Article

Klebsiella pneumoniae yggG Gene Product: A Zinc-Dependent Metalloprotease

Chee Sian Kuan [1], Mun Teng Wong [1], Sy Bing Choi [2], Ching Ching Chang [1], Yoke Hiang Yee [1], Habibah A. Wahab [2],*, Yahaya Mohd Normi [3], Wei Cun See Too [1] and Ling Ling Few [1],*

[1] School of Health Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia; E-Mails: kcs10_psk046@student.usm.my (C.S.K.); wendywong84@gmail.com (M.T.W.); jinjinde_86@yahoo.com (C.C.C.); yee_hiang85@yahoo.com.ph (Y.H.Y.); stweicun@kb.usm.my (W.C.S.T.)

[2] Pharmaceutical Design and Simulation (PhDS) Laboratory, School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800 Minden, Pulau Pinang, Malaysia; E-Mail: sybing@gmail.com

[3] Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia; E-Mail: normi@biotech.upm.edu.my

* Authors to whom correspondence should be addressed; E-Mails: habibahw@usm.my (H.A.W.); fewling@kck.usm.my (L.L.F.); Tel.: +604-6532238 (H.A.W.); +609-7677536 (L.L.F.); Fax: +604-6570017 (H.A.W.); +609-7677515(L.L.F.).

Abstract: *Klebsiella pneumoniae* causes neonatal sepsis and nosocomial infections. One of the strains, *K. pneumoniae* MGH 78578, shows high level of resistance to multiple microbial agents. In this study, domain family, amino acid sequence and topology analyses were performed on one of its hypothetical protein, YggG (KPN_03358). Structural bioinformatics approaches were used to predict the structure and functionality of YggG protein. The open reading frame (ORF) of *yggG*, which was a putative metalloprotease gene, was also cloned, expressed and characterized. The ORF was PCR amplified from *K. pneumoniae* MGH 78578 genomic DNA and cloned into a pET14-b vector for heterologous expression in *Escherichia coli*. The purified YggG protein was subsequently assayed for casein hydrolysis under different conditions. This protein was classified as peptidase M48 family and subclan gluzincin. It was predicted to contain one transmembrane domain by TMpred. Optimal protein expression was achieved by induction with 0.6 mM isopropyl thiogalactoside (IPTG) at 25 °C for six hours. YggG was purified as soluble

protein and confirmed to be proteolytically active under the presence of 1.25 mM zinc acetate and showed optimum activity at 37 °C and pH 7.4. We confirmed for the first time that the *yggG* gene product is a zinc-dependent metalloprotease.

Keywords: *Klebsiella pneumoniae* MGH 78578; *yggG* gene; KPN_03358; metalloprotease; HEXXH motif

1. Introduction

Klebsiella pneumoniae is a Gram-negative, cylindrical rod shaped opportunistic pathogen that is found in the environment as well as a normal flora in mammalian mucosal surfaces like mouth, skin, and intestines. It is clinically the most important member of the family of Enterobacteriaceae that causes neonatal sepsis, nosocomial infections [1] and pneumonia [2]. Studies conducted in Asia estimated the incidence rate in elderly persons to be 15 to 40% [3,4] which is equal to or greater than that of *Haemophilus influenzae* [5].

K. pneumoniae strain MGH 78578 is one of the strains that shows high level of resistance to multiple antimicrobial agents including ampicillin, oxacillin, kanamycin, and chloramphenicol [6]. This strain was originally isolated from the sputum of a male patient in 1994 [6] and its genome has been sequenced by the Genome Sequencing Center of Washington University in Saint Louis in 2007. It was estimated that 20% of the total predicted open reading frames (ORFs) in the genome encode for hypothetical proteins, whose expressions and functions have not been proven experimentally. One of the hypothetical proteins is KPN_03358.

KPN_03358 has 231 residues of amino acids and codes for *yggG* gene. It was analyzed preliminarily using Uniprot. Uniprot [7] is an integrated database which performs retrieval of information from other databases such as metabolic database (KEGG) [8], protein–protein interaction (SPRING), family and motif databases (Pfam, InterProScan, PROSITE, HAMAP, *etc.*) homology database (SMR), dan Gene ontology (GO). Although *Klebsiella* YggG is classified as a putative uncharacterized protein, the result of sequence similarities annotation by Uniprot revealed that it belongs to peptidase M48 family. The gene ontologies (GO) indicated that the molecular function might be a hydrolase, metalloprotease or a protease based on the electronic annotation from InterPro scan database.

Metalloprotease, the most diverse of the six main types of proteases, has drawn much of our interest as it plays an important role in host-pathogen interactions by promoting enteropathogenicity, vascular permeability, host tissue damage and cytotoxicity [9]. Metalloproteases expressed by pathogens such as *Legionella pneumophila*, *Vibrio cholerae* and *Plasmodium vivax* involve in pathogenesis of the disease by degrading a wide range of host molecules [10–12]. The open reading frame of *yggG* gene (KPN_03358) from *K. pneumoniae* MGH 78578 was selected for cloning, expression and characterization in this study. This gene is highly conserved and its homologues could be found in various pathogenic microorganisms such as *Shigella dysenteriae*, *Shigella flexneri*, *Salmonella typhimurium* and *Salmonella enterica*. The gene was discovered in *Escherichia coli* when an open reading frame was found on the strand complementary to *speB* gene encoding agmatine

ureohyrolase [13]. YggG is up regulated by heat shock and it interacts with Era protein, a membrane associated GTPase that is essential for *E. coli* viability [14]. Despite its proposed function as a heat shock protein [15] and its importance for cell response to stress [16], the protease activity of YggG has never been reported and thus, it is still being designated as a hypothetical metalloprotease. The *yggG* gene product from organisms other than *E. coli* has also never been investigated. Most of the proteases contain HEXXH site, however there are certain proteins with the HEXXH site that do not possess the protease activity [17]. Besides, previous expressions of proteases under the M48 family in *E. coli* are generally toxic to the host cells [18,19]. Thus, this study aims to heterologously express YggG and to confirm the proteolytic activity of purified YggG. In addition, computational bioinformatics approaches were also utilized in order to predict the possible structure and function of this YggG protein from *K. pneumonia* strain MGH 78578.

2. Results

2.1. Homology Modeling of YggG Protein and Model Assessment

Selected hypothetical protein YggG (KPN_03358) was subjected to BLAST (Basic Local Alignment Search Tool) search against NCBI non-redundant (NR) database. Putative conserved domain was detected as Peptidase M48 superfamily during the BLAST search. More than 100 hits were found with above the threshold of 0.001 Expected-value (E-value) and majority of them were either conserved hypothetical protein or metalloprotease. Subsequently, KPN_03358 underwent another round of BLAST search with PDB (Protein Data Bank) for potential template for homology modeling. Only one available PDB structure, 3C37 has the E-value above the threshold of 0.0001. 3C37 is the X-ray structure of putative Zn-dependent peptidase from *Geobacter sulfurreducens* with the length of 253 amino acid residues. It belongs to the M48 family of peptidase. Besides having similar length of amino acid residues, both KPN_03358 and 3C37 also share the same conserved domain. The sequence identity of KPN_03358 and 3C37 is 28% with the coverage of 88% of the whole sequence length. Hence, 3C37 was selected as the template for homology modeling of KPN_03358.

The best Discrete Optimized Potential Energy (DOPE) scoring built model was selected out of the 20 randomly generated models by MODELLER. In the 3 dimensional (3D) homology model of KPN_03358 (Figure 1), the four beta strands are located in the center of the structure and they are sandwiched in between two bundles of alpha helices (Figure 1). The conserved secondary structure regions found in the multiple sequence alignment are located in the center of the structure.

The best built model was then validated using PROCHECK [20] and the Ramachandran plot showed that 98.5% of the total residues fell within the most favorable and additional allowed regions. One residue however, fell within the disallowed region. Nevertheless, based on the Ramachandran plot, the model can be accepted as the best potential model representing KPN_03358 hypothetical protein.

Figure 1. The model with the best Discrete Optimized Potential Energy (DOPE) scoring selected as our built model for KPN_03358. Alpha helix and beta sheet secondary structures are represented in purple and yellow, respectively. The graphic was generated using Visual Molecular Dynamics (VMD) visualization tool [21].
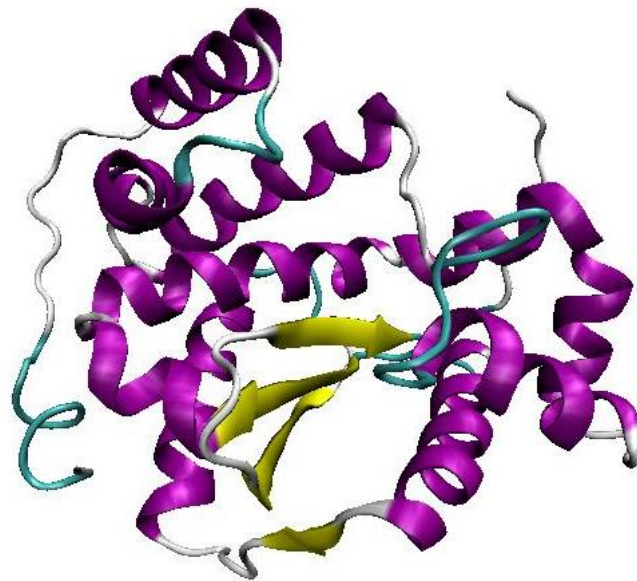
2.2. Domain Family, Amino Acid and Membrane Topology Analyses of YggG Protein

The ORF of *yggG* gene codes for a total of 231 amino acids with a calculated molecular weight of 24.7 kDa and a theoretical pI value of 5.76. Amino acids number 19 to 220 were aligned to the peptidase M48 family by Pfam sequence search. Figure 2 shows the amino acid sequence alignment between *K. pneumoniae* MGH 78578 YggG and its homologues from other species. There is one conserved putative domain found in all hits and is identified as HEXXH motif. The HEXXH motif and the third glutamic acid (55 amino acids C-terminal of the HEXXH motif) that are involved in the binding of zinc ion are marked with asterisks. *K. pneumoniae* YggG showed around 90% sequence identity with YggG proteins from *E. coli*, *S. enterica* and *Cronobacter turicensis*. In contrast, *K. pneumoniae* YggG showed only 18 and 19% sequence identities with *E. coli* HtpX [22,23] and *G. sulfurreducens* 3C37, respectively. Based on the identified conserved motif, the *K. pneumoniae* YggG protein can be classified under peptidase M48 family that belongs to subclan MA(E) or gluzincin of clan MA metalloprotease.

Figure 2. Amino acid sequence alignment of putative peptidase M48 family members of *Klebsiella pneumoniae* MGH 78578, Kp-YggG (YP_001336985); *Cronobacter turicensis*, Ct-YggG (YP_003211836); *Escherichia coli*, Ec-YggG (YP_854153); *Salmonella enterica*, Se-YggG (NP_461993); *E. coli* HtpX, Ec-HtpX (AAA62779); *Geobacter sulfurreducens* PCA, Gs-3C37 (3C37_A). GenBank accession numbers are given in parentheses. Asterisks indicate amino acids involved in zinc binding.

```
Kp-YggG    1  ----------------------------------------------------------
Ct-YggG    1  --------------------------MMKTSMERGSITMKMRAIVLSMTAAALLSG
Ec-YggG    1  MLACCAIHSPVRCAFPYTETERRFTCKRRIFRIILTLCGRENMKIRALLVAMSVATVLTG
Se-YggG    1  -----------------------------------MKIRALLIAIGMATVLTG
Ec-HtpX    1  -------------------------------MMRIALFLLTNLAVMVVFGLVLSLTG
Gs-3C37    1  --------------------------------------------------XATSXTD

Kp-YggG    1  ---MDSNGLLSSGAEAFQAYSLSDAQVKTLSDQACKEMDAKAKIAPANSEYSQRLNKIAA
Ct-YggG   31  CQNMNSNGLLSSGAEAFQAYTLSDAQVKALSDEACKSQDSKAKIAPANSQYTQRLNKIAA
Ec-YggG   61  CQNMDSSGLLSSGAEAFQAYSLSDAQVKKLSDQACQEMDSKATIAPANSEYARRLTTISR
Se-YggG   19  CQNMDSNGLLSSGAEAFQAYTLSDAQVKALSDQSCQELDSKAKIAPASSEYTKRLAKIAA
Ec-HtpX   27  IQSSSVQGLMIM--ALLFGFGGSFVSLLMSKWMALRSVGGEVIEQPRN-ERERWLVNTVA
Gs-3C37    8  IKGFNX---------------ISIEQEKELGNKFAVEIE-KQQQPVNDPEVQRYVDKVGK

Kp-YggG   58  ALGDNINGQPVNYKVYETKD--VNAFAMAN----GCIRVYSGLMDLMN-DNEVEAVIGHE
Ct-YggG   91  ALGDNINGQPVNYKVYEEKD--VNAFAMAN----GCIRVYSGLMDMMS-DNEVEAVIGHE
Ec-YggG  121  ALGDNINGQPVNYKVYMAKD--VNAFAMAN----GCIRVYSGLMDMMT-DNEVEAVLGHE
Se-YggG   79  ALGDNINGQPVNYKVYETKD--VNAFAMAN----GCIRVYSGLMDMMT-DNEVEAVIGHE
Ec-HtpX   84  TQARQAGIAMPQVAIYHAPD--INAFATGARRDASLVAVSTGLLQNMS-PDEAEAVIAHE
Gs-3C37   52  RLLSGARAVEFDYVFKVVKDDSVNAFAIPG----GRVYVHTGLLKAADNETELAGVLAHE

Kp-YggG  111  MGHVALG-----HVKKGMQVALGTNAVRAAAASAG-GIVGSLSQSQLGDLG---------
Ct-YggG  144  MGHVALG-----HVKRGMQVALGTNAVRGAAASAG-GIVGSLSQSDLGELG---------
Ec-YggG  174  MGHVALG-----HVKKGMQVALGTNAIRVAAASAG-GIVGSLSQSQLGDLG---------
Se-YggG  132  MGHVALG-----HVKKGMQVALGTNAVRVAAASAG-GVVGSLSQSQLGDLG---------
Ec-HtpX  141  ISHIANGDMVTMTLIQGVVNTFVIFISRILAQLAA-GFMGGNRDEGEESNCNPLIYFAVA
Gs-3C37  108  INHAVAR-----HGTRQXTQEYGYSLVLSLVLGDNPNXLAQLAGQLFGKAG---------

Kp-YggG  156  ----------EKLVNSQFSQRQESEADDYSYDLLRKRGINPS--GLATSFEK------LA
Ct-YggG  189  ----------EKLVNAQFSQRQESEADDYSYDLLRKRGINPS--GLATSFEK------LA
Ec-YggG  219  ----------EKLVNSQFTQRQESEADDYSYDLLRQRGISPA--GLATSFEK-----LA
Se-YggG  177  ----------EKLVNSQFSQRQESEADDYSYDLLRKRGISPA--GLATSFEK------LA
Ec-HtpX  200  TVLELVFGILASIITMWFSRHREFHADAGSAKLVGREKMIAALQRLKTSYEPQEATSMVA
Gs-3C37  154  --------------XXSYSREYENQADFLGVETXYKAGYNPN--GLTSFEQK------LN

Kp-YggG  198  KLEAGRQS---SMFDDHPASEARAQHIRDRMKADGIK------------------
Ct-YggG  231  KLEKGRQS---SMFDDHPASEARAQHIRDRMKADGIK------------------
Ec-YggG  261  KLEEGRQS---SMFDDHPASAERAQHIRDRMSADGVK------------------
Se-YggG  219  KLEAGRQS---SMFDDHPASAARAQHVRDRMSADGIK------------------
Ec-HtpX  260  LCINGKSK---SLSELFMTHPPLDKRIEALRTGEYLK------------------
Gs-3C37  192  AXDGGTQSNVARFFSTHPLTSERIQRVQAEIAKLPPQRYLTDETEFKKIKGRLKLE
```
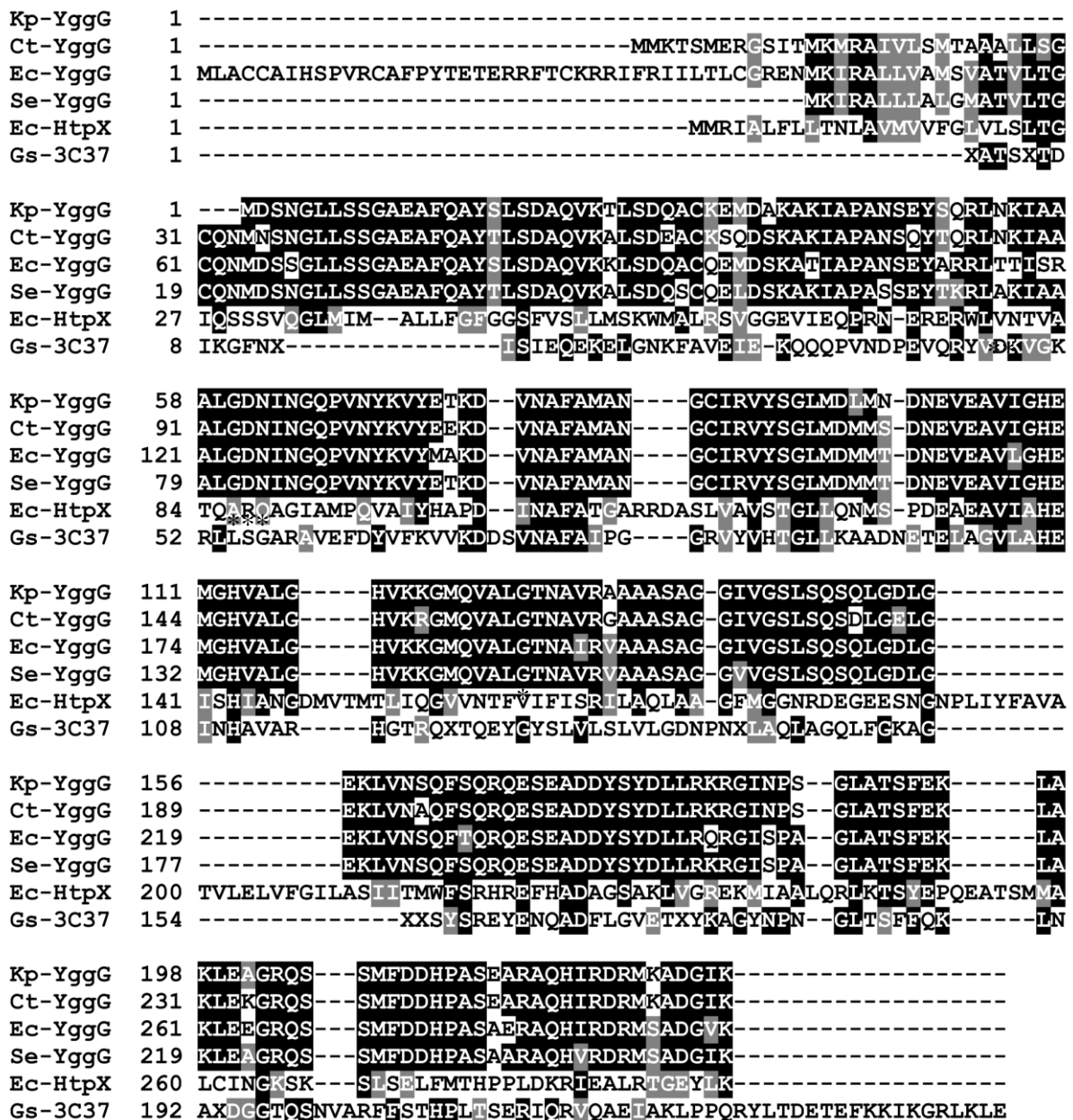
Figure 3 shows the membrane topology analysis of *K. pneumoniae* YggG by TMpred, which predicted a single membrane spanning domain from amino acid 125 to 146. The zinc binding motif (HEXXH and Glu-168) was predicted to be located in the cytosolic domain. In comparison, *E. coli* YggG, which is a membrane associated protein according to Huang *et al*. [15], contains two TMpred predicted transmembrane regions at amino acid number 3 to 20 (score = 1004) and number 152 to 170 (score = 702). In contrast, *E. coli* HtpX and *S. cerevisiae* Ste24p (Ste24p data not shown)

were predicted to contain four and six transmembrane domains, respectively, their zinc binding motif was also predicted to be in the cytosolic domain. The hydrophobicity score of *K. pneumoniae* YggG predicted transmembrane domain was 601 (transmembrane helix score >500 predicts membrane-spanning domains with high probability) compared to transmembrane regions of HtpX and Ste24p with scores of up to 2000. As shown in Figure 3, no transmembrane domain was predicted for *G. sulfurreducens* 3C37 protein by TMpred. It is interesting to note that potential transmembrane hydrophobic region was not predicted in *K. pneumoniae* YggG by other programs like SOSUI [24], hmmtop [25,26], TMHMM [27] and PhDhtm [28].

Figure 3. Comparison of transmembrane domains of *K. pneumoniae* YggG, *E. coli* YggG, *E. coli* HtpX and *G. sulfurreducens* 3C37 as predicted by TMpred. The horizontal line represents the level of hydrophobicity (score ≥500) that predicts membrane-spanning domains with high probability. Predicted transmembrane domains are indicated with Roman numerals. The position of metalloprotease motif (HEXXH) is also indicated. The X axis is the amino acid sequences of the enzyme and the Y axis is the hydrophobicity of the residues.
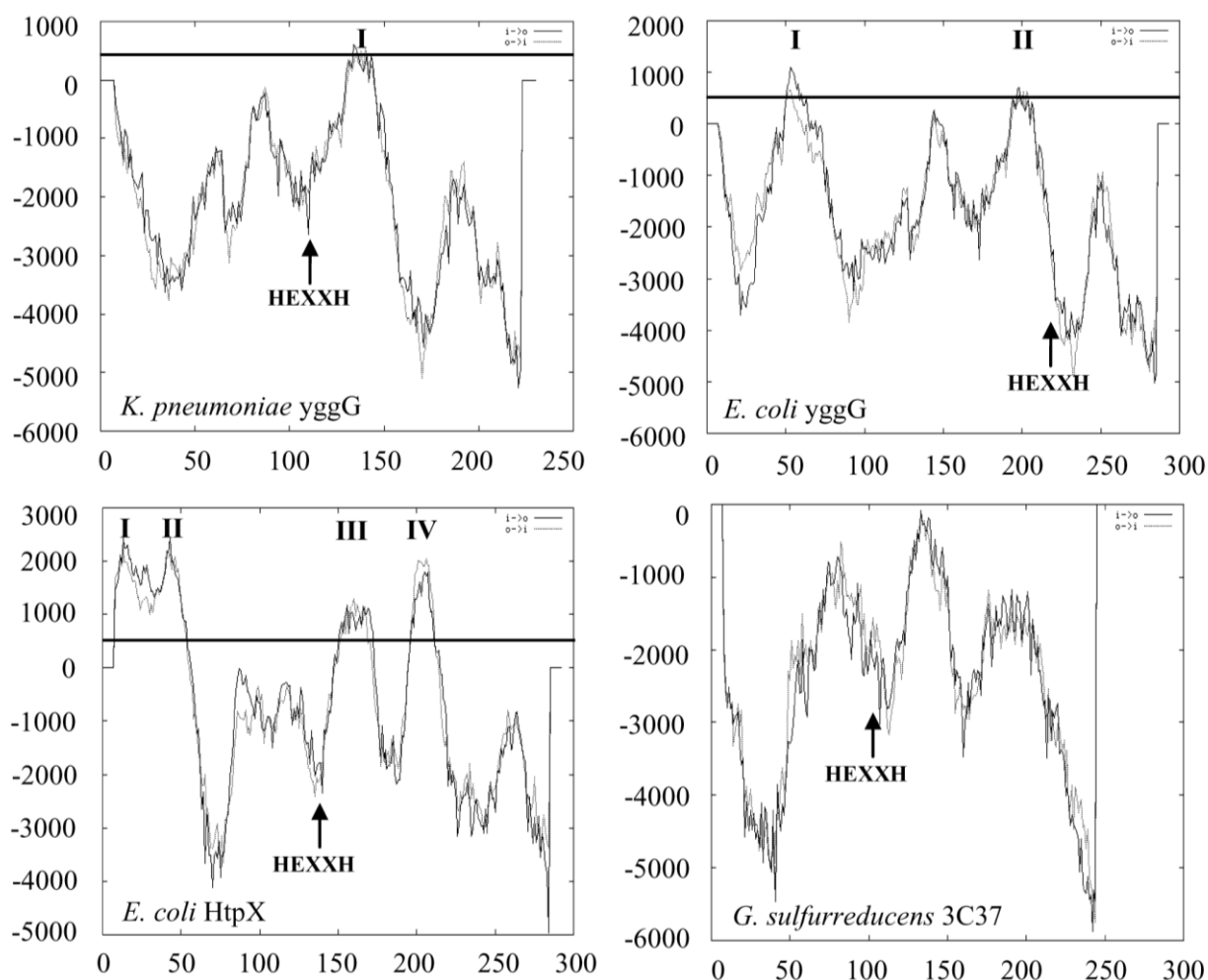
2.3. Cloning and Heterologous Expression of YggG in E. coli

A 696 bp ORF of *yggG* gene was amplified from *K. pneumoniae* MGH 78578 genomic DNA (Figure 4, lane 1) by PCR and it was cloned into a pET14-b vector for the expression of YggG as a 6×histidine tagged (His$_6$-tagged) protein in *E. coli*. The optimal induction time and temperature were 6 h and 25 °C. More than 80% of the His$_6$-tagged YggG was present as soluble fraction under the purification procedure used in this study. The histidine tag was removed in the final step and the typical yield of purified YggG (Figure 4, lane 3) per liter culture was around five milligrams.

Figure 4. PCR amplification and expression of *K. pneumoniae* MGH 78578 *yggG* ORF. A 696 bp PCR product was amplified from genomic DNA (lane 1) or cDNA (lane 2). The PCR product from lane 1 was purified and cloned into a pET-14b vector for protein expression. "−" lanes indicate no-template controls. Lane 3: Purified *K. pneumoniae* YggG protein on 12% SDS-PAGE. M1: GeneRuler 1 kbp DNA ladder (Fermentas); M2: BenchMark™ protein ladder.
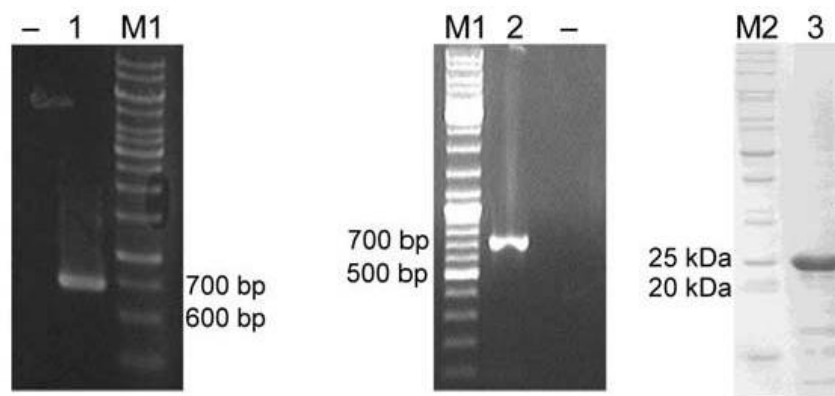
2.4. Detection of yggG mRNA Expression in K. pneumoniae

The *yggG* mRNA was detected by RT-PCR as shown in Figure 4 (lane 2). A PCR product corresponding to the size of *yggG* ORF (696 bp) was successfully amplified from *K. pneumoniae* cDNAs. The result confirms that *yggG* is not a pseudogene.

2.5. Proteolytic Activity of Purified K. pneumoniae YggG

Proteolytic activity of purified YggG was confirmed by casein hydrolysis assay. As shown in Figure 5a, five folds higher activity was detected in the reaction containing 1.25 mM zinc acetate compared to reactions without zinc and negative control. There was no significant activity without the presence of zinc ion ($p < 0.05$). YggG showed optimum proteolysis at 37 °C and pH 7.4 (Figure 5b,c). The proteolytic activity of YggG increased hyperbolically with increasing enzyme and substrate concentrations and reached its maximum at about 100 μg/assay of YggG protein and 300 μg/assay of casein (Figure 5d,e).

Figure 5. Effects of Zn^{2+} (**a**), temperature (**b**), pH (**c**), enzyme concentration (**d**), and substrate concentration (**e**) on the proteolytic activity of *K. pneumoniae* YggG. The protease activity was determined as the rate of succinylated casein hydrolysis. All the reactions were performed in triplicate and the data were plotted as their mean ± standard deviation. Asterisk indicates statistically significant difference ($p < 0.05$).
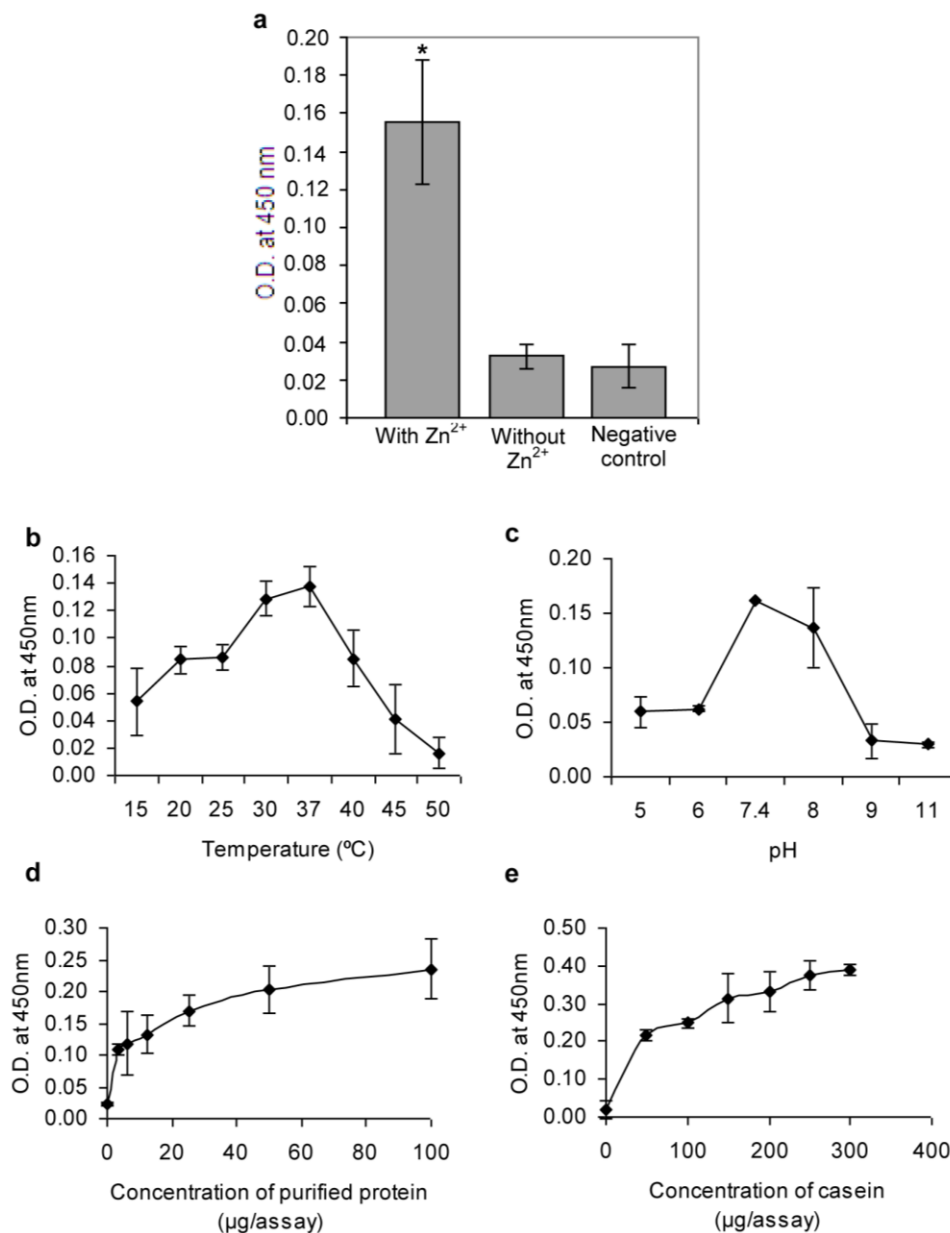
3. Discussion

In this study, the *yggG* ORF encoding a putative metalloprotease in *K. pneumoniae* was cloned and expressed in *E. coli* as 6× histidine fusion protein. The protein was purified to apparent homogeneity and confirmed to be catalytically active under the presence of zinc ion. Protein family database search predicted YggG as a metalloprotease from the peptidase M48 family. The conserved HEXXH motif and a third glutamic acid residue responsible for metal binding were identified. Over the past decade, several microbial metalloproteases in the peptidase M48 family have been cloned and characterized,

most prominent examples are Ste24p from *S. cerevisiae* [29], HtpX from *E. coli* [22] and HtpX-like heat shock metalloprotease from an unknown organism related to *Methylobacillus flagellatus* [19]. The HEXXH motif in this group of protease is critical for their catalytic activity since the proteolytic activities of Ste24p and HtpX were lost when the motif was mutated [29,30].

K. pneumoniae YggG has a smaller molecular mass (~25 kDa) compared to Ste24p (52.3 kDa) and HtpX (32 kDa) proteins. Transmembrane helices prediction with TMpred (Figure 3) also showed contrasting topologies between these proteins. The distinctive features of *K. pneumoniae* YggG suggest that this protein could be functionally different from Ste24p or HtpX. Comparison with its putative ortholog from *E. coli* [13,15] also revealed some differences in terms of the number of putative transmembrane domain and solubility. *E. coli* YggG was predicted to have an extra transmembrane domain at the N-terminus (Figure 3) and it was reported to be membrane bound [15] whereas we have successfully expressed the *K. pneumoniae* YggG as a soluble protein.

Heterologous expressions of metalloproteases usually encounter the problem of low solubility due to their incorporation into the membrane of the host organism and their associations with insoluble inclusion bodies. Laborious refolding step was required to obtain the proteases in their active and stable forms [19,23,31]. The full length and 21 amino acids N-terminal truncated versions of YggG protein from *E. coli* were also membrane-associated [15]. On the contrary, the *K. pneumoniae* YggG produced in this study was soluble when expressed in *E. coli*. Recently, an astacin metalloprotease from a parasitic nematode, *Steinernema carpocapsae* was also successfully expressed as a soluble protein in *E. coli* [32]. Expression of *K. pneumoniae* YggG as a soluble protein was in agreement with the results from topology analysis that predicted a single weak hydrophobic transmembrane domain in this protein.

Most metalloproteases are characterized by a catalytic zinc ion although in some enzymes, manganese, cobalt, nickel or even copper ions can also undertake this function [33]. The purified *K. pneumoniae* YggG exhibited significant casein proteolytic activity when zinc ion was supplied as a metal cofactor. The presence of a zinc-binding motif, HEXXH, in this enzyme suggests that the catalytic activity of YggG is modulated by zinc through its binding to this site [22,23,30]. Specific point mutation on this motif can be done to confirm the importance of zinc binding to YggG's activity.

The sequence of *yggG* ORF cloned in this study was based on the completed GenBank genome sequence of *K. pneumoniae* strain MGH 78578. Recently, completed genome sequences of N_2-fixing *K. pneumoniae* 342 (KPK_0741) [34] and liver abscess and meningitis causing *K. pneumoniae* NTUH-K2044 (KP1_4646) [35] strains have been published. The deduced YggG protein sequences from these two strains have an additional stretch of 21 amino acids that are highly similar to the *E. coli* YggG protein. This stretch of amino acids would also constitute a putative hydrophobic transmembrane domain. However, based on bioinformatics analysis and molecular mass of immunodetected *E. coli* YggG protein, it was suggested that the first 17 to 19 amino acids were removed in the mature YggG protein [15]. Therefore, the catalytically active YggG of *K. pneumonia* MGH 78578 produced in this study has a high likelihood of resembling the mature YggG in the cell.

4. Experimental Section

4.1. Computational Methodology for Homology Modelling

YggG protein sequence was subjected to a series of BLAST search against non-redundant database (NR) and PDB for template selection. Its structure prediction was done using MODELLER 9v8 [36]. 20 models were generated randomly and the best Discrete Optimized Potential Energy (DOPE) scoring model was selected. Verification of the built model was done using PROCHECK.

The domain family analysis was performed using Pfam protein families database [37]. ClustalW2 [38] was used to perform the multiple sequence alignment with related metalloproteases. Transmembrane region of YggG protein was predicted and compared to other metalloproteases under the same peptidase M48 family by using TMpred [39].

4.2. Bacterial Strain, Growth and Culture Conditions

Klebsiella pneumoniae subsp. *pneumoniae* MGH 78578 (ATCC number 700721) was used for this study. The bacterial strain was routinely cultured in Luria-Bertani [18] medium at 37 ℃.

4.3. Total Genomic DNA and RNA Extractions from K. pneumoniae *MGH 78578*

Genomic DNA and total RNA were isolated from a five mililiter overnight culture of *K. pneumoniae* MGH 78578 using QIAmp DNA Mini Kit (Qiagen) and RNeasy Mini Kit (Qiagen), respectively, according to the manufacturer's protocol. The integrity and size distribution of total purified RNA was visualized by ethidium bromide staining after electrophoresis on a 1% agarose gel.

4.4. PCR Cloning of K. pneumoniae yggG *Open Reading Frame*

The *yggG* ORF was PCR amplified from *K. pneumoniae* genomic DNA in a 50 μL reaction consisted of $10 \times$ Thermopol buffer (New England Biolabs), 1 μM each of forward (5'-GAATTCCAT ATGGACTCCAACGGTCTGCTCAGC-3') and reverse (5'-CGCGGATCCTTATTTAATGCCGTCG GCCTTCATGC-3') primers, 5 mM dNTPs, 1 unit of Taq polymerase and 100 ng genomic DNA. The PCR was performed for 30 cycles of 95 °C for 30 s, 66 °C for 30 s and 72 °C for 60 s.

The PCR product was gel purified by Qiaquick gel extraction kit (Qiagen), digested with *Nde*I and *BamH*I (New England Biolabs) and ligated into a pET-14b vector (Novagen) precut with the same restriction enzymes. The resulting pET-14b-yggG was confirmed by sequencing.

4.5. Expression and Purification of YggG Protein

For protein expression, the pET-14b-yggG plasmid was transformed into the *E. coli* BL21 (DE3) strain. The culture was grown in LB medium (with 100 μg/mL ampicillin) at 37 °C, 200 rpm to an $OD_{600 \ nm}$ of 1.8. Subsequently, the expression of $6 \times$ histidine tag YggG was induced with 0.6 mM isopropyl thiogalactoside (IPTG) After the induction period, the cells were pelleted and re-suspended in 5 mL pre-cooled buffer (50 mM Tris-HCI, pH 7.2, 300 mM sodium chloride and 10 mM imidazole). The cells were then sonicated and centrifuged at $2000 \times$ g for 20 min. Ni^{2+}-NTA resin (Qiagen) was added to the supernatant and binding of the His-YggG was carried out by gentle mixing for 2 h. After

the binding step, the resin was spun down at 440 × g for 5 min and the supernatant was discarded. The resin was washed eight times with 10 mL of the same buffer for 30 min. After washing, the resin was resuspended in 3 mL of buffer and six NIH units of bovine thrombin (MP Biomedicals) were added to cleave off the His-tag from the fusion protein. The purified protein was eluted as un-tagged YggG after overnight thrombin cleavage and quantified by using Bradford reagent (Bio-Rad).

4.6. Reverse Transcription PCR of yggG Gene

RevertAid H Minus first strand cDNA synthesis kit (Fermentas) was used to synthesize the cDNA from the extracted total RNA. One microgram of total RNA was mixed with 0.2 μg random hexamer primer and DEPC-treated water. Subsequently, the mixture was preheated at 70 °C for 5 min, chilled on ice and followed by the addition of 4 μL of 5 × RT buffer, 1 mM of dNTP mix and 20 units of Ribolock Ribonuclease inhibitor. The mixture was incubated at 25 °C for 5 min followed by incubation at 37 °C for another 5 min. 200 units of Revertaid H Minus M-MuLVRT was added to make up a total volume of 20 μL. The mixture was then incubated at 42 °C for 1 h and heated at 70 °C for 10 min for the termination of the reverse transcription process. PCR was performed as described above except that the template was replaced with one microliter of cDNA.

4.7. Casein Hydrolysis Assay

The protease activity of the purified YggG protein was measured by casein hydrolysis assay using a QuantiCleave Protease Assay Kit (Pierce). Briefly, 100 μL of succinylated casein substrate solution (2 mg/mL in assay buffer: 40 mM disodium hydrogen orthophosphate, pH 7.4) were added into one set of microtiter plate wells. Another duplicate set of wells were added with the assay buffer to serve as blanks. Forty microliters of purified YggG (1 mg/mL) were added into both the substrate and blank wells and incubated at 37 °C for 150 min. Subsequently, 100 μL of diluted trinitrobenzene sulfonic acid (TNBSA), 0.033% (w/v) were added to each well and incubated for 35 min at 37 °C. The absorbance at 450 nm was determined for every well and the proteolytic activity was represented by the change in absorbance (ΔA_{450}), which was calculated by subtracting the absorbance of the blank from that of the corresponding casein well. Data from a triplicate experiment were analyzed with one way ANOVA using SPSS version 15.0. The level of significance was set at $p = 0.05$. The assay was also repeated with assay buffer containing 1.25 mM zinc acetate. Reaction without purified YggG was used as the negative control.

The effects of purified YggG (0 to 100 μg/assay) and substrate (0 to 300 μg/assay) concentrations on the protease activity were also determined under the presence of 1.25 mM zinc acetate and the assay conditions as described above.

4.8. Determination of Temperature and pH Optima of the Activity of Purified YggG

Forty micrograms of purified protein and 100 μg of succinylated casein were used per assay for the determination of optimum temperature and pH for YggG casein hydrolysis activity. The buffers used were 100 mM acetate buffer for pH 5.0 and 6.0, and 40 mM disodium hydrogen orthophosphate buffer

for pH 7.4, 8.0, 9.0 and 11.0. The effect of temperature on the protease activity of YggG was studied by incubating the standard reaction mixture at temperatures ranging from 15 ℃ to 50 ℃ for 150 min.

5. Conclusions

Here, we have presented the predicted model as well as described the cloning, expression, purification and characterization of YggG metalloprotease from *K. pneumoniae* MGH 78578. This enzyme was soluble in *E. coli* and required zinc as cofactor for catalysis. The soluble and active YggG protein can promote further biochemical studies such as the identification of natural substrate(s) and the search for inhibitors of this enzyme [40]. It is interesting to see whether *yggG* knockout of *K. pneumoniae* strain is also not lethal as in the case of *E. coli* [14] and whether it would have any effect on the virulence of *K. pneumoniae* similar to the previously reported peptidase M48 family of *Porphyromonas gingivalis* [41]. In addition, the feasibility of YggG heterologous expression in *E. coli* is hoped to encourage structural studies of this enzyme.

Acknowledgements

References

1. Alcantar-Curiel, M.D.; Garcia-Latorre, E.; Santos, J.I. *Klebsiella pneumoniae* 35 and 36 kDa porins are common antigens in different serotypes and induce opsonizing antibodies. *Arch. Med. Res.* **2000**, *31*, 28–36.
2. Ishida, T.; Hashimoto, T.; Arita, M.; Ito, I.; Osawa, M. Etiology of community-acquired pneumonia in hospitalized patients: A 3-year prospective study in Japan. *Chest* **1998**, *114*, 1588–1593.
3. Lee, K.H.; Hui, K.P.; Tan, W.C.; Lim, T.K. *Klebsiella* bacteraemia: A report of 101 cases from National University Hospital, Singapore. *J. Hosp. Infect.* **1994**, *27*, 299–305.
4. Tang, L.M.; Chen, S.T.; Hsu, W.C.; Chen, C.M. *Klebsiella* meningitis in Taiwan: An overview. *Epidemiol. Infect.* **1997**, *119*, 135–142.
5. Ko, W.C.; Paterson, D.L.; Sagnimeni, A.J.; Hansen, D.S.; Von Gottberg, A.; Mohapatra, S.; Casellas, J.M.; Goossens, H.; Mulazimoglu, L.; Trenholme, G.; *et al.* Community-acquired *Klebsiella pneumoniae* bacteremia: global differences in clinical patterns. *Emerg. Infect. Dis.* **2002**, *8*, 160–166.
6. Ogawa, W.; Li, D.W.; Yu, P.; Begum, A.; Mizushima, T.; Kuroda, T.; Tsuchiya, T. Multidrug resistance in *Klebsiella pneumoniae* MGH78578 and cloning of genes responsible for the resistance. *Biol. Pharm. Bull.* **2005**, *28*, 1505–1508.
7. Jain, E.; Bairoch, A.; Duvaud, S.; Phan, I.; Redaschi, N.; Suzek, B.E.; Martin, M.J.; McGarvey, P.; Gasteiger, E. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinform.* **2009**, *10*, 136.

8. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360.

9. Miyoshi, S.; Shinoda, S. Microbial metalloproteases and pathogenesis. *Microbes. Infect.* **2000**, *2*, 91–98.

10. Dreyfus, L.A.; Iglewski, B.H. Purification and characterization of an extracellular protease of *Legionella pneumophila*. *Infect. Immun.* **1986**, *51*, 736–743.

11. Hase, C.C.; Finkelstein, R.A. Cloning and nucleotide sequence of the *Vibrio cholerae* hemagglutinin/protease (HA/protease) gene and construction of an HA/protease-negative strain. *J. Bacteriol.* **1991**, *173*, 3311–3317.

12. Fakruddin, J.M.; Biswas, S.; Sharma, Y.D. Metalloprotease activity in a small heat shock protein of the human malaria parasite *Plasmodium vivax*. *Infect. Immun.* **2000**, *68*, 1202–1206.

13. Szumanski, M.B.; Boyle, S.M. Influence of cyclic AMP, agmatine, and a novel protein encoded by a flanking gene on speB (agmatine ureohydrolase) in *Escherichia coli*. *J. Bacteriol.* **1992**, *174*, 758–764.

14. Huang, Y.; Zhang, B.; Dong, K.; Zhang, X.; Hou, L.; Wang, T.; Chen, N.; Chen, S. Up-regulation of *ygg*G promotes the survival of *Escherichia coli* cells containing era-1 mutant protein. *FEMS Microbiol. Lett.* **2007**, *275*, 8–15.

15. Huang, Y.; Dong, K.; Zhang, X.; Zhang, B.; Hou, L.; Chen, N.; Chen, S. Expression and regulation of the *ygg*G gene of *Escherichia coli*. *Curr. Microbiol.* **2008**, *56*, 14–20.

16. Ojima, Y.; Komaki, M.; Nishioka, M.; Iwatani, S.; Tsujimoto, N.; Taya, M. Introduction of a stress-responsive gene, *ygg*G, enhances the yield of L-phenylalanine with decreased acetic acid production in a recombinant *Escherichia coli*. *Biotechnol. Lett.* **2009**, *31*, 525–530.

17. Rawlings, N.D.; Barrett, A.J. Evolutionary families of metallopeptidases. *Methods Enzymol.* **1995**, *248*, 183–228.

18. Halbhuber, Z.; Petrmichlova, Z.; Alexciev, K.; Thulin, E.; Stys, D. Overexpression and purification of recombinant membrane PsbH protein in *Escherichia coli*. *Protein Expr. Purif.* **2003**, *32*, 18–27.

19. Siddiqui, A.A.; Jalah, R.; Sharma, Y.D. Expression and purification of HtpX-like small heat shock integral membrane protease of an unknown organism related to *Methylobacillus flagellatus*. *J. Biochem. Biophys. Methods* **2007**, *70*, 539–546.

20. Laskowski, R.A.; Macarthur, M.W.; Moss, D.S.; Thornton, J.M. Procheck—a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Cryst.* **1993**, *26*, 283–291.

21. Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

22. Kornitzer, D.; Teff, D.; Altuvia, S.; Oppenheim, A.B. Isolation, characterization, and sequence of an *Escherichia coli* heat shock gene, *htp*X. *J. Bacteriol.* **1991**, *173*, 2944–2953.

23. Sakoh, M.; Ito, K.; Akiyama, Y. Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *J. Biol. Chem.* **2005**, *280*, 33305–33310.

24. Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **1998**, *14*, 378–379.

25. Tusnady, G.E.; Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **2001**, *17*, 849–850.

26. Tusnady, G.E.; Simon, I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **1998**, *283*, 489–506.

27. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.

28. Combet, C.; Blanchet, C.; Geourjon, C.; Deleage, G. NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* **2000**, *25*, 147–150.

29. Fujimura-Kamada, K.; Nouvet, F.J.; Michaelis, S. A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH2-terminal processing of the yeast a-factor precursor. *J. Cell Biol.* **1997**, *136*, 271–285.

30. Shimohata, N.; Chiba, S.; Saikawa, N.; Ito, K.; Akiyama, Y. The Cpx stress response system of *Escherichia coli* senses plasma membrane proteins and controls HtpX, a membrane protease with a cytosolic active site. *Genes Cells* **2002**, *7*, 653–662.

31. da Silveira, R.B.; Wille, A.C.; Chaim, O.M.; Appel, M.H.; Silva, D.T.; Franco, C.R.; Toma, L.; Mangili, O.C.; Gremski, W.; Dietrich, C.P.; *et al*. Identification, cloning, expression and functional characterization of an astacin-like metalloprotease toxin from *Loxosceles intermedia* (brown spider) venom. *Biochem. J.* **2007**, *406*, 355–363.

32. Jing, Y.; Toubarro, D.; Hao, Y.; Simoes, N. Cloning, characterisation and heterologous expression of an astacin metalloprotease, Sc-AST, from the entomoparasitic nematode *Steinernema carpocapsae. Mol. Biochem. Parasitol.* **2010**, *174*, 101–108.

33. Mansfeld, J. Metalloproteases. In *Industrial Enzymes: Structure, Function and Applications*; Polaina, J., MacCabe, A.P., Eds.; Springer: Amsterdam, The Netherlands, 2007; pp. 221–242.

34. Fouts, D.E.; Tyler, H.L.; DeBoy, R.T.; Daugherty, S.; Ren, Q.; Badger, J.H.; Durkin, A.S.; Huot, H.; Shrivastava, S.; Kothari, S.; *et al*. Complete genome sequence of the N$_2$-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet.* **2008**, *4*, e1000141.

35. Wu, K.M.; Li, L.H.; Yan, J.J.; Tsao, N.; Liao, T.L.; Tsai, H.C.; Fung, C.P.; Chen, H.J.; Liu, Y.M.; Wang, J.T.; *et al*. Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. *J. Bacteriol.* **2009**, *191*, 4492–4501.

36. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

37. Finn, R.D.; Mistry, J.; Tate, J.; Coggill, P.; Heger, A.; Pollington, J.E.; Gavin, O.L.; Gunasekaran, P.; Ceric, G.; Forslund, K.; *et al*. The Pfam protein families database. *Nucleic Acids Res.* **2010**, *38*, D211–D222.

38. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.

39. Hofmann, K.; Stoel, W. Tmbase—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **1993**, *374*, 166.

40. Enoksson, M.; Zhu, W.; Salvesen, G. Identification of Protease Substracts by Mass Spectrometry Approaches-1. In *The Cancer Degradome*; Edwards, D., Hoyer-Hansen, G., Blasi, F., Sloane, B.F., Eds.; Springer Science + Business Media, LLC: New York, NY, USA, 2008; pp. 67–82.

41. Walters, S.; Belanger, M.; Rodrigues, P.; Whitlock, J.; Progulske-Fos, A. A member of the peptidase M48 superfamily of *Porphyromonas gingivalis* is associated with virulence *in vitro* and *in vivo*. *J. Oral Microbiol.* **2009**, *1*, DOI:10.3402/jom.v1i0.2021.