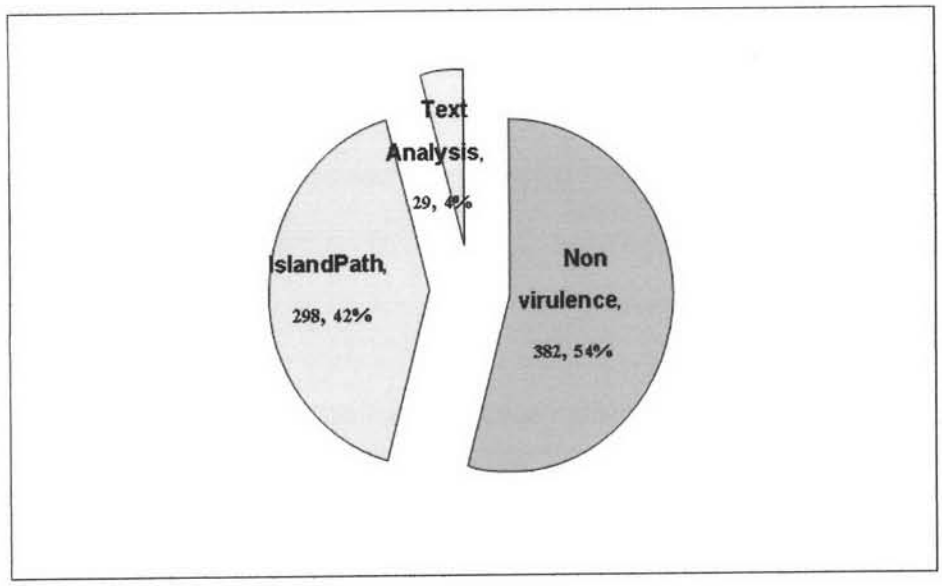
CHAPTER IV

RESULTS

1. Virulence factor identification

The totals of 327 putative virulence factors were derived from 689 open reading frame of *M. pneumoniae* genome by *PAIs* prediction tools and data mining (Figure 2A). Among these, 298 were predicted by *Islandpath*. For those predicted by *Islandpath* 113 genes were un-annotated (Figure 2B). The results from data mining showed that they were classified into Sec group, Fst group, cytadherence, hemolysin, enolase, toxin, and permease. Subsequently, un-annotated proteins were analyzed by protein feature prediction and homology motif alignment with the tools shown in Table 2. They were classified into three groups, transmembrane protein, secretory protein, and un-annotated in Table 4. Those 64 un-annotated genes which could not be identified by these tools agree with the report of *M. pneumonia* proteomics by Berkeley Structural Genomics Center (BSGC; http://www.strgen.org/status/) in March, 2006 (Figure 3). Based on results from bioinformatic tools and data mining, the characteristic of all obtained virulence factor genes were analyzed in order to screen and select most feasible genes which involved in *M. pneumoniae* invasion to host cell. The hypothetical and un-annotated genes were ignored as their functions were remained unclear. P1 adhesin has been commonly recognized for a virulence factor for *M. pneumoniae*. Enolase had been reported as a virulence factor of several microorganisms that involve in brain invasion. Therefore, P1 adhesin and enolase were selected for further study in detail by structural modeling. In order to model P1 adhesin protein structure, a modeling template had to be investigated. Unfortunately a suitable template for the whole molecule of P1 adhesin protein could not be identified by amino acid sequence similarity search. These agreed with *M. pneumoniae* proteome database at BSCG that there were no proteins with known structure matched with P1 adhesin, in Figure 4.

(A)



(B)

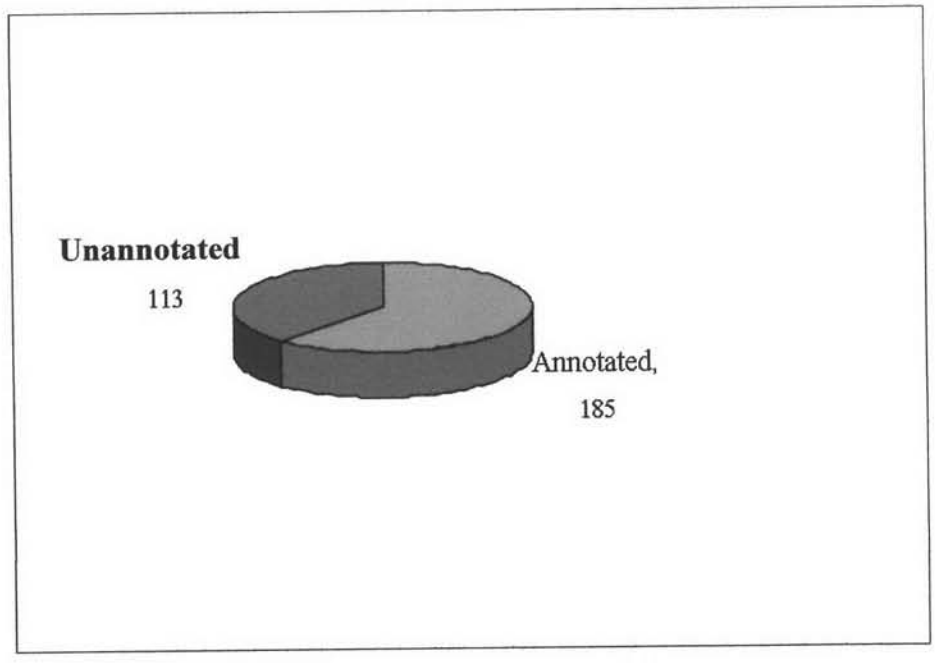


Figure 2 Summary of putative virulence factors from entire genome of *M. pneumoniae*. (A) The total number of putative virulence factors was 327 genes, 298 genes were derived by *IslandPath*, and the rest (29) genes were obtained by text-mining. (B) 113 genes of 298 virulence factors genes predicted by *IslandPath* were un-annotated.

Table 4 The protein property prediction of un-annotated gene that found in *PAIs* performed by sequence annotation tool.

Types	Gi Number	Total
Transmembrane proteins	gi_13508078,gi_13508079, gi_13508089, gi_13508098, gi_13508109, gi_13507816, gi_13507902, gi_13507842, gi_13508052, gi_13508020, gi_13508150, gi_13508187, gi_13508190, gi_13508193, gi_13508194, gi_13508208, gi_13508332, gi_13508333, gi_13508378, gi_13508379, gi_13508078, gi_13508079, gi_13508089, gi_13508098, gi_13508109, gi_13508150, gi_13508187, gi_13508190, gi_13508193, gi_13508194, gi_13508208, gi_13508332, gi_13508333, gi_13508378, gi_13508379, gi_13508419	21
Secretory proteins	gi_13507750, gi_13507751, gi_13507902, gi_13507998, gi_13508010, gi_13508020, gi_13508023, gi_13508047, gi_13508052, gi_13508102, gi_13508108, gi_13508115, gi_13508139, gi_13508195, gi_13508198, gi_13508244, gi_13508245, gi_13508326, gi_13508327, gi_13508331, gi_13508380, gi_13508381, gi_13508382, gi_13508383, gi_13508384, gi_13508385, gi_13508386, gi_13508389	28
Un-annotate proteins	gi_13507827, gi_13507749, gi_13507830,gi_13507839, gi_13507840, gi_13507843,gi_13507846, gi_13507847, gi_13507849, gi_13507860, gi_13507867, gi_13507870, gi_13507871, gi_13507883, gi_13507888, gi_13507889, gi_13507940, gi_13507941, gi_13507942, gi_13508021, gi_13508022, gi_13508025, gi_13508026, gi_13508034, gi_13508048, gi_13508050, gi_13508051, gi_13508053, gi_13508077, gi_13508082, gi_13508083, gi_13508088, gi_13508097, gi_13508103, gi_13508104, gi_13508107, gi_13508110, gi_13508116, gi_13508148, gi_13508149, gi_13508196, gi_13508201, gi_13508202, gi_13508207, gi_13508225,gi_13508238, gi_13508239, gi_13508240, gi_13508241, gi_13508243, gi_13508263, gi_13508264, gi_13508265, gi_13508269, gi_13508326, gi_13508327, gi_13508328, gi_13508330, gi_13508372, gi_13508387, gi_13508388,gi_13508401,gi_13508407, gi_13508423,	64

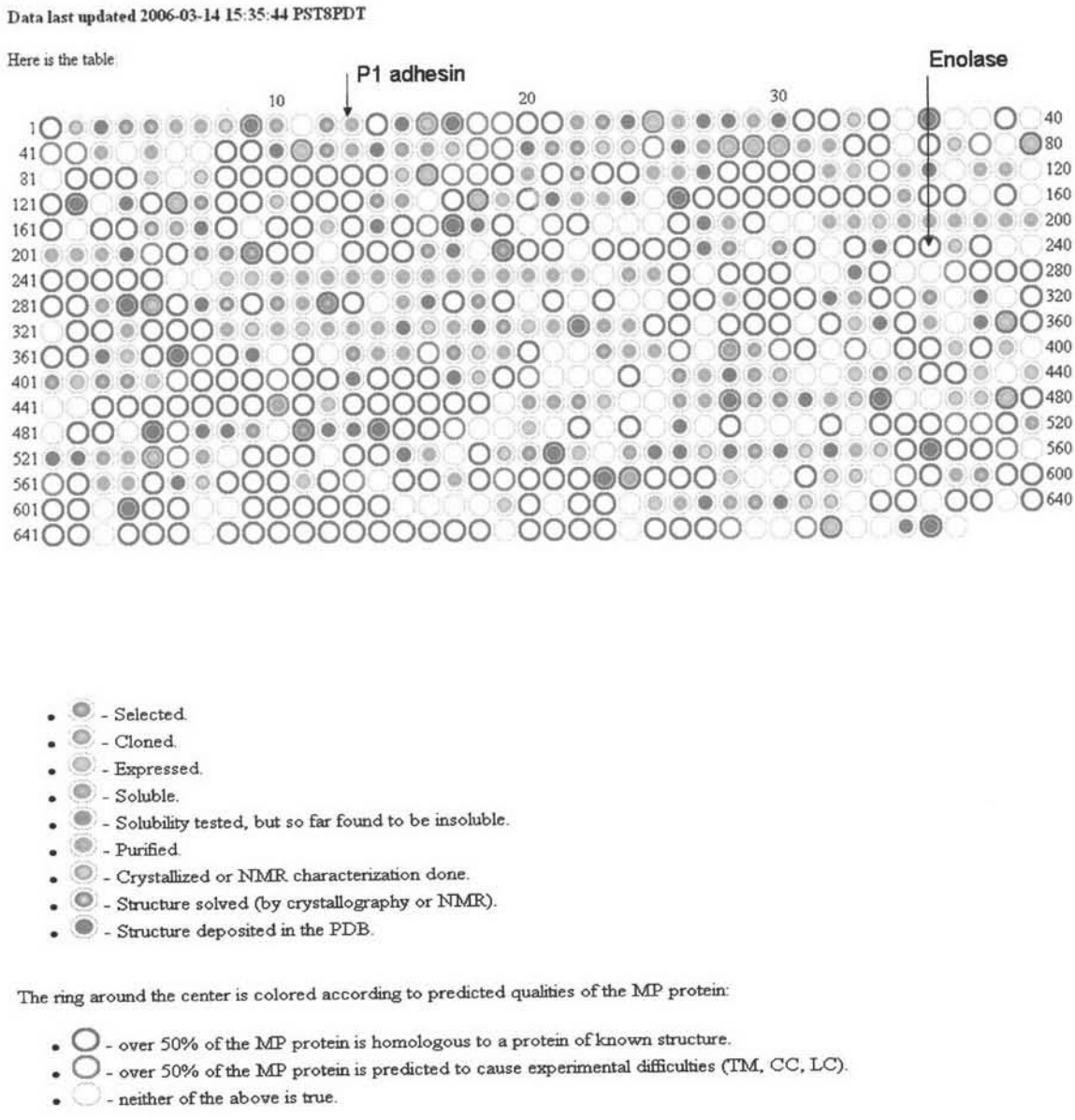


Figure 3 The proteomics profile of entire *M. pneumoniae* protein. P1 adhesin is located on MPN013 and enolase is located on MPN 606. The green ring of enolase spot indicated that it shared more than 50 % sequence identity with a protein of known structure.

BERKELEY STRUCTURAL GENOMICS CENTER

Mycoplasma pneumoniae Proteome Overview

This page shows all *Mycoplasma pneumoniae* proteins from the Proteome database (now Integr8) at EBI. Proteins with significant similarity to known structure are hilighted in green, and the structures listed in the PDB column. The include structures solved at the BSGC as well as elsewhere. Those for which the *Mycoplasma* protein itself (rather than a homolog) has been solved are indicated in darker green. Those for which the *Mycoplasma* protein is thought to be experimentally intractable for high throughput study (e.g., due to predicted transmembrane helices) are hilighted in red.

For this table, similarity was identified through a PSI-BLAST E-value of 10^-4 or better over 50 residues, or by a full-length match against a Pfam family that also matches the PDB entry. Only the 3 most significant matches are shown.

Locus	Accession	Protein	gi	Annotation	PDB
ACKA_MYCPN	P75245	MP309	2492616	Acetate kinase (EC 2.7.2.1) (Acetokinase)	1g99_A 1esz
ACPD_MYCPN	P75305	MP362	2498100	Putative acyl carrier protein phosphodiesterase (EC 3.1.4.14) (ACP phosphodiesterase)	1t5b 1ord_A 1dzq_A
ACPP_MYCPN	P75378	MP432	2492656	Acyl carrier protein homolog (ACP)	1l0b_A 1acp 1t8k
ACPS_MYCPN	P75480	MP538	15828584	Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase) (4'-phosphopantetheinyl transferase acpS)	1fte_A 1ff_A 1fth_A
ADH_MYCPN	P75214	MP278	2492770	Probable NADP-dependent alcohol dehydrogenase (EC 1.1.1.2)	1mc5_A 1mg5_A 1n6w_A
ADP1_MYCPN	P11311	MP013	13507880	Adhesin P1 precursor (Cytadhesin P1) (Attachment protein)	
ALF_MYCPN	P75089	MP129	15828581	Fructose-bisphosphate aldolase (EC 4.1.2.13)	1tv8_A 1rvr_A 1gvf_A

Figure 4 The *M. pneumoniae* proteome overview. P1 adhesin protein (pink box) showed no similarity to known protein structures in PDB.

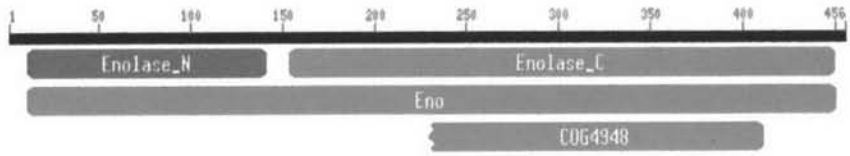
It has been known that P1 adhesin is a membrane protein. Modeling only the extracellular part of the P1 adhesin was reasonable. Membrane bound region of P1 adhesin was identified by TMHMM. It was shown that this region was still too long and a suitable modeling template was not available. Alternatively, this extracellular region was divided into 5 overlapping segments, containing approximately 300 amino acid residues of each overlapping region. They were used for searching suitable modeling templates by ProDom. An overlapping region was matched with adeno-virus knob domain (PDB code, 1KNB). It shared only 16% amino acid identity with P1 adhesin protein fragment at residue 533-704. This region was only a small part of P1 adhesin and its sequence identity with template found was very low. Theoretically, it is rather difficult to obtain a good quality model of the reality of P1 adhesin protein by comparative modeling. Practically, an alternative approach which is more appropriate should be exploited for modeling P1 adhesin structure.

Therefore enolase was a selected candidate virulence factor and would be further studied in detail of its structure as well as the feasible interaction to its counterpart.

2. *Mycoplasma pneumoniae* enolase (MpnE) structural model

The amino acid sequence of MpnE was analyzed to predict its domain by NCBI conserve domain survey (Figure 5A). N-terminal domain, ***gnl/CDD/26162 pfam03952***, was located on amino acid sequence between residue 11 and 142 (Figure 5B), 133 amino acids in length, and the region between residue 153 and 450 (red arrow), 295 residues, Figure 5B showed conserved with C-terminal TIM barrel domain, ***gnl/CDD/25403 pfam00113***. C-terminal region between residue 229 and 412, totally 372 residues, was identified "L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily" ***gnl/CDD/14081 COG4948***.

(A)



(B)

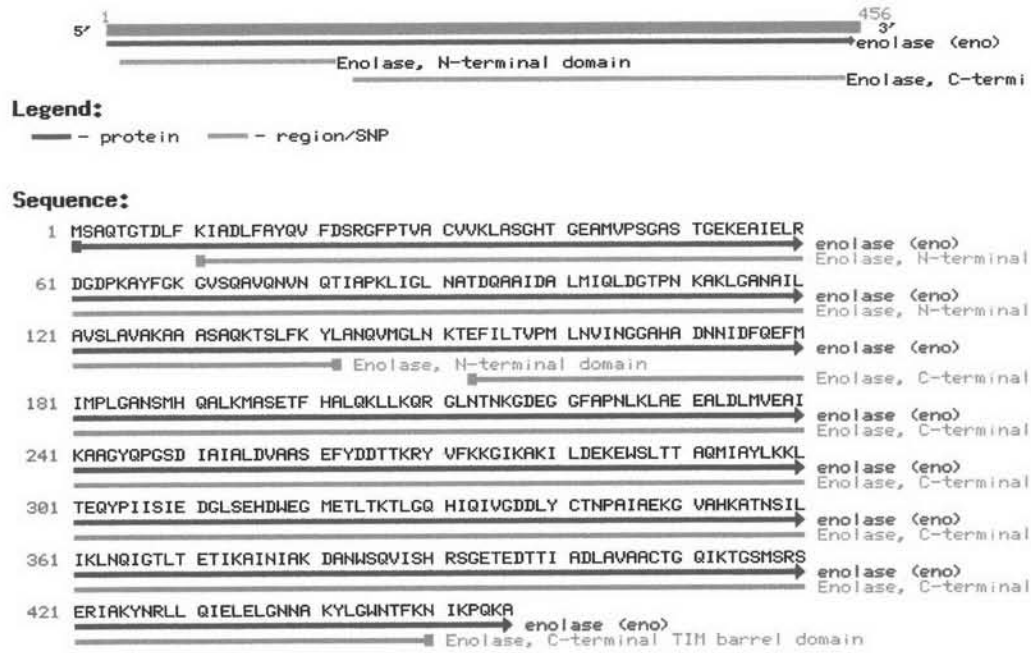


Figure 5 Protein domain predictions. (A) Two main elements, N-terminal and C-terminal domain, of MpnE which were identified by **NCBI Conserved Domain Survey**. (B) Amino acid sequence correspond N- and C- terminal domain.

Sequence comparison of microbial enolases was performed as shown in Figure 6. They were thoroughly conserved along sequences. Notably multiple sequence alignment of enolase among species represented N-terminal domain ends with highly conserved elements. Highly conserved regions were found at including $_{122}$VSLAVALAAA$_{131}$ and $_{210}$SEFY$_{213}$ which involve in quaternary structure formation.

2.1 Model evaluation of *M. pneumoniae* α- enolase

MpnE structural modeling was performed by MODELLER 8v1, a comparative modeling program, using the homologous α- enolase from *E. coli* (PDB code, 1E9I) as a template. The atomic coordinates of 1E9I was selected because, it shared 56% amino acid sequence identity with MpnE sequence. This is the best match as searched by threading web server. Two thousand models were generated. After loop modeling, the best model with lowest objective function was chosen. Subsequently, the rough selected model was then verified by validating program as following, PROCHECK, Verify3D, and ERRAT. PROCHECK evaluation showed 98.8% of the residues are in both the most favored and in the additionally allowed regions of the Ramachandran plot (Table 5).

The overall G-factor from PROCHECK was -0.11 (accepted overall G-factor a score should be above -0.5). The Verified-3D displayed 87.53% of the residues within an averaged criteria score > 0.2. The ERRAT plots showed 81.982% for overall quality factor. The refinement by energy minimization for model with 1st refinement and MD simulation for model with 2nd refinement showed better quality comparing to rough model (Table 5). Ramachandran plot for 2nd refinement model showing 99.1% of the residues are in both the most favored and in the additionally allowed regions. The overall G-factor from PROCHECK was -0.33 as shown in Table 6. The evaluation score from all structure evaluation tools suggests that implicit solvent and MD simulation can improve the structure quality.

MGAS: Musser group A Streptococcus., human bacterial pathogen group A streptococci (GAS)

Figure 6 Multiple sequence alignment of microbial enolases performed by CLUSTAL W. Gray color highlight represents the residues that located on the variant region loop of each species. Black highlighted represents the unique sequence of MpnE. The internal motif residues for quaternary structure formation are yellow highlighted.

Table 5 Model evaluations the model quality between crude and refinement model

Model	Validation tools		
	PROCHECK (%)	Verify 3D (%)	ERRAT (%)
Rough model	98.8	87.53	81.982
Model with 1st refinement	99	88.42	88.943
Model with 2nd refinement	99.1	89.31	88.943

Table 6 The model refinement quality evaluated by PROCHECK.

Model	The quality of Ramachandran plot (%)					
	Core	Allowed	General	Disallowed	Overall	G-factor
Rough model	89.5	7.8	1.5	1.2	98.8	-0.11
Model with 1st refinement	82.1	15.3	1.6	1.0	99	-0.33
Model with 2nd refinement	81.9	15.7	1.5	0.9	99.1	-0.33

Notably, the compatibility assessment of each amino acid residue qualified by Verify-3D displayed 89.31% the most residue in 2nd model refinement were best within an averaged 3D-1D score > 0.2 (overall acceptable is over 80%). The ERRAT plots showed 88.94% for overall quality factor. Based on model evaluation the selected model after second refinement is of good quality and suitable for further study.

2.2 Molecular dynamics simulation for MpnE with Mg^{2+} ion

The Mg^{2+} ion was added into the selected model and then taken to further refined by MD simulation in the canonical ensemble (NVT) at 300 K for structural equilibration. The rough model was energy minimized using 2000 steps. The state simulations reveal an increase in the root meaning square deviation (RMSD) over the first 0.5 ns, after which the RMSD state change slightly and reach structural equilibrium (Figure 7). The system reached to structure stability after 2.5 ns. The data shows that trajectories were stable with the average value of 3 Å. The rms deviation for the two superimposed structure between rough model and refined model by MD is 0.42 Å

2.3 Characteristic of MpnE model

The obtained MpnE model consists of two main domains, a smaller domain was N-terminal and a larger one was C-terminal domain. The N-terminal domain was from residue 1 to residue 142. It contains 3 β-strands forming antiparallel sheet, following by 4 α-helices as shown in Figure 8A. The C-terminal domain begins from residue 153 to 456 which contains $(\alpha/\beta)_8$ structure of TIM barrel fold with small protruding loop. This agrees with common topology of α-enolase enzyme family. Notably, the first α-helix within C-terminal domain, H5, and the second β-strand, S5, of the barrel domain are arranged anti-parallel to all other α-helices and β-strands respectively, as show the topology structure in Figure 8B. Based on visual inspection on constructed model, the metal ion cofactor, Mg^{2+}, is surrounded by D_{256}-E_{310}-D_{337} that stabilizes the active site of enolase activity, Figure 8C. This agrees with the predicted result on UniProt from http://www.expasy.org/uniprot/P75189 (see

appendix A). The result represented the enolase model sharing three conserved caboxylate ligands for the necessary divalent metal cofactor binding at the end of D_{256}, E_{310}, and D_{337}, the third, the fourth, and the fifth β-strand of the TIM barrel domain, respectively. This is consistent with the experimental study of enolase enzymatic activities evolution (Derbise *et al.*, 2004).

The active site of MpnE model is located in a pocket formed at the C-terminal ends of the barrel strands. An overall structure of MpnE model is structurally similar to those of α-enolase from *E. coli* (1E9I) and *S. pneumoniae* (1W6T) (Figure 9A). However, the superposition among α–enolase structure from *M. pneumoniae*, *S. pneumoniae*, and *E. coli* has shown the significant structural difference at variant loop, Figure 9B. The variant loop of MpnE is 13 amino acid residues long, $_{264}$DDTTKRYVFKKGI$_{276}$ as shown in gray area in Figure 6.

The variant element of *S. pneumoniae* enolase contains 20 residues as $_{252}$RKVYDYTKFEGEGAAVRTSA$_{272}$. The yellow highlighted SEFY (Figure 6), which conserved for all species indicates an internal motif for quaternary formation. These characteristics suggest that variant loop should not affect on quaternary formation and enolase activity. It can be indicated that MpnE model generally shares the same fold of overall structure with other enolases.

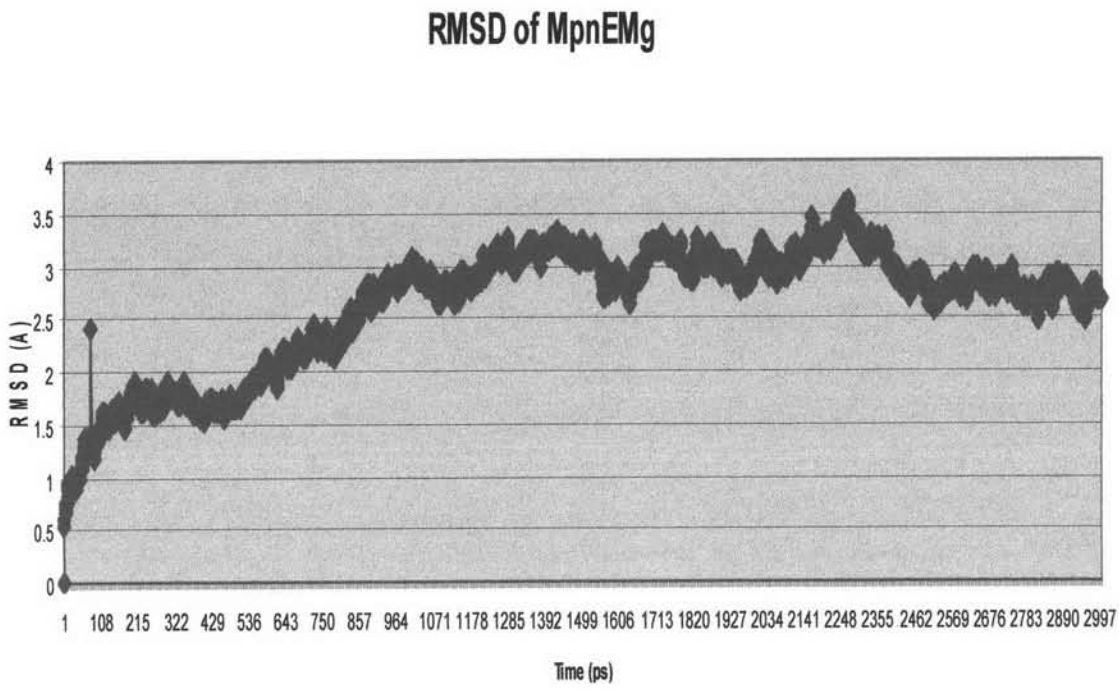


Figure 7 RMSD of MpnE with Mg^{2+} model from MD simulation in NVT ensemble versus time in picoseconds. Structure becomes to stable by 2.5 ns

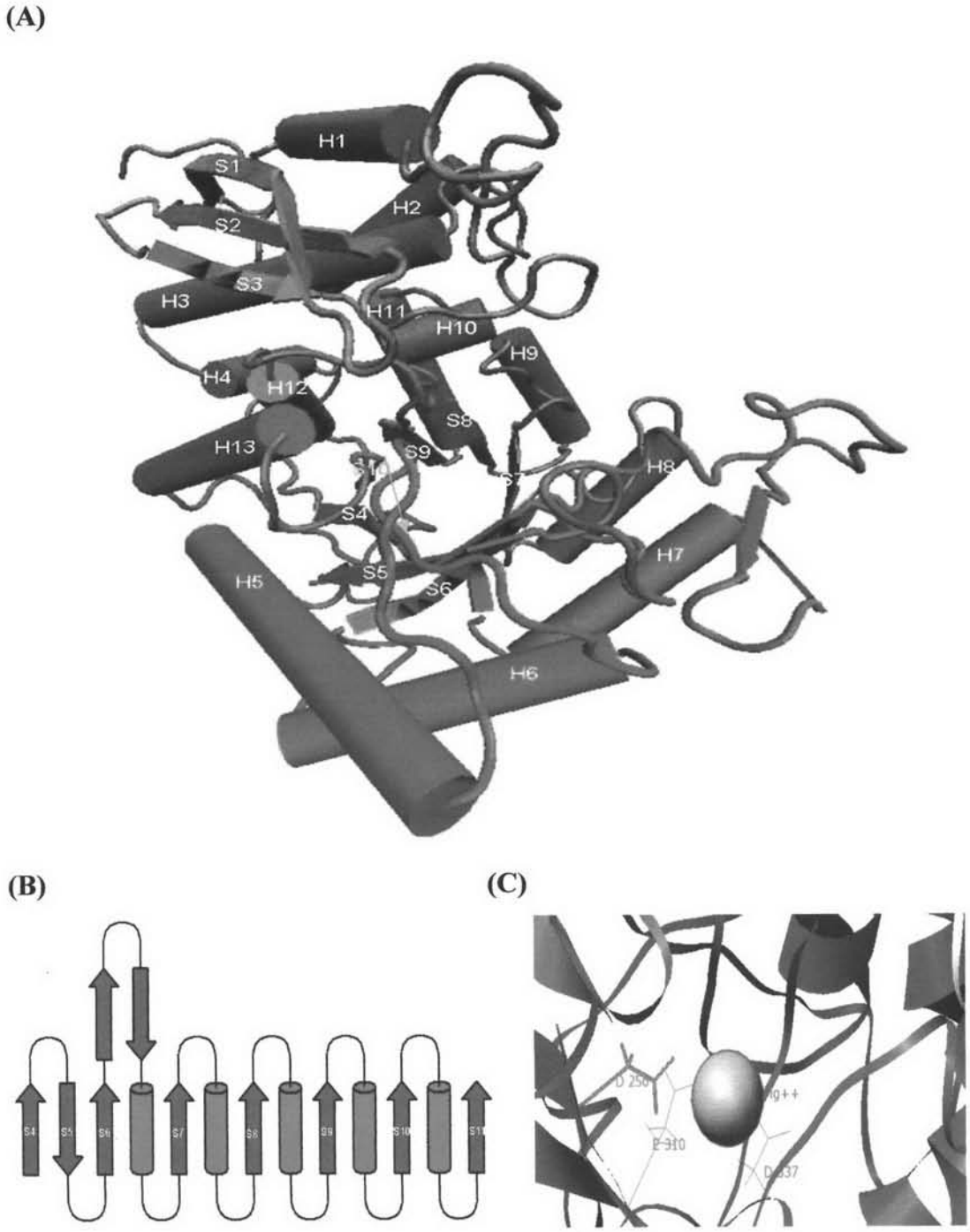
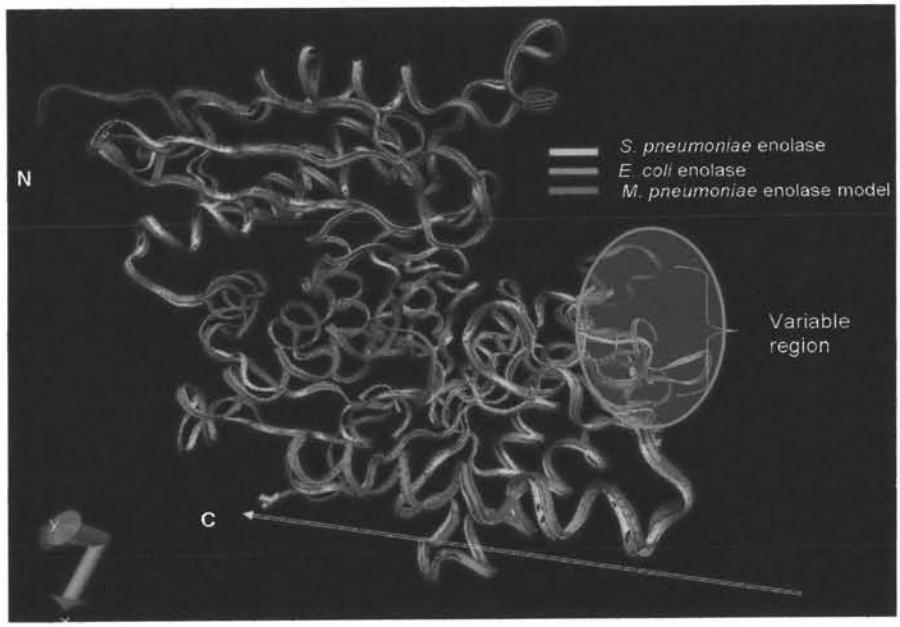


Figure 8 MpnE model consists of (A) C–terminal TIM barrel domain. The first α-helix within C- terminal domain (H5) and the second β-strand (S5) of the barrel domain (yellow arrow) are arranged anti-parallel to all other α-helices and β-strands, respectively. (B) Topology feature of α/β TIM barrel from MpnE model. (C) Mg^{2+}ion cofactor is encircled by Asp-Glu-Asp within TIM barrel domain.

(A)



(B)

Figure 9 The structural alignment of enolase among 3 organisms as *E.coli* (red), *S. pneumoniae* (yellow), and MpnE model (blue) that (A) represent variant loop region at C-terminal domain (arrow). (B) Variant loop region of MpnE model exhibits 13 residues between residue 264 and 276.

2.4 Interaction model between MpnE and human plasminogen kringle domain 2

Reliable docking experiment could provide substantial knowledge of structural protein complexes and infer functional constituents' information (Gray *et al.*, 2003). The attempts to characterize binding site of MpnE-plasminogen interaction required further analyses by molecular docking. This is an alternative approach when experimental 3D structures protein complex is not available. Interaction of MpnE-plasminogen was refined by MD simulation at 300 K with 1fs/stage (total 500,000 stages). MD simulation reveals structural equilibrium after 3 ns. The lowest global structure obtained was used to compute intermolecular binding energy. The E_{total} and binding energy compared between rigid docking and dynamic scheme showed this interaction models can be improved by MD simulation as observed by E_{total} and binding energy calculation as shown in Table 7.

This result significantly indicated that MpnE has binding affinity to human plasminogen kringle domain 2. Based on interaction from computational simulation it was pointed out that binding between MpnE and human plasminogen kringle domain 2 would be rationally occurred as shown in Figure 10A.

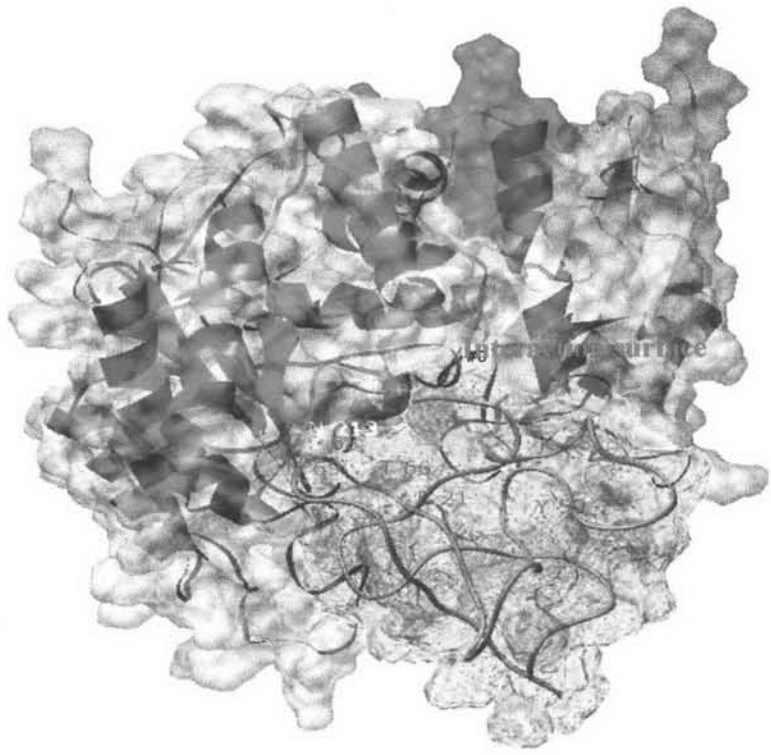
The distance of all hydrogen bonds are shorter than 3.4 Å (Table 8). This strongly indicates that hydrogen bonding can be formed. Based on hydrogen bond formation, the key residues involved in the interaction between enolase (e) and plasminogen (plg) are eR_{24}-$plgK_{48}$, eK_{70}-$plgY_{50}$, eN_{165}-$plgT_{66}$, eA_{168}-$plgE_{21}$, eD_{171}-$plgK_{70}$, and eN_{213}-$plgP_{68}$/$plgN_{69}$ (Figure 10B). From interaction result, hydrophobic and ionic interactions could be observed by electrostatic potential surface areas as shown in Figure 11. The properties of each amino acid that form complex were shown in Table 10. The result showed that 10 amino acid residues located in electrostatic potential surface area of MpnE (Figure 12A) display by PyMol. These amino acid residues, mostly located in C- terminal domain of MpnE, agree with residues involved in the interaction between human plasminogen and others plasminogen binding protein (Rios-Steiner, *et al.*, 2001; Ge, *et al.*, 2004; Gulick, *et al.*, 2001). On human plasminogen kringle domain 2 molecule, 13 residues, which involve in MpnE binding, were observed in electrostatic potential surface area (Figure 13A). These

residues included E_{21}, K_{48}, D_{67}, and K_{70} that match with corresponded residues in other microbial enolase (Jong *et al.*, 2003; Rios-Steiner *et al.*, 2001).

It has been reported that lysine residues play significant role in binding of Streptococcal surface enolase (SEN) to plasminogen (Derbise, *et al.*, 2004). It is noted that several lysine residues from both MpnE and human plasminogen kringle domain 2 contribute significantly on the interaction. The changes in accessible surface area (ASA) of MpnE model, with and without human plasminogen and also those of human plasminogen, with and without MpnE model were shown in Table 9A and 9B, respectively.

The result strongly indicated interaction of MpnE-human plasminogen is feasible. The result showed good agreement with binding experimental data on binding of human plasminogen to plasminogen-binding protein in *C. albican* (Jong, *et al.*, 2003) and *Streptococcus* species (Rios-Steiner, *et al.*, 2001). Notably the present investigation suggested that MpnE plasminogen binding structure might have a biochemistry important function and putative role as *M. pneumoniae* virulence factor

(A)



(B)

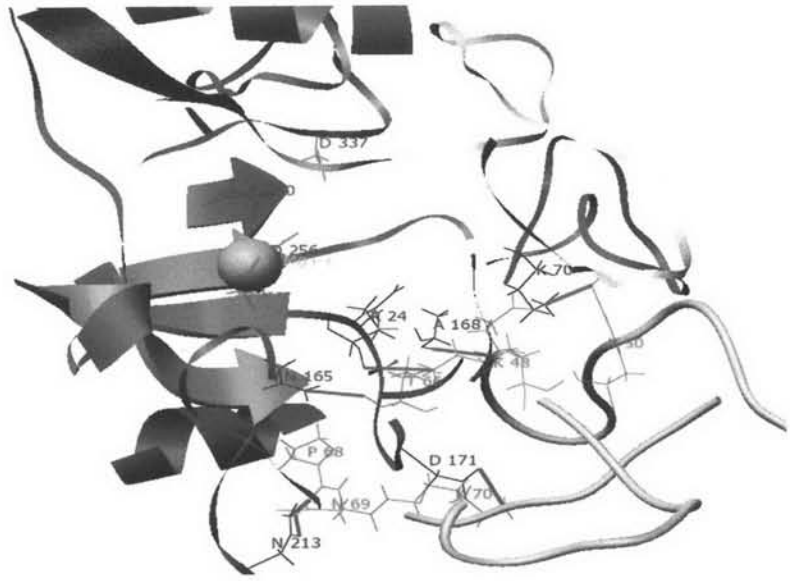


Figure 10 Plausible interactions model of (A) MpnE model (gray) and human plasminogen kringle domain 2 (1B2I) (green) was determined by molecular docking approach. (B) This picture represents 7 hydrogen bonding (red line) between MpnE model and human plasminogen kringle domain 2 (1B2I). This picture was obtained from Chimera software.

Table 7 The calculated energies (kcal/mol) of MpnE-Plasminogen interaction comparison between rigid-docking and MD simulation.

Docking method	E_{total} (kcal/mol)	Binding energy (kcal/mol)
Rigid-dcocking by Hex4.5	-324.69	-200.25
MD simulation by NAMD	-315.495	-260.312

Table 8 The distance between MpnE model and human plasminogen that makes contact by hydrogen bonding

MpnE model	Human plasminogen kringle domain 2	Distance (Å)
R 24-O	K 48-NZ	3.347
K 70-NZ	Y 50-OH	2.689
N 165-ND2	T 66-O	3.212
A 168-O	E 21-N	3.215
D 171-OD1	K 70-NZ	3.243
N 213-ND2	P 68-O	1.934
N 213-OD1	N 69-ND2	2.945

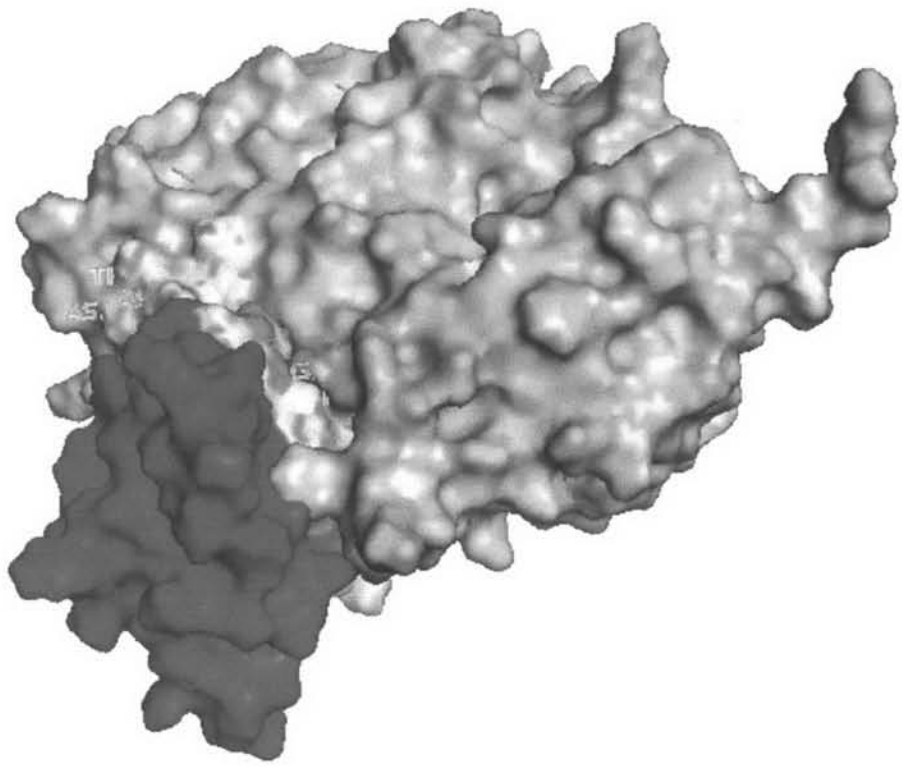
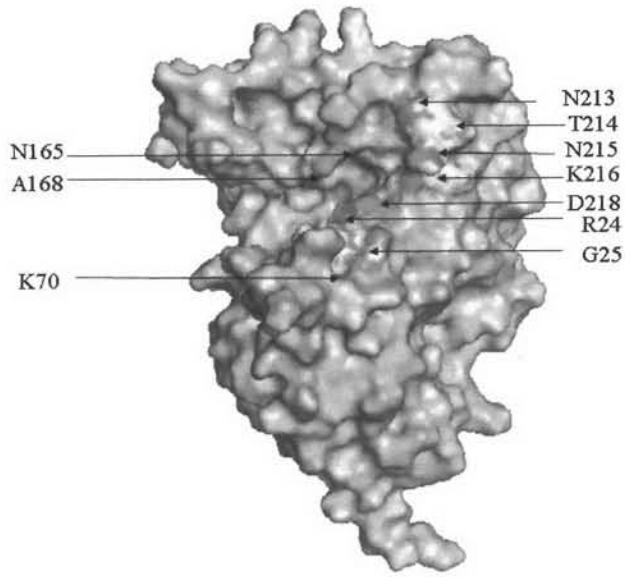


Figure 11 PyMol representation of electrostatic potential surface area of interaction between the MpnE model (pink) and human plasminogen kringle domain 2 (green)

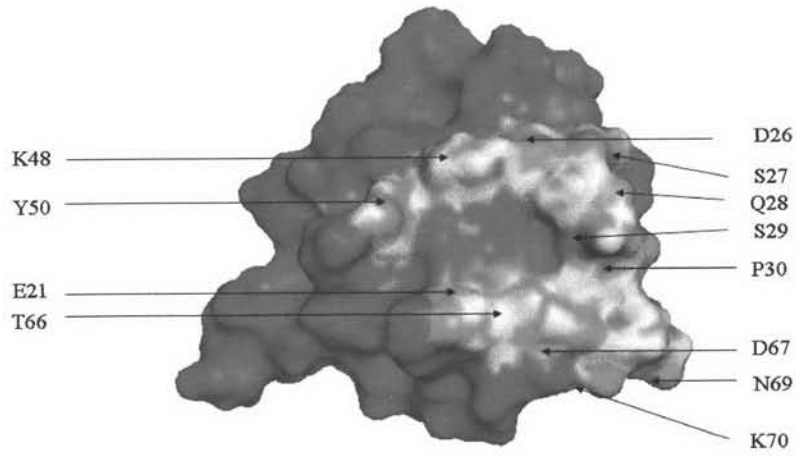
(A)



(B)

Figure 12 PyMol represents electrostatic potential surface (A) 10 particular residues that make contact with human plasminogen kringle domain 2 has been defined by ΔASA. (B) 6 of 10 residues that make contact with human plasminogen kringle domain 2 by hydrogen bonding are R_{24}, K_{70}, N_{165}, A_{168}, D_{171}, and N_{213}.

(A)



(B)

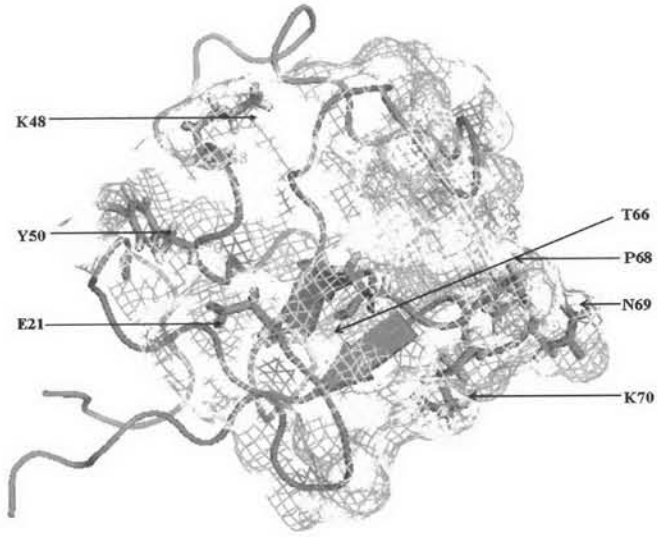


Figure 13 PyMol representation of the electrostatic potential surface of (A) human plasminogen kringle domain 2 that make contact with MpnE model has been defined by ΔASA. (B) 7 residues of human plasminogen kringle domain 2 that found hydrogen interaction with MpnE has been defined by Chimera program are K_{48}, Y_{50}, E_{21}, T_{66}, P_{68}, N_{69}, and K_{70}.

Table 9A Different Accessible Molecular Surface Accessibility (ASA) values of the amino acid of *M. pneumoniae* enolase model that makes contact or produces hydrogen bonding with human plasminogen kringle domain 2 that evaluated by *WHATIF*

MpnE model Amino acid residue	ASA of free MpnE model	ASA of MpnE model with bound plasminogen
R 24	37.1398	19.4094 [a]
F26	24.1145	9.6138
P27	0.6990	0.6990
K 70	22.9636	4.0806 [a]
N 165	0.9352	0.9352 [a]
A 168	21.2658	3.1718 [a]
D171	38.3355	38.3355 [a]
N 213	23.4285	23.4285 [a]
T214	25.0540	25.0540
N215	33.4864	33.4864
K216	19.4417	19.4417
D218	5.8420	5.6306

[a]Hydrogen bond produced

Table 9B Different Accessible Molecular Surface Accessibility (ASA) values of the amino acid of human plasminogen kringle domain 2 that makes contact or produces hydrogen bonding with MpnE model that evaluated by *WHATIF*

Plasminogen Amino acid residue	ASA of free Plasminogen	ASA of Plasminogen with bound MpnE model
E21	21.8051	3.4474 [a]
D26	9.4944	3.4258
S27	16.1071	14.5538
Q28	25.9223	7.5900
S29	20.2141	11.3453
P30	15.5025	6.6632
K48	34.7039	0.7982 [a]
Y50	18.7622	4.1334 [a]
T66	21.3467	21.0986 [a]
D67	12.6591	0.7033
P68	30.7833	7.7512 [a]
N69	40.0474	20.2065 [a]
K70	27.3235	1.2723 [a]

[a]Hydrogen bond produced

Table 10 Amino acid characterized in the interface between MpnE and Human plasminogen kringle domain 2

Amino acids property	MpnE	Plasminogen
Hydrophobic	A168, F26, P27	P30, P68
Ring containing	F26	Y50
Hydrophilic	K70, K216, R24, D171,D218, N165, N213, T214, N215	K48, K70, Q28, N69, S27, S29, T66
Basic	K70, R24	K48, K70
Acidic	D171, D218	D26, D67,E21