https://doi.org/10.48048/tis.2023.6377

In silico Analysis and Characterization of a Putative Aspartic Proteinase Inhibitor, IA_3 from *Lachancea kluyveri*

Wichuda Jankangram[1] and Sunthorn Chooluck[2,*]

[1]Faculty of Science and Social Sciences, Burapha University, Sakaeo Campus, Sakaeo 27160, Thailand
[2]Department of Biochemistry, Faculty of Science, Burapha University, Chonburi 20131, Thailand

([*]Corresponding author's e-mail: sunthornc@buu.ac.th)

Received: 21 November 2022, Revised: 17 January 2023, Accepted: 18 January 2023, Published: 19 January 2023

Abstract

Aspartic proteinases play important role in various physiological and biological processes. Understanding catalytic mechanisms of these enzymes could lead to crucial medical and biological applications. Two types of aspartic proteinase inhibitors have been identified i.e. small molecule inhibitor and naturally occurring peptides. Most of aspartic proteinases are highly susceptible to inhibition by a series of small, non-proteinaceous natural products; pepstatin. However, only limited number of naturally-occurring polypeptide inhibitors of aspartic proteinases has so far been discovered. One among these inhibitors is *Saccharomyces cerevisiae* IA_3. The *S. cerevisiae* Proteinase A (ScPrA) is solely and potently inhibited by this small polypeptide at sub-nanomolar level. It was proven that, not only IA_3 has no detectable effect against a wide range of aspartic proteinases, but it was also shown to be cleaved as a substrate of these non-target enzymes. Bioinformatics analysis of the IA_3 structure has been undertaken in this study. Database searching for sequence homology from available fungal genome data bank using the Basic Local Alignment Search Tool (BLAST) revealed that 4 structurally related, IA_3-like proteins have successfully been identified. The amino acid sequence of IA_3-like proteins from *Lachancea kluyveri* share highest degree of similarity toward wild-type IA_3. The Lk-IA_3-like gene was synthesized using a PCR-based gene synthesis method. Protein expression in *E. coli*, protein purification and characterization of Lk-IA_3-like by enzyme kinetic assays were performed. The results indicated that Lk-IA_3-like protein inhibits ScPrA at the Ki value of 190 ± 0.11 nM and possesses no inhibitory effect toward AfPrA (*Aspergillus fumigatus* proteinase A) or HuCatD (Human cathepsin D).

Keywords: Bioinformatics, BLAST, QSAR, Aspartic proteinase, IA_3

Introduction

Proteinases are classified into groups or families base on their similarity in their primary structures and/or their biological activities. Some families are also divided into subfamilies based on each unique property. The proteinase enzymes involved in a number of biological and physiological process and some of these enzymes also have medical and biological applications.

Proteinases

Proteinases are a group of enzymes that catalyze the proteolysis reaction, which results in the digestion of proteins into smaller peptide fragments and amino acids. Although the hydrolysis reaction is known to be a spontaneous process, this step occurs at an extremely slow rate in the absence of the catalyst. In this case, the half-life of the typical peptide hydrolysis at neutral pH has been determined to be in the range of 10 to 1000 years which makes it clearly ineligible for any biological process [1]. Living organisms can overcome this problem by utilizing a biological catalyst, a peptidase enzyme. These enzymes can act as endopeptidases that cleave the proteins within the polypeptide chain or exopeptidases that remove residues from the end of the peptide chain. The catalytic mechanisms of peptidase enzymes enable us to classify these enzymes generally into 4 categories based on their modes of action [1,2]. There are metal, serine, cysteine, and aspartic peptidase families. To date, following a completion of genome projects from various organisms other peptidases with the modes of action differ from those of the 4 well-defined groups have been discovered, suggesting that the model might need further refinement [2].

The MEROPS database provides updated data on structures and properties of the peptidases from various organisms ranging from bacteria, viruses, fungi, plants and human as well as the information on 3 subgroups that are classified as endopeptidases, exopeptidases and aminopeptidases [3,4]. The database

also provides resources of available peptidase inhibitors to date. The MEROPS database classifies the peptidases and their inhibitors based on structures and activities. So far, at least 7 types of peptidases from various organisms have been identified, with the addition of the threonine, glutamic (formerly known as carboxyl peptidases) and unknown peptidases, to the 4 groups already mentioned above [4].

Aspartic proteinases

Aspartic proteinases are characterized by acidic pH optima and the utilization of 2 aspartic acid residues in their catalytic mechanism. Widely known members of the aspartic proteinases include renin, cathepsin D, pepsin, and HIV proteinase. All enzymes within this family are initially produced as precursors, known as pro-enzymes or zymogens [5]. Activation of the zymogen that gives rise to the fully active proteinase is achieved by limited proteolysis, sometimes of a single bond, and the removal of an activation segment that can occur by several mechanisms ranging from enzymatic or non-enzymatic cofactors that initiate the activation, to a simple pH change that results in conversion by an autocatalytic mechanism but all of these mechanisms lead to a molecular rearrangement resulting in the exposure of the full catalytic activity of the enzyme [5]. The fungal aspartic proteinases have been utilized successfully in the fermentation and food processing industries. These enzymes, along with mammalian aspartic proteinases, have long been studied and are well characterized. The enzymes of the aspartic proteinase family normally contain approximately 330 amino acid residues that adopt a crescent-shaped structure divided into 2 lobes, generating the deep active site cleft. Two catalytic aspartic residues are located at the center of the active site cleft, forming the catalytic dyad that interacts with a hydrogen bonding solvent molecule. This enzyme family has the characteristic that is identified by the possession of the amino acid sequence Asp-Thr/Ser-Gly within the primary structure. These residues play a crucial role in the catalytic function of aspartic proteinases [6-8]. A flexible β-hairpin loop that is normally referred to as the "flap", is located over the active site cleft and closes in order to trap substrates or inhibitors. In most monomeric aspartic proteinases, an interaction between a conserved tryptophan residue (Trp39, porcine pepsin numbering) located just after the first catalytic aspartic acid (Asp32), and the tyrosine located near the tip of the flap (Tyr75) controls the movement of the flap allowing access to the cleft. It has been indicated that the tip of the flap also plays a role in the capture prior to cleavage of substrates [8]. Another commonly found structural feature of aspartic proteinases known as the "polyproline loop" can also be observed in monomeric aspartic proteinases. The common features found in aspartic proteinases are demonstrated in **Figure 1** using aspartic proteinases from the fungi, *Saccharomyces cerevisiae*.

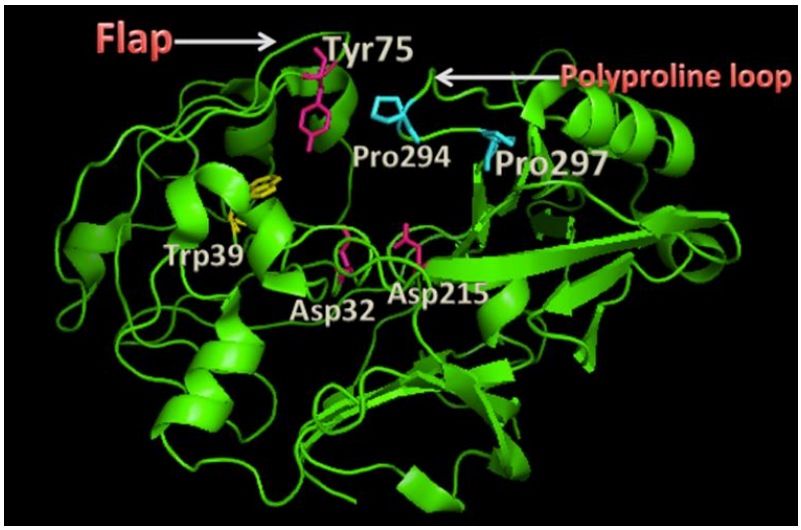

Figure 1 The 3-dimensional structure demonstrating the common structural features of aspartic proteinases.

The proteinase A from *S. cerevisiae* (PDB entry 1GOV) is shown here as a model. The β-strand-turn-β-strand that has been defined as the "flap region", which plays a role in trapping substrate is indicated. Tyrosine residue (Tyr75) at the 'flap' region functions in controlling flap movement and 2 catalytic aspartic acid residues are shown in pink. The conserved tryptophan residue that interacts with Tyr75 during flap movement is shown in yellow. The polyproline loop is indicated and the 2 key proline residues (Pro294 and Pro297) are also demonstrated. The image was generated by PyMol software.

Aspartic proteinase inhibitors

During the last decade, the importance of aspartic proteinases to various living organisms has become more apparent. Inhibitors of aspartic proteinases are not only important for human disease treatment but are also vital to plants and other economically important animals [9]. Two types of aspartic proteinase inhibitors have been identified i.e. small molecule inhibitors and naturally occurring peptides. Most aspartic proteinases are highly susceptible to inhibition by pepstatins, a series of non-proteinaceous natural products [10,11]. However, only a small number of naturally occurring polypeptide inhibitors of aspartic proteinases has so far been discovered and characterized [3,4,12] such as renin inhibitor [13,14], equistatin [15-19] and pepsin-inhibiting protein, PI-3 [20-23].

Another protein inhibitor of aspartic proteinases is *S. cerevisiae* IA_3 which consists of 68 amino acid residues and is classified as a cytosolic-heat-stable protein [10,24,25]. The IA_3 inhibitor was first isolated and characterized from *S. cerevisiae* in 1974 [26]. The amino acid sequence of *S. cerevisiae* IA_3 was then determined by using automated Edman degradation [27] and the gene coding for IA_3 was successfully cloned and characterized [28]. The N-terminal half (residues 2-34) of the molecule has been shown to possess the inhibitory activity [25,28]. Further evidence on the location of IA_3 inhibitory activity was later revealed by Phylip *et al*.[12]. A synthetic IA_3 peptide containing residues 2-34 was shown to be as potent as the wild-type IA_3. Moreover, the synthetic IA_3 residues 2-34, like the naturally occurring IA_3, exhibit no significant inhibitory effect toward any of other aspartic proteinases from a wide range of various species despite considerable sequence and structural similarities to its sole target substrate, the *S. cerevisiae* Proteinase A (ScPrA) [12]. The interaction between the IA3 and its target enzyme is demonstrated in **Figure 2**.

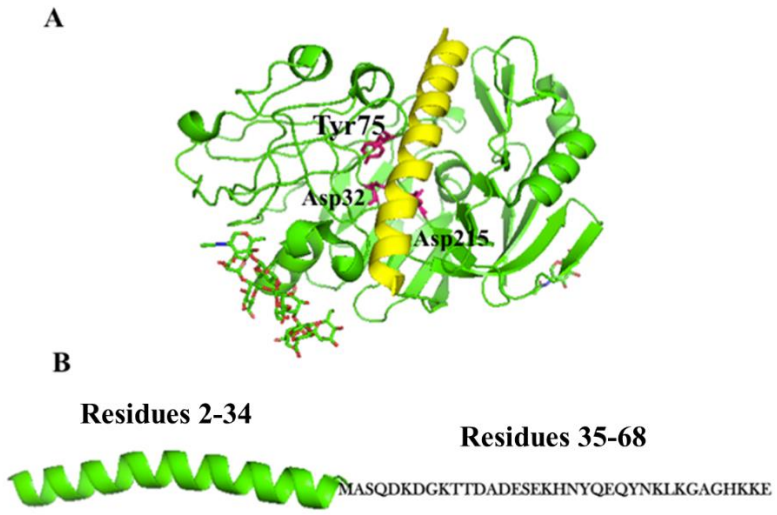

Figure 2 The interaction between *S. cerevisiae* IA_3 and its target enzyme, ScPrA. A) The crystal structure of ScPrA-IA_3 complex in which the IA_3 adopts a nearly perfect α-helix structure upon contact with ScPrA. Two catalytic aspartic acid residues and Tyr75 that control the movement of the flap region are also indicated. B) Demonstrating the secondary structure of IA_3 in the presence of ScPrA (not shown). The residues 2-34 form α-helix structure, whereas residues 35-68 remain unstructured.

In this study, bioinformatics analysis was performed in order to identify putative IA_3 orthologs from *Saccharomyces* and other fungi species. Database sequence similarity searching was repeatedly performed to obtain sequences resembling that of wild-type IA_3. The identified IA_3-like proteins were synthesized and overexpressed in *E. coli*. Characterization of putative IA_3 orthologs was also carried out.

Materials and methods

Bioinformatics analysis

Sequence homology searching was repeatedly performed on newly updated information on the fungal databases using basic local alignment search tools (BLAST) [29] to identify the novel IA_3-like proteins. Multiple alignment was carried out using CLUSTAL X and GENEDOC software [30].

PCR-based gene synthesis of the synthetic IA$_3$-like gene

Six oligonucleotide primers, including Lk-IA$_3$-1 to Lk-IA$_3$-6, were designed corresponding to a nucleotide sequence for a synthetic Lk-IA$_3$-like gene, generated by a reverse translation of Lk-IA$_3$-like protein sequence utilizing an *E. coli* codon usage database to generate the optimum coding sequence for *E. coli* expression. This sequence manipulation suite (SMS) contains various programs for generating, formatting and analyzing short DNA and protein sequences developed by Stothard [31]. The basic principles for primer design were followed [32]. Each primer is approximately 45 nucleotides in length with 20 overlapping nucleotides. The primers contain no inverted repeat or palindromic sequences to minimize the non-specific binding and secondary structure formation, respectively. All oligonucleotide primers used in this experiment were synthesized and provided by MWG-Biotech. The nucleotide and amino acid sequences for the Lk-IA$_3$-like protein are illustrated in **Figure 3**. The restriction recognition sites of available enzymes found within its sequence are also shown.

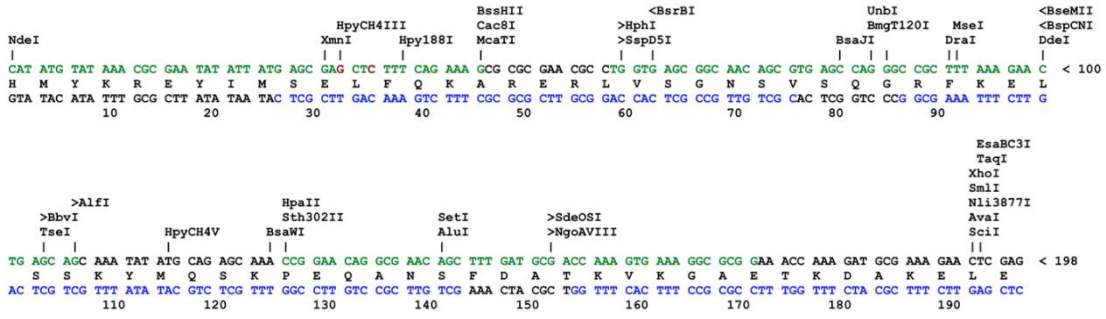

Figure 3 Primers designed for the synthesis of Lk-IA$_3$-like aspartic proteinase inhibitor.

Designed nucleotide and amino acid sequences for Lk-IA$_3$-like protein with restriction recognition sites. Six primers were designed and used to synthesize the Lk-IA$_3$-like gene. Codons 11 and 12 were changed from the optimum *E. coli* codons to generate a *Xho*I recognition site. These mutations (as indicated in red) did not change the encoded amino acid. The oligonucleotide primers used in this experiment are as follows:

Lk-IA$_3$-1 CATATGTATAAACGCGAATATATTATGAGCGAGCTCTTTCAGAAAG
Lk-IA$_3$-2 CGCTGTTGCCGCTCACCAGGCGTTCGCGCGCTTTCTGAAAGAGCTCGCTC
Lk-IA$_3$-3 TGGTGAGCGGCAACAGCGTGAGCCAGGGCCGCTTTAAAGAACTGAGCAG
Lk-IA$_3$-4 GCTGTTCGCCTGTTCCGGTTTGCTCTGCATATATTTGCTGCTCAGTTCTTTAAAGCGG
Lk-IA$_3$-5 CCGGAACAGGCGAACAGCTTTGATGCGACCAAAGTGAAAGGCGCGG
Lk-IA$_3$-6 CTCGAGTTCTTTCGCATCTTTGGTTTCCGCGCCTTTCACTTTG

Gene assembly and amplification

To facilitate the gene assembly, all 6 oligonucleotide primers (Lk-IA$_3$-1 to Lk-IA$_3$-6) were prepared as a primer mixture by combining each primer (25 μM each primer) and mixing, then an aliquot of 5 μL of the primer solution was used to set up the PCR-based gene assembly reaction. The 50 μL PCR reaction mixture contained 160 nM final concentration each primer, 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1 mg/mL nuclease-free BSA, 0.1 % Triton® X-100, 200 μM each dATP, dCTP, dGTP, dTTP and 2.5 units of *Pfu* DNA polymerase (Promega). Amplification was performed in a DNA Thermal Cycler (T3 Thermocycler, Biometra) with initial denaturation at 94 °C for 1 min and 25 cycles of 94 °C for 30 s, 52 °C for 30 s and 72 °C for 2 min, followed by a 10-min incubation at 72 °C as a final extension. The amplification products were analyzed by 1.2 % agarose gel electrophoresis.

An aliquot of 5 μL of the gene assembly product was used as a template for gene amplification. The reaction using *Pfu* DNA polymerase was carried out as described previously in the gene assembly process except that 1 μM of each of the outer primers (Lk-IA$_3$-1 and Lk-IA$_3$-6) were substituted for the 6 combined primers. The PCR program employed in this process consisted of an initial denaturation step at 94 °C for 60 s, followed by 25 cycles at 94 °C for 45 s, 65 °C for 45 s, 72 °C for 5 min and a final incubation step at 72 °C for 10 min. The amplification products were analyzed by 1.2 % agarose gel electrophoresis and the PCR product of expected size (approx.198 bp) was then purified by using QIAquick Gel Extraction Kit (QIAGEN) and the purified products was also analyzed on 1.2 % agarose gel electrophoresis.

Molecular cloning of Lk-IA$_3$-like into pGEM-T vector

The amplified Lk-IA$_3$ products were cloned into pGEM-T vector using the manufacturer's protocol provided with pGEM®-T vector system (Promega). The ligation mix was transformed into competent *E. coli* DH5α cells. The transformants, selected on LB agar supplemented with 100 µg/mL ampicillin, were subjected to plasmid extraction and screened for the presence of insert by PCR using Lk-IA$_3$-1 and Lk-IA$_3$-6 primers and confirmed by restriction digestion with the enzyme *Eco*RI. The pGEM®-T plasmids harboring the Lk-IA$_3$ gene were submitted to nucleotide sequencing to verify the sequence.

Molecular cloning of Lk-IA$_3$ into pET22b expression vector for protein production

After nucleotide sequence verification, the pGEM-T plasmid containing Lk-IA$_3$ gene was digested with restriction enzymes *Nde*I and *Xho*I. The digested products were analyzed on agarose gel electrophoresis and purified using QIAquick Gel Extraction Kit (QIAGEN). The purified Lk-IA$_3$ gene was ligated into pET22b vector that had previously been digested with *Nde*I and *Xho*I. The transformants were selected on LB agar plates containing 100 µg/mL ampicillin and screened by PCR using Lk-IA$_3$-1/Lk-IA$_3$-6 as primers and confirmed by restriction enzyme (*Nde*I and *Xho*I) digestion. The recombinant plasmids were also submitted to nucleotide sequencing for sequence verification.

Expression of Lk-IA$_3$ gene in BL21(DE3)pLysS

The pET22b plasmid harboring verified Lk-IA$_3$ gene was transformed into *E. coli* BL21(DE3)pLysS (Promega) using the conventional $CaCl_2$ methods and the production of Lk-IA$_3$ protein was then carried out. The host cells containing the interested gene was grown in LB broth supplemented with 100 µg/mL of ampicillin and 34 µg/mL of chloramphenicol at 37 °C until the $O.D._{600}$ = 0.7, when IPTG was added to the final concentration of 1 mM and incubated at 37 °C for 4 h. The culture was then harvested by centrifugation at 7,354×*g* for 15 min at 4 °C. The protein content of the cells was analyzed by 20 % SDS-PAGE.

Protein purification using IMAC

Purification of IA$_3$-like protein was performed using immobilized metal affinity chromatography (IMAC). The concentration of Lk-IA$_3$ protein was determined spectrophotometrically using the Bradford assay (Expedeon) [33].

Inhibitory assays of Lk-IA$_3$ protein against its potential targets

The inhibitory activity of the Lk-IA$_3$ protein was investigated enzymatically against its potential targets ScPrA, AfPrA and human cathepsin D (HuCatD) using the synthetic RS6 as a substrate. All assays were performed in semi-micro quartz cuvettes (1 cm pathlength) in a final volume of 800 µL in 100 mM sodium acetate buffer, pH 4.7. The kinetic parameters were determined using non-linear estimation on STATISTICA software (Statsoft) version 8.0.

Results and discussion

Identification of novel IA$_3$-like proteins using bioinformatics tools

In this study, 4 novel putative IA$_3$-like inhibitors have been found from bioinformatics analysis. These are putative IA$_3$ proteins derived from fungal species *Lachancea kluyveri* (Lk-IA$_3$-like), *Lachancea waltii* (Lw-IA$_3$-like), *Zygosaccharomyces rouxii* (Zr-IA$_3$-like) and *Kluyveromyces thermotolerans* (Kt-IA$_3$-like). The amino acid sequence alignment of all 4 IA$_3$-like proteins showed a similarity in length and some key sequence features compared to the wild-type IA$_3$ (**Figure 4**).

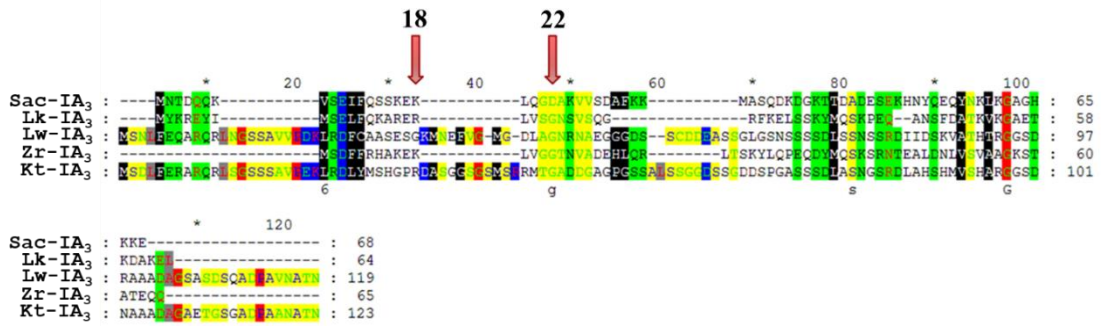

Figure 4 Amino acid alignment of *S. cerevisiae* IA$_3$ with 4 putative novel IA$_3$ orthologs.

Amino acid sequence alignment of these IA_3-like proteins against the wild-type *S. cerevisiae* IA_3 (Sac-IA_3) using CLUSTAL X software revealed high degree of similarity (47 - 70 %). Key residues (K18/D22 in *S.cerevisiae* IA_3) essential for its inhibitory activity are indicated by arrows. The identical residues are highlighted in red or yellow, conserved residues are highlighted in green and residues contain similar side chain group are highlighted in blue or black.

The proteins Kt-IA_3-like (123 amino acid residues) and Lw-IA_3-like (119 amino acid residues) are slightly larger than *S. cerevisiae* IA_3 and contain apparent inserts in the region expected to contribute in the inhibition process. The other orthologs, Lk-IA_3-like (64 amino acid residues) and Zr-IA_3-like (65 amino acid residues) are similar in length to *S. cerevisiae* IA_3 (**Figure 4**). Analysis of the primary structure of these IA_3-like inhibitors revealed that the

C-terminal halves of these proteins are highly variable but a small region aligning with the sequence NKLKGA in *S. cerevisiae* IA_3 shows some conservation of properties in most of the sequences (polar, basic, hydrophobic, basic, Gly, small side chain) but the Zr-IA_3, among these 4 orthologs, is least well conserved. However, the significance of this conserved sequence, if any, is yet to be clarified. This variation at the C-terminus is, however, unlikely to generate any effects on the inhibitory activity of the protein since it has been well demonstrated that the inhibitory region of the IA_3 is located within residues 2-34 of the *S. cerevisiae* IA_3. The crucial $K^{16}E^{17}K^{18}L^{19}$ motif that is present in wild-type IA_3 and all previously identified IA_3 orthologs, and has been shown to play an important role in the inhibition process is conserved in the Zr-IA_3-like protein and has also shares similar features in Lk-IA_3-like sequence. Interestingly, none of these 4 new sequences contain the well-conserved charged residue, D22 located on the hydrophobic face of the amphipathic helix that has previously been shown to play a crucial role on the IA_3 target selectivity by interacting with other key residues within the inhibitor and potentially crucial residues within ScPrA but possesses a small side chain residue G22 instead. The importance of charged residue D22 in the IA_3 sequence has previously been elucidated [34,35].

The IA_3-like aspartic proteinase inhibitor from *L. kluyveri* (Lk-IA_3-like) has revealed the highest degree of identity compared to wild-type IA_3 from *S. cerevisiae*. The Lk-IA_3-like protein contains 64 amino acid residues with important characteristics of the IA_3 family (**Figure 5**). Hence, this putative 64 residue Lk-IA_3-like was selected as an excellent candidate for further analysis in this study.

Figure 5 An alignment of the *S. cerevisiae* IA_3 and Lk-IA_3-like proteins.

The Lk-IA_3-like protein shares the highest degree of identity to the *S. cerevisiae* IA_3 compared to that of other new IA_3-like proteins identified. An alignment was performed using CLUSTAL X software.

PCR-based gene synthesis of the synthetic Lk-IA_3-like gene

The Lk -IA_3-like protein contains 64 amino acid residues with key amino acid residues conserved within the IA_3 family. As a result, attempts have been made to synthesize Lk-IA_3-like gene using a PCR-based gene synthesis method. The oligonucleotide primers were designed corresponding to the synthetic Lk-IA_3-like gene with overlapping regions to facilitate the gene assembly process. Six primers (Lk-IA_3-1 to Lk-IA_3-6) were employed in the gene assembly and the full-length Lk-IA_3-like gene was subjected to PCR-amplification using primers Lk-IA_3- 1 and Lk-IA_3-6. The amplification products were analyzed by 1.2 % agarose gel electrophoresis (**Figure 6**) and the PCR product of expected size (approx. 198 bp) was then purified and characterized.

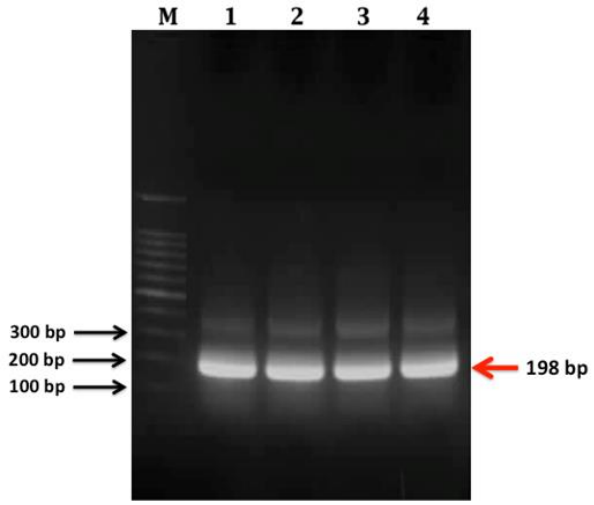

Figure 6 Gene amplification products of Lk-IA$_3$-like.

The amplification products were analyzed on 1.2 % agarose gel electrophoresis using TAE buffer. The PCR products of approximately 198 bp (indicated by red arrow) were detected compared to 100 bp ladder DNA size standard size marker. M = 100 bp ladder; 1-4 = amplification products of Lk-IA$_3$-like gene

Expression of Lk-IA$_3$ -like gene in BL21(DE3)pLysS

After verification, the amplified Lk-IA$_3$ products were cloned into pET22b expression vector and protein production was carried out in BL21(DE3)pLysS cells. The protein content of the cells was analyzed by 20 % SDS-PAGE and subsequently purified by IMAC (**Figure 7**) and the protein concentration of Lk-IA$_3$ determined from the BSA standard calibration curve was 0.55 mg/mL.

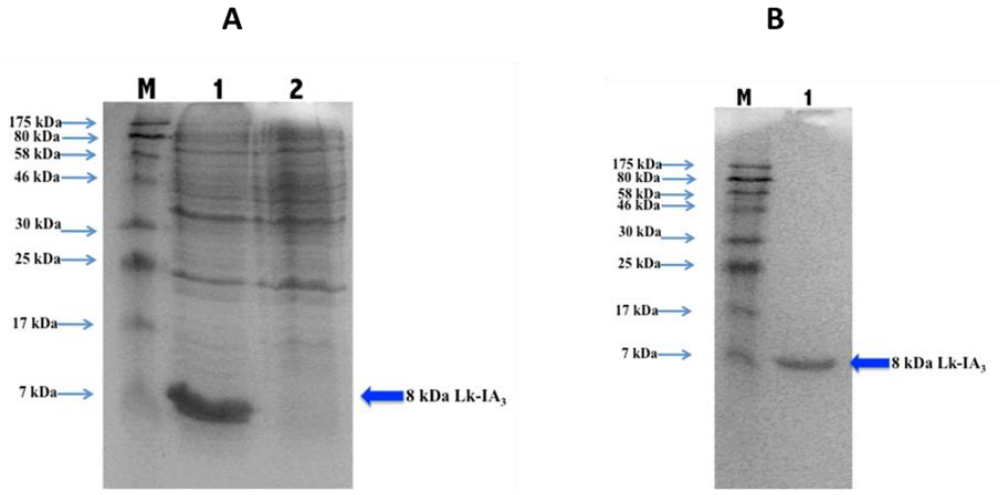

Figure 7 Production of Lk-IA$_3$ protein in *E. coli* BL21(DE3) pLysS. A) Expression of the Lk-IA$_3$ was carried out in BL21(DE3)pLysS upon IPTG induction. The cells were analyzed on 20% SDS-PAGE. M = Broad range (7-175 kDa) protein molecular mass marker (NEB); 1 = BL21(DE3)pLysS containing Lk-IA$_3$ 4 h after induction; 2 = BL21(DE3)pLysS without Lk-IA$_3$ 3 h after induction. B) The Lk-IA$_3$-like protein produced in *E. coli* was purified using IMAC and analyzed on 20 % SDS-PAGE. M = Broad range protein molecular mass marker; 1 = Purified Lk-IA$_3$-like protein.

Inhibitory assays of Lk-IA$_3$ protein against its potential targets

The inhibitory activity of the Lk-IA$_3$ protein was determined enzymatically against its potential targets. The Ki (inhibition constant) values of Lk-IA$_3$-like protein as well as that of the wild-type IA$_3$ against these target enzymes are shown in **Table 1**.

Table 1 Inhibitory assay of the Lk-IA_3-like protein against some of its potential targets.

Name	Amino acid sequence	Ki vs ScPrA (nM)	Ki vs AfPrA (nM)	Ki vs HuCatD (nM)
Sc-IA_3	MNTDQQKVSE IFQSSKEKLQ GDAKVVSDAFKKMASQDKD GKTTDADESEK HNYQEQYN KLKGAGHKKELE(6xHis)	0.06 ± 0.02	NI	NI
Lk-IA_3	MYKREYIMSELFQKARERLV SGNSVSQGRFKELSSKYMQS KPEQANSFDATKVKGAETKD AKELLE(6xHis)	190 ± 0.01	NI	NI

NI = Not inhibited

The sequence of the 64 residues Lk-IA_3 protein is shown and the first 7 amino acids resulting from the putative exon identified in this study are highlighted. The Lk-IA_3 recombinant proteins were expressed in *E. coli* with an LE-6xHis tag at the C-terminus (shown in blue). The inhibitory assays were performed against the potential target enzymes, i.e. ScPrA, AfPrA and HuCatD. The inhibitory activity of the wild-type IA_3 from *S. cerevisiae* (Sc-IA_3) against these enzymes is also shown. The Ki values were calculated using STATISTICA software (Statsoft) version 8.0.

Conclusions

A small protein aspartic proteinase inhibitor, IA_3, from *S. cerevisiae* has proved to be an extremely potent and highly specific inhibitor of its target enzyme ScPrA. Five IA_3 orthologs from *Saccharomyces* species have previously been reported using bioinformatics analysis [35]. In this study, 4 potentially novel IA_3 orthologs have been discovered.

The Lk-IA_3-like sequence initially discovered lacked the first 7 amino acids shown in **Figure 5**. A putative exon was found upstream, coding for these 7 amino acids and an intervening (intron) sequence (87 bp) separates the first 7 residues from the remaining 57 residues. The size of the putative intron found within the Lk-IA_3-like sequence corresponds to those found in various fungal species previously reported [36]. The finding of an intron in the Lk-IA_3-like has proved to be unique among all IA_3 sequences that have thus far been identified. It is possible that these new IA_3-like proteins lack the N terminus of wild-type IA_3 or that it exists as a separate exon in these species.

The preliminary work described in this study indicates that the Lk-IA_3-like protein may indeed be an ortholog of the wild-type IA_3 sequences previously identified since it shows inhibitory properties against ScPrA. While it shows no inhibition of AfPrA, or human cathepsin D, this is also true for the *S. cerevisiae* IA_3. Weak inhibition of ScPrA may be an indication that this new inhibitor is not well adapted to this target. The expected target of Lk-IA_3-like protein would be the vacuolar proteinase from *L. kluyveri*. To date, the sequence of this enzyme cannot be found in the *L. kluyveri* genome. There is clearly scope to produce the proteinase A protein from *L. kluyveri* in the future for assay with the Lk-IA_3-like protein. The Lk-IA_3-like sequence has a number of features that distinguish it from other IA_3 proteins. The substitution K18R appears to have little effect on the inhibitory potency of IA_3 against ScPrA but allows it to inhibit proteinase A from *Pichia pastoris* (PpPrA; Ki = 6 nM) and, weakly, to inhibit AfPrA (Ki = 620 nM) [34,35]. In addition, the Ala found at position 15 of the Lk-IA_3-like sequence is also found at this position in the *S. castellii* IA_3 protein [35] and, thus, also appears to be well tolerated in these inhibitors. It would also be interesting to introduce Lk-IA_3 sequences into the wild-type IA_3 sequence and vice versa to observe effects on inhibition.

Acknowledgements

This work is partially supported by the National Research Council of Thailand (NRCT).

References

[1] JM Berg, JL Tymoczko and L Stryer. *Biochemistry*. W.H. Freeman, Basingstoke, 2012.

[2] AMD Castro, AFD Santos, V Kachrimanidou, AA Koutinas, DMG Freire. *Solid-state fermentation for the production of proteases and amylases and their application in nutrient medium production*. *In*: A Pandey, C Larroche and CR Soccol (Eds.). Current developments in biotechnology and bioengineering, Elsevier, Cambridge, 2018, p. 185-210.

[3] ND Rawlings. Peptidase inhibitors in the MEROPS database. *Biochimie* 2010; **92**, 1463-83.

[4] ND Rawlings, AJ Barrett, PD Thomas, X Huang, A Bateman and RD Finn. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 2018; **46**, D624-D632.

[5] AR Khan amd MN James. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* 1998; **7**, 815-36.

[6] BM Dunn. *Aspartic proteases*. *In*: WJ Lennarz and MD Lane (Eds.). Encyclopedia of biological chemistry. 2nd ed. Academic Press, Waltham, 2013, p. 137-40.

[7] ND Rawlings and AJ Barrett. Introduction: Aspartic and glutamic peptidases and their clans. *In*: ND Rawlings and G Salvesen (Eds.). *Handbook of proteolytic enzymes.* 3rd ed. Academic Press, Cambridge, 2013, p. 3-19.

[8] A Wlodawer, A Gustchina and MNG James. Catalytic pathways of aspartic peptidases. *In*: ND Rawlings and G Salvesen (Eds.). *Handbook of proteolytic enzymes* 3rd ed. Academic Press, Cambridge, 2013, p. 3-19.

[9] JB Cooper. Aspartic proteinases in disease: A structural perspective. *Curr. Drug Targets* 2002; **3**, 155-73.

[10] J Kay. Aspartic proteinases and their inhibitors. *Biochem. Soc. Trans.* 1985; **13**, 1027-9.

[11] MJ Valler, J Kay, T Aoyagi and BM Dunn. The interaction of aspartic proteinases with naturally-occurring inhibitors from actinomycetes and *Ascaris lumbricoides*. *J. Enzym. Inhib.* 1985; **1**, 77-82.

[12] LH Phylip, WE Lees, BG Brownsey, D Bur, BM Dunn, JR Winther, A Gustchina, M Li, T Copeland, A Wlodawer and J Kay. The potency and specificity of the interaction between the IA3 inhibitor and its target aspartic proteinase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 2001; **276**, 2023-30.

[13] S Takahashi, K Takahashi, T Kaneko, H Ogasawara, S Shindo, K Saito and Y Kawamura. Identification of functionally important cysteine residues of the human renin-binding protein as the enzyme N-acetyl-D-glucosamine 2-epimerase. *J. Biochem*. 2001; **129**, 529-35.

[14] S Takahashi, H Ogasawara, K Takahashi, K Hori, K Saito and K Mori. Identification of a domain conferring nucleotide binding to the N-acetyl-d-glucosamine 2-epimerase (Renin binding protein). *J. Biochem*. 2002; **131**, 605-10.

[15] G Dubin. Proteinaceous cysteine protease inhibitors. *Cell Mol. Life Sci*. 2005; **62**, 653-69.

[16] K Galesa, B Strukelj, S Bavec, V Turk and B Lenarcic. Cloning and expression of functional equistatin. *Biol. Chem.* 2000; **381**, 85-8.

[17] B Strukelj, B Lenarcic, K Gruden, J Pungercar, B Rogelj, V Turk, D Bosch and MA Jongsma. Equistatin, a protease inhibitor from the sea anemone *Actinia equina*, is composed of three structural and functional domains. *Biochem. Biophys. Res. Commun*. 2000; **269**, 732-6.

[18] B Lenarcic and V Turk. Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J. Biol. Chem.* 1999; **274**, 563-6.

[19] B Lenarcic, A Ritonja, B Strukelj, B Turk and V Turk. Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. *J. Biol. Chem.* 1997; **272**, 13899-903.

[20] T Kageyama. Molecular cloning, expression and characterization of an Ascaris inhibitor for pepsin and cathepsin E. *Eur. J. Biochem.* 1998; **253**, 804-9.

[21] T Kageyama. Pepsinogens, progastricsins, and prochymosins: Structure, function, evolution, and development. *Cell Mol. Life Sci.* 2002; **59**, 288-306.

[22] K Girdwood and C Berry. The disulphide bond arrangement in the major pepsin inhibitor PI-3 of *Ascaris suum*. *FEBS Lett.* 2000; **474**, 253-4.

[23] K Ng, JFW Petersen, M Cherney, C Garen, J Zalatoris, C Rao-Naik, B Dunn, M Martzen, R Peanasky and M James. Structural basis for the inhibition of porcine pepsin by *Ascaris* pepsin inhibitor-3. *Nat. Struct. Biol.* 2000; **7**, 653-7.

[24] T Dreyer, MJ Valler, J Kay, P Charlton and BM Dunn. The selectivity of action of the aspartic-proteinase inhibitor IA3 from yeast (*Saccharomyces cerevisiae*). *Biochem. J.* 1985; **231**, 777-9.

[25] M Li, LH Phylip, WE Lees, JR Winther, BM Dunn, A Wlodawer, J Kay and A Gustchina. The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. *Nat. Struct. Biol*. 2000; **7**, 113-7.
[26] T Saheki, Y Matsuda and H Holzer. Urification and characterization of macromolecular inhibitors of proteinase A from yeast. *Eur. J. Biochem*. 1974; **47**, 325-32.
[27] K Biedermann, U Montali, B Martin, I Svendsen and M Ottesen. The amino acid sequence of proteinase a inhibitor 3 from baker's yeast. *Carlsberg Res. Commun.* 1980; **45**, 225.
[28] P Schu and DH Wolf. The proteinase yscA-inhibitor, IA3, gene. Studies of cytoplasmic proteinase inhibitor deficiency on yeast physiology. *FEBS Lett.* 1991; **283**, 78-84.
[29] SF Altschul, W Gish, W Miller, EW Myers and DJ Lipman. Basic local alignment search tool. *J. Mol. Biol*. 1990; **215**, 403-10.
[30] JD Thompson, TJ Gibson, F Plewniak, F Jeanmougin and DG Higgins. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997; **25**, 4876-82.
[31] P Stothard. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 2000; **28**, 1102.
[32] C Withers-Martinez, EP Carpenter, F Hackett, B Ely, M Sajid, M Grainger and MJ Blackman. PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome. *Protein Eng*. 1999; **12**, 1113-20.
[33] MM Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976; **72**, 248-54.
[34] TJ Winterburn, DM Wyatt, LH Phylip, C Berry, D Bur and J Kay. Adaptation of the behaviour of an aspartic proteinase inhibitor by relocation of a lysine residue by one helical turn. *Biol. Chem.* 2006; **387**, 1139-42.
[35] TJ Winterburn, DM Wyatt, LH Phylip, D Bur, RJ Harrison, C Berry and J Kay. Key features determining the specificity of aspartic proteinase inhibition by the helix-forming IA3 polypeptide. *J. Biol. Chem.* 2007; **282**, 6508-16.
[36] DM Kupfer, SD Drabenstot, KL Buchanan, H Lai, H Zhu, DW Dyer, BA Roe and JW Murphy. Introns and splicing elements of five diverse fungi. *Eukaryot Cell* 2004; **3**, 1088-100.