CHAPTER III

MATERIALS AND METHODS

1. Materials

Hardware and software

The visualization tools used in this study are composed of Rasmol, VMD, PyMol, and Chimera. The weblinks of those tools were showed in Table 1A. VMD is designed for the visualization and analysis of biological systems such as proteins. It may be used to view general structure of the molecules, as VMD is able to read standard Protein Data Bank (PDB) files and demonstrate the contained structure. Additionally, VMD provides a broad variety of methods for rendering and coloring a molecule. It can be also used to animate and analyze the trajectory of a MD simulation (Eargle *et al.*, 2006). Pymol v.0.99 is a molecular graphics system with contained Python interpreter designed for real-time visualization and rapid generation of high-quality molecular graphics images. Chimera software (Pettersen *et al.*, 2004) was used to analyze the molecular modeling and docking complex on Window XP professional operating system. Target-template investigation was performed by the THREADING of which web servers were shown in Table 1B. MODELLER 8v1 (Marti-Renom *et al.*, 2000) was used to perform molecular modeling on an Intel P4-base Suse 9.3 Table 1C. Validation tools for this present study compose of PROCHECK, *WHATIF* web interface, Verify 3D and ERRAT, of which weblink were shown in Table 1D. Notably Verify3D works best on proteins with at least 100 residues. Hex 4.5 was used to calculate molecular docking, Table 1E. The MD simulation, CHARMM force filed in this study was run by NAMD2 program, Table 1F (Kal'e *et al.*, 1999) with 7 processors of Solaris on Sparc processors and 10 processors of MacIntosh operating system those provided by National Center for Genetic Engineering and Biotechnology, Thailand.

Table 1 Selected structure annotation tools

PROGRAMS	WEBLINK	TYPES
Chimera	http://www.cgl.ucsf.edu/chimera/	executable
Rasmol	http://www.umass.edu/microbio/rasmol/	executable
PyMol	http://pymol.sourceforge.net/	executable
VMD	http://www.ks.uiuc.edu/Research/vmd/	executable
B. THREADING		
3D-PSSM	http://www.sbg.bio.ic.ac.uk/**3dpssm**/html/ffrecog_simple.html	server
Phyre	http://www.sbg.bio.ic.ac.uk/**phyre/**	server
UCLA-DOE	http://www.doe.mbi.ucla.edu/people/frsvr.html	server
GenThreader	http://insulin.brunel.ac.uk/psiform.html	server
C. MODELLING		
MODELLER	http://salilab.org/modeller/	executable
D. PROTEIN ASSESSMENT		
PROCHECK	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html	executable
WHATIF	http://swift.cmbi.kun.nl/WIWWWI/	server
Verify 3D	http://nihserver.mbi.ucla.edu/Verify_3D/	server
ERRAT	http://nihserver.mbi.ucla.edu/ERRATv2/	server
E. MOLECULAR DOCKING		
Hex 4.5	http://www.csd.abdn.ac.uk/hex/	executable
F. MOLECULAR DYNAMICS		
NAMD	http://www.ks.uiuc.edu/Research/namd/	executable

2.2 Selection of protein candidate that makes contact with *M. pneumoniae* α-enolase

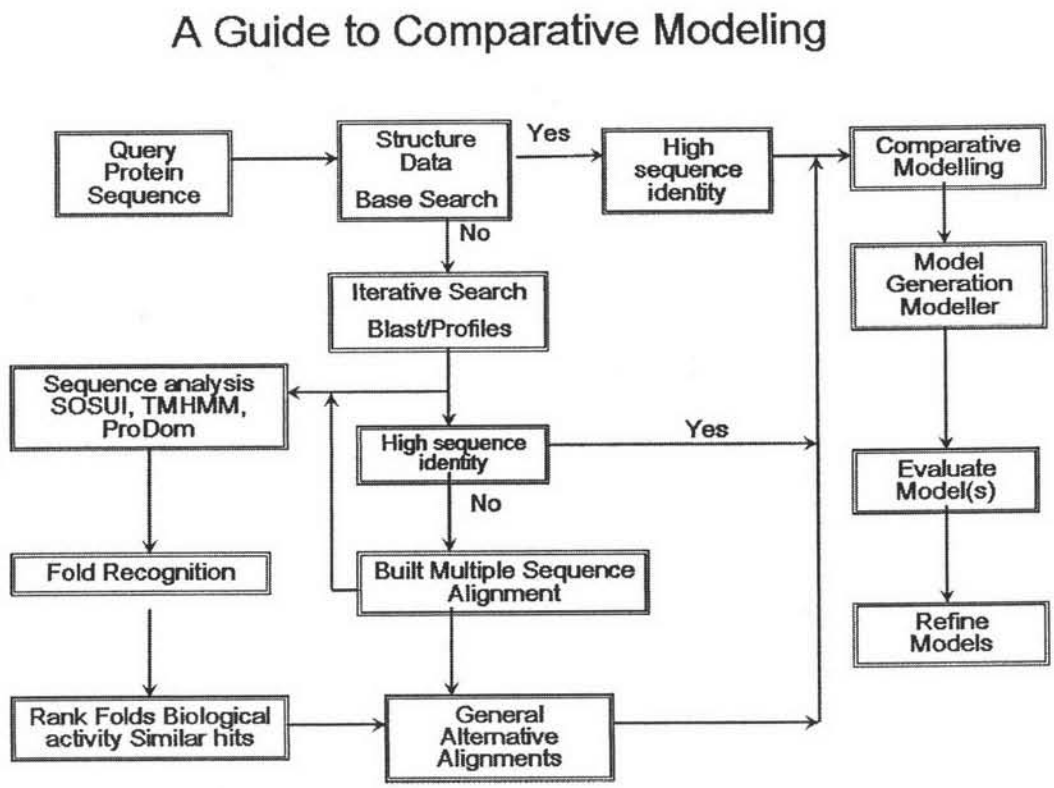
Human plasminogen kringle domain 2 was selected from data-mining of enolase binding proteins with hBBB domain (Jong *et al.*, 2003; Pancholi *et al.*, 1998; Redlitz *et al.*, 1995).

2.3 Comparative modeling

2.3.1 Template selection

Threading web servers were used to match the predicted secondary structure of query sequence, MpnE and P1 adhesin, with those of the proteins of known structure. These templates were further screened by prioritizing to the highest degree of sequence identity. The screening was done by BLAST search with *P* value <0.00001, global degree of sequence identity > 50%, and minimal projected model length is 431 amino acid residues for α-enolase. Then enolase sequence was analyzed using the program THREADER, 3D-PSSM, and Phyre which compared primary sequences with entire of the known 3D-structure in PDB. The output is composed of the optimally aligned and 3D-structures that are similar to MpnE. P1 adhesin protein which contains 1,627 amino acid residues was too long to submit to template finding web servers, 3D-PSSM and Phyre. Therefore it was then searched for putative membrane region by TMHMM tool. Using TMHMM, a putative membrane region of amino residue 32 to residue 1522 was identified, but it is still too long to be submitted to those web servers. That region was divided into 5 overlapping segments, contains approximate 300 amino acid residues per each overlapping region. Each region was then used in a search for potential domain by ProDom tool to identify potential template. Nevertheless the best template, adeno virus knob domain (1KNB), has 16% amino acid identity to residue 533-704 of P1 adhesin protein fragment,. This finding pointed to impossible to generate a good quality of 3D structure model for P1 adhesin protein by present available computation modeling tools that will be described in detail in result and discussion. Therefore MpnE was intensively studied to the next

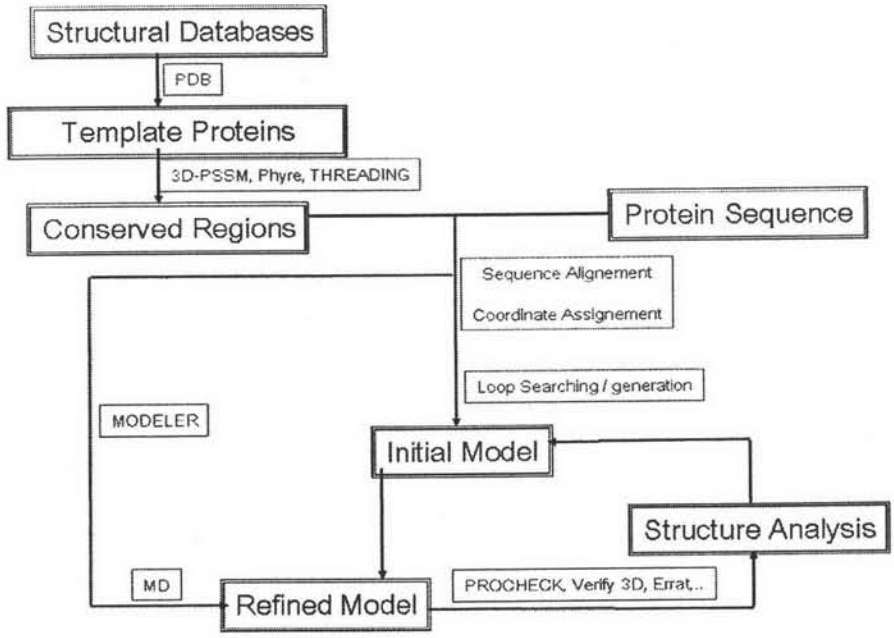
structural modeling. A guide to model 3d structure by comparative modeling approach is shown in diagram as follows.



2.3.2 Model building, by MODELLER and Model evaluation

MODELLER program package was used to generate 3D structure with satisfaction of spatial restraints. The modeling step is shown in diagram as follows.

3D structure of enolase were generated by Comparative Modeling



2.3.2.1 Target-template alignment

The amino acid sequence of target protein was prepared in *.PIR format and then aligned with template sequence run by MODELLER program under parameter script"ALIGN2D" (see appendix B)

2.3.2.2 Model building

After sequence alignment was produced, it was taken to derive 2000 homology 3D structure models by MODELLER mod 8v1 program package. They were run under parameter script (see appendix B).

Table 2 Selected sequence annotation tools

A. VIRULENCE FACTOR SEARCH TOOLS		
PROGRAMS	WEBLINKS	TYPES
IslandPath	http://www.pathogenomics.sfu.ca/islandpath/update/IPindex.pl	server
VFDB	http://zdsys.chgb.org.cn/cgibin/VFs/	server
Virulence search tool	http://193.129.245.227/pise/virfactfind_small.html	server
B. PROTEIN CLASSIFICATION OF PREDICTION		
HMMTOP	http://www.enzim.hu/hmmtop/server/hmmtop.cgi	server
TMHMM	http://www.cbs.dtu.dk/services/TMHMM/	server
SignalP	http://www.cbs.dtu.dk./services/SignalP/	server
SOSUI	http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html	server
TopPred2	http://www.sbc.su.se/~erikw/toppred2/	server
PSORT	http://psort.nibb.ac.jp	server
C. SEQUENCE FAMILY SEARCH		
PRINTS	http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/	server
BLOCKS	http://www.blocks.fhcrc.org	server
ProDom	http://protein.toulouse.inra.fr/prodom.html	server
PROSITE	http://www.expasy.ch/prosite/	server
D. PROTEOMICS DATABASE		
BSGC	http://www.strgen.org/3/05/06	server

2.3.2.3 Loop modeling

The top 20 best models of MpnE were selected by lowest objective function by minimum energy which implement in program. They were then further optimized loop by loop using refinement package which was implemented in MODELLER mod 8v1. They were run by loop refinement parameter call "MODELLOOP#" (#; number of model) (see appendix). A hundred loops refinement were optimized and loop conformation repeated by 2000 steps for each model. The chosen model was selected by the lowest objective function.

2.4 Model evaluation

An acceptable model was validated by input PDB file of chosen model to PROCHECK, Verify 3D, and ERRAT validation program. Then the chosen model was taken to further refine by energy minimization and MD simulation with implicit solvent.

2.5 Model refinement by molecular dynamics simulation with implicit solvent model

Molecular dynamics simulations in this present study compute atomic trajectory via equation solving of motion numerically using empirical force fields as CHARMM force field, which approximate the actual atomic force in biomolecule systems to improve the model. The CHARMM program is the most advanced research program developed by Harvard University MD research group for the energy minimization and dynamics simulation of not only for proteins but also nucleic acids and lipids in vacuum, solution or crystal environments (Harvard CHARMM Web Page http://yuri.harvard.edu/). In this present work, each complex structure is subjected to refine by repeating 2000-3000 steps of energy minimization with conjugate gradient approach using NAMD2 suit program which include constant energy dynamics, constant temperature dynamics, constant pressure dynamics, fix atoms, rigid waters, rigid bonds to hydrogen, harmonic restraint, and spherical or

cylindrical restraints (NVT ensemble) (Kal'e *et al.*, 1999). Each atom is assigned an initial velocity by temperature-dependent from Boltzmann distribution and Newton's laws of motion.

The obtained model was submitted to *WHATIF* web interface to rebuild missing atom and hydrogen including generated protein structure file, psf file by VMD. Then the rebuild model was embedding in 10 Å thick water layers generated by solvate parameter script (see appendix B) in VMD program.

The implicit water solvation was conducted with periodic boundary conditions, the enolase model was solvated in a 55×61×77 $Å^3$ cubic box with 9,256 water molecules. Energy minimization was generated at constant volume for all the water molecules with molecular mechanics force field, CHARMM22, using the following cutoffs NAMD2 program: 12 Å, dielectric value and 1 * r. The energy minimization was performed for 2,000 steps of conjugate gradient until minimum energy was obtained, then followed by 30,000 steps of heat optimization until the temperature reached 300 K. The MD simulation was run by NAMD2 program with 7 processors, "namd2 +7p(procs) (configfile)", of Solaris on Sparc processors. This program is based on the Charm++ messaging system and the Converse communication layer which distributed to a wide variety of parallel platform. The time step of 1fs was used through out in the MD simulation. This MD simulation was done under NVT ensemble, fix number of atoms, fix volume, and constant temperature at 300 K for 4,000,000 steps or 3ns (parameter detail was collected in appendix). Ultimately, the refined model was then validated again by PROCHECK, Verify 3D, and ERRAT protein verify server. The model was then visualized by VMD program. Then Mg^{2+}, metal ion cofactor were added to the best obtained model and refinement was performed using CHARMM22 force field in implicit water by NAMD2. The properties of molecule refinement by MD simulation were described in detail in Table 3. Then the refined model was taken to the next procedure to docking with human plasminogen kringle domain 2 procedure.

2.6 Molecular Docking

Hex 4.5, macromolecular docking program, was used to calculate feasible docking between MpnE model and human plasminogen kringle domain 2 (1B2I),

which obtained from the Protein Data Bank at www.rcsb.org/pdb/. 1B2I was also prepared for docking with the same procedure as MpnE model refinement. From docking step, the 500 complexes were produced by full rotation search mode implement with shape and electrostatics correlation type that further analyzed. The best complex was selected by best docking parameter.

2.7 Molecular dynamics simulation with implicit solvation models for complex protein between MpnE model and human plasminogen kringle domain 2

The obtained complex model was embedded in 20 Å thick water layers that programmed by VMD. The generated protein structure file (psf file) and solvate parameter script were presented in appendix B.

The implicit water solvation was simulated with periodic boundary conditions; the enolase model was solvated in a cubic box with numerous water molecules (Table 3). Energy minimization was generated at constant volume for all the water molecules with molecular mechanics force field, CHARMM22, by NAMD2 program, using the following cutoffs: 12 Å, dielectric value: 1 * r. The energy minimization was performed for 2,000 steps using conjugate gradient. After minimum energy was obtained, heat was optimized for 30,000 steps until the temperature reached 300 K. The MD simulation was run by NAMD2 program with "namd2 +10p(procs) (configfile)", by 10 processors MacIntosch. The time step of 1fs was used through out in the MD simulation. This MD simulation was done under NVT ensemble at 300 K for 4,000,000 steps or 4,000 ps. The MD simulation parameter script is shown in the appendix C.

2.9 Complex model analysis

The chimera program was used to define hydrogen bonding of interaction pairs of the protein complex. The decrease in accessible surface area (ΔASA) was determined by Mark Gerstein's calc-surface program on protein complex by probe 1.4 Å implement in chimera program. *WHATIF* server was used to define the overall of residues that possessed an interface solvent accessible surface of complex model.

Electrostatic potentials surfaces that make contact between protein complexes were shown by PyMol program.

Table 3 Properties of molecular system used in investigation of protein complexes by MD simulation

Proteins	hPlg	MpnE+Mg^{2+}	MpnEMg^{2+}+hPlg
PDB code	1B2I	N/A	N/A
Method	NMR	Modeling	Docking
No. amino acid residues	81	456	537
Number of atoms	1,272	6,965	8,228
Number of water molecules	9,255	9,250	12,118
Box size $Å^3$	62x57x50	96x102x118	109x102x122