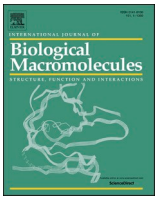
International Journal of Biological Macromolecules 249 (2023) 126004

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac



Characterization of the common bed bug's eggshell and egg glue proteins

Yeo Won Sim [a,1], Susie Cho [b,1], Si Hyeock Lee [b,c], Ju Hyeon Kim [d,*], Julian Lee [a,*], SangYoun Park [a,*]

[a] School of Systems Biomedical Science and Integrative Institute of Basic Sciences, Soongsil University, Seoul 06978, Republic of Korea
[b] Entomology Division, Department of Agricultural Biotechnology, College of Agriculture and Life Science, Seoul National University, Seoul 08826, Republic of Korea
[c] Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea
[d] Department of Tropical Medicine and Parasitology, College of Medicine, Seoul National University, Seoul 03080, Republic of Korea

ARTICLE INFO

Keywords:
Bed bug
Cimex lectularius
Eggshell protein
Egg glue protein
Repeat protein

ABSTRACT

An insect egg is one of the most vulnerable stages of insect life, and the evolutionary success of a species depends on the eggshell protecting the embryo and the egg glue securing the attachment. The common bed bug (*Cimex lectularius*), notorious for its painful and itchy bites, infests human dwellings to feed on blood. They are easier to find these days as they adapt to develop resistance against commonly used insecticides. In this study, we identify and characterize the eggshell protein and the probable egg glue protein (*i.e.* keratin associated protein 5–10 like protein) of the bed bug by using mass spectrometry and bioinformatics analysis. Furthermore, by using transcription profiling and *in vivo* RNA interference, we show evidences that the keratin associated protein 5–10 like protein functions as the glue protein. Finally, structural characterizations on the two proteins are performed using recombinant proteins. Amino acid sequences of various insect eggshell and egg glue proteins support their independent evolution among different insect groups. Hence, inhibiting the function of these proteins related to the earliest stage of life can achieve species-specific population control. In this respect, our results would be a starting point in developing new ways to control bed bug population.

1. Introduction

A female insect lays its eggs covered with a soft eggshell. The egg is both the start and the most vulnerable stage during an insect's life cycle, and the eggshell acts as the first line of defense in protecting the developing embryo. Generally, an insect's egg (oocyte) from the embryo is layered by the vitelline membrane, the endochorion, and the outermost exochorion. Usually, the term eggshell includes all these regions outside the oocyte, and its main function is to protect the developing embryo [1–3]. This eggshell must allow gas exchange (respiration) and the entry of sperm (fertilization) as well as to provide protection against desiccation and micro-organisms (especially *via* secreted anti-microbial proteins) [1–5]. The insect eggshells are usually made of proteinaceous materials, and several eggshell proteins have been identified from various insects including the fruit fly (*Drosophila melanogaster* (M.)) [4,5], the yellow fever mosquito (*Aedes aegypti* (L.)) [6], the kissing bug (*Rhodnius prolixus* (S.)) [7], and the German cockroach (*Blattella germanica* (L.)) [8].

The female insects in most cases also secret adhesive materials along with the egg for its secure attachment to a selected preferred environment. These so-called egg glues are also made of proteins and are usually secreted from the female accessory gland during oviposition. So far, the egg glues from several Australian insect species [9], the human head and body louse (termed the louse nit sheath protein) [10,11], and the silk moth (*Bombyx mandarina* (M.)) [12] have been characterized. The results of these various insect eggshell and egg glue proteins indicate that their identities (*i.e.* protein sequences) are seemingly diverse and unique suggesting that independent evolution likely resulted in the specialized sets of eggshell and egg glue proteins in each insect group. From a human perspective, inhibiting the function of these proteins in a harmful insect would allow species-specific population control.

The common bed bug (*Cimex lectularius* (L.)) is an insect that infests houses and feeds on human blood. While they are not known to directly transmit diseases, they come out at night from their hidings in small cracks and crevices (of the bed mattresses, wallpapers, and furniture) and their bites are of extreme annoyance. Moreover, bed bugs are

* Corresponding authors.
E-mail addresses: biomyst5@snu.ac.kr (J.H. Kim), jul@ssu.ac.kr (J. Lee), psy@ssu.ac.kr (S. Park).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.ijbiomac.2023.126004
Received 6 June 2023; Received in revised form 24 July 2023; Accepted 25 July 2023
Available online 28 July 2023

considered as a public health issue in many countries because they can easily spread and are becoming increasingly resistant to various pesticides (DDT, organophosphates, and pyrethroids) [13–15]. In this study, we identify and characterize the eggshell protein and the egg glue protein of the common bed bug. In the future, this information may help find new ways for species-specific control of bed bug population.

2. Methods and materials

2.1. Bed bug rearing and eggshell collection

The Florida strain of bed bug (*C. lectularius*) was maintained on *in vitro* membrane feeding systems [16] under environmental conditions of 30 °C, 70 % relative humidity, and 16/8 h light/dark in a rearing chamber (reviewed and approved by the Institutional Review Board of Seoul National University, IRB No. E2211/001-003). The Korean Red Cross approved the supply of blood that fell below the human blood transfusion threshold for our research use. The hatched eggs (eggshells) attached on filter paper (Whatman, Maidstone, UK) were harvested under a stereomicroscope (SZ61, Olympus, Tokyo, Japan), washed five times using ddH_2O to remove contaminants such as bed bug feces and used for the following experiments.

2.2. Dissolving the empty bed bug eggs

Either a single egg or ~50 eggs were treated with 8 M urea, 100 mM Tris pH 7.5 and 100 mM dithiothreitol (DTT) solution. After overnight treatment at 25 °C, the supernatant and the residual which remained undissolved were separated by centrifugation. The undissolved residual was further washed with ddH_2O three times to remove any leftover urea.

2.3. Mass spectrometry analysis of the dissolved eggshell supernatant

The supernatant of the dissolved eggs was run on SDS-PAGE to identify its protein content. First, 10 μL of the supernatant was mixed with 20 μL of commercial SDS-PAGE sample buffer (Invitrogen, USA) containing 100 mM DTT, and boiled overnight. Next, 5 μL of 1 M DTT was added to 5 μL of the boiled mixture, incubated for 30 min and gel-electrophoresed on Nu-PAGE 4–12 % Bis-Tris gel (Invitrogen, Carlsbad, CA, USA) using a MOPS running buffer (100 V, 60 min). The gel was stained using Coomassie Blue, and the region of the major protein band was cut into a plug and their tryptic digestion fragments analyzed using LC-MS/MS at a mass spectrometry facility of the National Instrumentation Center for Environmental Management (NICEM, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea). The peptide fragments of the mass spectrum were searched against the whole genomic proteins of *C. lectularius* in the public database [17].

2.4. Amino acid composition analysis of the residual

The residuals remaining after the urea-treatment of the bed bug eggs were analyzed for the amino acid composition at the Korea Basic Science Institute (KBSI, Seoul, Korea). The residual left after dissolving ~50 empty bed bug eggs with urea-based solution was completely air dried, and then hydrolyzed using 1 mL of 6 N HCl at 110 °C for 24 h. The hydrolyzed amino acids were modified with phenylisothiocyanate (PITC) using 20 μL of PITC-solution [methanol:H_2O:trimethylamine (TEA):PITC = 7:1:1:1] at RT for 30 min. After removing the solvents with vacuum rotary evaporator, the remaining PITC-modified amino acids were re-dissolved in 200 μL of 140 mM sodium acetate (pH 6.1), 0.15 % TEA, 0.03 % EDTA, and 6 % CH_3CN in water, followed by filtration using a 0.45 μm filter (Millipore, Burlington, MA, USA). After further centrifugation, the supernatant was removed and loaded onto the auto-sampler of HPLC (Hewlett Packard 1100) (Agilent, Santa Clara, CA, USA) for injection into a Nova-Pak C18 (4 μm, 3.9 × 300 mm) column (Waters, Milford, MA, USA). The elution was performed using a linear gradient of 60 % CH_3CN and 0.015 % EDTA for 30 min (1.0 mL/min flow rate). The PITC-derivatized amino acid standards were also run using the same condition. The individual areas under the peaks of the sample chromatogram were compared to that of the amino acid standards to identify and determine the amino acid type and their relative mole percent (Table 1).

2.5. Bioinformatics search for the candidate egg glue protein of the undissolved residual

If the residual remaining after the urea solution treatment of the bed bug egg were to dissolve, the use of mass spectrometry would be the method of choice for the identification of the composing proteins as done in the case of the dissolved eggshell. Instead, the amino acid composition result of the residual was analyzed by using a bioinformatics-based search. First, the complete protein list obtained from the whole genome sequencing of *C. lectularius* was downloaded from NCBI [17] in order to search for the protein whose amino acid composition corresponds to the amino acid analysis result of the residual. Among the 24,194 bed bug whole-genome encoding proteins, 2353 proteins containing a signal peptide which allows the protein to be secreted were selected using SignalP-5.0 [18]. Furthermore, proteins predicted by TMHMM [19] to contain an internal transmembrane spanning α-helix (or helices) within the 2353 proteins were excluded from the final protein set for they are membrane proteins. As a result, 1428 non-redundant protein sequences were included in the final protein set.

Next, the amino acid compositions from these 1428 sequences were extracted and compared with that of the residual. Since Cys and Trp cannot be detected in the amino acid composition analysis due to the harsh HCl hydrolysis condition which degrades these residues, we excluded them when computing the amino acid compositions. Also, since Asp and Asn, as well as Glu and Gln, are not experimentally distinguishable, each pair was combined into one class. Therefore, comparisons of the fractions of 16 amino acids composing the individual 1428 non-redundant protein sequences against those of the experimentally determined residual were performed, as detailed below.

2.5.1. Fit using one protein sequence

In order to compare the experimental amino acid composition of the residual with that obtained from each of the 1428 non-redundant sequences, the number $n_{\alpha i}$ of the i-th amino acid ($i = 1,\ldots16$) in the sequence α ($\alpha = 1,\ldots,1428$) is first counted and divided by the sequence

Table 1
Experimentally determined amino acid composition of the residual after the urea treatment of the bed bug eggs.

Amino acid[a]	% moles
Asx (Asp *or* Asn)	7.4
Glx (Glu *or* Gln)	6.0
Ser	3.2
Gly	23.3
His	3.5
Arg	3.9
Thr	10.9
Ala	7.7
Pro	6.6
Tyr	3.5
Val	13.5
Met	0.0
Ile	0.0
Leu	4.9
Phe	3.5
Lys	2.0

[a] There were no information on the contents of cysteine and tryptophan residues due to the harsh HCl hydrolysis condition which degrades those residues.

length N_α, to obtain the fraction $f_{\alpha i} \equiv \frac{n_{\alpha i}}{N_\alpha}$. The amino acid composition of the sequence α is defined by the 16-component vector formed by $f_{\alpha i}$, and the protein sequence whose composition is closest to the experimental composition of the residual (f_i^{exp}, similarly defined by the 16-component vector formed by the amino acid fractions) is selected. The closeness of the compositions was measured in terms of the usual Euclidean distance in 16 dimensions,

$$R_\alpha = \sqrt{\sum_{i=1}^{16} \left(f_i^{exp} - f_{\alpha i}\right)^2} \quad (\alpha = 1, \cdots 1428).$$

The protein sequence with the lowest value of R was selected among 1428 sequences as the one having the composition most similar to that of the residual.

2.5.2. Fit using two protein sequences (two-protein mix model)

The amino composition in the solution was also fitted using a complete pair set of two sequences within the 1428 selected proteins. For each pair of 1428 proteins labelled by the indices α and β, we combined their compositions with weights r_α and r_β with $r_\alpha + r_\beta = 1$, and minimized the distance from the experimental composition of the egg residual,

$$R_{\alpha\beta} = \sqrt{\sum_{i=1}^{16} \left(f_i^{exp} - r_\alpha f_{\alpha i} - (1 - r_\alpha) f_{\beta i}\right)^2}$$
$$= \sqrt{\sum_{i=1}^{16} \left(f_i^{exp} - f_{\beta i} - r_\alpha \left(f_{\alpha i} - f_{\beta i}\right)\right)^2},$$

with respect to r_α. The minimization is straightforward since it is just a quadratic function. We then selected the pair α and β with minimal value of $R_{\alpha\beta}$ already minimized with respect to r_α. We inferred that the residual was mainly composed of the pair of protein sequences α and β selected in this manner, their relative compositions being r_α and r_β. We also performed a similar analysis with one protein of the pair fixed as the eggshell protein (ESP).

2.6. Temporal and spatial transcription profiles of ESP and KAP5-10L

Whole female reproductive organs were dissected from 12-day old 25 virgin females or gravid females in ice-cold, nuclease-free PBS (pH 7.4). Previously, the reproductive organs of female bed bugs were found to be mainly divided into three parts: ovary, oviduct and spermatheca [20]. For this study, the tissues were dissected into two parts, ovaries and oviduct with spermatheca (SFig. 1), and stored in RNA later (Invitrogen, Carlsbad, CA, USA) until all bed bugs were dissected. Total RNA was extracted with TRIzol reagent (MRC, Cincinnati, OH, USA) according to the manufacturer's protocol, and the mRNA was further purified by Oligo (dT) magnetic beads (Qiagen, Germany). The subsequent cDNA library construction and adapter ligation were conducted with a TruSeq™ Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA).

To quantify the temporal transcription levels of *ESP* and *KAP5–10L*, total RNA was extracted from the dissected ovaries and oviducts with spermatheca at different ages of bed bugs using TRIzol reagent (MRC) as described above. For the spatial expression pattern of *KAP5–10L*, dissected ovaries, oviducts, brains, malpighian tubules, midguts and cuticles were used to extract total RNA. First-strand cDNA was synthesized from DNaseI (Takara, Shiga, Japan)-treated total RNA using SuperScript IV reverse transcriptase (Invitrogen) according to the manufacturer's protocol. The reaction mixtures contained 1× TB Green® Premix Ex Taq (Takara) and 0.5 μM primers for *ESP* or *KAP5–10L* (sequences are shown in STable 1). *Elongation Factor 1-alpha* (*EF1a*) and *60S Ribosomal protein L18* (*60SRpL18*) were used as reference genes. Quantitative real-time PCR (qRT-PCR) was conducted in the Roche LightCycler 96 system (Roche, Basel, Swiss) using the following cycle conditions: 95 °C for 10 min, 40 cycles of 95 °C for 30 s, 57 °C for 30 s, 72 °C for 30 s and serial increase of 0.2 °C per 1 s from 45 to 95 °C for melting curve analysis. The relative expression level of a gene was calculated based on the original concept of $2^{-\Delta\Delta Ct}$ [21]. All qRT-PCR was repeated 3 times with total RNA extracted independently. One-way ANOVA with Tukey's *post hoc* test was performed to determine significant differences by interpreting *P*-values using the GraphPad Prism 6.01 software (GraphPad Software, San Diego, CA, USA).

2.7. dsRNA synthesis

The transcription template of *KAP5–10L* from the reared bed bug cDNA was amplified by PCR using gene-specific primers (STable 1) which targeted the 428-bp region that nearly spanned the whole 450-bp length of the reported *KAP5–10L* sequence in NCBI (SFig. 2). The amplified sequence was cloned into pGEM-T easy vector (Promega, Madison, WI, USA), and the inserted gene sequence confirmed (Macrogen, Daejeon, Korea). The *KAP5–10L* fragment-containing plasmid was employed as a template for the *in vitro* production of double-stranded RNA (dsRNA) using a set of gene-specific primers with a T7 promoter sequence attached to their 5′ ends. The dsRNA sequences were transcribed with the Megascript T7 transcription kit (Invitrogen) according to the manufacturer's protocol. Nuclease-free water was used for dsRNA elution, and the synthesis was verified by 1.5 % agarose gel electrophoresis and quantified on a NanoDrop ND-1000 instrument (Thermo Fisher Scientific, Waltham, MA, USA). As a negative control, dsRNA for *pQE30* was generated from the pQE-30 UA vector (Qiagen Korea, Osong, Korea) using the exact same method.

2.8. Administration of the dsRNA by direct injection into the female bed bug

Bed bugs are gonochorists with obligate traumatic insemination, in which the male injects sperm into the female's hemocoel by puncturing her abdomen cuticle during copulation [22]. The male injects his paramere to a unique female anatomical organ called a spermalege to pierce the female abdomen. The evolutionary emergence of spermalege is reported to be the outcome of protection against the inevitable infection during the mating [23,24]. Since spermalege injections of dsRNA result in a lower death rate than injections in other abdominal areas, the synthesized dsRNAs were injected into the female bed bug through the spermalege.

The *KAP5–10L* RNA interference was most efficient when the dsRNA was administered two times (24 and 48 h) after mating. To produce virgin female adults, bed bugs at the 5th instar were blood-fed and separated individually in a 96-well plate. After molting, the female adults were blood-fed and freely mated with male bed bugs. Once copulation was confirmed by the scar on the spermalege, the mated female was placed in a petri dish for 24 h. As mentioned above, the dsRNA injection was repeated two times (24 and 48 h) after the mating. *KAP5–10L* dsRNA (2 μg, 200 nL per bed bug) was injected individually using a nanoinjector (NanoLiter 2000, World Precision Instruments, FL, USA) and a needle prepared with glass capillary tubes. Same dose of *pQE30* dsRNA was injected into a female as a negative control. A total of 15 bed bugs were injected per treatment, and three replicates were performed. After injection, five bed bugs were placed into a 5 mL glass vial with a piece of black smooth paper (1.3 cm × 3.5 cm, 52 mg/cm^2). The paper was changed every 24 h to assess the daily oviposition and hatching rates.

Proper gene knockdown was verified by qRT-PCR. After 48, 72, and 96 h of the first dsRNA injection, three bed bugs per replicate were collected, and total RNA was extracted from the dissected abdomens including the reproductive organs. cDNA synthesis and qRT-PCR were conducted as described above. Data on gene knockdown, oviposition, and egg hatchability were analyzed using multiple *t*-tests with the Holm-Sidak method ($P < 0.05$). All analyses were done with GraphPad Prism 6.01.

2.9. Recombinant ESP and KAP5–10L expression and purification using Escherichia coli

The DNA encoding ESP (19–365) and KAP5–10L (23–149) both without the N-terminal signal sequence were synthesized with codon-optimization for *E. coli* expression and cloned into the pET28a vector (Merck, Rahway, NJ, USA). Cloning was performed using the *Nde*I and *BamH*I restriction enzyme sites to allow the expression of ESP with an N-terminal His_6-tag for Ni-NTA affinity purification. Of note, the codons of two terminally located cysteine residues (Cys18 and Cys366) in ESP were not included in the synthesized gene to prevent complication from possible Cys-Cys disulfide cross-linking. Also, all the cysteine residues of KAP5–10L (30 % of native KAP5–10L are cysteine residues.) were mutated to encode serines for the same reason. The generated plasmids were sequence verified (Enzynomics) surrounding the insert regions. Plasmids were used to transform *E. coli* BL21 (DE3) (Merck) competent cells using heat shock at 42 °C (45 s). The transformed cells were grown at 37 °C in 1 L of Luria-Bertani (LB) medium to an OD_{600} of ~0.8 in the presence of 25 μg/mL kanamycin. Expression of the recombinant proteins were induced by addition of 0.5 mM isopropyl-D-thiogalactopyranoside (IPTG) at 22 °C, and cells were allowed to grow for extra 16 h. Cells were harvested using centrifugation at 4500 ×*g* for 10 min (4 °C).

The expressed ESP and KAP5–10L with the N-terminal His_6-tag were both soluble. For protein purification, the bacterial cell pellets were re-suspended in 50 mL ice-cold lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, and 5 mM imidazole) and lyzed on ice by sonication. The homogenates were centrifuged at 70,000 ×*g* for 30 min (4 °C), and supernatants poured over 5 mL Ni-nitrilotriacetic acid agarose (Ni-NTA) (Qiagen, Hilden, Germany) gravity column. The columns were washed with five column volumes of wash buffer (20 mM Tris pH 7.5, 20 mM imidazole, and 500 mM NaCl), and the proteins were eluted with the elution buffer (20 mM Tris pH 7.5, 200 mM imidazole, and 500 mM NaCl). For the ESP, the elution fractions containing the protein were checked using Bradford assay (Bio-Rad, Hercules, CA, USA), combined, and added with 50 μL of 0.25 U/μL bovine thrombin (Invitrogen, Waltham, MA, USA). After proteolysis of the His_6-tag for 16 h incubation at 4 °C, the protein samples were further purified using HiLoad® 26/60 Superdex® 200 size-exclusion column (SEC) pre-equilibrated with SEC buffer (50 mM Tris pH 7.5, 150 mM NaCl, and 2 mM DTT). The ESP: thrombin proteolysis mixture was loaded into the column connected to an ÄKTA FPLC system (GE Healthcare, Chicago, IL, USA). The elution profile showed one major peak, and the fractions were concentrated by Amicon® ultracentrifugation filtration system (Merck). The ESP protein concentration was estimated by using the extinction coefficient [0.6 $(mg/mL)^{-1}$ cm^{-1}] calculated from the ESP sequence in ProtParam tool [25]. For the KAP5–10L, the Ni-column eluted fractions containing the protein were pooled and concentrated with the His_6-tag left intact. Further purification of KAP5–10L using size-exclusion was not performed because the protein sequence lacks tyrosine and tryptophan residues necessary for UV-mediated detection. However, concentration for the circular dichroism (CD) experiment was determined using Bradford assay with trypsin as the standard.

2.10. CD analysis

The secondary structure contents of the purified ESP and KAP5–10L were analyzed using CD. Ellipticity was scanned over 190–240 nm wavelength on a JASCO spectropolarimeter (Model J-810, Tokyo, Japan) where a concentration of 0.1 mg/mL protein in 0.1 cm path-length cuvette was used. The ellipticity data were fitted using secondary structure estimation program BESTSEL [26] to approximate the secondary structure content.

3. Results

3.1. Treating the empty bed bug egg with a urea-based solution

When a single hatched bed bug egg (Fig. 1A) was treated with 8 M urea, 100 mM Tris pH 7.5 and 100 mM dithiothreitol (DTT) solution and observed under the microscope, it began to shrink and dissolve immediately. However, some undissolved residuals still remained even after the overnight treatment (Fig. 1B). We reasoned that the supernatant contains the dissolved eggshell proteins and that the undissolved residual contains the egg glue proteins, and pursued to identify the exact proteinaceous contents. Approximately 50 hatched bed bug eggs were treated overnight using the same solution, and the supernatant and the residual separated for further analysis (Fig. 1C).

3.2. Proteomic analysis of the supernatant for identification of the eggshell protein

In order to analyze the contents consisting the dissolved bedbug egg (presumed to contain the eggshell protein), the supernatant was gel-electrophoresed (Fig. 1C). The stained gel indicated that the urea-dissolved part of the egg was made up of multiple proteins although one major band between the size markers of 41 kDa and 53 kDa was significantly more abundant than others. When this major band was excised and analyzed using LC-MS/MS, the uncharacterized protein LOC106669538 (NCBI ref. seq. XP_014254554.1) of the bed bug whole genome was correctly matched to the LC-MS/MS peptide fragment mass data.

The uncharacterized LOC106669538 protein is a 366-residue protein with calculated molecular weight of 39.2 kDa (Fig. 2). As expected for an eggshell protein which is secreted out of the cell, LOC106669538 contains a predicted N-terminal signal peptide sequence (residues 1–17). The remaining parts of LOC106669538 can be characterized by two domains of repeating sequences, in which the first repeat domain contains ten 7-residue (mostly FNGVATG) sequences and the second repeat domain holds eight 16-residue repeat sequences. As with a protein having repeat sequences, the LOC106669538 as a whole contains unusually high contents of Val (18 %), Thr (11 %), Ala (9 %), Gly (8 %), and Pro (7 %). It is also worthy to note that 2 % of the protein sequence is Cys. A similar study by Gordon et al. [27,28] have identified LOC106669538 as the eggshell protein of the bed bug in the past (See Section 4. Discussion). Since those authors refer the protein as the eggshell protein (ESP), we will also hereafter use this name for the LOC106669538.

3.3. Amino acid composition analysis of the residual

Unlike in the case of the supernatant containing the dissolved eggshell, the residual remaining after the urea solution treatment failed to dissolve in SDS or any other detergents, and the use of mass spectrometry for the identification of the composing proteins couldn't be performed. Hence, the residual was instead subjected to amino acid composition analysis by HCl hydrolysis, phenylisothiocyanate (PITC)-derivatization of the amino acids, and subsequent HPLC analysis (SFig. 3). Due to the harsh HCl hydrolysis condition degrading the Cys and Trp residues as well as hydrolyzing Gln to Glu, the content of 16 amino acid residues were detected. The result indicated that the residual has characteristically high proportions of Gly (23 %), Val (14 %), and Thr (11 %) residues (Table 1).

3.4. Bioinformatics search for the candidate egg glue protein using the whole bed bug genome

Since we believed that the undissolved residual contains the egg glue protein(s), a bioinformatics-based mapping and searching for the identification of protein was performed using the sequences of the proteins

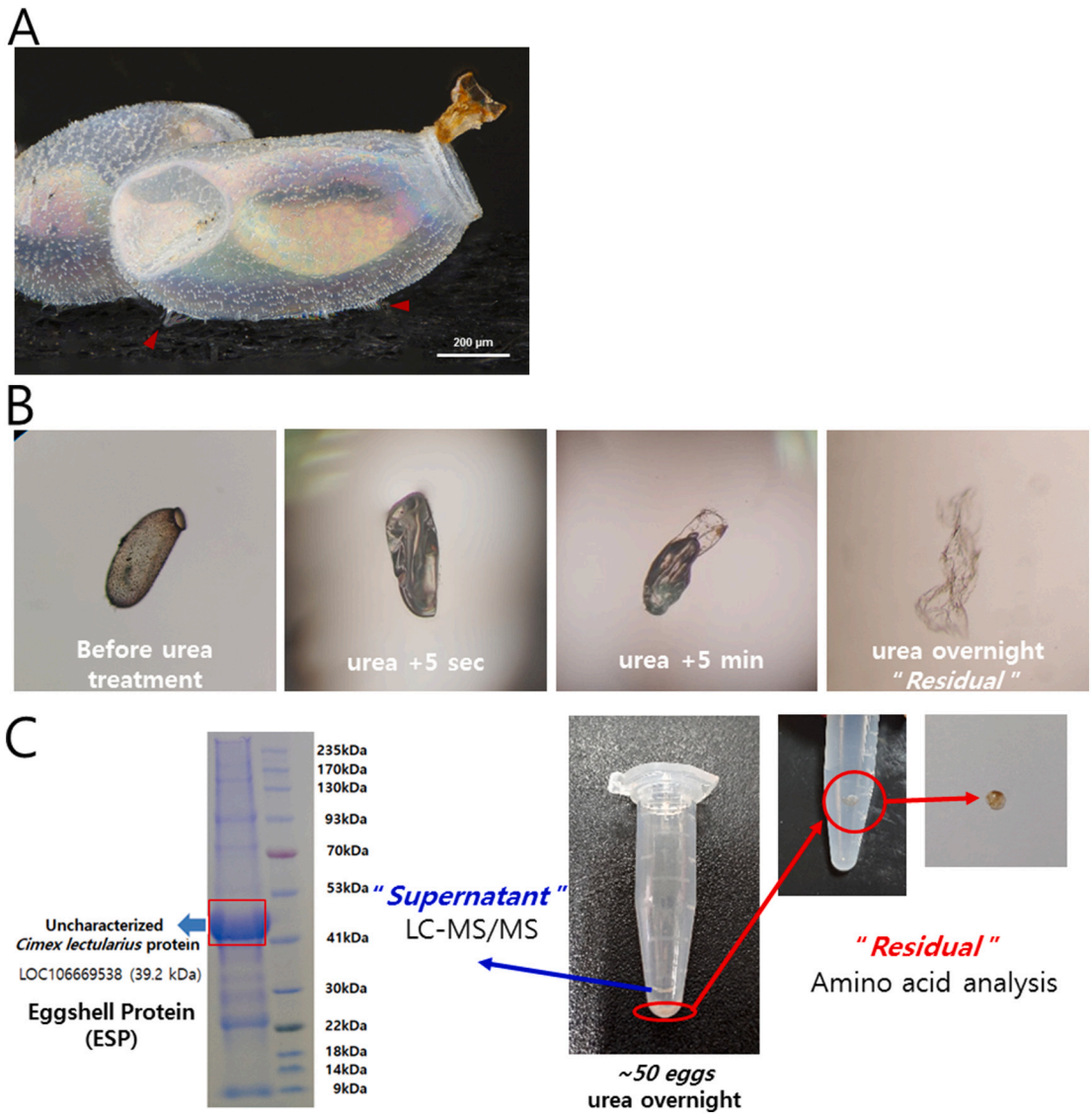


Fig. 1. An empty bed bug egg (A) was treated with urea-based solution (B) and the supernatant content analyzed using mass spectrometry (C). (A) Empty eggshells of bed bugs are shown with the glue material covering the outside. The arrowheads indicate the glue attaching the egg and the paper. Photographed using a DMC 5400 digital camera with a Leica Z16 APO motorized macroscope (Leica, Wetzlar, Germany). (B) Time-lined image of the bed bug egg is shown after urea treatment, in which an undissolved "residual" remained after overnight. (C) When ~50 eggs were treated overnight with the identical urea-based solution, a hard substance was left after washing and drying. The supernatant which is expected to contain the eggshell protein was run on SDS-PAGE and the major band analyzed for its content using LC-MS/MS. The undissolved "residual" was analyzed using amino acid composition analysis.

encoded from the whole genome of bed bug *C. lectularius* [17]. A similar approach has been successfully applied in identifying the human louse nit sheath proteins in the past from our lab [10]. Of the 24,194 bed bug whole-genome encoding proteins, 1428 proteins were selected in the final set used for comparison against the experimental amino acid content of the residual. The selected final list only contained proteins with a predicted signal sequence necessary for secretion, and also without any internal transmembrane α-helix indicative of being a membrane protein. The amino acid compositions of the 1428 proteins were deduced from the bed bug louse genome, and when the distances (R, see Methods and materials Section 2.5) between the amino acid (excluding Cys and Trp residues) composition deduced *versus* the experimental amino acid composition were calculated, the best amino acid content fit was obtained for the superoxide dismutase [Cu-Zn]-like isoform X1 (NCBI ref. seq. XP_014255310.1) with $R = 0.134$. However, in this case, significant discrepancies were observed even for the highly abundant residues (23 % Gly; 14 % Val; 11 % Thr) of the residual (SFig. 4).

Since the fit was not satisfactory, we assumed that the residual obtained from the egg consists of a mixture of two proteins (two-protein mix model, see Methods and materials Section 2.5.2) and compared the experimental amino acid composition against that of a complete pair of two sequences selected from the 1428 proteins. When this was done, the best matched fits were obtained mostly by a two protein mix containing the ESP (STable 2). At this point, it was only probable to reason that the ESP was not completely dissolved (or separated) out from the residual, complicating our bioinformatics analysis of using a single protein model.

Accordingly, the amino acid content mapping was performed to specifically include ESP as one component and any other of the remaining 1427 proteins to be the second component. When such analysis with ESP fixed as one element was done, the mapping results showed much less discrepancies between the amino acid contents reflected by the low R values. The lowest R-value ($R = 0.038$) was given when the mix model was of 36 % ESP and 64 % keratin associated protein 5–10like (KAP5–10L, XP_014254545.1) (Fig. 3). As expected, a much better satisfactory match between the amino acid residue contents

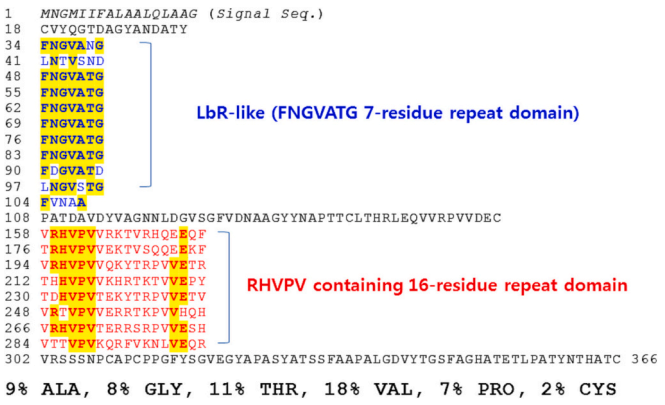


Fig. 2. The amino acid sequence of the eggshell protein. The uncharacterized protein LOC106669538 of the bed bug (NCBI reference sequence XP_014254554.1) with repeating amino acid sequences is the major component of the eggshell. The first repeat domain is classified as LbR-like domain which contains the ten 7-residue (FNGVATG) repeat. The second repeat domain is characterized by having eight mostly RHVPV-containing 16-residue repeat.

was observed (SFig. 5) and the second lowest R-valued protein had significantly higher R-value compared to the KAP5–10L (Fig. 3 *inset* and STable 3). Hence, the overall results implied that KAP5–10L along with the ESP is the component that comprises the urea undissolved residual.

3.5. *Temporal and spatial transcription profiles of ESP and KAP5–10L in various developmental bed bug stages*

The temporal transcription levels of *ESP* and *KAP5–10L* genes were investigated in 3rd instar nymphs, males, virgin females and gravid females using quantitative real-time PCR (qRT-PCR) (Fig. 4A). *ESP* transcribed mostly in gravid females and the relative level was 50,400, 302,000, and 386,000-fold higher than those in virgin females, 3rd instar nymphs, and males, respectively ($P < 0.0001$). Similarly, the highest expression level of *KAP5–10L* was observed in gravid females and was 33,700, 36,000, and 77,000-fold higher than those in virgin females, 3rd instar nymphs, and males, respectively ($P = 0.0001$).

In case of the spatial expression profile, *ESP* and *KAP5–10L* were most predominantly transcribed in the ovaries of gravid females and then followed by the oviducts of gravid females (Fig. 4B). The relative transcription levels of *ESP* and *KAP5–10L* in ovaries of gravid females were 263 and 52-fold higher compared to that in oviducts of gravid females and the fold difference was more than three orders of magnitude

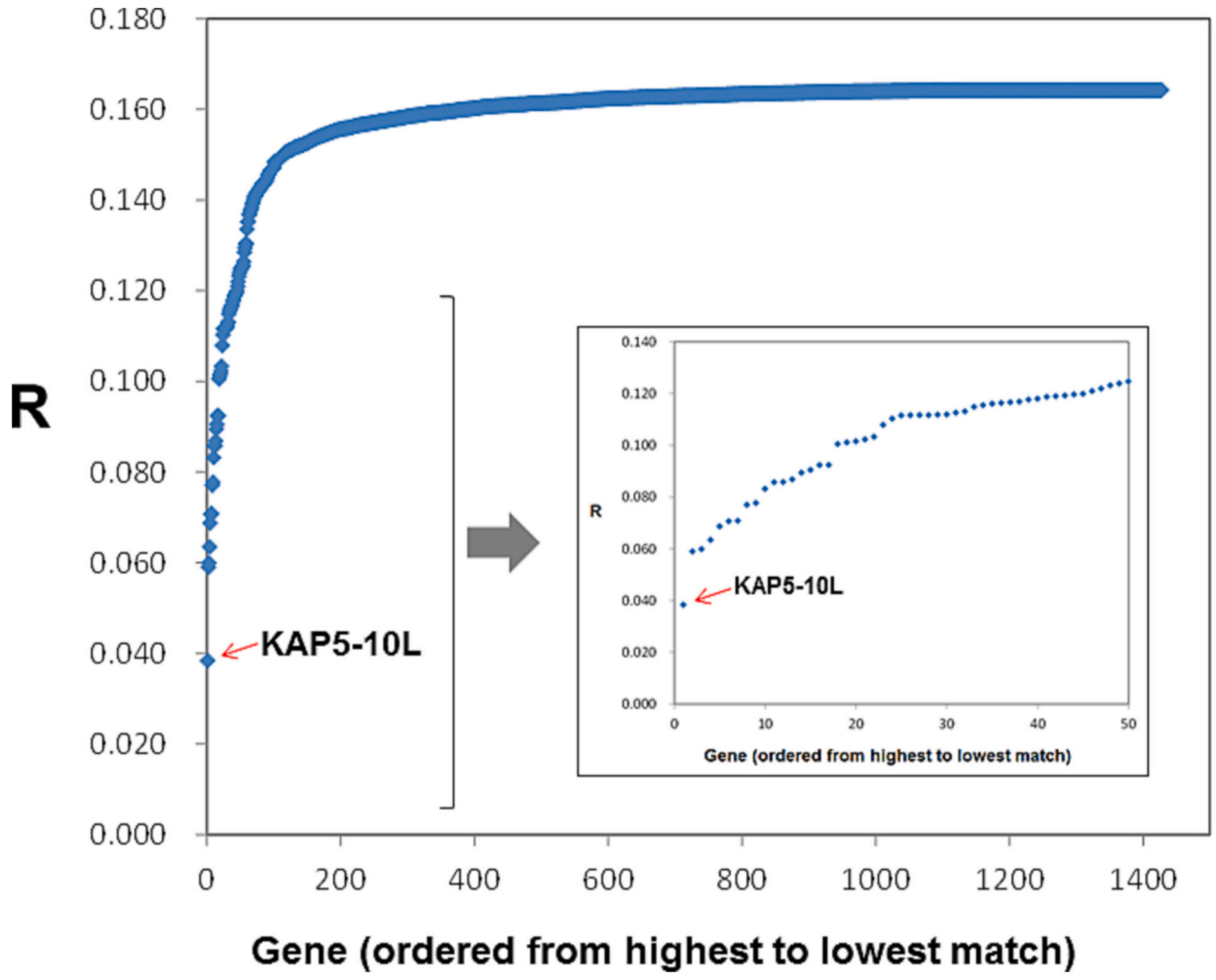


Fig. 3. Bioinformatics search of candidate egg glue protein. The distances (R, see Methods and materials Section 2.5 for definition) of the 16 amino acid (excluding cysteine and tryptophan) offsets were calculated between the experimental amino acid composition of the egg residual and the amino acids of 1428 proteins encoded by the bed bug genome. The mapping was performed using a two protein mix model where the ESP was assigned to be one component of the residual. The lowest R (shown in arrows) corresponds to a KAP5–10L, hence being the most probable protein constituting the egg glue. The genes on the horizontal axis were sorted according to the R values. The region with low R values is shown in the inset.

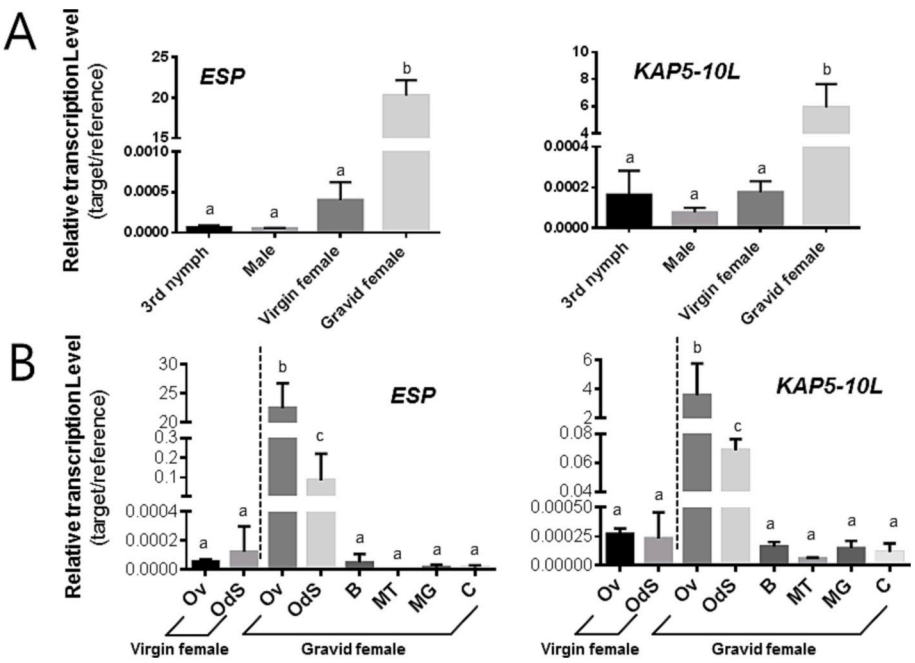


Fig. 4. Temporal (A) and spatial (B) transcription profiles of *ESP* and *KAP5–10L*. Transcription profiles of *ESP* (*egg shell protein*) and *KAP5–10L* (*keratin associated protein 5–10-like*) in various developmental stages (A) of *C. lecturarius* and in different organs (B) from 12-day old virgin or gravid females were measured (Ov, ovary; OdS, oviduct with spermatheca; B, brain; MT, Malpighian tubule; MG, midgut; C, cuticle). The transcription level of *ESP* in Malpighian tubules of gravid females was too low (1×10^{-6}) to be seen in the graph. The lowercase letters (a, b, and c) above the error bars indicate data sets that are statistically different ($P < 0.05$).

compared to those in other tissues.

3.6. *In vivo knockdown of KAP5–10L in female gravid bed bug by RNA interference*

Encouraged by the fact that the *KAP5–10L* is associated mostly with the reproductive organs of gravid females, we sought to knockdown the gene in the bed bug using RNA interference to gain more direct evidence that KAP5–10L is the egg glue protein. The *KAP5–10L* dsRNA was administered two times (24 and 48 h) after mating (SFig. 6A). To confirm that the targeted gene was indeed knocked down, the relative transcription levels of the *KAP5–10L* gene were assessed at 48, 72, and 96 h after the first dsRNA injection (respectively 72, 96, and 120 h after mating). Gene expression data analyzed from the bed bugs injected with *KAP5–10L* dsRNA showed a significant reduction in the *KAP5–10L* mRNA levels, in which were lowered by 5.9 fold (72 h after first injection; $P = 0.025$) and 7.0 fold (96 h after first injection; $P = 0.0025$) compared to the control *pQE30* dsRNA-injected bed bugs (SFig. 6B). No mortality was observed until day 20 after the first injection.

The knockdown of *KAP5–10L* significantly affected the oviposition (Fig. 5A) and the egg hatching rates (Fig. 5B). The average daily number of eggs laid by bed bugs injected with *KAP5–10L* dsRNA was significantly lower than that of the eggs laid by the control ($P < 0.02$) (Fig. 5A). The difference between the two groups was the highest at 9 days post first injection (9.0 fold higher for the control than the *KAP5–10L* dsRNA treated-bed bugs; $P < 0.02$), which was followed by 7 days (7.0 fold; $P < 0.02$), 8 days (6.4 fold; $P < 0.005$), and 4 days (3.5 fold; $P < 0.002$). In both cases, no eggs were laid after 10 days of the injection. When the eggs laid by the bed bugs were monitored for their hatchability, the percentage of nymphs emerging from the eggs laid by bed bugs with the *KAP5–10L* knockdown was significantly lower than that of the control (Fig. 5B). The hatchabilities of the eggs from the *KAP5–10L* knockdown bed bugs were below 17 %, while those from the control were higher than 85 %. The unhatched eggs appeared to be badly developed based on their overall shape and color, possibly due to water loss (*see below*).

Most of the eggs laid by the control were tightly bound to the paper by the egg glue (Fig. 6A) and were laid in a clustered form (Fig. 6C), whereas the eggs laid by the *KAP5–10* knockdown bed bugs were

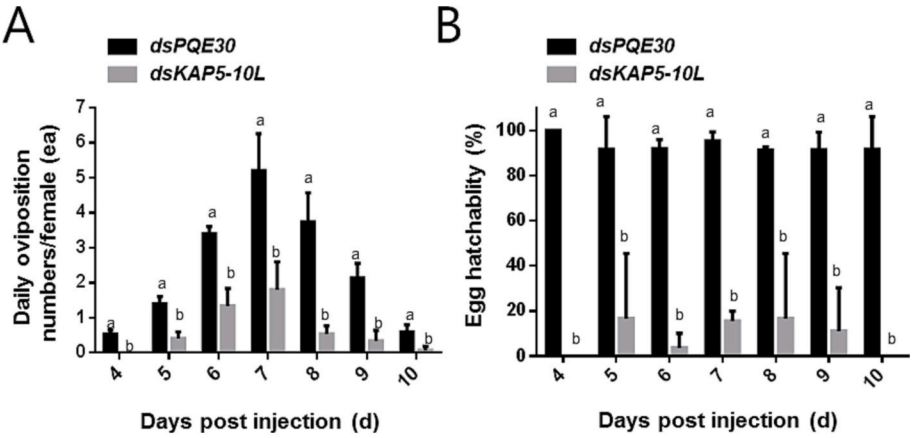


Fig. 5. The effects of KAP5–10L RNAi in bed bug (A) oviposition and (B) egg hatchability. The female bed bug's oviposition and the egg hatchability were analyzed using the multiple *t*-test with the Holm-Sidak method. The lowercase letters (a and b) above the error bars indicate data sets that are statistically significant ($P < 0.05$).

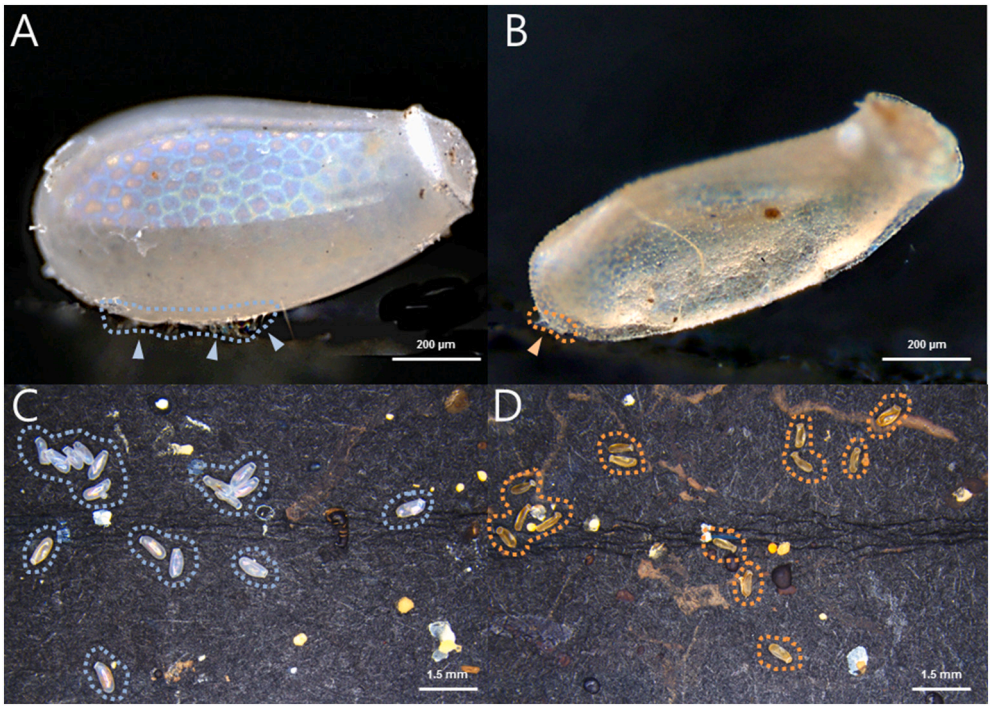


Fig. 6. RNAi phenotypes of eggs laid by bed bugs injected with *pQE30* dsRNA control (A and C) and *KAP5–10L* dsRNA (B and D). The RNAi phenotypes of eggs for *KAP5–10L* knockdown females (B) and control (A) are shown where the arrowheads and the boundaries indicate the glue between the egg and the paper. The eggs of the *KAP5–10L* knockdown females were found to be more scattered (D) than the clustered eggs of the control (C). The laid egg(s) is circled for visual clarity.

shrunken with a characteristic orange color (Fig. 6B). There were significant losses of glue material attaching the egg to the paper in the eggs laid by the *KAP5–10L* knockdown bed bugs (Fig. 6 A and B). All the unhatched eggs exhibited compound eyes, suggesting the death of embryos at a late developmental stage (Fig. 6B and D). Furthermore, the eggs were loosely attached to the paper, and were found to be scattered around (Fig. 6D). Unlike in the control, most of the laid eggs fell off from the paper with a few slight taps.

3.7. Protein sequence of the putative egg glue protein, KAP5–10L

The KAP5–10L (keratin associated protein 5–10 like) protein is a 149-residue protein containing a predicted N-terminal signal peptide sequence (residues 1–22) necessary as expected for a secreted glue protein (Fig. 7). Moreover, as for the ESP, there are also repeating sequences in the KAP5–10L where ten of the GNCGCG(*CG*)L-containing 7–9-residue repeat is characteristic. Also, two KTCCVTVPC sequences flank the N- and C-terminal regions of the GNCGCG(*CG*)L repeat (Fig. 7). It is important to note that due to such repeating sequences Cys (30 %) and Gly (33 %) are very abundant in the KAP5–10L protein. BLASTp search against all known NCBI protein sequences failed to find any sequence similar to KAP5–10L.

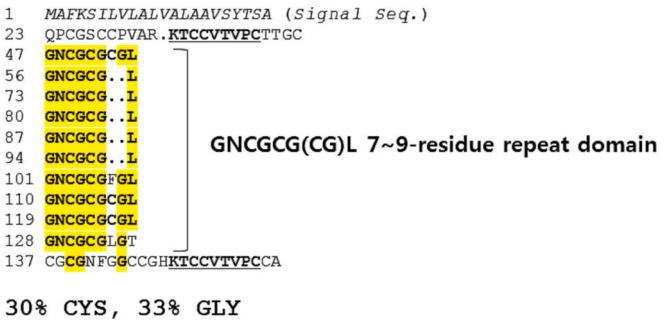


Fig. 7. The amino acid sequence of the KAP5–10L, the putative egg glue protein. The KAP5–10L is a 149-residue protein which contains a predicted signal peptide sequence at the N-terminus. Also, a ten repeating sequence of GNCGCG (*CG*)L is characteristic along with two KTCCVTVPC sequences (*underlined*). The protein is high in Cys (30 %) and Gly (33 %) content.

3.8. The secondary structure contents of recombinant bed bug ESP and KAP5–10L

In order to characterize the secondary structure content of bed bug ESP, the gene encoding ESP (residue 19–365) without the signal peptide (residue 1–17) and without the two terminal cysteines (Cys18 and Cys366, *i.e.* to avoid unwanted disulfide cross-linking) was cloned into an *E. coli* expression vector, and the ESP purified as a soluble protein (SFig. 7A). However, SDS-PAGE analysis indicated that the recombinant protein is quite unstable and gets continuously degraded (SFig. 7A *inset*). Nevertheless, CD was further used on this recombinant ESP to estimate the secondary structure contents (SFig. 7B). When the experimental ellipticity values between the 190 to 240 nm range were measured and fitted for secondary structure estimation using BESTSEL [26], the result indicated that ESP is composed of ~41 % anti-parallel β-sheets and ~ 14 % turns (and the remaining being unstructured).

Similarly, KAP5–10L (23–149) was over-expressed and purified using the *E. coli* expression system as well. As with the ESP, the N-terminal signal peptide (residue 1–22) was not included and all the cysteine residues (~30 % of the protein) of KAP5–10L were switched to encode serine residues instead to avoid complication *via* uncontrolled Cys-Cys disulfide cross-linkings. Moreover, because KAP5–10L does not contain any Tyr or Trp residues necessary for UV detection, the FPLC purification step was omitted, and the concentration estimated by using Bradford assay (SFig. 7C *inset*). The CD result indicated that KAP5–10L is estimated to be ~45 % anti-parallel β-sheets and ~16 % turns, which is very similar to the secondary composition of the ESP (SFig. 7C).

4. Discussion

We have dissolved the egg shells of empty bed bug egg cases by using a urea-based solution (8 M urea, 100 mM Tris pH 7.5 and 100 mM DTT), and analyzed the dissolved protein content by LC-MS/MS. A similar approach using a slightly different condition - 8 M urea, 2 M thiourea, 25 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA and 1 % Triton X-100 - was previously attempted by Gordon et al. [27,28], which resulted in identifying that the same LOC106669538 protein was the main protein content of the egg shell hence naming it as ESP (Eggshell protein). As Gordon et al.'s purpose of the study (written as a patent and an abstract only report) was in finding ways to develop antibodies against bed bugs at the earliest stage of its development, detailed analysis on the ESP has not been made before. In general, the ESP can be divided into two domains with characteristic repeat sequences. The first domain with ten 7-residue (mostly FNGVATG) repeats has been found in proteins with a LbR (left-handed β-roll)-like domain exemplified in VgrG (virulence factor of type VI secretion tip protein) from β-proteobacteria of the genus *Chromobacterium* and *Burkholderiabacteria*, and in the proteins of malaria parasite *Plasmodium*. The second domain is characterized by having eight repeats of mostly RHVPV sequence as part of the extended 16-residue sequence. These repeat sequences of the second domain have been also found in the eggshell proteins of stink bug (*Plautia stali* (S.)) and the kissing bug (*R. prolixus*) [7] as expected from the fact that bed bugs and they are closely related insects of the Hemiptera order.

The undissolved proteins of the residual leftover from the urea treatment were HCl-hydrolyzed into individual amino acids, and the contents analyzed. The component of this residual which is expected to contain the egg glue protein were identified using a bioinformatics approach of matching the experimental amino acid contents with the protein sequence information derived from the whole-genome of the bed bug. A similar approach has been proven successful before in identifying the human head/body louse glue proteins (termed the louse nit sheath proteins) [10] in our group. We believe that this bioinformatics approach using the experimental amino acid content and the whole genome data is useful and valid in specifically identifying mostly fibrous proteins containing high percentages of selected amino acids concentrated from the repeating sequences. Especially in the case of proteins that are unable to be solubilized, we do not know of any other method in identifying the content. The only difference in the case of the bed bug compared to that of the previous human louse one was that the undissolved residual could not be matched to a single protein, and hence a two-protein mix model was used to identify ESP (36 %) and KAP5–10L protein (64 %) as its major content. Although several species (including human and fruit fly) have proteins called the keratin associated protein 5–10 (also known as KRTAP5–10) in their genome, they have no resemblance to the bed bug KAP5–10L other than being proteins with repeat sequences high in Cys and Gly contents [29].

As expected, the transcription levels of *ESP* and *KAP5–10L* genes were found to be exceptionally high in the ovaries of the gravid female bed bug. Unlike most insects having specified female accessory glands secreting the egg glue, the bed bugs appear to have a unique simplified reproductive system consisting of three parts (*i.e.* ovaries, extremely shorten oviduct and spermatheca; SFig. 1), in which the glue-like substances may be secreted from the fused tissues of ovaries/accessory glands/*etc*. *In vivo* RNA interference experiments to knockdown *KAP5–10L* further indicated that the absence of KAP5–10L affects both the number of eggs that the bed bug lays and the hatchability of the egg itself as well. Knocking down *KAP5–10L* resulted not only in decreasing the amount of the glue protein around the egg but also had a negative effect on the healthiness of the egg as judged by their shriveled shape and orange color. Although the less amount of glue on the egg and the less clustered eggs laid by the *KAP5–10L* knockdown female support the glue-like function of the gene, the overall phenotypic results (number of eggs laid by the female, hatchability, and appearance of the egg) indicate a much more complex role of the gene, perhaps including a function of lubrication during oviposition and protection against desiccation. The effect of KAP5–10L on the fertilization rate cannot be completely excluded because fertilization takes place during oviposition, which usually occurs around a week after mating. Nonetheless, since all the unhatched eggs almost developed to the late embryonic stage, as judged by the presence of red compound eyes in the dead embryo, we presume that KAP5–10L is unlikely to affect the fertilization rate.

Both the final hardening of the eggshell and the curing of the egg glue are known to be completed only after oviposition by further protein cross-linking events. From the fact that the ESP and the KAP5–10L share similar folding structures as determined from the CD results, it is tempting to suppose that the anti-parallel β-sheets of the two proteins pack against each other to form multiple packing layers (Of note, various physiological conditions may render the actual native proteins into a more ordered form). Transglutaminases secreted along with the glue proteins in human head/body lice [11] and in silkworms [12] have provided a certain mechanism of glue curing by allowing multiple covalent bonds to form between the abundant Gln and Lys residues in the glue protein. Although this curing mechanism might not apply for the bed bug KAP5–10L with low existence of Gln and Lys residues, the extremely high content of Cys points to a possibility that complex Cys-Cys disulfide bonds may form assisted by the oxygen-mediated oxidation of Cys happening only after the oviposition. Through such mechanism, further cross-linking between KAP5–10L and ESP is also plausible given the fact that 2 % of ESP sequence is Cys of which their locations are at the N/C-terminal ends and the linker connecting the two repeat domains of ESP (Fig. 2). In retrospect, this state of the bed bug eggshell and egg glue aggregating *via* complex intermolecular and intramolecular covalent Cys-Cys disulfide bondings could have been the reason why complete separation of ESP and KAP5–10L using the urea-based solution was not possible and thus the experimental amino acid composition of the bed bug egg residual could only be matched by using a two protein (ESP and KAP5–10L) mix model.

It is also important to note that as observed from the multiple minor protein bands in SDS-PAGE of the dissolved eggshell, the bed bug eggshell is clearly not made up of a single protein, but is likely a complex mixture of various proteins along with the major ESP. In a similar manner, the egg glue may also contain additional proteins other than the identified KAP5–10L.

List of abbreviations

CD	circular dichroism
Glx	glutamic acid or glutamine
ESP	Eggshell protein
KAP5–10L	keratin associated protein 5–10 like

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A10044154). J.L. was also supported by the NRF, funded by Ministry of Science and ICT (2020R1A2C1005956).

CRediT authorship contribution statement

S.H. Lee and SY. Park designed the research. Y.W. Sim, S. Cho, J.H. Kim and S.Y. Park performed the experiments. S. Cho, J.H. Kim, S.H. Lee, J. Lee and S.Y. Park analyzed the data. S. Cho, J.H. Kim, S.H. Lee, J. Lee and S.Y. Park discussed the results and co-wrote the manuscript.

Declaration of competing interest

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2023.126004.

References

[1] J.W.L. Beament, The formation and structure of the chorion of the egg in an Hemipteran, *Rhodnius prolixus*, Q. J. Microsc. Sci. 87 (1946) 393–439.
[2] L.H. Margaritis, The egg-shell of *Drosophila melanogaster* III. Covalent crosslinking of the chorion proteins involves endogenous hydrogen peroxide, Tissue Cell 17 (1985) 553–559.
[3] L.H. Margaritis, Structure and physiology of the eggshell, in: L.I. Gilbert, G. A. Kercut (Eds.), Comprehensive Insect Physiology, Biochemistry and Pharmacology vol. 1, Pergamon Press, Oxford, 1985, pp. 153–230.
[4] G.L. Waring, Morphogenesis of the eggshell in Drosophila, Int. Rev. Cytol. 198 (2000) 67–108.
[5] V. Cavaliere, F. Bernardi, P. Romani, S. Duchi, G. Gargiulo, Building up the Drosophila eggshell: first of all the eggshell genes must be transcribed, Dev. Dyn. 237 (2008) 2061–2072.
[6] J.S. Li, J. Li, Major chorion proteins and their crosslinking during chorion hardening in *Aedes aegypti* mosquitoes, Insect Biochem. Mol. Biol. 36 (12) (2006) 954–964.
[7] D.M. Bouts, A.C. Melo, A.L. Andrade, M.A. Silva-Neto, O. Paiva-Silva Gde, M. H. Sorgine, L.S. da Cunha Gomes, H.S. Coelho, A.P. Furtado, E.C. Aguiar, L.N. de Medeiros, E. Kurtenbach, S. Rozental, N.L. Cunha-E-Silva, W. de Souza, H. Masuda, Biochemical properties of the major proteins from *Rhodnius prolixus* eggshell, Insect Biochem. Mol. Biol. 7 (11) (2007) 1207–1221.
[8] P. Irles, M.D. Piulachs, Citrus, a key insect eggshell protein, Insect Biochem. Mol. Biol. 41 (2) (2011) 101–108.
[9] D. Li, M.G. Huson, L.D. Graham, Proteinaceous adhesive secretions from insects, and in particular the egg attachment glue of Opodiphthera sp. moths, Arch. Insect Biochem. Physiol. 69 (2) (2008) 85–105.
[10] J.K. Park, Y.J. Han, J. Lee, S.-W. Joo, J.H. Kim, S.H. Lee, S.Y. Park, Characterization of the human head louse nit sheath reveals proteins with adhesive property that show no resemblance to known proteins, Sci. Rep. 9 (2019) 48.
[11] J. Kim, D. Lee, S. Park, J.M. Clark, S.H. Lee, Characterization of nit sheath protein functions and transglutaminase-mediated cross-linking in the human head louse, *Pediculus humanus capitis*, Parasites Vectors 14 (2021) 425.
[12] Y. Lei, K. Guo, Y. Zhang, X. Zhang, L. Qin, X. Wang, H. Zhu, Y. Guo, W. Yang, B. Li, Q. Xia, P. Zhao, Z. Dong, Adhesive property and mechanism of silkworm egg glue protein, Acta Biomater. 134 (2021) 499–512.
[13] C.D. Steelman, A.L. Szalanski, R. Trout, J.A. McKern, C. Solorzano, J.W. Austin, Susceptibility of the bed bug *Cimex lectularius* L. (Hemiptera: Cimicidae) to selected insecticides, J. Agric. Urban Entomol. 25 (1) (2008) 45–51.
[14] K. Dang, S.L. Doggett, G. Veera Singham, C.-Y. Lee, Insecticide resistance and resistance mechanisms in bed bugs, Cimex spp. (Hemiptera: Cimicidae), Parasites Vectors 10 (2017) 318.
[15] K.F. Haynes, Sleeping with the enemy, Sci. Am. 306 (2) (2012) 50–55.
[16] K.S. Yoon, D.H. Kwon, J.P. Strycharz, C.S. Hollingsworth, S.H. Lee, J.M. Clark, Biochemical and molecular analysis of deltamethrin resistance in the common bed bug (Hemiptera: Cimicidae), J. Med. Entomol. 45 (6) (2008) 1092–1101.
[17] J.B. Benoit, Z.N. Adelman, K. Reinhardt, A. Dolan, M. Poelchau, E.C. Jennings, E. M. Szuter, R.W. Hagan, H. Gujar, J.N. Shukla, F. Zhu, M. Mohan, D.R. Nelson, A. J. Rosendale, C. Derst, V. Resnik, S. Wernig, P. Menegazzi, C. Wegener, N. Peschel, J.M. Hendershot, W. Blenau, R. Predel, P.R. Johnston, P. Ioannidis, R. M. Waterhouse, R. Nauen, C. Schorn, M.-C. Ott, F. Maiwald, J.S. Johnston, A. D. Gondhalekar, M.E. Scharf, B.F. Peterson, K.R. Raje, B.A. Hottel, D. Armisén, A.J. J. Crumière, P.N. Refki, M.E. Santos, E. Sghaier, S. Viala, A. Khila, S.-J. Ahn, C. Childers, C.-Y. Lee, H. Lin, D.S.T. Hughes, E.J. Duncan, S.C. Murali, J. Qu, S. Dugan, S.L. Lee, H. Chao, H. Dinh, Y. Han, H. Doddapaneni, K.C. Worley, D. M. Muzny, D. Wheeler, K.A. Panfilio, I.M. Vargas Jentzsch, E.L. Vargo, W. Booth, M. Friedrich, M.T. Weirauch, M.A.E. Anderson, J.W. Jones, O. Mittapalli, C. Zhao, J.-J. Zhou, J.D. Evans, G.M. Attardo, H.M. Robertson, E.M. Zdobnov, J.M. C. Ribeiro, R.A. Gibbs, J.H. Werren, S.R. Palli, C. Schal, S. Richards, Unique features of a global human ectoparasite identified through sequencing of the bed bug genome, Nat. Commun. 7 (2016) 10165.
[18] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (2019) 420–423.
[19] A. Krogh, B. Larsson, G. von Heijne, E.L.L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J. Mol. Biol. 305 (3) (2001) 567–580.
[20] K. Yoshitaka, H. Mitsumoto, C.-Y. Lee, Duplicated female receptacle organs for traumatic insemination in the tropical bed bug Cimex hemipterus: adaptive variation or malformation? PLoS One 9 (2) (2014), e89265.
[21] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT–PCR, Nucleic Acids Res. 29 (9) (2001), e45.
[22] J. Carayon, Paragenital system, in: R. Usinger (Ed.), Monograph of the Cimicidae, Entomological Society of America, Philadelphia, Pennsylvania, 1966, pp. 81–167.
[23] E.H. Morrow, G. Arnqvist, Costly traumatic insemination and a female counter-adaptation in bed bugs, Proc. Biol. Sci. 270 (1531) (2003) 2377–2381.
[24] K. Reinhardt, R. Naylor, M.T. Siva-Jothy, Reducing a cost of traumatic insemination: female bedbugs evolve a unique organ, Proc. Biol. Sci. 270 (1531) (2003) 2371–2375.
[25] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J. M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, New Jersey, 2005, pp. 571–607.
[26] A. Micsonai, F. Wien, L. Kernya, Y.H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 112 (24) (2015) E3095–E3103.
[27] N. Gordon, L. O'Shaughnessy, D. Fitzpatrick, S. Doyle, B. Mitchell, Two pronged approach to the detection of *Cimex Lectularius* (common bed bug) using novel bed bug proteins, J. Allergy Clin. Immunol. 137 (2) (2016) (Suppl AB32 (Abstract only)).
[28] Patent WO2017037126A1, Novel proteins and detection methods, N. Gordon, L. O'Shaughnessy, B. Mitchell.
[29] Y. Shimomura, M. Ito, Human hair keratin-associated proteins, J. Investig. Dermatol. Symp. Proc. 10 (3) (2005) 230–233.