could be observed after induced with IPTG for 1 h. The highest of recombinant MLL expression was at 2 h of inducting time and was maintained at the same levels with 3-4 h. Therefore, 2 h induction time for recombinant protein expression would be used in any further experiment.

2.3.2 IPTG concentration

The recombinant cells were cultured and induced with IPTG to a final concentration of 0.2, 0.5, and 1.0 mM, respectively. The expressed protein was then submitted to 12% SDS–PAGE. The Figure 13 shows that the recombinant MLL protein was expressed at high level when 0.5 mM or 1.0 mM IPTG was used but less amount when induced by 0.2 mM IPTG. Therefore, in the further expression, 0.5 mM of IPTG would be used to induce the recombinant MLL protein.

2.3.3 Temperature for expression

In this experiment, the inducing temperature was varied. After induced with 1 mM IPTG, cells were induced at 25 and 37 °C, respectively. The expressed protein was then submitted to 12% SDS–PAGE. The Figure 14 shows that the recombinant MLL protein mainly appeared in insoluble fraction of 37 °C cells culture and the protein band was stronger than induced at 25 °C.

2.3.4 OD of cell culture in initial induction

The recombinant cells were cultured at 37 °C and 1 mM IPTG was added when OD_{600} reached about 0.3, 0.6, 0.9 and 1.1 respectively. The culture

was harvested after 4 h. The expressed protein was submitted to 12% SDS–PAGE. The Figure 15 shows that the recombinant MLL protein was found to be induced at OD_{600} of 0.3 and high amount when induced at OD_{600} of 0.6, 0.9 and 1.1. In the further expression, to add IPTG for induction of the recombinant MLL protein would be started at OD_{600} about 0.6 because the amount of recombinant MLL protein was relatively high and less contaminant was observed.

From the result, the optimal expressed condition to achieve high amount of recombinant protein was induced cell by 0.5 mM IPTG at OD_{600} of 0.6 for 2 h at 37 °C. When induced cell by optimal condition, the recombinant protein was mostly expressed in the inclusion bodies which were inactive form and hard to purified because it like to aggregate. In order to avoid the expression of recombinant MLL as inclusion bodies, the factor to decrease the expression level was considered as adding catabolite repressor in inducing cultured.

2.3.5 The basal expression level

The recombinant cells were cultured at 37 °C, 0.4% glucose and 1 mM IPTG was added into the culture when OD_{600} reached about 0.6. The culture was harvested after cultured at 25 °C for overnight. Cells were lysed and screened for MLL. As expected, most protein was contained in the form of soluble fraction (data not shown).

For protein purification, the recombinant cells would be induced by 0.5 mM IPTG and 0.4% glucose when OD_{600} reached 0.6 for overnight at 25 °C. This was optimal condition to yield the high amount of MLL protein and mostly soluble protein was obtained.

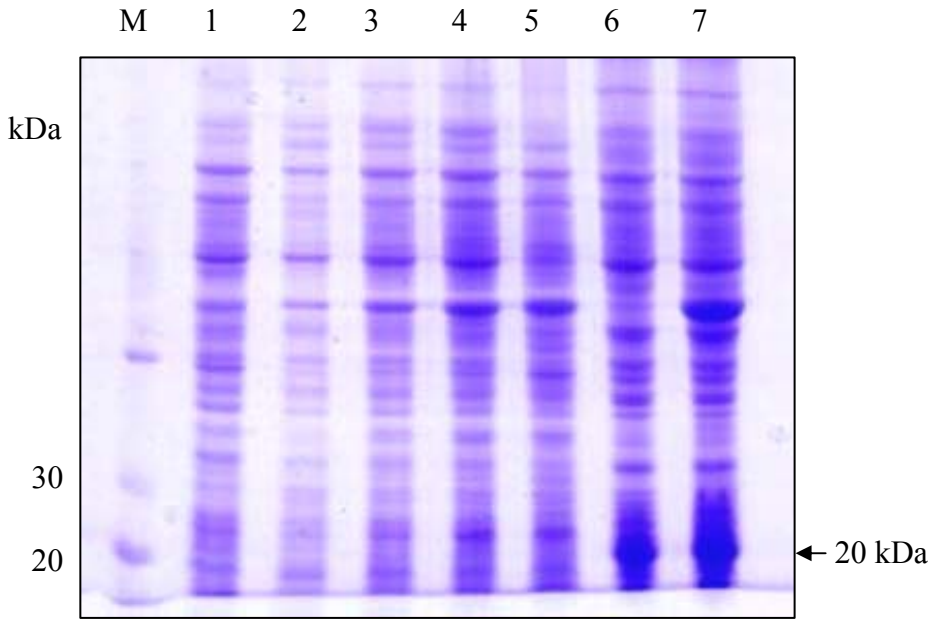


Figure 11 Analysis of recombinant MLL. Cell extract of BL21 (DE3) after transformation with pET33b plus insert and induced by IPTG for 4 h was submitted to 12% SDS–PAGE. pET33b (lane 1), pET-MLL at OD_{600} reached 0.5,0.9, 2.3 and was cultured overnight (lane 2-5), Soluble protein from pET-MLL (induced with 1 mM IPTG for 4 h at OD_{600} = 0.6) (lane 6), Inclusion body protein from pET-MLL (induced with IPTG 4 h at OD_{600} = 0.6) (lane 7) are compare with low molecular weight protein marker (lane M)

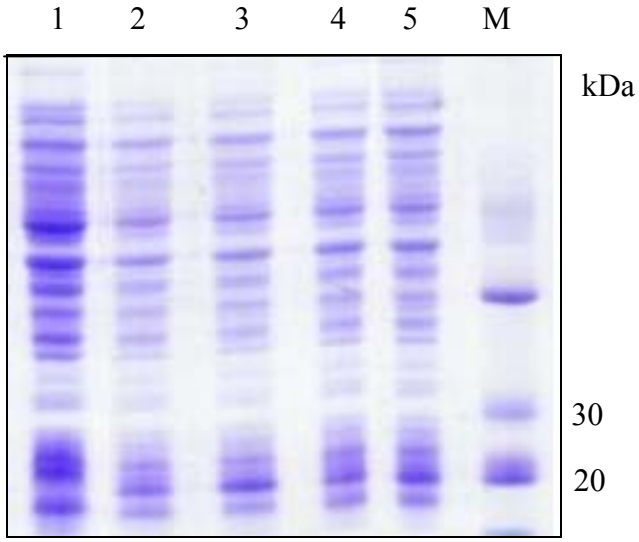


Figure 12 Induction of MLL expression by various inducing times. The recombinant cells were harvested after induced by 1 mM IPTG 1, 2, 3, and 4 h (lane 2-5, respectively). pET-MLL non induced by IPTG (lane 1) are compared with low molecular weight protein marker (lane M)

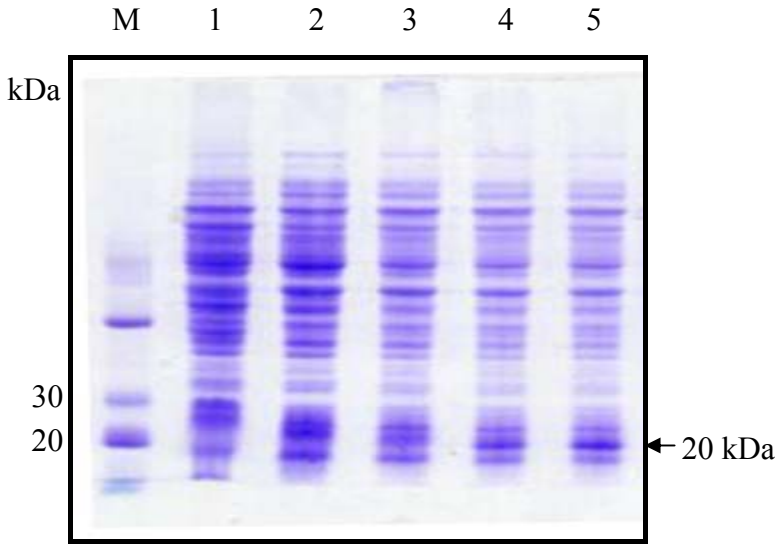


Figure 13 Induction of MLL expression by various IPTG concentrations. pET33b (lane 1), pET-MLL with 0 mM IPTG (lane 2), pET-MLL were induced with IPTG to a final concentration of 0.2, 0.5 and 1.0 mM (lane 3-5) are compare with low molecular weight protein marker (lane M)

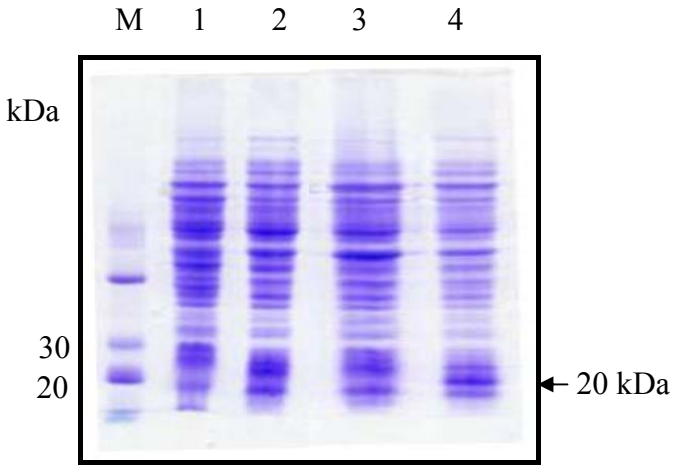


Figure 14 Temperature for induction of MLL expression. pET33b was cultured at 37°C (lane 1), non- induced pET-MLL was cultured at 37°c (lane 2), induced pET-MLL were cultured at 25 and 37 °C respectively after induced by 1 mM IPTG (lane 3-4) are compare with low molecular weight protein marker

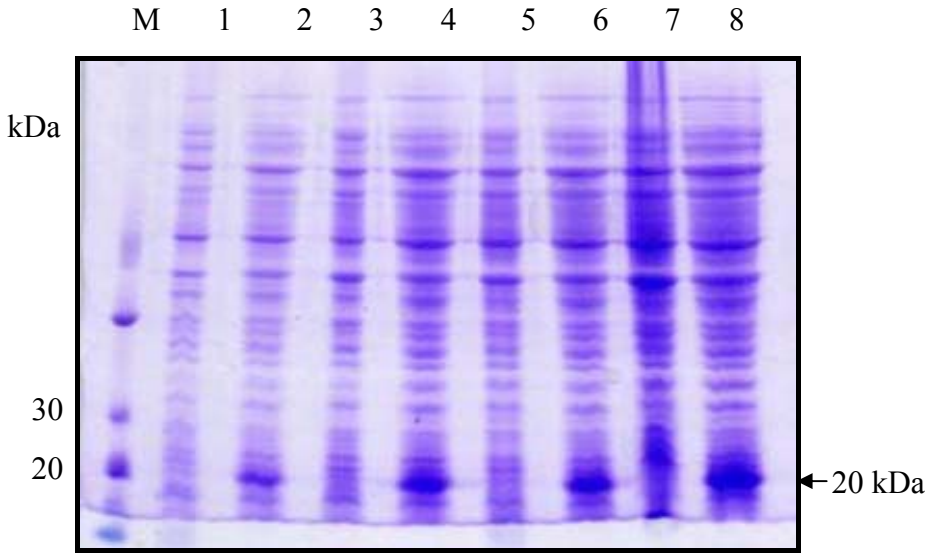


Figure 15 Various OD of cell cultures in initial induction. Non-induced pET-MLL at OD_{600} nm reached about 0.3, 0.6, 0.9 and 1.1 (lane 1, 3, 5 and 7, respectively), pET-MLL were induced by 1 mM IPTG at OD_{600} nm reached about 0.3, 0.6, 0.9 and 1.1 (lane 2, 4, 6 and 8, respectively) are compare with low molecular weight protein marker (lane M).

3. Purification and characterization of recombinant MLL protein

3.1 Purification of MLL by anion exchange column chromatography

The induced cells were disrupted by sonicator, the inclusion bodies and the soluble proteins were obtained. Since the inclusion bodies were solubilized by 8 M urea was needed before subjecting to DEAE - cellulose column. Proteins were eluted with the stepwise of 0-0.9 M NaCl. The elution profile (Figure 16) was monitored for protein by measuring the absorbance at 280 nm. The profile is shown that the most bound protein was eluted by 0.3 and 0.4 m NaCl. The fraction of each peak was submitted to 12%SDS–PAGE. The Figure 17 is shown that the protein fraction eluted with 0.3 M NaCl displayed a strong band of protein of approximately 20 kDa with some impurities. Therefore, the 0.3 M NaCl eluted fractions were pooled for further Ni^{2+}-NTA affinity column chromatography purification.

3.2 Purification of MLL by affinity column chromatography

The pooled fractions (fraction no.35-45) from DEAE-cellulose column containing the recombinant MLL were dialyzed against binding buffer (50 mM phosphate buffer pH 7.4, 0.15 M NaCl, 10 mM imidazole) and subjected to Ni^{2+}-NTA affinity column. The bound protein was eluted by the PBS buffer containing 50-350 mM imidazole and low pH buffer, pH 6.0 and pH 4.0. The protein purity was investigated by 12% SDS–PAGE. The purified proteins fractions eluted by 200, 250 and 300 mM imidazole imply the single band of about 20 kDa but the excellent band of purified protein was obtained when eluted by 250 mM imidazole as shown in Figure 18.

The purified protein could be obtained a single protein band of 20 kDa. The result indicated that advantage of using DEAE-cellulose column and Ni^{2+}-NTA affinity column for the purification process.

The recovery of the recombinant protein was average 3 mg from total 1 l of cultured cells protein extraction (Table 2)

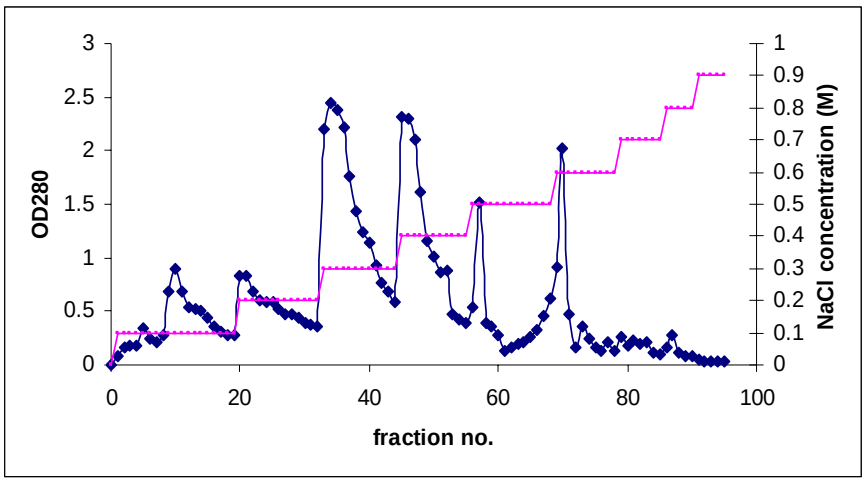


Figure 16 Purification profiles of the recombinant MLL protein by DEAE-cellulose column. The protein was applied to the column and washed with 50 mM phosphate buffer, pH 7.4 and eluted with 0- 0.9 M NaCl.

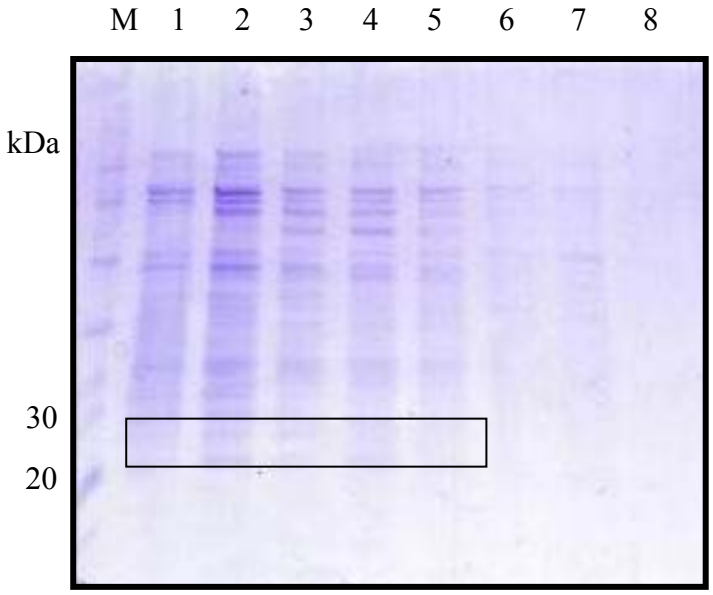


Figure 17 Purification of the recombinant MLL by DEAE- cellulose column was subjected to tricine-SDS-PAGE. The protein was applied to the column and washed with 50 mM phosphate buffer, pH 7.4 and eluted with 0.2- 0.9 M. NaCl (lane1-8, respectively) were compared with low molecular weight protein marker (lane M)

.

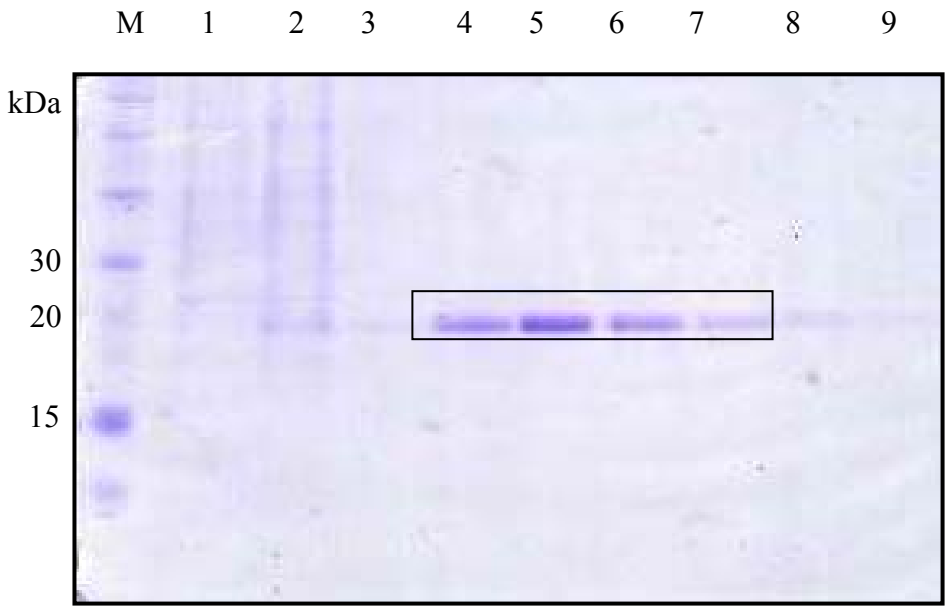


Figure 18 Purification of the recombinant MLL by Ni^{2+}-NTA affinity column was subjected to tricine-SDS-PAGE. The protein which eluted by 0.3 M NaCl form DEAE–cellulose was applied to the column and washed with 30 mM phosphate buffer, pH 7.4 and eluted with buffer containing 50, 100, 150, 200, 250, 300 and 350 mM imidazole (lane 1-7, respectively) and low pH buffer, pH 6.0 and pH 4.0 (lane 8-9, respectively) compare with low molecular weight protein marker (lane M).

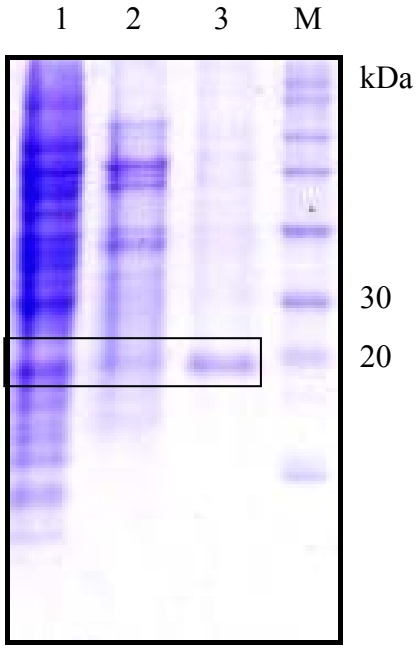


Figure 19 The Tricine-SDS –PAGE protein pattern from different step of purification. Soluble crude extract (lane 1), a fraction eluted with 0.3 M NaCl from DEAE-cellulose column (lane 2) and a fraction eluted with 250 mM imidazole from Ni^{2+}-NTA affinity column (lane 3) were compared with low molecular weight protein marker (lane M).

Table 2 Purification of recombinant MLL. Protein was obtained from 1 L of *E. coli* BL21 (DE3) cultured cells. Protein concentration was estimated by Lowry's method.

Protein	Total protein (mg)	Protein yield (%)
Crude recombinant protein	240	100
Recombinant protein from DEAE column	18	7.5
Recombinant MLL from His-tag column	~3	1.25

4 Characterization of MLL

4.1 Hemagglutination activity

The hemagglutination activity of twofold dilution of purified recombinant MLL was performed. The purified recombinant MLL showed no hemagglutination activity on a 2% suspension of rabbit erythrocytes (data not shown). This assay demonstrated that the recombinant MLL could not induced agglutination of rabbit erythrocytes.

This recombinant protein could not show strict specificity of the mulberry leaf lectins for Neu5Gc. The homology structure of this lectin and computational analysis were used to indicate direct interactions between specific sugars and residues of this lectin.

5. Computational structure analysis

5.1 The homology structure of MLL

The deduced amino-acid sequence of the MLL showed a similarity with jacalin related lectins which were galactose-binding lectin such as jackfruit *Artocarpus integrifolia* and *Maclura pomifera* agglutinin (MPA). Both lectins are tetramers composed of α chains (133 residues) which are non-covalently associated to small β chain (20 residues) subunits **(**Lee *et al*., 1998). The percentages of similarity of jackfruit *Artocarpus integrifolia* and *Maclura pomifera* agglutinin (MPA) compared to MLL from ClustalW program (http://www.ebi.ac.uk/clustalw/) are 88.9 % (data not shown) and 78.2 % identity, respectively. The region of 63 to 82 residues serve as β-chain and the region of 86 to 218 residues serve as α- chain (Figure 20).

To automated homology modeling of MLL monomer, the MLL sequence was submitted to the GENO3D server (http://geno3d-pbil.ibcp.fr/cgi bin/geno3d_automat.pl? page=/GENO3D /geno3d_home.html). MPA (PDB accession

code 1JOT) with a resolution of 2.2 Å was used as the template to model the MLL lectin monomer. The homology modeling of MLL was visualized by SPDBV program. Figure 21 shows that the overall folding of MLL exhibited a α- chain and exhibited the same β-sheet structure of the MPA monomer.The MLL have 5 amino acids insertion in the N-terminus causing a shifted of 5 residues compared with the MPA.

5.2 Binding site of MLL

The binding site of the crystal structure of MPA was considered to be prerequisite for binding site of MLL. The binding pocket site was formed by the N-terminal residue starting with Gly6 in the MLL or the first residue in MPA, and the turns of residues 51–64, 80–87, and 127–130. Interestingly, there were two residues in the binding site of MLL which were different from MPA, Glu76 and Thr79 in MPA binding site were replaced with Ser and Ile in MLL, respectively (Figure 22).

The electrostatic distribution on molecular surface of MLL was computed to compare with MPA. Figure 23 is shown that MPA has more negative charge distribution while MLL surface is neutral nearly positive charge. Therefore, these may cause MLL to likely bind with acidic sugar than MPA.

5.3 Interaction of MLL

AutoDock was used to validate for the interactions between MLL and ligands (sialic acid derivertives, GalNAc, Gal, Man, Glu and Gal-GalNAc). The grid map volume plus a shell of 40x40x40 Å was large enough to allow the ligand to rotate freely around the binding site. A final docked conformation was produced within a fixed protein structure. This proved successful after 100 runs of conformational space search. The final docked energy was computed by final intermolecular energy combined final internal energy of ligand. The best docking result can be considered to be the conformation with the lowest (docked) energy which was used for further analysis. Table 3 is shown that MLL preferred to bind with Neu5Gc, a sialic acid

derivative, better than GalNac, and Gal. The result is in good agreement with carbohydrate-binding specificity of the mulberry leaf lectins which showed the MLL1 activity was most effectively inhibited by Neu5Gc higher than GalNAc and Gal.

To study the binding characteristics, the model of the interaction between GalNAc and MPA was firstly examined and compared with the interaction between GalNAc and MLL. Both models based on the structures obtained from the best docking results. Figure 24 shows the comparison between the orientation of GalNAc with MPA and with MLL. These assumed that the NAc arrangement would be closely bound to the amino acid residues of MLL in the opposite position in MPA. The charge group of Asp-125 in MPA as Asp-130 in MLL forms the strong H-bonds to the GalNAc. It suggested that NAc moiety plays an important role in the specific binding to these lectins.

To consider the specificity of MLL binding, naturally specific to Neu5Gc which is a derivative of sialic acid, the best structure of Neu5Gc was used as a model to determine the interaction with the residues of this lectin. Figure 25 shows that the Neu5Gc fit in a groove formed by the N-terminal residue (Gly-6) and the three turn regions which are the residues 51to 64, 80 to 87, and 127 to 130. This sugar was expected to face the amino acids which the distance from the sugar was less than 6 Å. They were Gly-6, Gly- 82, Tyr-83, Ile 84, Val 85, Ser 124, Val 125, Gly 126, Tyr-127, Trp-128, Asp-130 and Tyr 131. The amino group of the Gly-6 formed the strong H-bonds to C2 of glycolyl group of Neu5Gc. The Ile- 84 and Trp-128 significantly contributed hydrophobic interactions to the overall interaction between Neu5Gc and MLL.

```
MPA                 ----------------------------------------------------
MLL                 MLMASSSFLSLSFLVLLFSISSANTRKWSLSNVLDQKPISIIEAAIGVSEDLLNL 55

                              β-chain                              α-chain
MPA           -------NEQSGISQTVIVGPWGAKVS---GVTFDDGAYTGIREINFEYNSETAIGGLRV
MLL           NGMEAKNNQQSGKSQTIVVGTWGAQATSSNGVAFDDGSYTGIREINFEYNNETAIGSIQV 115
                     *:***.***::**.***:.:   **:****:************.*****.::*

MPA           TYDLNGMPFVAEDHKSFITGFKPVKISLEFPSEYIVEVSGYVGKVEGYTVIRSLTFKTNK
MLL           TYDVNGTPFEAKKHASFITGFTPVKISLDFPSEYIVEVSGYTGKVSGYIVVRSLTFKTNK 175
              ***:** ** *:.* ******.******:************.***.** *:*********

MPA           QTYGPYGVTNGTPFSLPIENGLIVGFKGSIGYWLDYFSIYLSL
MLL           ETYGPYGVTSGTHFKLPIQNGLIVGFKGSVGYWLDYIGFHLAL 218
               :********.** *.***:**********:******:.::*:*
```

Figure 20 Sequence alignment between MLL and MPA. The region of 63 to 82 residues serve as β-chain and the region of 86 to 218 serve as α- chain.

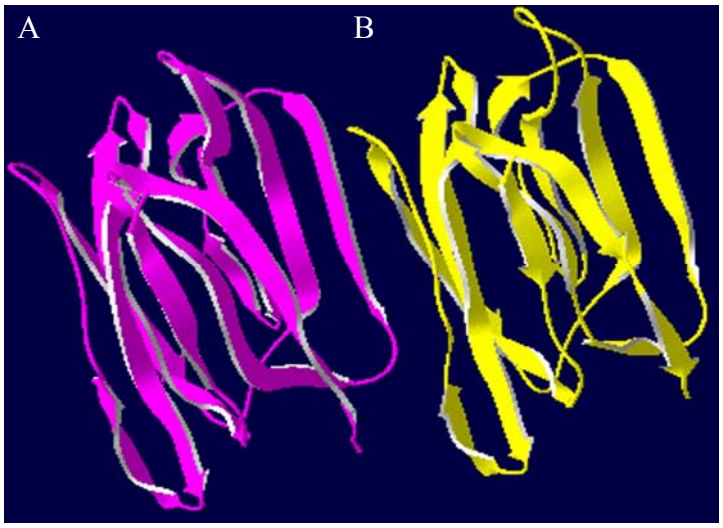


Figure 21 The crystal structure of MPA (A) and the homology modeling of MLL (B) show that the overall fold of MLL is very similar and exhibits the same β-sheet structure as the MPA monomer.

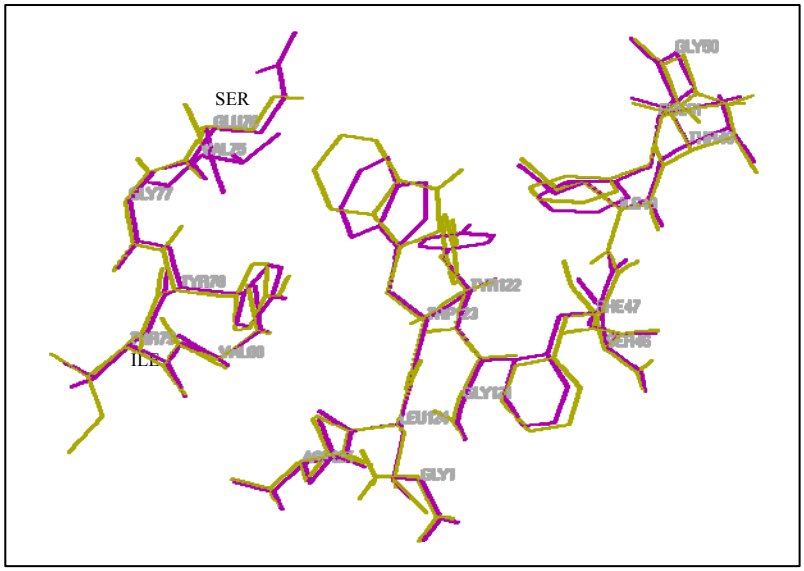


Figure 22 Binding site of MPA and MLL. There are two residues in binding site of MLL (yellow) are different from MPA (pink). Glu 76 and Thr 79 in MPA binding site is substituted by Ser and Ile, respectively, in MLL.

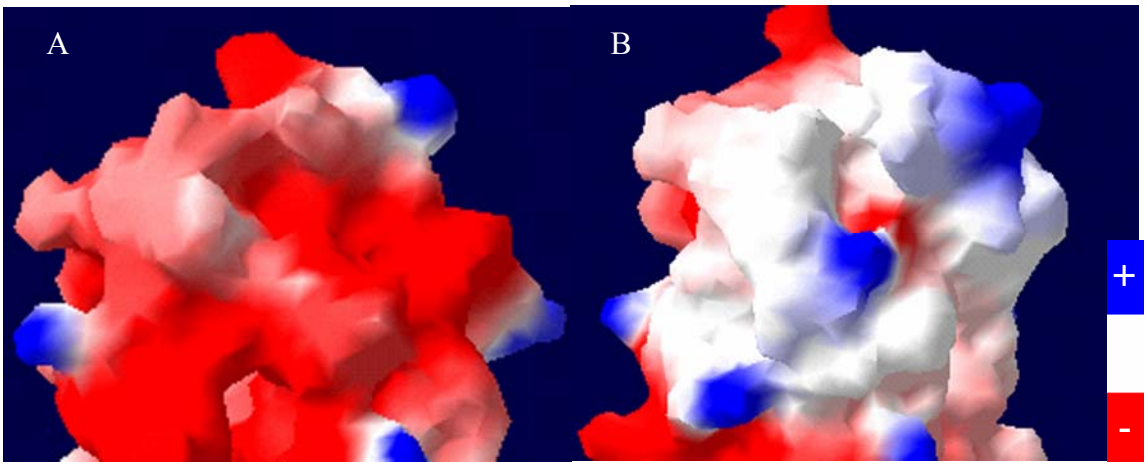


Figure 23 The electrostatic distribution of molecular surface of MPA (A) and MLL (B). MPA has more negative charge distribution on the surface higher than MLL

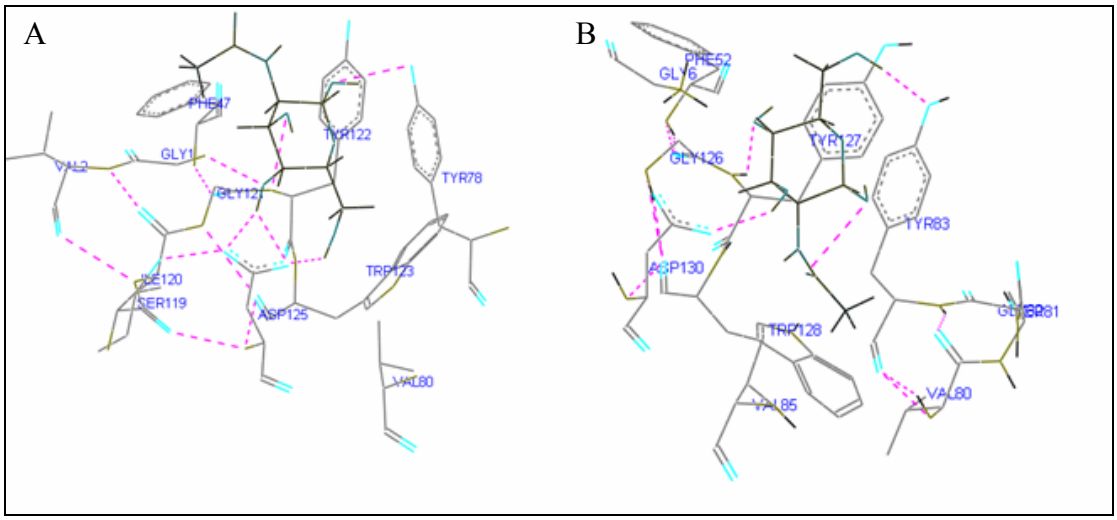


Figure 24 Docking of GalNAc into the carbohydrate-binding sites of MPA and MLL, respectively. The orientation of GalNAc was performed that the Gal moiety arrangement would be closely bound to the amino-acid residues of MPA (A) in the opposite position and orientation in MLL, the NAc moiety arrangement would be closely bound to the amino-acid residues of MLL (B)

Table 3 Experimental data compare with prediction data.

Sugar	Experimental Data	Prediction data	
	Purified MLL*	Modeling MLL	Modeling MPA
	mM	Final Dock Energy (kcal/mol)	Final Dock Energy (kcal/mol)
Neu5Gc	0.78	-12.39	-9.77
GalNac	100	-8.22	-8.53
Gal	100	-7.73	-6.40

*Purified lectin from Sephadex G-75 (Ratanapo *et al*., 1998)

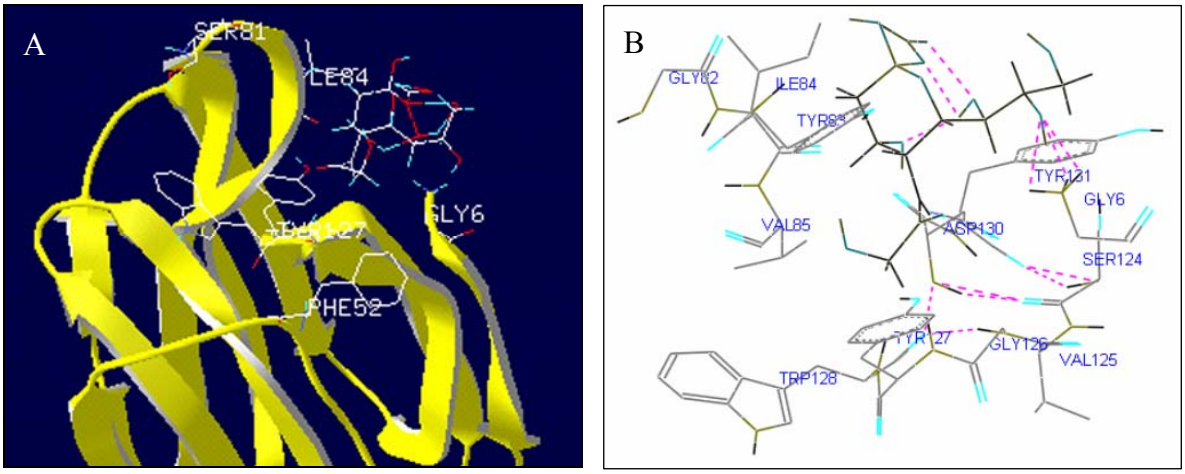


Figure 25 Binding of Neu5Gc moiety to MLL. Backbone structure of the MLL monomer with Neu5Gc (A) and interactions around the Neu5Gc moiety (B). H-bond form atoms in a pocket around the Neu5Gc contributed to the strong interaction between the Neu5Gc moiety and the MLL binding site.

DISCUSSION

MLL1, Neu5Gc -specific lectin from mulberry (*Morus rotunbiloba*) leaves, was previously purified, characterized and sequenced. In this study, MLL gene was subcloned and expressed in *E.coli*. The purpose of high mass MLL production in this study is for production of antibody against MLL to follow of MLL in mulberry and eukaryotic expression system. Therefore, active protein is not required for this application. Several host systems are available for expression of a target protein. The choice of host depends upon the specific requirements and applications for the recombinant protein. In this study, the bacterial expression system, *E.coli*, was used because its cultivation is inexpensive and easy to grow with high yields up to 50% of total cell protein. There are several reports of using *E.Coli* to express lectin proteins from plants. . For example, Tateno, *et al*. (2003) were successfully cloned and expressed of novel members of jacalin-related lectins from rhizomes of the true fern *Phlebodium aureu.*

The MLL coding region was amplified by PCR using a MLL20 clone from Kankamol et al. (2003) as a template. The gene was cloned into *Nhe*I and *Xho*I sites of pET33b expression vector. Proteins were expressed as a fusion protein with his-tag on the N and C-terminus which were easily purified by metal chelating chromatography and detection. The MLL could be expressed by the pET vector system containing strong T7 promoter which produces many RNA molecule that increase an opportunity of MLL production. Mass calculated by computing program is 28 kDa. It was computed by deduced amino acid plus the existence of additional residues at the N and C-terminus, including 6x his-tag, thrombin and kinase site as described to pET- 33b vector. The recombinant MLL showed an apparent molecular mass about 20 kDa in tricine-SDS-PAGE analysis. This could be due to post-translational processing at the C-terminus, which cleavage and removed the residues of protein at the C-terminus. The SignalP-NN prediction result indicated that the first 25 residues of MLL act as signal peptide domain and the most likely cleavage site between positions 25 and 26: ANT-RK (Figure 26). Although, the molecular weight of the native MLL is 16.5 kDa. When consider the deduce amino sequence, the region

of 63 to 82 residues serve as β-chain and the region of 86 to 218 serve as mature peptide for α-chain named Jacalin-like lectin domain. Proteins containing this domain are lectins. Both mature polypeptides originate from a single precursor through a complex post-translational processing, which involves the removal of an N-terminal peptide, an internal peptide and a C-terminal peptide.

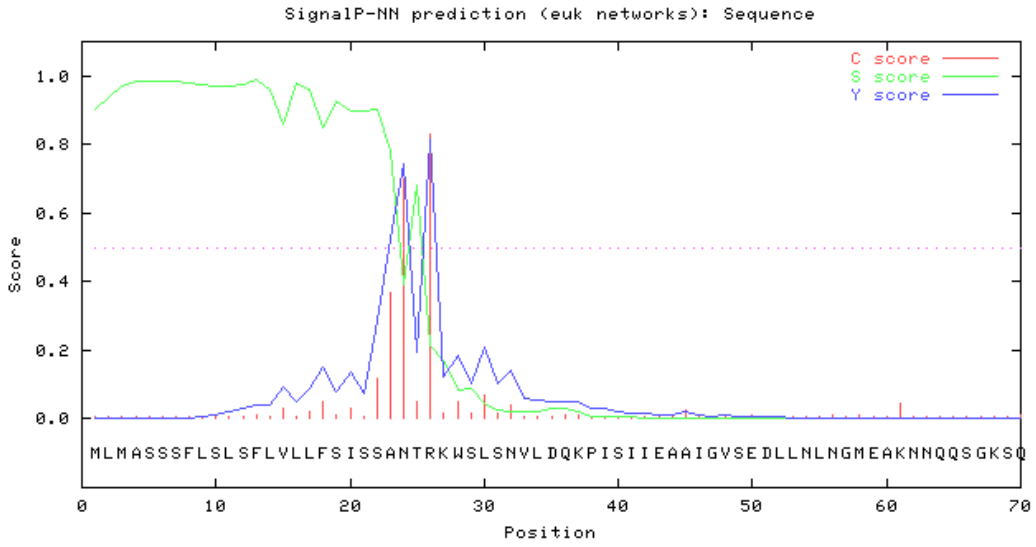


Figure 26 Signal peptide and cleavage site prediction. The first 25 residues of MLL act as signal peptide domain and the most likely cleavage site between pos. 25 and 26: ANT-RK.

Moreover, the major expression protein level presented as inclusion bodies in the cells, it suggested that the basal expression level was high. Lilie *et al* (1998) was demonstrated where *E. coli* is applied as host cells. Inclusion bodies are aggregates of inactive protein when expressed in high level at fast rate. Intermolecular hydrophobic and ionic interactions of these partially folded proteins with expose hydrophobic moieties caused the aggregation. In this study, the induction condition was varied by induction time, IPTG concentration, temperature and OD of cell culture for expression to decrease the expression in the inclusion body form. However, the recombinant MLL was still in inclusion bodies form after varying those conditions. Therefore the other factors have been considered to decrease the expression level. The basal level expression can be minimized by catabolite repression (Amershem Phamacia Biotech). Inada *et al* (1996) had been reported that glucose can decrease the basal expression level. Hence, high concentration of glucose was used in this study.

The effect of glucose is mediate by cAMP as a co-activator and an activator know as CRP. CRP binds to DNA most avidly when cAMP concentrations are high. In the present of glucose, the synthesis of cAMP is inhibited and efflux of cAMP from the cell is stimulated. As [cAMP] declines, CRP binding to DNA declines, thereby decreasing the expression of *lac* operon and one of its products, β-galactosidase, to decrease the basal expression level. (Nelson and Cox, 2005)

The recombinant protein was purified using the DEAE-cellulose ion exchange chromatography and a nickel affinity column. Since the MLL is an acidic protein with a calculated PI of 6.5, the protein should bind to the DEAE- cellulose column at pH 7.4. For nickel affinity column, MLL expressed in pET-33b with His-Tag fusion protein can be chelated by Ni^{2+} ions on the beads of column. His-Tag fusion protein can be easily desorbed with buffer containing imidazole. From this purified process could be easily obtained high mass of recombinant MLL when compared with purified from mulberry leaves as Ratanapo, *et al*., 1998 reported. Crude extract from mulberry leaves 2,263 mg could be obtained MLL1 4 mg while crude extract from recombinant cell 240 mg could be obtained recombinant protein 3 mg. The other more advantage was the recombinant MLL could be purified by two steps of purified process while MLL from mulberry leaves could be purified fives steps of purified process and more complicate than purification from recombinant cells.

The purified protein was tested for the hemagglutination activity. The purified recombinant lectins did not show any hemagglutination activity. It can be suggested that

1. The recombinant MLL was not correctly post-tranlational modified in the bacterial cytoplasm. Some post-translational processes which do not occur in *E. coli* may be necessary for proper folding and function of MLL such as glycosylation. A way of taking advantage of the host cell's equipment to deal with protein folding is the co-overexpression of the recombinant protein with molecular chaperones.

2. His-Tag might be interfered the protein structure and affect folding and biological activity. For this reason biological activity might differ from native MLL.

3. The recombinant protein did not present as a correct conformation. The active lectins in the nature form oligomers. It is possible that, oligomeric form is necessary for sugar binding and hemagglutination. The presence without formed structural instability of oligomers configuration may be abolished the activity (Kakiuchi *et al*, 2001; Kassab *et al*, 2004 and Nelson and Cox, 2005)

The fact that, in this study, the recombinant MLL is no hemagglutination activity, the sugar specific binding of MLL can not be observed. Also, there was no strict specificity of the mulberry leaf lectin for Neu5Gc.

When the crystal structure of the protein had not been solved, the *in silico* computational analysis is an excellent way to built the protein structure model for studying the protein-ligand interaction. The three dimensional structure of this lectin and computational analysis were used to indicate the direct interactions between specific sugars and residues of this lectin.

From the MLL modeling, MLL revealed a striking sequence similarity with α-chain of MPA. However, sugar specific binding of MLL was different from MPA. When considered to the binding site, there were two residues MLL which differ from MPA. Ser and Ile were replaced Glu and Thr in MPA, respectively. These replacements may lead be to negatively charge decreasing on MLL binding surface. In electrostatic distribution investigating, the surface of MPA binding site showed more negative charge than MLL. These differences had an effect on carbohydrate specificity of lectins.

The carbohydrate specificity of Moraceae lectins is various. It could be considered that the functional diversities among MLL and other lectins have apparent presence of evolutionary in Moraceae family.

In the MLL-sugar interaction study, the sugar is bound with MLL through hydrogen bonds. The carboxylate moiety interacts with main-chain amide groups and polar side chains (especially Serine) in what are essentially hydrogen-bond interactions.

The common occurrence of acidic side chains as hydrogen bond acceptors from sugar OH and the frequent use of the serine OH play an important role in the recognition of the sialic acid carboxylate. The carbon backbone of the glycerol moiety of sialic acid can also present an apolar surface to the protein. Moreover, a cluster of Ile and Trp residues form hydrophobic interactions with sialic acid (Weis, 1996). The reasons mentioned above may be the important recognition process for the strong affinity and specificity of MLL for the sialic acid derivatives.

CONCLUSIONS

1. The molecular weight of the native MLL is 16.5 kDa, the recombinant MLL showed an apparent molecular mass of about 20 kDa in Tricine-SDS-PAGE analysis while calculated mass by computing program is 28 kDa.

2. The recombinant MLL accumulated mostly in an insoluble form in inclusion bodies. In order to obtain the high amount of soluble MLL protein, the recombinant cells were induced with 0.5 mM IPTG and 0.4% glucose at 25 °C.

4. Purified MLL was obtained from DEAE-ion exchange and Ni^{2+}-NTA affinity column chromatography**.** This protein did not show any hemagglutination activity.

5 .The result from computational analysis showed that MLL preferred to bind with Neu5Gc a sialic acid derivative better than GalNac and Gal, respectively. The result is in good agreement with carbohydrate-binding specificity of the mulberry leaf lectins. (Ratanapo, *et al*., 1998)

6. There were two residues in the binding site of MLL which were different from MPA, Glu76 and Thr79 in MPA binding site were replaced with Ser and Ile in MLL, respectively. These residues effect to binding of MLL is different from MPA.

7. The electrostatic distribution on molecular surface of MLL is neutral nearly positive charge. Therefore, MLL like to bind with acidic sugar.

8. In molecular docking, Gly-6 forms strong H-bond with glycolyl group of Neu5Gc. The Ile- 84 and Trp-128 contributed hydrophobic interactions to the sugar. Serine in binding site of MLL may be played an important role in the recognition of the sialic acid carboxylate.

LITERATURE CITED

Amersham Phamacia biotech. The Recombinant Protein Handbook; Protein Amplification and Simple Purification. www. apbiotech.com.

Baxevanis, A.D. and B.F.F. Ouellette. 2005. **Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins.** John Wieley and Sons, Inc., New York.

Bush, C.A. and M. Martin-Pastor. Structure and conformation of complex carbohydrates of complex carbohydrates of glycoproteins, glycolipids and bacterial polysaccharides. **Annu. Rev. Biophys. Biomol. Struct**. 28: 269-293.

Chrispeels, M.J. and N.V. Raikhelb. 1991. Lectins, Lectin Genes, and Their Role in Plant Defense. **Plant Cell** 3: 1-9.

Colangeli, R., J.S. Spencer, P. Bifani, A. Williams, K. Lyashchenko, M.A. Keen, P.J. Hill, J. Belisle and M.L. Gennaro. 2000. MTSA-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis-Specific, delayed-type hypersensitivity in guinea pigs. **Infect. Immun.** 68: 990–993.

Huey, R. and G.M. Morris. 2003. Using AutoDock with AutoDockTools: A Tutorial. The Scripps Research Institute Molecular Graphics Laboratory. California, USA.

Inada, T., K. Kimata and H. Aiba. 1996. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. **Genes Cells** 1: 293-301.

IUPAC-IUB. 1984. Joint commission onbiochemical nomenclature; Nomenclature and symbolism for amino acids and peptides. **Eur. J. Biochem.** 138: 9-37.

Kakiuchi,M., N. Okino, N. Sueyoshi, S. Ichinose, A. Omori, S. Kowabata, K. Yamaguchi and M. Ito. 2002. Purification, charecterrization and cDNA cloning of α-*N*- acetylgalactosamine-specific lectin from starfish, *Asterina pectinifera*. **Glycobiology** 12: 85-94.

Kankamol C. 2003. Molecular cloning and expression of mulberry leaf lectin (MLL) genes. Ph.D. Thesis, Kasetsart University.

Kassab, B.H., D.D. de Carvalho, M.A. Oliveira, G.R. Baptista, G.A.G. Pereira and J. C. Novelloa. 2004. Cloning, expression, and structural analysis of recombinant BJcuL, a c-type lectin from the *Bothrops jararacussu* snake venom. **Protein. Expr. Purif**. 35: 344–352.

Kobayashi, Y., K. Kobayashi, K. Umehara, H. Dohra, T. Murata, T. Usui, and H. Kawagishi. 2004. Purification, characterization, and sugar binding specificity of an *N*-Glycolylneuraminic acid-specific lectin from the mushroom *Chlorophyllum molybdites*. **J. Biol. Chem.** 279: 53048–53055.

Lee, X., A. Thompson, Z. Zhang, H. Ton-that, J. Biesterfeldt, C. Ogatai, L. Xu, R.A.Z. Johnston and N.M. Young. 1998. Structure of the complex of *Maclura pomifera* agglutinin and the T-antigen disaccharide, Galb1,3GalNAc. **J. Biol. Chem**. 273: 6312-6318.

Lehmann,F., E. Tiralongob and J. Tiralongo. 2006. Sialic acid-specific lectins: occurrence, specificity and function. **Cell. Mol. Life Sci**. 63: 1331–1354.

Lilie, H., E. Schwarz and R. Rudolph. 1998. Advances in refolding of proteins produced in *E. coli*. **Curr. Opin. Biotechnol.** 9: 497-501.

Lowry, O.H., N.J. Rosbrough, A.L. Farr, R.J. Randall. 1951. Protein measurement with the Folin phenol reagent. **J. Biol. Chem**. 193: 265–275.

Nelson D.L. and M.M.Cox. 2005. Lehninger Principles of biochemistry. W.H. Freeman
and company. England.

Novagen. 1998. pET-33b Vector. http://www.clinalfa.com/Products/ProductDisplay.asp?catNO=69054.

Peumans, W.J., and E.J.M. Van Damme. 1995. Lectins as plant defense proteins. **Plant Physiol.** 109: 347-352.

Promega. 2003. Technical Manual; pGEM-T and pGEM-T Easy Vector Systems. Promega Corperation. USA.

Ratanapo, S., W. Ngamjunyaporn and M. Chulavatnatol. 1998. Sialic acid binding phytopathogenic bacterium, *P. syringae pv mori*. **Plant Sci.** 160: 739-744.

_____, _____ and _____. 2001. Interaction of a mulberry leaf lectin with a lectins from leaf of mulberry (*Morus alba*). **Plant Sci.** 139: 141-148.

Ratanapo, S., Y. Thawalvichit, and A. Promboon. 2005. Inhibitory effects of mulberry leaf lectins to silkworm proteases. KU J. (Nat. Sci.) 39, 125-135.

Saito, T., M. Hatada, S. Iwanaga and S. Kawabata. 1997. A newly identified horseshoe crab lectin with binding specificity to O-antigen of bacterial lipopolysaccharides. **J. Biol. Chem**. 272: 30703–30708.

Sambrook, J. and D.W. Russell. 2000. **Molecular Cloning: A Laboratory Manual.** 3rd ed**.** Cold Spring Harbor Lab. Press, Planview, New York.

Schagger H.. 2006. Protocol tricine–SDS-PAGE. **Nature Protocol** 1: 16-18.

Sharon, N. and H. Lis. 2004. Review history of lectins: from hemagglutinins to biological recognition molecules. **Glycobiology** 14 : 53R–62R.

Singh,T., J.H. Wu, W.J. Peumans, P. Roug, El.J.M. Van Damme and A.M. Wu. 2007. Recognition profile of *Morus nigra* agglutinin (Morniga G) expressed by monomeric ligands, simple clusters and mammalian polyvalent glycotopes. **Mol. Immunol**. 44: 451–462.

Tanczos, A.C., D.A. Faux, D.C. Povey and B.J. Howlin. 2004. The influence of sequence variability and dimerization on mannose binding in monocot mannose binding lectins. **Internet Electronic Journal of Molecular Design** 3: 560–571.

Tateno H., H.C. Winter, J. Petryniak, and I.J. Goldstein. 2003. Purification, characterization, molecular cloning and expression of novel members of jacalin-related lectins from rhizomes of the true fern *Phlebodium aureum* (L) J. Smith (Polypodiaceae). **J. Biol. Chem**. : in Press.

Van Damme, E.J.M., B. Hause, J. Hu, A. Barre, P. Rouge´, P. Proost, and W.J. Peumans. 2002. Two distinct jacalin-related lectins with a different specificity and subcellular location are major vegetative storage proteins in the bark of the black mulberry tree. **Plant. Physiol**. 130: 757–769.

Van Damme, E.J.M., W.J. Peumans, A. Pusztai and S. Bardocz. 1998. Plant lectins: A special class of plant proteins, pp. 3-25. **Handbook of Plant Lectins: Properties and Biomedical Applications.** John Wieley and Sons, Inc., New York.

Weis W.I.. 1996. Structure basis of recognition lictin-carbohydrate. **Annual Reviews Inc.** 65:441-473.

Wu A.M., J.H. Wu, T. Singh, K. Chu, W.J. Peumans, P. Rougé, and Els J.M. Van Damme. 2004. A novel lectin (Morniga M) from mulberry *(Morus nigra)* bark recognizes oligomannosyl residues in *N*-glycans. **J. Biomed. Sci**. 11: 874-885.

Yeasmin, T., M.A.K. Tang, A. Razzaque and N. Absar. 2001. Purification and characterization of three galactose specific lectins from mulberry seeds (Morus sp.). **Eur. J. Biochem**. 268: 6005–6010.

APPENDIX

General reagent

0.5 M EDTA, pH 8.0

EDTA 18.61 g

The chemical was dissolved in distilled water, adjusted to pH 8.0 with 1 M NaOH. Distilled water was added to the final volume 100 ml. The solution was sterilized by autoclaving for 15 minutes at 121°C.

1 M $CaCl_2$

Dissolved 54 g of $CaCl_2$ $6H_2O$ in 200 ml of distilled water.
The solution was sterilized by pass through a 0.2 μm filter and stored at –20 °C until use.

2 M NaCl

NaCl 29.2 g

Distilled water was added to final volume 100 ml

Dispensed into aliquots and sterilized by autoclaving.

10 % SDS

Sodium dodecyl sulfate 10 g

Distilled water was added to final volume 90 ml

The chemical was dissolved at 68 °C and adjusted to pH 7.2 by adding 2-3 drops of concentrated HCl. The solution was adjusted to the volume 100 ml.

Reagent for molecular cloning

TE, pH 8.0 (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0)

1 M Tris-HCl pH 8.0	1.0 ml
0.5 M EDTA pH 8.0	0.2 ml

The mixture was mixed thoroughly and adjusted to the final volume of 100 ml with distilled water.

50x TAE buffer (stock)

Tris-base	240.2 g
Glacial acetic acid	57.0 ml
0.5 M EDTA pH 8.0	100.0 ml

The chemicals were dissolved in distilled water, and adjust the final volume to 1 liter with distilled water. (Working solution is 1 X)

Chloroform: Isoamyl alcohol (24: 1, v/v)

Chloroform (Merck)	24.0 ml
Isoamyl alcohol	1.0 ml

Both reagents were mixed together and stored in a dark bottle at room temperature.

1M IPTG

Isopropylthio-β-D-galactoside	2.38 g
Distilled water	100 ml

The solution was sterilized by filtration through a 0.2 μm filter and dispensed the solution into 1 ml aliquot tube and stored at –20 °C.

X-gal

5-bromo-4-chloro-3-indolyl-β-D-galactoside 100 mg

The chemical was dissolved in 2 ml of dimethyl-formamide. The solution was stored in a tube covered with aluminum foil and stored at -20°C.

Ethidium bromide (10 mg/ml)

Ethidium bromide	1.0 g
Distilled water	10.0 ml

The solution was stored in a dark bottle at room temperature.

6x gel-loading dye buffer

0.25 % bromophenol blue

0.25 % xylene cyanol FF

30 % glycerol in water

The chemicals were dissolved and adjusted the final volume with distilled water.

RNase A (10 mg/ml)

RNaseA 10.0 mg

RNase was dissolved in 10 mM Tris-HCl pH 7.5, 15 mM NaCl and stored at -20°C.

5x DNA ligation buffer

250 mM Tris-HCl (pH 7.8)

50 mM $MgCl_2$,

5 mM DTT

5 mM ATP

25 % w/v polyethylene glycol (MW 8000)

Reagent for alkaline lysis method

Solution I (50 mM glucose, 25 mM Tris-Cl, and 10 mM EDTA, pH 8.0)

1.0 M glucose	5.0 ml
1.0 M Tris-HCl (pH 8.0)	2.5 ml
0.5 M EDTA (pH 8.0)	2.0 ml
Distilled water	90.5 ml

The mixture was sterilized by autoclaving for 15 minutes and stored at 4 °C.

Solution II (0.2 M NaOH, 1 % SDS)

1.0 M NaOH	2.0 ml
10 % SDS	1.0 ml
Distilled water	7.0 ml

Solution III (3 M Potassium acetate, Glacial acetic acid)

5 M Potassium acetate	90.0 ml
Glacial acetic acid	11.5 ml
Distilled water	28.5 ml

Media for bacterial culture

Luria-Bertani medium (LB medium per liter)

Tryptone	10 g
Yeast extract	5.0 g
NaCl	5.0 g

Adjust pH to 7.0 with NaOH. Then the solution was adjusted to the final volume of 1000 ml with distilled water and sterilized by autoclaving.

LB plates with ampicillin/kanamycin

Fifteen grams of agar was added to 1 l of LB medium then the media was sterilized by autoclaving. The medium was allowed to cool to 50°C before adding ampicillin to a final concentration of 100μg/ml, or kanamycin to a final concentration 30μg/ml The medium (30-35 ml) was poured into 85 mm petri dishes. The agar was allowed to harden. Agar plates were stored at 4°C for up to 1 month or room temperature for up to 1 week.

LB plates with ampicillin / IPTG / X-gal

The LB media with ampicillin was prepared then supplement with 0.5 mM IPTG and 80 μg/ml X-Gal and pour the plates. Alternatively, 100μl of 100 mM IPTG and 20 μl
of 50 mg/ml X-gal may be spread over the surface of an LB ampicillin plate and allowed to absorb for 30 minutes at 37°C prior to use.

SOC medium (100 ml)

Bacto tryptone	2.0 g
Bacto yeast extract	0.5 g
1 M NaCl	1.0 ml
1 M KCl	0.25 ml
$MgCl_2$ stock, filter steriled	1.0 ml
2 M glucose, filter steriled	1.0 ml

Bacto tryptone, bacto yeast extract, NaCl, and KCl were added to 97 ml distilled water. The media was sterilized by autoclaving. The medium was allowed to cool to room temperature before adding 2 M glucose and 2 M $MgCl_2$ stock, each to a final concentration of 20 mM. The solution was adjusted to the final volume of 100 ml with sterile distilled water. The complete medium was filtered through a 0.2 μm filter unit. The final pH should be 7.0.

Lowry's reagent; Alkaline Copper Reagent

1% (w/v) $CuSO_4$	1 ml
2 % (w/v) Na- Tartrate	1 ml
2 % (w/v) Na_2CO_3 in 0.1 M NaOH	98 ml

Tricine–SDS-PAGE

Gel monomer (30% acrylamide, 29 : 1)

Acrylamide	29.0 g
Bisacrylamide	1.0 g

Acrylamide and Bisacrylamide are dissolved in a total volume of 100 ml H_2O. Store at 4 °C and protect from light.

Gel buffer

Tris-base	181.7 g
SDS	1.5 g

Adjust the pH to 8.45 using HCl. Bring to final volume 500 ml and stored at 4 °C.

2x sample buffer

3.0 M Tris-HCl, pH 8.45	3.0 ml
Glycerol	2.4 ml
SDS	0.8 g
0.1 % Coomassie Blue G	1.5 ml
0.1% Phenol Red	0.5 ml

Distilled water to final volume 10 ml.

Anode buffer (Lower buffer)

Tris base	12.11 g

Is dissolve in distilled water in a total volume of 500 ml, adjust the pH to 8.9 using HCl and store at 4 °C.

Cathode buffer (Upper buffer)

Tris-base	6.055 g
Tricine	8.96 g
SDS	0.5 g

Are dissolved in a total volume of 500 ml distilled water.

Fixing solution

50% methanol
10% acetic acid
Bring to final volume 100 ml.

Staining solution

0.025% Coomassie dye
45% methanol
10% acetic acid
Bring to final volume 100 ml with distilled water.

Destaining solution

10% methanol
10% acetic acid
Bring to final volume 100 ml with distilled water.

Table 4 Abbreviation for amino acid residues

Amino acid	Three-letter	One-letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Source: IUPAC-IUB

CURRICULUM VITAE

NAME	Ms.Tuanta Getsang
BIRHT DATE	May 16th, 1981
BIRTH PLACE	Phattalung, Thailand

EDUCATION	YEAR	INSTITUTE	DEGREE
	2003	Kasetsart University	B.S. (Biochemistry)

OTHER TRAINING

March – May 2003	Trainee in Thailand Institute of Scientific and Technological Research (TISTR) -DNA extraction, PCR and sequencing techniques
May16-19, 2006	Trainee in Genetic engineering, Interdisciplinary Graduate Program,Kasetsart University - Protein Bioinformatics

PUBLICATION

February 2006	"Three Dimentional Structure of Mulberry Leaf Lectin" The Proceedings of 44th Kasetsart University Annual Conference (Subject: Science) pp. 513-519.
December 15-18, 2006	"Mulberry leaf lectin: purification, cloning and modeling". The 20thCongress of the International Sericulture Commission Conference. Papers Volume III. Bangalore, India, pp. 147-149. (co- author)