

Identification and characterization of ovary development-related protein EJO1 (Eri s 2) from the ovary of *Eriocheir sinensis* as a new food allergen

Lina Zhu<sup>1</sup>, Tiantian She<sup>1</sup>, Yingying Zhang<sup>1</sup>, Shaoshen Li<sup>2</sup>, Zixian Xu<sup>3</sup>, Juanjuan Yan<sup>1</sup>, Huiqiang Li<sup>1</sup>

<sup>1</sup> School of Medical Laboratory, Tianjin Medical University, Tianjin, China

<sup>2</sup> Department of Medical Laboratory, Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital, Tianjin, China

<sup>3</sup> Department of Medical Laboratory, Tianjin Chest Hospital, Tianjin, China

#### Keywords

crab ovary / *Eriocheir sinensis* / food allergen/ ovary development-related protein EJO1 / recombinant expression

Corresponding author: Prof. Huiqiang Li, School of Medical Laboratory, Tianjin Medical University, No.1 Guangdong Road, Hexi District, Tianjin 300203, China

Email: lhq@tmu.edu.cn

Fax: +86-13520031642

#### Abbreviations

Received: 02 14, 2016; Revised: 04 26, 2016; Accepted: 05 19, 2016

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/mnfr.201600144](#).

This article is protected by copyright. All rights reserved.

TBST, tris buffered saline containing 0.05% tween 20; PB, phosphate-buffer; PBS, phosphate-buffered saline; PBST, PBS containing 0.05% tween 20; MW, molecular weight; RT, room temperature;

Scope: Crab is a major source of shellfish allergens. Most studies have focused on allergens in crab muscle (CM) rather than on allergens in crab ovary (CO). This study aimed to identify potential allergens in CO from *Eriocheir sinensis*.

Methods and results: Dot blot and immunoblotting results revealed the differential reactivity of CM and CO extracts to positive sera from patients allergic to crabs. Four CO proteins showed good specific IgE-binding activities in 2-DE/immunoblotting analysis; mass spectrometry identified the proteins as hemocyanin, vitellogenin, ovary development-related protein EJO1 and EJO2. The recently identified allergen EJO1 is named 'Eri s 2' in the World Health Organization and International Union of Immunological Societies (WHO/IUIS) allergen nomenclature database. Recombinant Eri s 2 exhibited good reactivity to positive sera, and pre-incubation with recombinant Eri s 2 abrogated the reactivity of positive sera from two patients to CO in a dose-dependent manner. Moreover, co-incubation of recombinant Eri s 2 with patient basophils dose-dependently promoted basophil activation, confirming the biological activity of Eri s 2.

Conclusion: CO tissue is an important allergen source, and Eri s 2 is a new crab allergen. This study provides insights that will be useful for component-resolved diagnostics for crab allergy.

## 1 Introduction

Shellfish allergy affects as many as 0.6% of the global population [1], and it is one of the 8 major types of food allergies. The overall pooled estimates for all age groups of the self-reported lifetime prevalence of shellfish allergy is 1.3% (0.9–1.7), and the prevalence of food-challenge-defined shellfish allergy is 0.1% (0.06–0.3) in Europe[2]. A telephone survey found that the prevalence of shellfish allergy in the United States is as high as 2.0%[3], and shellfish allergy is also found in people in many Asian countries[4]. A cross-sectional population-based questionnaire survey in Hong Kong reported that shellfish was the top allergen, accounting for more than a third of all allergic reactions in Chinese children and teenagers[5]. A recent comparative study of shellfish allergy in Singapore and the Philippines showed that the prevalence rates of shellfish allergy were 1.19% in preschoolers and 5.23% in adolescents in Singapore and 5.12% in adolescents in the Philippines[6]. Crab allergy, a common form of shellfish allergy, is now a prominent food

allergy in Asia. In southern China, crab(14.85%) was one of the five leading allergic foods on self-reported in the children aged 7-12 years, and the most common causative food were shrimp and crab on doctor's diagnosis[7]. In Korea, crab is also one of the top 5 allergen sources for doctor-diagnosed and self-reported food allergy[8]. As the major type of shellfish allergy, crab allergy is also a long-lasting disorder, typically persisting throughout life[3,9]. The symptoms of crab allergy usually include cutaneous reactions, with throat and lip pruritus, flushing, urticaria, and localized angioedema. There may also be severe and life-threatening reactions such as laryngeal edema, asthma, and fatal anaphylaxis[10, 11]. It is thus important for crab allergic individuals to avoid eating crab, and early diagnosis of crab allergy can help them avoid future exposure.

Until now, serum-specific IgE (sIgE) tests and SPTs have been the principal in vitro methods for diagnosing allergy[12-14]. However, the specificity and sensitivity of food allergy diagnosis depend primarily on the quality of the antigens. Van Kampen et al. reported that the protein and antigen content and patterns of SPT solutions varied remarkably depending on the manufacturer, and assays using solutions with higher antigen levels were more sensitive and efficient[15]. Improving the allergen preparation process is essential for standardizing sIgE tests and SPTs. Notably, food allergies differ in different regions[16, 17], and factors such as food processing practices may also impact food protein allergenicity and oral tolerance[18]. In the Asian-Pacific region, there is a relatively unique clinical manifestation of shellfish allergy in which a single individual can have reactions of varying severity. It is hypothesized that shellfish allergenicity can vary according to the consumed species and that allergic reactions may depend on which part of the shellfish was eaten as well as on the method of preparation[19].

Many studies have investigated the identity and the biological, chemical, immunological, and structural properties of allergens in crab. However, most studies have focused on crab muscle (CM) tissues. Known CM-derived allergens include tropomyosin, arginine kinase, troponin (23 kDa), sarcoplasmic calcium-binding protein (20 kDa),  $\alpha$ -actin (42 kDa), and the smooth endoplasmic reticulum  $\text{Ca}^{2+}$  ATPase protein (113 kDa)[20-22]. Of these, only tropomyosin (Cha f 1, Por p 1, Chi o 1, Scy pa 1), arginine kinase (Chi o 2, Scy pa 2), and sarcoplasmic calcium-binding protein (Chi o 4, Scy pa 4) are officially named in the WHO/IUIS allergen nomenclature database (<http://www.allergen.org/>). Crab ovary (CO), which contains masses of eggs, is a structure unique to female crabs that is processed into a variety of food products due to its high nutritional value and delicious taste. Studies have identified allergens in the muscle tissue and various organs of shellfish, and vitellogenin, ovarian peritrophin 1 precursor,  $\beta$ -actin, and the 14-3-3 protein are suggested to be allergens that are present in the banana shrimp ovary at different stages of ovarian development[23]. However, no studies have investigated possible allergens in CO, although the identification of new allergens from CO will greatly contribute to the effective diagnosis and treatment of crab allergy

*Eriocheir sinensis*, also known as the Chinese mitten crab, is a popular aquatic food in Asia, especially in China[24, 25]. In this study, we analyzed extracts derived from the *E. sinensis* ovary and demonstrated that the extracts contained allergenic proteins. These proteins were identified as hemocyanin, vitellogenin, the ovary development-related protein EJO1 and EJO2. Hemocyanin is a copper-containing multi-subunit respiratory protein that is abundant in CO but not in CM[26], and it is known as an important shellfish allergen[23]. Vitellogenin, a high molecular weight glycolipid protein complex, is a precursor of the yolk protein. It is enriched in the hemolymph, fat bodies, and ovaries of non-mammal egg-laying animals after sexual maturity, and it was recently found in CO and in crab hepatopancreas[27]. Vitellogenin is a major allergen in fish ovary, shellfish ovary, and even in bees[23, 28-30]. Ovary development-related protein is never reported as an allergen. This study demonstrated that the ovary development-related protein EJO1, a recently identified protein, is a new crab allergen. According to the nomenclature series of *E. sinensis* allergens, a submission was made to the WHO/IUIS Allergen Nomenclature, which granted the name Eri s 2 to this protein

## 2 Materials and methods

### 2.1 Patients and blood collection

For this study, all individuals were recruited from the Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital between August 2014 and December 2014 (Table 1). Of these, sera from 36 patients (median IgE level, 6.54 kUA/l to crab; range, 0.35–37 kUA/l) were used individually to test IgE binding to proteins from crab ovary and crab muscle by dot blots. The patients were selected on the basis of IgE to crab and crab-induced allergy symptoms and a history of crab allergy. (Table S1). The sera used individually or as a pool to test IgE binding to proteins by 2-DE/ immunoblotting or ELISA were chosen from these. In addition, 6 non-crab-sensitive individuals were used as the control group and had no crab allergy-related symptoms or history plus low levels of serum total IgE (<100 kUA/L) and crab-specific IgE (<0.1 kUA/L).

Whole blood samples (2 mL blood was drawn into EDTA tubes). were separated by centrifugation (3000 xg for 5 min) to remove blood cells unless the sample was to be used for the basophil activation test. The sera were stored at -20°C. The protocol was approved by the Ethics Committee, and patients provided written informed consent before they were included in the study.

### 2.2 Preparation of the CO extract and CM extract from *E. sinensis*

*E. sinensis* were purchased from the local supermarket (Tianjin, China) and boiled in water for 10 min. Next, the ovary and muscle tissues were collected. Crab ovaries can turn yellowish red after heating (Fig. S1). Then, CO and CM tissue were each ground up with a mortar in liquid nitrogen environment and the cells were lysed in lysis buffer (40 mM Tris base, 7 M urea, 2 M thiourea, 4%

CHAPS [3-(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate], 65 mM DTT, and 1× protease inhibitor cocktail from Roche, Mannheim, Germany) by a ratio of 1mg/4ml [31]. After sonication on ice for 30 min, the tissue lysate was incubated at 4°C for 1 h before being centrifuged at 10,000×g for 1 h at 4°C. The supernatant fraction was recovered and sterilized by filtration through a 0.22-µm filter. Aliquots were stored at -80°C. The protein concentration was determined using the 2D quant kit from GE healthcare (USA). Lysis buffer were used to adjust the concentration of CO and CM before each test.

### 2.3 Dot blot

For dot blot analysis, approximately 10 µg of CO or CM proteins was spotted directly onto nitrocellulose (NC) membrane. Each spots was 2 µl (diameter , 0.5 mm). After blocking with 5% (w/v) skim milk in Tris-buffered saline containing 0.05% Tween 20 (TBST) for 1 h at room temperature (RT), the membranes were washed (3×10 min) with TBST. The membrane was probed with 36 positive serum samples and with 6 negative serum samples diluted 1:10 in 0.9 mL TBST for 2 h at RT. After washing 3× with TBST, Alkaline Phosphatase (AP)-labeled goat anti-human IgE antibodies (A3525, Sigma-Aldrich, St. Louis, MO, USA) diluted 1:1000 in 1 mL TBST was were added to the membranes and incubated for 1 h at RT. After washing 3× with TBST, 200 µl AP substrate (Sigma-Aldrich) were added to produce measurable purple blue signals. ImageJ software was utilized for quantification of all dots.

### 2.4 SDS-PAGE

SDS-PAGE analysis was performed using the Bio-Rad Mini Protean II (Bio-Rad, Hercules, CA, USA). 25 µg of CO proteins and 30 µg of CM proteins ( ) were boiled for 5 min in 1× loading buffer (0.1 M Tris, pH 6.8, containing 4% w/v SDS, 20% w/v glycerol, 0.02% w/v bromophenol blue, and 10% w/v 2-mercaptoethanol) before loading onto 12% polyacrylamide gels. After the proteins were separated by electrophoresis, they were visualized directly by CBB R250 staining (Sigma-Aldrich,).

### 2.5 Western blotting

30 µg of CO or CM proteins ( ) was boiled in 1× loading buffer for 5 min before separation on 12% polyacrylamide gels. The separated proteins were electro-transferred onto polyvinylidene fluoride (PVDF) membrane[32]. After blocking with 5% (w/v) skim milk in TBST overnight at 4°C, the membrane was washed (3×10 min) with TBST. The membrane was probed with 6 positive serum samples diluted 1:10 in 0.9 mL TBST overnight at 4°C; subsequent steps were performed as described for dot blot analysis.

### 2.6 Two-dimensional gel electrophoresis (2-DE) and 2-DE immunoblotting

Using one-dimensional PAGE, CO proteins were separated by isoelectric point using the Bio-Rad PROTEAN i12 IEF system (Bio-Rad)[33]. Briefly, about 0.2 mg of CO proteins was added to 125-μL rehydration solution containing 40 mM Tris-base, 7 M urea, 2 M thiourea, 4% CHAPS, 0.001% (w/v) bromophenol blue, 65 mM DTT, and 0.2% Pharmalyte pH 3-10 (IPG buffer, GE Healthcare, USA). Next, the mixture was covered with a 7-cm IPG strip (pH 3-10 Non-linear, GE Healthcare) and loaded into the IEF system. The parameters for isoelectric focusing (IEF) were as follows: 18°C, 50 V, 14 h; S1: 250 V for 30 min; S2: 500 V for 30 min; S3: 4000 V for 3 h; S4: 4000 V for 20000 volt-hours (Vh); S5: 500 V for any length of time. After rinsing the IPG strip with equilibration solution containing 6 M urea, 20% (v/v) glycerol, 2% (w/v) SDS, 0.002% (w/v) bromophenol blue, and 0.375 M Tris-HCL (pH 8.8) (2×10 min), second dimension electrophoresis was performed using the Mini Protean II. The IPG strip ran on 12% gels vertically, with the top sealed with a mixture of 0.5% (w/v) agarose and 0.002% (w/v) bromophenol blue. The proteins on the gels were visualized by CBB staining.

For 2-DE immunoblotting, about 0.2 mg of CO proteins were separated using 2-DE gels and then transferred onto PVDF membrane. Subsequent steps were as described for Western blotting. The membrane was probed using pooled sera from 10 positive individuals (P. 3,4,5,6,7,8,10,11,12,13), pooled sera from 6 negative serum and using unpooled sera from 2 positive individuals (all diluted 1:10 in 6-mL TBST) by overnight incubation at 4°C. Subsequent steps were performed as described for the dot blot analysis.

## 2.7 Mass spectrometry (MS) identification of allergenic proteins

Proteins with high content in 2-DE and could easy to separate were chose to MS/MS analysis. The proteins isolated by 2-DE were digested in-gel into smaller peptides using trypsin (Promega, Madison, WI, USA) prior to MS/MS analysis. An ABI-4800 matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI TOF/TOF) analyzer was used for peptide sequencing, and the acquired data was further analyzed using ABI GPS Explorer and Mascot software (www.matrixscience.com). Databases, including the NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), were searched to characterize the proteins

## 2.8 Expression and purification of recombinant Eri s 2 (rEri s 2)

Total RNA was extracted from 100 mg of CO tissue using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit from Thermo Scientific (Waltham, MA, USA). The Eri s 2 cDNA was amplified by PCR using specific primers (Table S2) and Phusion Hot Start II DNA Polymerase from Thermo Scientific. The PCR parameters were as follows: 98°C for 30 s, 35 cycles of 98°C for 10 s, 70°C for 30 s, 72°C for 55 s, and finally 72°C for 10 min. The PCR product was excised from the agarose gel and purified using the Universal DNA

Purification Kit (TIANGEN Biotech, China). After purification, the Eri s 2 cDNA sequence was inserted between the Nde I and Xho I restriction cleavage sites in the pET-28b (+) expression vector (Novagen, Darmstadt, Germany) using T4 DNA ligase (Thermo Scientific). Top 10 Competent Cells (TIANGEN Biotech) were transformed with the recombinant plasmid pET28b(+)-Eri s 2 and grown on luria broth (LB) plates containing 50 µg/mL kanamycin at 37°C overnight. Clones were selected by Colony PCR using the specific primers 'Colony for' and 'Colony back' (Table S2), and positive colonies were identified by sequencing. After obtaining a pET28b(+)-Eri s 2 plasmid with the correct sequence, the plasmid was transformed into *E. coli* Rosetta (DE3) cells (TIANGEN Biotech). Protein expression was induced in transformed cells using 1 mM isopropyl β-D-1-thiogalactoside (IPTG) when the cells grew to an OD600 of 0.6–0.8. Expression was induced at 37°C for 4 h.

The rEri s 2 protein was subsequently purified using a Ni-NTA column according to the manufacturer's instructions (Qiagen, Dusseldorf, Germany). Purified rEri s 2, which contained an N-terminal His-tag, was verified using either His-tag antibodies or positive sera and Western blotting as described above.

## 2.9 ELISA and the ELISA inhibition

96-well EIA/RIA plates (CoStar, St. Louis, MO, USA) were coated with 100 ng/well of rEri s 2 diluted in coating buffer (0.05 M carbonate-bicarbonate buffer, pH 9.6) by overnight incubation at 4°C. The plates were then washed 3× with PBS containing 0.05% PBST. The plates were blocked with PBS containing 0.05% Tween 20 and 2% BSA (PBSTB) overnight at 4°C. After blocking, the plates were washed again and incubated with 100-µl of positive sera (diluted 1:20 with PB) at 37°C for 2 h. After washing, the plates were incubated with 100-µl HRP-labeled mouse anti-human IgE antibodies (diluted 1:20 with PBS) at 37°C for 1 h. After washing, TMB (3, 3', 5, 5'-tetramethylbenzidine; Invitrogen, Carlsbad, CA, USA) was used for color development, and 1 M HCl was added to terminate the reaction. Absorbance values were measured at 450 nm using an Epoch microplate reader (BioTek, VT, USA). Negative sera from 6 non-crab allergic patients were used to exclude the possibility that nonspecific binding led to false-positive results.

For the ELISA inhibition assay, 96-well plates were coated with 100 ng/well of CO diluted in coating buffer overnight at 4°C. Serum samples from patients who were found to be allergic to the Eri s 2 protein by ELISA assay were preincubated with 1 µg/mL of the inhibitor (rEri s 2) in 1 mL serum for 1 hour at RT. In addition, sera from 2 patients who were allergic to the Eri s 2 protein were preincubated with different amounts (0.01-10 µg) of rEri s 2 or CO in 1 mL of serum.

## 2.10 IgE immunoblotting inhibition

For IgE immunoblotting inhibition, approximately 10µg CO were spotted directly on NC membrane and blocked with 5% skim milk in TBST. Sera from Eri s 2-allergic patients were preincubated with 10µg/mL of rEri s 2 in 1 mL of serum for 1 hour at RT and then tested with CO by dot blot as described above.

### 2.11 Basophil activation test

Upregulated CD63 or/and CD203c expression on the surface of basophils was considered a marker of basophil activation[34-36], so we investigated the effect of rEri s 2 on CD63 and CD 203c expression in basophils. Briefly, whole blood from Patient 13 was collected and incubated with different concentrations (0.04, 0.2, 1, and 5 µg/mL) of rEri s 2 or 1 µg/mL of CO at 37°C for 15 min. Next, the blood was transferred to ice for 5 min to stop the degranulation of activated basophils before probing with a mixture of FITC-labeled CD63, PE-labeled CD123, and PerCP-labeled HLA-DR antibodies (BD Biosciences, NJ, USA). After lysis of red blood cells by 1× BD FACS lysing solution (BD Biosciences), the cells were recovered and resuspended in PBS before analysis on a FACSCalibur cell analyzer (BD Biosciences).

### 2.12 Statistical analysis

Data are reported as the mean ± SD (standard deviation) and were analyzed using SPSS 19.0 software. The two-tailed student's t-test or Analysis of Variance (ANOVA) was used to determine the statistical significance of differences between groups, and Spearman's rank correlation was used to assess the correlation between sIgE levels and integrated densities of dots in dot blotting. We considered *p*-values <0.05 to be statistically significant.

## 3 Results

### 3.1 Dot blot

The immunoreactivity of CO and CM extracts to positive sera from crab-allergic patients was subjected to dot blot analysis (Fig. 1). The integrated densities of all dots were determined using ImageJ software. Of the serum samples, 8 of 36 samples showed stronger reactivity to CO extracts than to CM extracts, and the remaining 11 serum samples showed stronger reactivity to CM extracts than to CO extracts (Fig. 1A). The median integrated density of the dots was substantially different between the CO and CM groups (Fig. 1B). These results suggested that the sIgE composition differed in different positive serum samples.

The sIgE binding reactivity can be evaluated by determining the integrated density of positive spot in the dot blot. Accordingly, we evaluated the correlations between serum sIgE levels and the integrated dot densities for the CO and CM groups using Spearman's rank correlation method. As expected, there was a better correlation ( $R^2 = 0.1558$ ,  $p < 0.05$ ) between sIgE levels and the integrated densities of dots in the CM group than the case in the CO group ( $R^2 = 0.0009743$ ,  $p = 0.85677$ ; Fig. 1C and 1D).

### 3.2 SDS-PAGE and western blotting

To confirm that the protein composition was different in the CO and CM extracts, SDS-PAGE was used to separate and visualize CO and CM extract proteins. As seen in Figure 2, CO extract had at least 15 protein bands, most of which were concentrated at molecular weight (MW) 85-110 kDa, 70 kDa, 43 kDa, 28 kDa, and 22 kDa. In contrast, the 17 major bands in CM were around MW 230 kDa, 80 kDa, 43 kDa, and 38 kDa. Furthermore, the CO and CM extracts showed different binding reactivity to 6 positive sera samples, as seen in Western blotting (Fig. 2, right panel). At least 10 IgE-binding proteins were detected in CO extract, with MWs ranging from 22 kDa to 250 kDa. The bands were mainly around MW 250 kDa, 130 kDa, 105-85 kDa, 70 kDa, and 28 kDa. Positive sera from Patients 21 and 26 showed higher binding reactivity to CM extract than to CO extract, while the opposite was true for positive serum from Patient 16. The CM and CO groups otherwise showed similar binding signals within the groups.

### 3.3 2-DE and 2-DE immunoblotting

2-DE and 2-DE immunoblotting were used to identify potential new allergens in CO extract (Fig. 3). Five spots on the 2-DE immunoblot showed good binding activity with a pool of 10 positive sera (Fig. 3B). Some of them showed good activity with unpooled sera from Patient 13 and 20 (Fig. 3D&E). Five positive proteins had MW 105 kDa and pI 7.5, MW 85 kDa and pI 8, MW 70 kDa and pI 5.5, and MW 28 kDa and pI 6, MW 28 kDa and pI 6.2 were chosen for MS analysis.

### 3.4 Protein identification and characterization by MS

To identify the five IgE binding proteins that were identified by 2-DE immunoblotting, we performed MS/MS and obtained the sequences of several internal peptides. The five proteins were identified by searching an NCBI database using the Mascot program (Table S3 & Table 2). Information from protein databases identified the 105-kDa and 85-kDa protein band as vitellogenin and the 70-kDa protein band as hemocyanin subunit 6, which are known shellfish allergens. Comparison of the sequences from the 4 and 5 band to other sequences in the NCBI database showed that the 28-kDa protein bands were most likely the ovary development-related protein EJO1 and EJO2, which were isomers of ovary development-related protein, and the two sequences are 97.37% identical (Fig. S2).

In this study, ovary development-related protein EJO1 was been Expression, purification and characterization as a new food allergen and was nominated as Eri s 2 in WHO/IUIS allergen nomenclature database

### 3.5 Expression and purification of rEri s 2

Expression and purification of the recombinant protein rEri s 2 was performed after confirming the Eri s 2 cDNA sequence(Figure 4), which is a 756-bp open reading frame, by PCR and agarose gel electrophoresis (Fig. S3A). The expressed protein was confirmed to be a 32-kDa protein using SDS-PAGE and CBB staining (Fig. S4).

### 3.6 ELISA , ELISA inhibition and IgE immunoblotting inhibition assays

Both ELISA and ELISA inhibition were used to check the reactivity of rEri s 2 to positive sera. As shown in Figure 5A, rEri s 2 exhibited significantly higher reactivity to all 29 positive serum samples than to the 6 negative serum samples ( $p=0.042$ ). In addition, preincubation with rEri s 2 greatly attenuated the binding of positive sera from 13 crab-allergic patients to CO extract; this was most evident for sera from Patients 11 and 16 (Fig.5B). Furthermore, rEri s 2 showed dose-dependent inhibition in an ELISA inhibition assay that used positive sera from Patients 11 and 16 (Fig. 5C and 5D). Notably, pre-incubation with rEri s 2 greatly attenuated the binding reactivity of positive sera from 9 crab-allergic patients to CO in the IgE immunoblotting inhibition assay (Fig. 6) . The binding of IgE from Patients 3, 4, 11, 16, and 25 was almost completely inhibited by rEri s 2.

### 3.7 Basophil activation test

To further confirm the allergenicity of rEri s 2, we analyzed its role in activating basophils. Activated basophils can be identified by upregulated CD63 expression and by high levels of cell surface-bound IgE (Fig. 7A and 7B). As expected, both rEri s 2 and CO extract induced the activation of basophils from Patient 13, with rEri s 2 having a much stronger effect at the same protein concentration. When using 1  $\mu\text{g/mL}$  of both in BAT assay, we observed that rEri s 2 and CO activated 25.49% and 10.35% of basophils respectively. In addition, rEri s 2 induced basophil activation in a dose-dependent manner (Fig. 7C). These results confirmed the basophil-activating ability of rEri s 2.

## 4 Discussion

Many studies have demonstrated that allergens in crab are extremely diverse. Most of these allergen studies focused on CM, ignoring the allergens in CO. In this study, we analyzed CO extracts from *E. sinensis* and demonstrated that CO also contains allergens. We identified and characterized a

new allergen- ovary development-related protein EJO1 for the first time, it was nominated as Eri s 2 in WHO/IUIS allergen nomenclature database. This information could be used to improve standardization of crab extracts for in vivo or in vitro diagnosis, for component-resolved diagnosis (CRD) of crab allergy and, perhaps, for immunotherapy treatment.

We compared the sIgE-binding proteins in CO and CM extracts using dot blot and Western blotting. The CM and CO extracts had very different binding reactivity toward the same positive serum samples, demonstrating that there was ovary-specific IgE in positive sera from crab-allergic patients. 8 of 36 of the positive serum samples (8 of 36) showed higher reactivity to CO extract than to CM extract. However, there was a correlation between clinical sIgE levels and the integrated dot densities in dot blotting in the CM group but not in the CO group. This may be because most studies researched the allergens in crab muscle [20], so clinical diagnostic kits mostly used crab muscle extracts or allergens in crab muscle as coating antigens for detecting serum sIgE . The addition of ovary allergens to clinical diagnostic kits would improve the sensitivity of the crab allergy diagnosis.

Regarding the IgE-binding protein profile in the CO extract, we detected at least 10 IgE-binding proteins with MWs ranging from 22 kDa to 250 kDa by Western blotting, plus we identified 5 spots in 2-DE immunoblotting that had good binding activity to positive sera and high content. MS/MS analysis identified these CO proteins as hemocyanin, vitellogenin, ovary development-related protein EJO1 and EJO2. The MW of spot 1 and 2 in fig 3B were about 105kDa and 85 kDa, but the theory MW of vitellogenin was 288kDa. Under denaturing conditions, vitellogenin aggregates released polypeptides with low molecular masses of 115, 105, and 85 kDa [37].The presence of vitellogenin and hemocyanin in CO of shellfish was known previously[23], but the present study was the first to identify ovary development-related protein as allergen in crab. This information is important in terms of identifying the allergen profile of a particular crab that can be used for CRD of crab allergy. The current trend in improving allergy diagnostics involves CRD, which takes advantage of single recombinant allergens to determine the patients' sensitization profile[38]. The use of high quality recombinant allergens rather than natural allergens could help standardize allergen preparations, thereby improving the sensitivity and effectiveness of the sIgE test[39-41] and ensuring the reliability and accuracy of CRD[42]. Ultimately, this could also advance desensitization therapy[43-46]. Therefore, identification of the specific components that cause symptoms in crab allergy is critical, since allergen identification provides a basis for allergy testing and treatment. Notably, we compared the sequences of ovary development-related protein isomers--EJO1(Genbank No. AY 185917.1) and EJO2 (AY185918.1), and the two sequences are 97.37% identical (Fig. S2). In this study, the reactivity of rEri s 2 was comparable to that of positive sera for CO, indicating that rEri s 2 can be used in the diagnosis of Eri s 2-induced or CO-induced allergy. And we will characterization the allergenicity of EJO2 in the future,EJO1 was nominated as Eri s 2 in WHO/IUIS allergen

nomenclature database. GenBank data show that the full-length EJO1 cDNA is 876 bp long with a 756-bp open reading frame. Fifteen peptide fragments of EJO1 were sequenced (Figure 4 and Table S3). The encoded EJO1 protein is theoretically 252 amino acids long and is predicted to have an isoelectric point of 6.18 and a MW of 28.180 kDa[47]. These values are consistent with our 2-DE immunoblotting results. Some research has indicated that EJO1 is expressed in the heart, intestine, and ovary, with high abundance in the ovary, but that it is nearly absent from muscle and from the hepatopancreas[25]. Other work has indicated that EJO1 is a secretory protein with a role in ovarian development in *E. sinensis*. Our ELISA results showed that rEri s 2 had excellent reactivity to sIgE from positive sera. Moreover, pre-incubation of the positive sera with rEri s 2 inhibited the binding of sIgE from positive sera to CO extracts. The inhibition assays did not lead to strong inhibition presumably because the amount of protein expressed was not high so the concentration of inhibitor we used was a little low. Furthermore, the ability of rEri s 2 to activate basophils from crab-allergic patients, which was stronger than the activation by CO confirmed that Eri s 2 functions as an allergen, when the BAT was performed using equal concentration of total protein. This stronger activity is due to the fact that CO contains many proteins, many of which are not allergens. Therefore the higher activation with Eri s 2 demonstrated that Eri s 2 accounts for a significant portion of the allergenic activity of CO. In our study, the reactivity of rEri s 2 to positive sera was comparable to the reactivity of CO, indicating that rEri s 2 can be used to diagnose Eri s 2- or CO-induced allergy.

In conclusion, this study confirmed that CO tissue is an important source of crab allergens. Our results revealed that Eri s 2, a new allergen from CO, had good sIgE-binding and basophil-activating activity and thus can be used in CRD of crab allergy. Our results not only enrich our knowledge of *E. sinensis*-derived allergens, but they also highlight the need to consider local eating habits and food species when studying food allergens. Food allergies are currently diagnosed at the molecular level [39], so our findings can be used to improve the accuracy of crab-specific IgE testing, and they provide an experimental basis for developing molecular diagnostic kits for crab allergy.

#### Author contributions

Study idea and design: Huiqiang Li, Lina Zhu; clinical samples: Shaoshen Li; experiment: Lina Zhu, Yingying Zhang, Juanjuan Yan; data analysis: Lina Zhu, Zixian Xu; manuscript: Huiqiang Li, Tiantian She, Lina Zhu.; critical review of the manuscript: Huiqiang Li, Tiantian She; final approval: all authors

#### Acknowledgments

This article is protected by copyright. All rights reserved.

We thank Zhi-Feng Zhu, Meng Zhao, and Yang Yu at the Department of Immunology, School of Basic Medicine, Tianjin Medical University, for their expert technical assistance. We also thank the personnel at the Pediatrics Research Institute, Tianjin Children's Hospital, for helping us perform the basophil activation tests.

Funding: This study was supported by the National Natural Science Foundation of China (#81371882).

#### Conflict of interest statement

The authors have declared no conflict of interests.

#### Ethical approval

Ethical approval was obtained from the medical ethics committee of Tianjin Medical University (TMuhMEC2013016).

#### References

- [1] Sicherer, S. H., Sampson, H. A., Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *The Journal of allergy and clinical immunology* 2014, 133, 291-307; quiz 308.
- [2] Nwaru, B. I., Hickstein, L., Panesar, S. S., Roberts, G., *et al.*, Prevalence of common food allergies in Europe: a systematic review and meta-analysis. *Allergy* 2014, 69, 992-1007.
- [3] Sicherer, S. H., Munoz-Furlong, A., Sampson, H. A., Prevalence of seafood allergy in the United States determined by a random telephone survey. *The Journal of allergy and clinical immunology* 2004, 114, 159-165.
- [4] Wong, L., Huang, C. H., Lee, B. W., Shellfish and House Dust Mite Allergies: Is the Link Tropomyosin? *Allergy, asthma & immunology research* 2016, 8, 101-106.
- [5] Ho, M. H., Lee, S. L., Wong, W. H., Ip, P., Lau, Y. L., Prevalence of self-reported food allergy in Hong Kong children and teens--a population survey. *Asian Pacific journal of allergy and immunology / launched by the Allergy and Immunology Society of Thailand* 2012, 30, 275-284.

- [6] Shek, L. P., Cabrera-Morales, E. A., Soh, S. E., Gerez, I., *et al.*, A population-based questionnaire survey on the prevalence of peanut, tree nut, and shellfish allergy in 2 Asian populations. *The Journal of allergy and clinical immunology* 2010, 126, 324-331, 331.e321-327.
- [7] Xian M, Li J. Allergic diseases of the skin and drug allergies – 2029. Prevalence and factors of food allergy among school-age children in Southern China[J]. *The World Allergy Organization Journal*, 2013,6(Suppl 1):P115.
- [8] Ju, S. Y., Park, J. H., Kwak, T. K., Kim, K. E., Attitudes and preferences of consumers toward food allergy labeling practices by diagnosis of food allergies. *Nutrition research and practice* 2015, 9, 517-522.
- [9] Ayuso, R., Sanchez-Garcia, S., Lin, J., Fu, Z., *et al.*, Greater epitope recognition of shrimp allergens by children than by adults suggests that shrimp sensitization decreases with age. *The Journal of allergy and clinical immunology* 2010, 125, 1286-1293.e1283.
- [10] Wu, A. Y., Williams, G. A., Clinical characteristics and pattern of skin test reactivities in shellfish allergy patients in Hong Kong. *Allergy and asthma proceedings : the official journal of regional and state allergy societies* 2004, 25, 237-242.
- [11] Turner, P., Ng, I., Kemp, A., Campbell, D., Seafood allergy in children: a descriptive study. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* 2011, 106, 494-501.
- [12] Muraro, A., Werfel, T., Hoffmann-Sommergruber, K., Roberts, G., *et al.*, EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. *Allergy* 2014, 69, 1008-1025.
- [13] Mehl, A., Niggemann, B., Keil, T., Wahn, U., Beyer, K., Skin prick test and specific serum IgE in the diagnostic evaluation of suspected cow's milk and hen's egg allergy in children: does one replace the other? *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2012, 42, 1266-1272.
- [14] Williams, P., Onell, A., Baldracchini, F., Hui, V., *et al.*, Evaluation of a novel automated allergy microarray platform compared with three other allergy test methods. *Clinical and experimental immunology* 2015.
- [15] van Kampen, V., de Blay, F., Folletti, I., Kobierski, P., *et al.*, Evaluation of commercial skin prick test solutions for selected occupational allergens. *Allergy* 2013, 68, 651-658.
- [16] Salo, P. M., Arbes, S. J., Jr., Jaramillo, R., Calatroni, A., *et al.*, Prevalence of allergic sensitization in the United States: results from the National Health and Nutrition Examination Survey (NHANES) 2005-2006. *The Journal of allergy and clinical immunology* 2014, 134, 350-359.

- [17] Lee, A. J., Thalayasingam, M., Lee, B. W., Food allergy in Asia: how does it compare? *Asia Pacific allergy* 2013, 3, 3-14.
- [18] Lepski, S., Brockmeyer, J., Impact of dietary factors and food processing on food allergy. *Molecular nutrition & food research* 2013, 57, 145-152.
- [19] Lee, A. J., Gerez, I., Shek, L. P., Lee, B. W., Shellfish allergy--an Asia-Pacific perspective. *Asian Pacific journal of allergy and immunology / launched by the Allergy and Immunology Society of Thailand* 2012, 30, 3-10.
- [20] Abramovitch, J. B., Kamath, S., Varese, N., Zubrinich, C., et al., IgE Reactivity of Blue Swimmer Crab (*Portunus pelagicus*) Tropomyosin, Por p 1, and Other Allergens; Cross-Reactivity with Black Tiger Prawn and Effects of Heating. *PLoS one* 2013, 8, e67487.
- [21] Misnan, R., Murad, S., Yadzir, Z. H., Abdullah, N., Identification of the major allergens of *Charybdis feriatus* (red crab) and its cross-reactivity with *Portunus pelagicus* (blue crab). *Asian Pacific journal of allergy and immunology / launched by the Allergy and Immunology Society of Thailand* 2012, 30, 285-293.
- [22] Abdel Rahman, A. M., Kamath, S. D., Lopata, A. L., Robinson, J. J., Helleur, R. J., Biomolecular characterization of allergenic proteins in snow crab (*Chionoecetes opilio*) and de novo sequencing of the second allergen arginine kinase using tandem mass spectrometry. *Journal of proteomics* 2011, 74, 231-241.
- [23] Khanaruksombat, S., Srisomsap, C., Chokchaichamnankit, D., Punyarit, P., Phiriyangkul, P., Identification of a novel allergen from muscle and various organs in banana shrimp (*Fenneropenaeus merguensis*). *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* 2014, 113, 301-306.
- [24] He, L., Li, Q., Liu, L., Wang, Y., et al., A Catalog of Proteins Expressed in the AG Secreted Fluid during the Mature Phase of the Chinese Mitten Crabs (*Eriocheir sinensis*). *PLoS one* 2015, 10, e0136266.
- [25] Gong, Y. N., Li, W. W., Sun, J. L., Ren, F., et al., Molecular cloning and tissue expression of the fatty acid-binding protein (Es-FABP) gene in female Chinese mitten crab (*Eriocheir sinensis*). *BMC molecular biology* 2010, 11, 71.
- [26] Yan, F., Zhang, Y., Jiang, R., Zhong, M., et al., Identification and agglutination properties of hemocyanin from the mud crab (*Scylla serrata*). *Fish & shellfish immunology* 2011, 30, 354-360.
- [27] Jia, X., Chen, Y., Zou, Z., Lin, P., et al., Characterization and expression profile of Vitellogenin gene from *Scylla paramamosain*. *Gene* 2013, 520, 119-130.

- [28] Pedrosa, M., Boyano-Martinez, T., Garcia-Ara, C., Quirce, S., Shellfish Allergy: a Comprehensive Review. *Clinical reviews in allergy & immunology* 2015, 49, 203-216.
- [29] Perez-Gordo, M., Sanchez-Garcia, S., Cases, B., Pastor, C., *et al.*, Identification of vitellogenin as an allergen in Beluga caviar allergy. *Allergy* 2008, 63, 479-480.
- [30] Blank, S., Seismann, H., McIntyre, M., Ollert, M., *et al.*, Vitellogenins are new high molecular weight components and allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* venom. *PloS one* 2013, 8, e62009.
- [31] Zhu L.N., Hou L.Y., Zhang Y.Y., Zhou Y.J., *et al.*, Analysis of allergic components in tissue protein of *Eriocheir sinensis* with different extraction and processing methods. *Chinese Journal of Immunology* 2014, 30, 1652-1657.
- [32] Towbin, H., Staehelin, T., Gordon, J., Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* 1979, 76, 4350-4354.
- [33] An, S., Shen, C., Liu, X., Chen, L., *et al.*, Alpha-actinin is a new type of house dust mite allergen. *PloS one* 2013, 8, e81377.
- [34] Khan, F. M., Ueno-Yamanouchi, A., Serushago, B., Bowen, T., *et al.*, Basophil activation test compared to skin prick test and fluorescence enzyme immunoassay for aeroallergen-specific Immunoglobulin-E. *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology* 2012, 8, 1.
- [35] MacGlashan, D., Jr., Expression of CD203c and CD63 in human basophils: relationship to differential regulation of piecemeal and anaphylactic degranulation processes. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2010, 40, 1365-1377.
- [36] Vila, L., Moreno, A., Gamboa, P. M., Martinez-Aranguren, R., Sanz, M. L., Decrease in antigen-specific CD63 basophil expression is associated with the development of tolerance to egg by SOTI in children. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2013, 24, 463-468.
- [37] Pateraki, L. E., Stratakis, E., Characterization of vitellogenin and vitellin from land crab *Potamon potamios*: identification of a precursor polypeptide in the molecule. *The Journal of experimental zoology* 1997, 279, 597-608.

- [38]Valenta, R., Lidholm, J., Niederberger, V., Hayek, B., *et al.*, The recombinant allergen-based concept of component-resolved diagnostics and immunotherapy (CRD and CRIT). *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 1999, 29, 896-904.
- [39] Aalberse, J. A., Meijer, Y., Derksen, N., van der Palen-Merkus, T., *et al.*, Moving from peanut extract to peanut components: towards validation of component-resolved IgE tests. *Allergy* 2013, 68, 748-756.
- [40] Schmid-Grendelmeier, P., Cramer, R., Recombinant allergens for skin testing. *International archives of allergy and immunology* 2001, 125, 96-111.
- [41] Aleksic, I., Popovic, M., Dimitrijevic, R., Andjelkovic, U., *et al.*, Molecular and immunological characterization of Mus a 5 allergen from banana fruit. *Molecular nutrition & food research* 2012, 56, 446-453.
- [42] Wangorsch, A., Jamin, A., Foetisch, K., Malczyk, A., *et al.*, Identification of Sola I 4 as Bet v 1 homologous pathogenesis related-10 allergen in tomato fruits. *Molecular nutrition & food research* 2015, 59, 582-592.
- [43] Hoffmann-Sommergruber, K., Pfeifer, S., Bublin, M., Applications of Molecular Diagnostic Testing in Food Allergy. *Current allergy and asthma reports* 2015, 15, 56.
- [44] Palomares, O., Cramer, R., Rhyner, C., The contribution of biotechnology toward progress in diagnosis, management, and treatment of allergic diseases. *Allergy* 2014, 69, 1588-1601.
- [45] Cromwell, O., Hafner, D., Nandy, A., Recombinant allergens for specific immunotherapy. *The Journal of allergy and clinical immunology* 2011, 127, 865-872.
- [46] Valenta, R., Linhart, B., Swoboda, I., Niederberger, V., Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. *Allergy* 2011, 66, 775-783.
- [47] Ma, C., Guo, Y., Zhou, K., Identification and characteristics of a novel gene, EJO1, in the Chinese mitten crab (*Eriocheir japonica sinensis*) ovary. *Molecular biology reports* 2010, 37, 2587-2592.

Figure 1. Dot blot analysis of crab ovary (CO) and crab muscle (CM) proteins. (A) A comparison of the reactivity of CO and CM with 36 positive serum samples using dot blotting. The integrated densities of all dots were determined using ImageJ software; the values are averaged values from two independent experiments. 1-36, patient serum samples; N, pooled sera from 6 non-crab-allergic individuals. Columns and bars represent the mean and SD. (B) Comparison of the median integrated density values in the CO and CM groups in (A). Box plots show the median integrated density values and the 5th and 75th interquartile ranges. Vertical bars indicate the 5%-95% range. P, patient sera; N, pooled sera from 6 non-crab-allergic individuals. (C, D) Assessment of correlations between the clinical sIgE levels and the integrated densities of the dot blots in (A) (C, CO group; D, CM group) using Spearman's rank correlation method. \* indicates a significant difference,  $p < 0.05$ .

Figure 2. SDS-PAGE and Western blotting analysis of crab ovary (CO) and crab muscle (CM) proteins. (A) SDS-PAGE analysis of the proteins in CO and CM extracts. (B) Western blotting analysis of the binding abilities of CO and CM to positive sera from Patients 5, 11, 13, 16, 21, and 26. M, molecular weight markers; P, positive sera.

Figure 3. 2-DE and 2-DE immunoblotting analysis of crab ovary (CO) proteins. (A) 2-DE analysis of CO proteins. (B,C,D,E) 2-DE immunoblotting analysis of CO proteins: (B) using pooled sera from 10 craballergic patients. (C) using pooled sera from 6 non-crab-allergic individuals. (D) using a positive serum samples from Patients 13. (E) using a positive serum samples from Patients 20. 1-5 indicate the allergen candidate proteins.

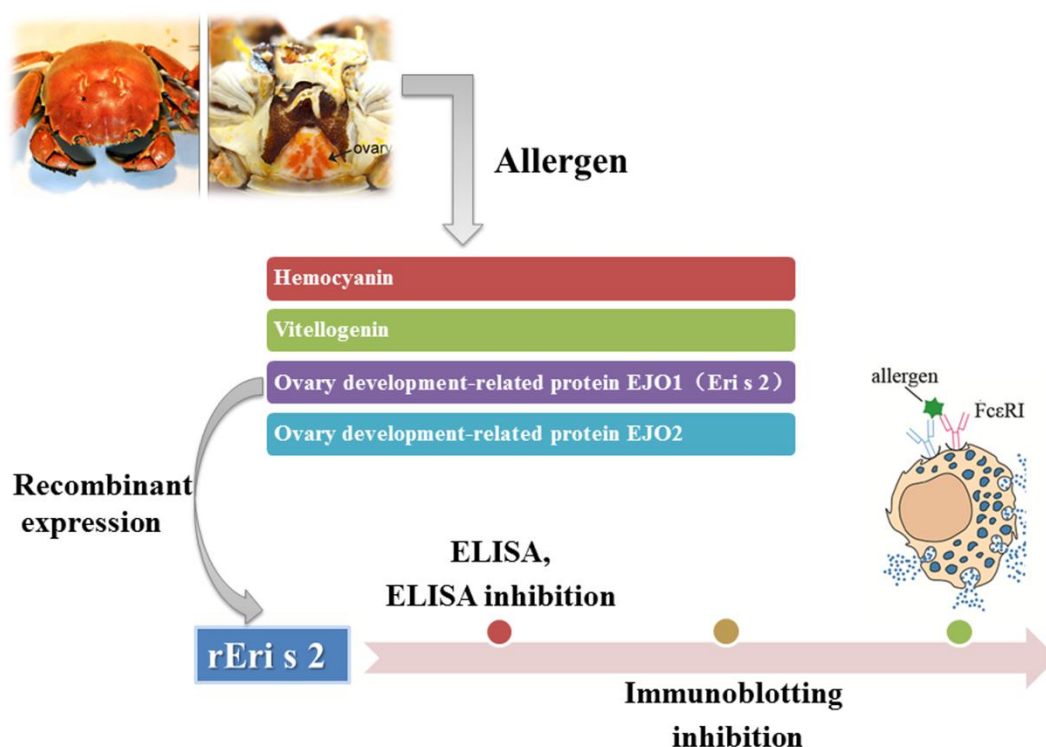
Figure 4. The cDNA sequence encoding Eri s 2 (upper panel) and the deduced amino acid sequence (lower panel). The underlined sequences were identified by MS/MS analysis and Edman degradation. \*, stop codon.

Figure 5. Positive sera shows binding reactivity with rEri s 2. (A) Evaluation of the reactivity of positive sera to rEri s 2 by ELISA. Two independent experiments with triplicate samples were conducted; a representative result is shown. P, positive sera from 29 crab-allergic patients; N, negative sera from 6 non-crab-allergic individuals. (B) Evaluation of the inhibitory effect of rEri s 2 on the binding of positive sera to crab ovary (CO) by ELISA inhibition assay. Averaged values of three independent experiments that were conducted in duplicate are shown. The control was sera that was not preincubated with Eri s 2. (C, D) Evaluation of the effects of rEri s 2 and crab ovary (CO)-induced inhibition on the binding of positive sera from Patients 11 and 16 to CO extracts by ELISA inhibition assay. The concentrations of Eri s 2 and CO ranged from 0 to 10  $\mu$ g/mL.

Figure 6. Evaluation of the inhibitory effect of rEri s 2 on the binding of positive sera to crab ovary (CO) by IgE immunoblotting inhibition.

Figure 7. rEri s 2 and crab ovary (CO) extract both activate basophils in vitro. (A) The gating strategy for basophils. CD123+/HLA-DR cells (R2) were identified as basophils, and CD63+ basophils were identified as activated basophils (R3). (B) Evaluation of the effect of Eri s 2 and CO on basophil activation by flow cytometry. Goat anti-human IgE antibodies were used as the positive control, and PBS was used as the negative control. The doses of CO and Eri s 2 were all 1 µg/mL. (C) Basophil activation test analysis of rEri s 2-induced dose-dependent (0.04, 0.2, 1, 5 µg/mL) activation of basophils from Patient 13. Shown were percentages of CD63 + basophils.

Crab ovary tissue is an important allergen source. Vitellogenin, hemocyanin, ovary development-related protein EJO1(Eri s 2) and ovary development-related protein EJO2 from the ovary of *Eriocheir sinensis* showed good specific IgE-binding activities. Recombinant Eri s 2 exhibited good reactivity to positive sera, and pre-incubation with recombinant Eri s 2 abrogated the reactivity of positive sera. Moreover, co-incubation of recombinant Eri s 2 with patient basophils dose-dependently promoted basophil activation.



## Tables

Table 1 Patient sera enrolled in the study

| Subjects               | Testing items             | Number of individuals | Gender: M/F | Age (years)  | IgE antibody levels to crab (range kU <sub>A</sub> /L) |
|------------------------|---------------------------|-----------------------|-------------|--------------|--------------------------------------------------------|
| crab-allergic patients | Dot blotting              | 36                    | 20/16       | 29.5 (1-68)  | 6.54(0.35-37)                                          |
|                        | Immunoblotting            | 6                     | 2/4         | 20.5 (12-35) | 16.27(1.2-37)                                          |
|                        | 2-DE immunoblotting       | 10                    | 5/5         | 23.5 (1-68)  | 6.84 (0.49-26)                                         |
|                        | ELISA                     | 29                    | 11/15       | 25.5 (1-68)  | 7.89(0.4-37)                                           |
|                        | ELISA Inhibition Assay    | 13                    | 6/7         | 22(1-43)     | 8.43(0.41-37)                                          |
| nonatopic individuals  | Inhibition Immunoblotting | 9                     | 5/4         | 25(1-36)     | 4.96(0.59-18.2)                                        |
|                        |                           | 6                     | 3/3         | 28.5 (3-57)  | < 0.1                                                  |

Table 2 *E.sinensis* proteins identified by MS/MS and Mascot database searches

| Spot | Protein      | Accession*   | Protein score | Nominal Mass | Queries matched |
|------|--------------|--------------|---------------|--------------|-----------------|
| 1    | vitellogenin | gi 507866656 | 385           | 288750       | 40              |
|      | ferritin 1   | gi 295291554 | 19            | 23648        | 5               |
|      | VASA protein | gi 305689985 | 17            | 68657        | 9               |
|      | ORF16        | gi 218156955 | 14            | 140441       | 11              |

|                                             |              |     |        |    |
|---------------------------------------------|--------------|-----|--------|----|
| ubiquitin-activating enzyme E1              | gi 388548526 | 13  | 115856 | 8  |
| cell division protein                       | gi 347597342 | 13  | 40158  | 5  |
| cell division protein                       | gi 347597340 | 13  | 40158  | 5  |
| cell division protein                       | gi 347597338 | 13  | 40158  | 5  |
| cell division protein                       | gi 347597336 | 13  | 40158  | 5  |
| cell division protein                       | gi 347597334 | 13  | 40158  | 5  |
| vitellogenin                                | gi 507866656 | 323 | 288750 | 37 |
| ubiquitin-activating enzyme E1              | gi 388548526 | 22  | 115856 | 9  |
| pacifastin                                  | gi 194246356 | 17  | 32074  | 4  |
| elongation factor 1-delta protein           | gi 375153528 | 16  | 29010  | 5  |
| VASA protein                                | gi 305689985 | 14  | 68657  | 7  |
| vitellogenin, partial                       | gi 59797477  | 13  | 18224  | 3  |
| farnesoic acid<br>O-methyltransferase       | gi 260060542 | 12  | 31730  | 4  |
| hemocyanin subunit 6                        | gi 334145788 | 435 | 78390  | 38 |
| cullin 4B                                   | gi 451354086 | 17  | 92663  | 11 |
| catalase                                    | gi 291170724 | 17  | 59279  | 7  |
| vacuolar ATP synthase subunit B,<br>partial | gi 6425067   | 16  | 14492  | 3  |
| elongation factor 1-delta protein           | gi 375153528 | 16  | 29010  | 6  |
| beta-actin                                  | gi 300247703 | 14  | 18396  | 3  |
| ubiquitin conjugating enzyme-1              | gi 294860888 | 14  | 11329  | 2  |
| VASA protein                                | gi 305689985 | 14  | 68657  | 11 |

|                                        |              |     |        |    |
|----------------------------------------|--------------|-----|--------|----|
| ovary development-related protein EJO2 | gi 37778440  | 267 | 28525  | 15 |
| ovary development-related protein EJO1 | gi 37778438  | 251 | 28575  | 14 |
| glutathione peroxidase                 | gi 257124473 | 20  | 24504  | 5  |
| peroxiredoxin 6                        | gi 209363473 | 20  | 24504  | 5  |
| RNA polymerase beta subunit            | gi 156628210 | 13  | 62752  | 7  |
| fatty acid binding protein             | gi 305690108 | 13  | 15557  | 4  |
| ecdysone receptor transcript variant S | gi 574486969 | 12  | 49128  | 4  |
| ecdysteroid receptor                   | gi 573462562 | 12  | 49100  | 4  |
| caspase                                | gi 304569876 | 12  | 43039  | 4  |
| glutamate dehydrogenase                | gi 347015079 | 11  | 62918  | 7  |
| ovary development-related protein EJO1 | gi 37778438  | 380 | 28575  | 15 |
| ovary development-related protein EJO2 | gi 37778440  | 219 | 28525  | 15 |
| vitellogenin, partial                  | gi 59797477  | 18  | 182244 | 4  |
| toll-like recepto                      | gi 283580031 | 14  | 20399  | 3  |
| vitellogenin                           | gi 507866656 | 13  | 288750 | 21 |
| metallothionein                        | gi 16555374  | 12  | 7209   | 2  |
| Rab-1                                  | gi 440548634 | 12  | 22899  | 3  |
| RNA polymerase beta subunit            | gi 156628210 | 11  | 62752  | 6  |
| glutamate dehydrogenase                | gi 347015079 | 10  | 62918  | 6  |

metallothionein-1

gi|317487564

10

7048

2

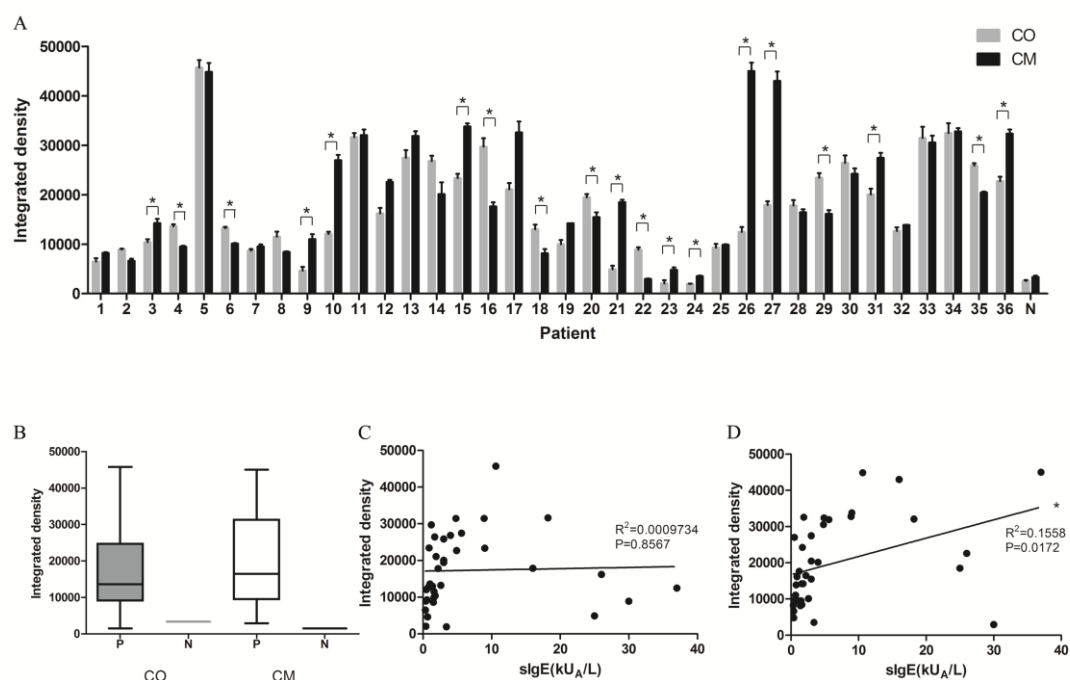


Figure 1

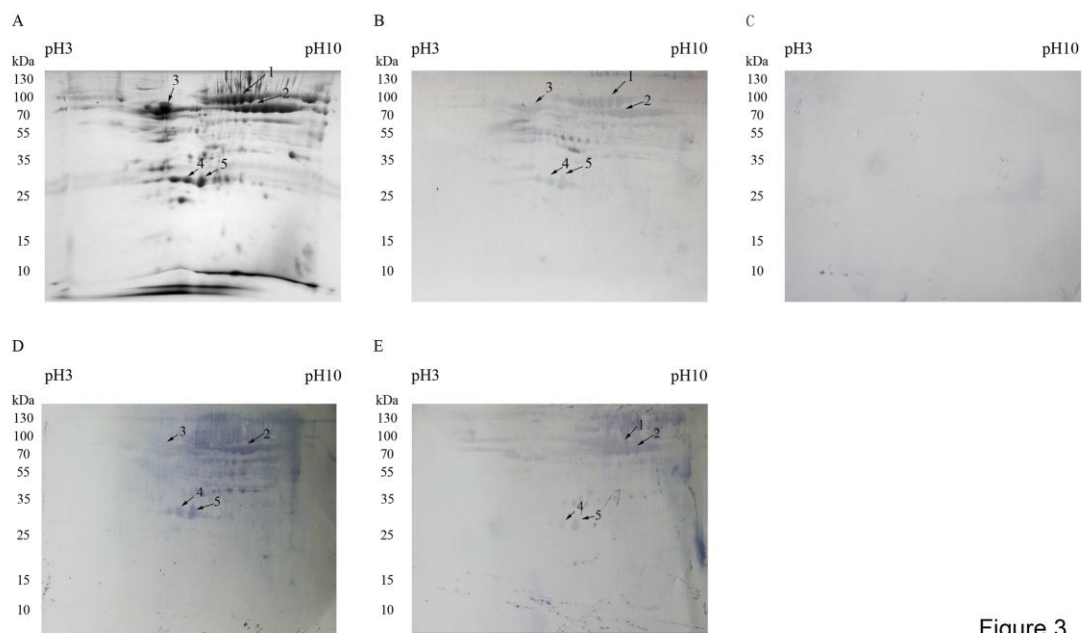
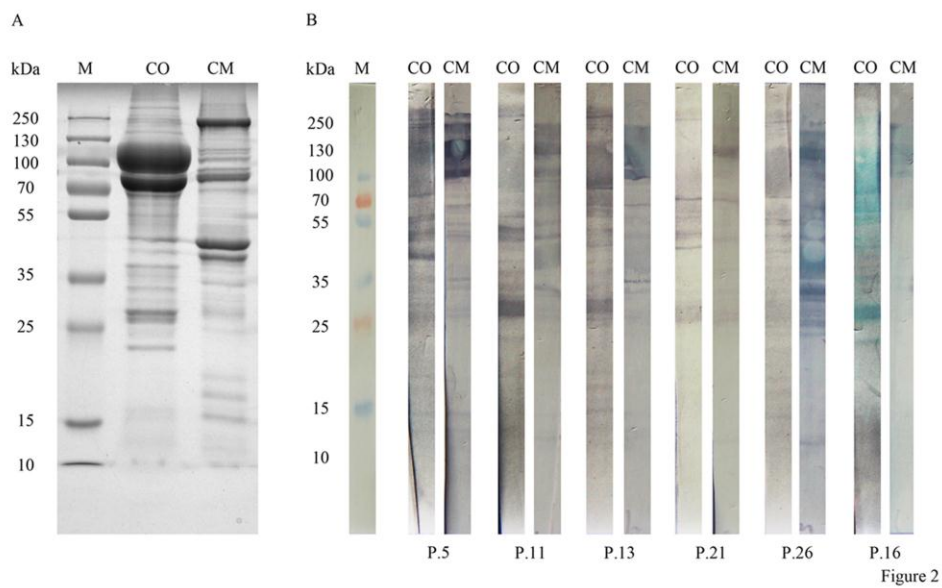


Figure 3

Translation of SEQUENCE(1-759)

Universal code

Total amino acid number: 252, MW=28144

Max ORF: 1-756, 252 AA, MW=28144

```

1      ATGTGGTGCAGTATCTTGCTGGCGGCAGTGATGGTGGCGGTGGCGGGCGGCCAGAGTCGA
1      M W C S I L L A A V M V A V A G G Q S R

61     TTTAGTTATGATGGGACGCCCTGGCTGACGACGTTGTAGTTGGGTGAGGACAGCGTCAG
21     F S Y D G T P L A D D V V V G S G Q R Q

121    CAGTTCTTTGCGTCCCCGACAGCCCCGCCACAACAGCTGTTTGTACAGGAGCGGCCCCAG
41     Q F F A S R Q P R P Q Q L F V Q E R P Q

181    CAGGTCTTCGTCCAGCAGCAGCCTCAGCCACCGGTGTTTGTGCAGCCTGAGCCACAGCAG
61     Q V F V Q Q Q P Q P P V F V Q P E P Q Q

241    GTGTTTGTGCAGCCCCAGCCACGCCCCATGTTCCAGATGCTCCCCTCGTTCTCACAGCAG
81     V F V Q P Q P R P M F Q M L P S F S Q Q

301    GCGGCGGTGATCCAGAGCGTCCAGCCGCAAGTGTCTTCCGCCTGTCCGGCCACTTACTCC
101    A A V I Q S V Q P Q V S S A C P A T Y S

361    CTGGTCCACACCACCTACCAAGGCCGAGTCTACCACTCCTCCTGGTGCAACCTGCCCGGC
121    L V H T T Y Q G R V Y H S S W C N L P G

421    CACCGTTTCTCCAGTCCGCCGCCAGAGACTACTGCCGCGGCCTCGGCGAGGTAACAGTG
141    H R F S Q S A A R D Y C R G L G E V T V

481    ATGAACCGACGCGGCAGCTTTGACCTATACAGCGTCGATCGCACTACTGATTTTCGATTAC
161    M N R R G S F D L Y S V D R T T D F D Y

541    TTCTTCGAGCTGCTCCTCCAGCACCAAAATCCCCTCCGTCTGGACCAGGGACCTTGCTCC
181    F F E L L L Q H Q I P S V W T R D L A S

601    ACCGCGCCCTATGAAATCCGCGGCTCCATCACCAGGTCCGGGGACTCTACGGCTGGCCAA
201    T A P Y E I R G S I T R S G D S T A G Q

661    GACTGCCTCTCGCTGGAGGCTTCCACAGACGGCCTCTTCTCCTCCAGGACTCTTGCTCC
221    D C L S L E A S T D G L F L L Q D S C S

721    GACCGGAAGGCCCGCTCTGCGTGGCTCATTGGCAGTAG
241    D R K A A V C V A H W Q *

```

Figure 4

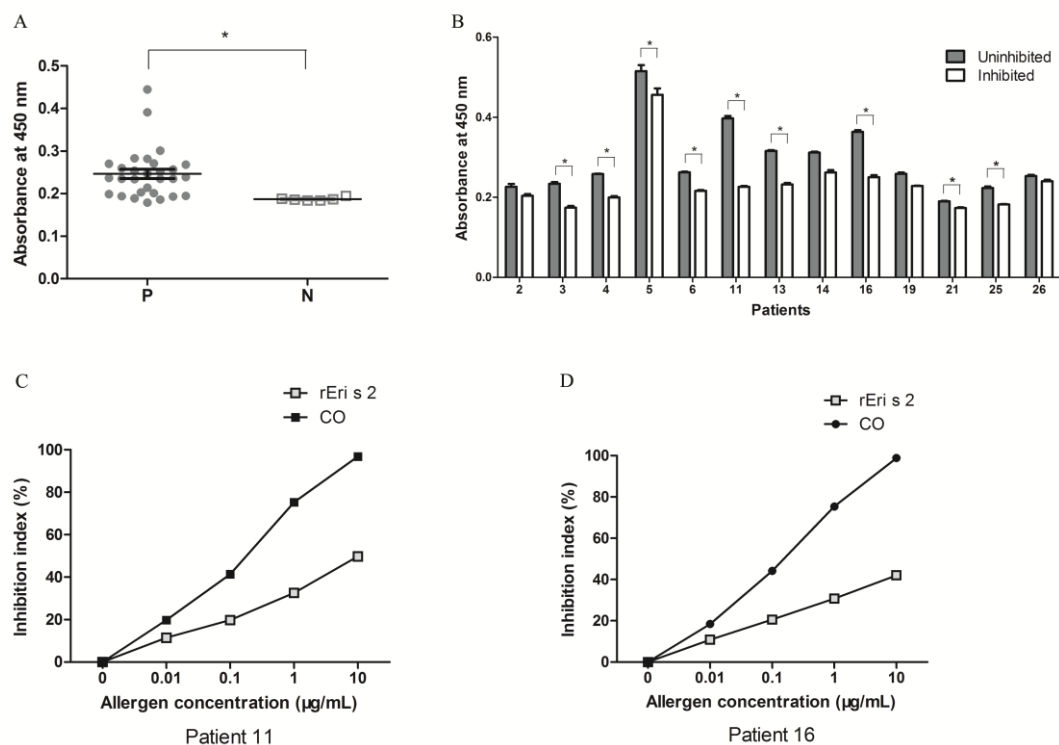


Figure 5

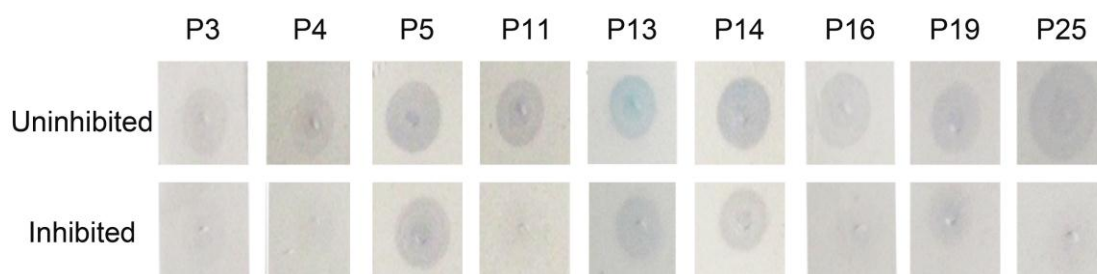


Figure 6

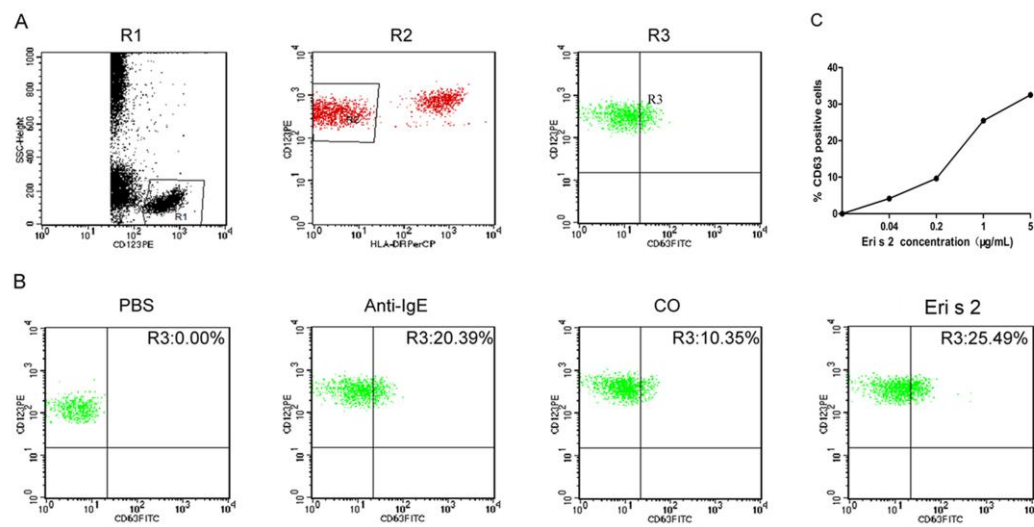


Figure 7

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.**

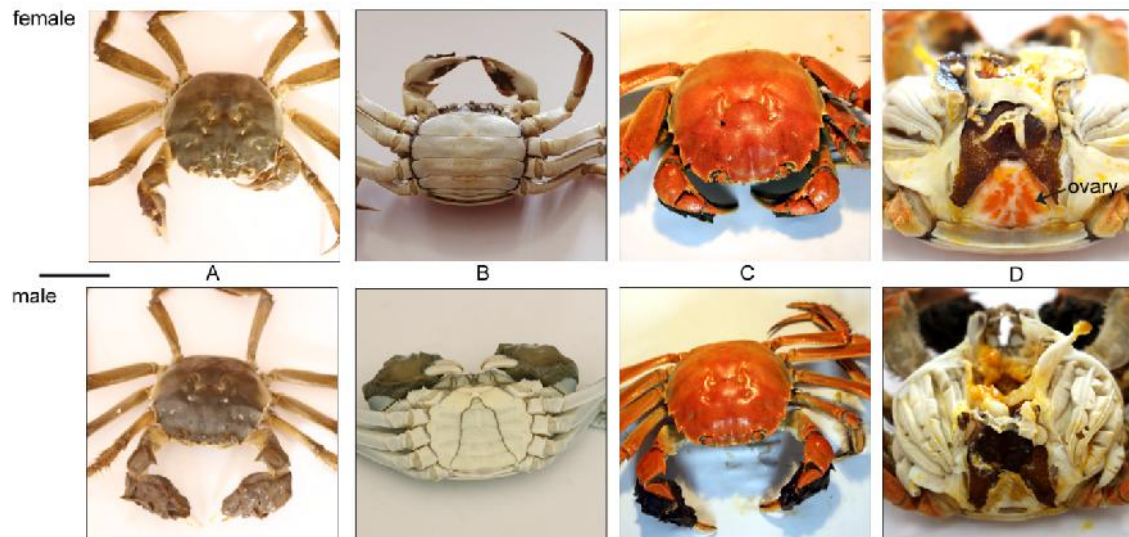


Figure S1

**Figure S1. Male and female *E. sinensis*.** A female crab is shown in the upper panel, and a male crab is shown in the lower panel. (A) dorsal and (B) ventral (B) views of live *E. sinensis*. Dorsal views of boiled *E. sinensis* before (C) and after (D) removal of the shell. The arrow in (D) indicates the crab ovary.

Figure S2

|             |                                            |     |
|-------------|--------------------------------------------|-----|
| gi 37778437 | CAGAGTATCTGGAGTCTTGGACCTTAACAAACGGACGAGA   | 40  |
| gi 37778439 | CAGAGTATCTGGAGTCTTGGACCTTAACAAACGGACGAGA   | 40  |
| Consensus   | agagtatctggagtctt gac t aaca acggacgaga    |     |
| gi 37778437 | TGTGGTGCAGTATCTTGTCTGGCGGCAGTGATGGTGGCGGT  | 80  |
| gi 37778439 | TGTGGTGCAGTATCTTGTCTGGCGGCAGTGATGGTGGCGGC  | 80  |
| Consensus   | tgtggtgcagtatcttgtctggcggcagtgatggtggcgg   |     |
| gi 37778437 | GGCGGGCGGCCAGAGTCGATTAGTTATGAGGGACGCC      | 120 |
| gi 37778439 | GGCGGGCGGCCAGAGTCGATTAGTTATGAGGGACGCC      | 120 |
| Consensus   | ggcgggcgggccagagtcgatttagttatga gggacgccc  |     |
| gi 37778437 | CTGCTGACGACGTGTGTGTTGGGTCAGGACAGCGTCAGC    | 160 |
| gi 37778439 | CTGCTGACGACGTGTGTGTTGGGTCAGGACAGCGTCAGC    | 160 |
| Consensus   | ct g tgacgacgt gt gttgggtcaggacagcgtcagc   |     |
| gi 37778437 | AGTTCTTTGCGTCCCGACAGCCCCGGCCACAACAGCTGTT   | 200 |
| gi 37778439 | AGTTCTTTGCGTCCCGACAGCCCCGGCCACAACAGCTGTT   | 200 |
| Consensus   | agttctttgcgccccacagccccggccacaacagctgtt    |     |
| gi 37778437 | TGTACAGGAGCGGCCCCAGCAGGCTTCGTCACGACGAG     | 240 |
| gi 37778439 | TGTACAGGAGCGGCCCCAGCAGGCTTCGTCACGACGAG     | 240 |
| Consensus   | tgtacaggagcggccccagcaggtcttcgtccagcagcag   |     |
| gi 37778437 | CCTCAGCCACCGGTGTTTGTTCAGCCTGAGCCACAGCAGG   | 280 |
| gi 37778439 | CCTCAGCCACCGGTGTTTGTTCAGCCTGAGCCACAGCAGG   | 280 |
| Consensus   | cctcagccac ggtgtttgt cagcctgagccacagcagg   |     |
| gi 37778437 | TGTTTGTGACGCCCCAGCCACGCCCCATGTTCCAGATGCT   | 320 |
| gi 37778439 | TGTTTGTGACGCCCCAGCCACGCCCCATGTTCCAGATGCT   | 320 |
| Consensus   | tgtttgtgcagccccagccacgccccatgttccagatgct   |     |
| gi 37778437 | CCCCTCGTTCTCACAGCAGGCGGCGGTGATCCAGAGCGTC   | 360 |
| gi 37778439 | CCCCTCGTTCTCACAGCAGGCGGCGGTGATCCAGAGCGTC   | 360 |
| Consensus   | cccctcgttctcacagcaggcggcggtgatccagagcgtc   |     |
| gi 37778437 | CAGCGCAAGTGTCTTCCGCCTGTCCGGCCACTTACTCCC    | 400 |
| gi 37778439 | CAGCGCAAGTGTCTTCCGCCTGTCCGGCCACTTACTCCC    | 400 |
| Consensus   | cagccgcaagtgtcttccgcctgtccggccacttactccc   |     |
| gi 37778437 | TGGTCCACACCACCTACCAAGGCGGAGTCTACCCTCTC     | 440 |
| gi 37778439 | TGGTCCACACCACCTACCAAGGCGGAGTCTACCCTCTC     | 440 |
| Consensus   | tgggtccacaccacctacca ggccgagctctaccact ct  |     |
| gi 37778437 | CTGGTGCACCTGCCCGGCCACCGTTTCTCCAGTCCGCC     | 480 |
| gi 37778439 | CTGGTGCACCTGCCCGGCCACCGTTTCTCCAGTCCGCC     | 480 |
| Consensus   | ctggtgcaacctgcccggccac gtttctccagtcggcc    |     |
| gi 37778437 | GCCAGAGACTACTGCCCGGGCTCGGCGAGGTAACAGTGA    | 520 |
| gi 37778439 | GCCAGAGACTACTGCCCGGGCTCGGCGAGGTAACAGTGA    | 520 |
| Consensus   | gccagagactactgcccgggcctcggcgaggtaaacagtga  |     |
| gi 37778437 | TGAACCGACGCGGCAGCTTTGACCTATACAGCGTCGATCG   | 560 |
| gi 37778439 | TGAACCGACGCGGCAGCTTTGACCTATACAGCGTCGATCG   | 560 |
| Consensus   | tgaaccgacgcggcagctttgacctatacagcgtcgatcg   |     |
| gi 37778437 | CACTACTGATTTCGATTACTTCTTCGAGCTGCTCCTCCAG   | 600 |
| gi 37778439 | CACTACTGATTTCGATTACTTCTTCGAGCTGCTCCTCCAG   | 600 |
| Consensus   | cactactgatttcgattacttcttcgagctgctcctccag   |     |
| gi 37778437 | CACCAAAATCCCCTCCGTCTGGACGAGGACCTTGCTCCA    | 640 |
| gi 37778439 | CACCAAAATCCCCTCCGTCTGGACGAGGACCTTGCTCCA    | 640 |
| Consensus   | caccaaatacccctccgtctggaccagggaaccttgctcca  |     |
| gi 37778437 | CGCGCCCTATGAAATCCGCGGCTCCATCACCAGGTCCGG    | 680 |
| gi 37778439 | CGCGCCCTATGAAATCCGCGGCTCCATCACCAGGTCCGG    | 680 |
| Consensus   | c gcgcccctatgaaatccgcggtccatcaccaggtccgg   |     |
| gi 37778437 | GGACTCTACGGCTGGCCAAGACTGCCTCTCGCTGGAGGCT   | 720 |
| gi 37778439 | GGACTCTACGGCTGGCCAAGACTGCCTCTCGCTGGAGGCT   | 720 |
| Consensus   | ggactctacggc ggccaagactgcctctcgctggaggct   |     |
| gi 37778437 | TCCACAGACGGCCTCTTCTCCTCCAGGACTCTTGCTCCG    | 760 |
| gi 37778439 | TCCACAGACGGCCTCTTCTCCTCCAGGACTCTTGCTCCG    | 760 |
| Consensus   | tccacagacggcctcttctcctcctccaggactcttgctccg |     |
| gi 37778437 | ACCGGAAGGCCGCCGCTGCGTGGCTCATTGGCAGTAGAT    | 800 |
| gi 37778439 | ACCGGAAGGCCGCCGCTGCGTGGCTCATTGGCAGTAGAT    | 800 |
| Consensus   | accggaaggccgccgctgctgctggtcattggcagtagat   |     |
| gi 37778437 | GAGTTTGTCAAAACCTGCTTATCATTGTGTAATAAATG     | 840 |
| gi 37778439 | GAGTTTGTCAAAACCTGCTTATCATTGTGTAATAAATG     | 840 |
| Consensus   | agttttgtcaa aacctgcttatcattgtgtaataaatg    |     |
| gi 37778437 | TGCTCCTACAAAAAAAAAAAAAAAAAAAAAAAAAAAA      | 875 |
| gi 37778439 | TGCTCCTACAAAAAAAAAAAAAAAAAAAAAAAAAAAA      | 875 |
| Consensus   | tgctcct aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa     |     |

Figure S2. A comparison shows the similarity between the EJO1 and EJO2 sequences. Dark blue indicates identical bases, while bases that differed are unshaded or light blue. GI:37778437: EJO1; GI:37778439: EJO2.

**Figure S3.**

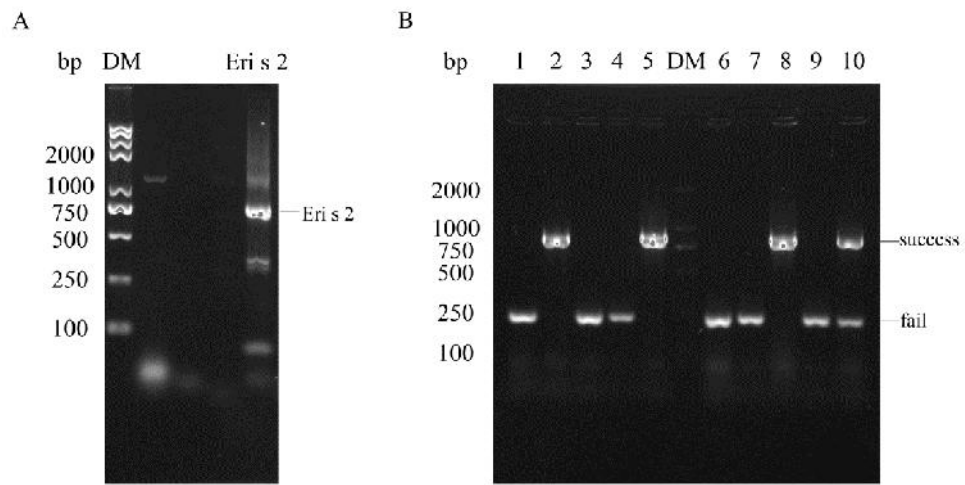


**Figure S3**

**Figure S3.** Verification of the sequence of the recombinant expression vector pET-28b(+)-Eri s 2. (A) PCR amplification of Eri s 2 sequences from *E. sinensis*. DM, DNA markers. (B) Identification of Eri s 2-positive colonies by colony PCR. DM, DNA markers; 1–10, 10 individual bacterial colonies. Success, positive colony, recombinant expression vector pET-28b(+)-Eri s 2; fail, negative colony.

**Figure S4.**

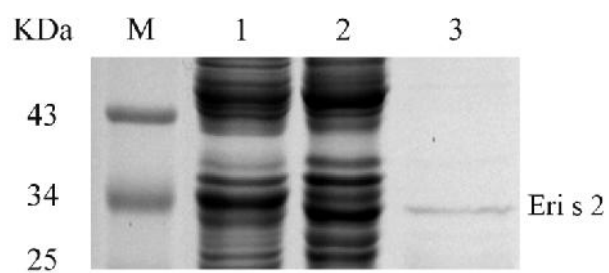


**Figure S4**

**Figure S4.** IPTG-induced expression of Eri s 2 in *E. coli* Rosetta cells was evaluated by SDS-PAGE. M, molecular weight markers; 1, whole cell lysate without IPTG induction(50μg); 2, whole cell lysate with IPTG induction(50μg); 3, purified rEri s 2(1μg).

**Table S1 crab-allergic patients enrolled in the study**

| No. | Gender | Age (years) | sIgE antibody levels to crab (range kUA/l) |
|-----|--------|-------------|--------------------------------------------|
| 1   | F      | 40          | 0.35                                       |
| 2   | M      | 1           | 0.41                                       |
| 3   | M      | 25          | 1.8                                        |
| 4   | M      | 1           | 1                                          |
| 5   | F      | 19          | 10.6                                       |
| 6   | F      | 43          | 2.6                                        |
| 7   | F      | 68          | 1.5                                        |
| 8   | M      | 54          | 1.6                                        |
| 9   | M      | 2           | 0.67                                       |
| 10  | M      | 4           | 0.49                                       |
| 11  | F      | 22          | 18.2                                       |
| 12  | M      | 8           | 26                                         |
| 13  | F      | 36          | 5.61                                       |
| 14  | F      | 37          | 4                                          |
| 15  | M      | 42          | 9                                          |
| 16  | M      | 35          | 1.2                                        |
| 17  | F      | 68          | 1.9                                        |
| 18  | F      | 60          | 1.4                                        |
| 19  | M      | 26          | 1.6                                        |
| 20  | M      | 16          | 3                                          |
| 21  | F      | 18          | 25                                         |
| 22  | F      | 27          | 30                                         |
| 23  | M      | 60          | 0.4                                        |
| 24  | M      | 42          | 3.34                                       |
| 25  | M      | 9           | 0.59                                       |
| 26  | M      | 12          | 37                                         |
| 27  | F      | 1           | 16                                         |
| 28  | F      | 3           | 2.2                                        |
| 29  | M      | 34          | 0.89                                       |
| 30  | M      | 37          | 1.7                                        |
| 31  | M      | 47          | 3                                          |
| 32  | F      | 54          | 0.76                                       |
| 33  | F      | 32          | 4.8                                        |
| 34  | F      | 41          | 8.9                                        |
| 35  | M      | 2           | 3                                          |
| 36  | M      | 5           | 4.9                                        |

**Table S2 Primer sequences used for cloning of Eri s 2**

| Primer Name     | Primer Sequence                           |
|-----------------|-------------------------------------------|
| Eri s 2 forward | 5'-GGGAATTCCATATGTGGTGCAGTATCTTGCTGGCG-3' |
| Eri s 2 reverse | 5'-CCGCTCGAGTCTGCCAATGAGCCACGCAGA-3'      |
| Colony forward  | 5'-TAA TAC GAC TCA CTA TAG GG-3'          |
| Colony reverse  | 5'-TGCTAGTTATTGCTCAGCGG-3'                |

**Table S3 Amino acid sequences of several peptide fragments of spot by MS/MS**

|   | Observed  | Mr(expt)  | Mr(calculated) | Delta   | Start | End  | Miss | Ions | Peptide                          |   |
|---|-----------|-----------|----------------|---------|-------|------|------|------|----------------------------------|---|
| 1 | 802.5123  | 801.5050  | 801.474        | 0.0302  | 335   | 341  | 0    | -    | R.IPLQGFK.E                      |   |
|   | 930.5702  | 929.5629  | 929.5334       | 0.0295  | 1507  | 1514 | 0    | -    | K.VEAPFLVR.Q                     |   |
|   | 932.5359  | 931.5286  | 931.4909       | 0.0377  | 430   | 438  | 0    | -    | K.LAAASMVNR.Y                    |   |
|   | 958.6184  | 957.6111  | 957.5759       | 0.0352  | 334   | 341  | 1    | 18   | R.RIPLQGFK.E                     |   |
|   | 958.6184  | 957.6111  | 957.5759       | 0.0352  | 334   | 341  | 1    | -    | R.RIPLQGFK.E                     |   |
|   | 978.6046  | 977.5973  | 977.5658       | 0.0315  | 275   | 283  | 0    | -    | R.ISQGHLVPK.R                    |   |
|   | 1002.5417 | 1001.5344 | 1001.5909      | -0.0565 | 1334  | 1342 | 0    | -    | K.LQGPVTFIK.T                    |   |
|   | 1016.5716 | 1015.5643 | 1015.5298      | 0.0346  | 61    | 70   | 0    | -    | K.GVEGGLVETR.W                   |   |
|   | 1023.5992 | 1022.5919 | 1022.5470      | 0.0449  | 791   | 798  | 0    | -    | R.YLIENLMK.M                     |   |
|   | 1080.5597 | 1079.5524 | 1079.5135      | 0.0389  | 420   | 429  | 0    | -    | R.ADEPYLSSAK.L                   |   |
|   | 1102.6327 | 1101.6254 | 1101.5818      | 0.0436  | 2050  | 2059 | 0    | -    | K.GVDFPAVEIR.K                   |   |
|   | 1141.6188 | 1140.6115 | 1140.5703      | 0.0413  | 285   | 293  | 0    | -    | R.LLYDYETPK.K                    |   |
|   | 1142.6060 | 1141.5987 | 1141.5114      | 0.0873  | 1669  | 1677 | 0    | -    | K.TWDIAFDMK.A<br>Oxidation (M)   | + |
|   | 1153.6757 | 1152.6684 | 1152.6291      | 0.0393  | 240   | 249  | 0    | -    | K.VVRPAYGTYK.F                   |   |
|   | 1173.6683 | 1172.6610 | 1172.5206      | 0.1405  | 1372  | 1380 | 1    | -    | K.DSCFMELKK.Q<br>Oxidation (M)   | + |
|   | 1190.5696 | 1189.5623 | 1189.7182      | -0.1559 | 1062  | 1071 | 0    | -    | K.VLLINKPEHK.S                   |   |
|   | 1221.6718 | 1220.6645 | 1220.6223      | 0.0422  | 109   | 119  | 0    | 103  | K.YPLVVAMTDGR.V                  |   |
|   | 1221.6718 | 1220.6645 | 1220.6223      | 0.0422  | 109   | 119  | 0    | -    | K.YPLVVAMTDGR.V                  |   |
|   | 1237.6655 | 1236.6582 | 1236.6172      | 0.0410  | 109   | 119  | 0    | -    | K.YPLVVAMTDGR.V<br>Oxidation (M) | + |
|   | 1297.7231 | 1296.7158 | 1296.6714      | 0.0445  | 284   | 293  | 1    | 20   | K.RLLYDYETPK.K                   |   |

|       |       |       |       |      |    |    |     |                                            |              |
|-------|-------|-------|-------|------|----|----|-----|--------------------------------------------|--------------|
| 1297. | 1296. | 1296. | 0.04  | 28   | 29 | 1  | -   | K.RLLYDYETPK.K                             |              |
| 7231  | 7158  | 6714  | 45    | 4    | 3  |    |     |                                            |              |
| 1308. | 1307. | 1307. | 0.04  | 25   | 26 | 0  | 43  | K.FVEAEQESTLR.L                            |              |
| 6865  | 6792  | 6357  | 35    | 0    | 0  |    |     |                                            |              |
| 1308. | 1307. | 1307. | 0.04  | 25   | 26 | 0  | -   | K.FVEAEQESTLR.L                            |              |
| 6865  | 6792  | 6357  | 35    | 0    | 0  |    |     |                                            |              |
| 1324. | 1323. | 1323. | 0.04  | 11   | 11 | 0  | -   | K.IFTAELINTFR.A                            |              |
| 7682  | 7609  | 7186  | 23    | 23   | 33 |    |     |                                            |              |
| 1342. | 1341. | 1341. | 0.04  | 22   | 22 | 0  | -   | R.IIIAPEPGTYR.V                            |              |
| 8138  | 8065  | 7656  | 09    | 54   | 65 |    |     |                                            |              |
| 1425. | 1424. | 1424. | 0.07  | 20   | 20 | 0  | -   | K.IDFEYKPEQTR.A                            |              |
| 7749  | 7676  | 6936  | 41    | 04   | 14 |    |     |                                            |              |
| 1441. | 1440. | 1440. | 0.15  | 20   | 20 | 1  | -   | R.AMKDFLMSEVDR.V                           |              |
| 8331  | 8258  | 6741  | 18    | 15   | 26 |    |     |                                            |              |
| 1441. | 1440. | 1440. | 0.15  | 20   | 20 | 1  | -   | R.AMKDFLMSEVDR.V                           |              |
| 8331  | 8258  | 6741  | 18    | 15   | 26 |    |     |                                            |              |
| 1486. | 1485. | 1485. | 0.03  | 72   | 73 | 1  | -   | K.IQEELDVSIRER.R                           |              |
| 8171  | 8098  | 7787  | 12    | 0    | 1  |    |     |                                            |              |
| 1509. | 1508. | 1508. | 0.04  | 29   | 30 | 0  | -   | K.EPSLVPQLEQTLR.H                          |              |
| 8679  | 8606  | 8198  | 08    | 5    | 7  |    |     |                                            |              |
| 1637. | 1636. | 1636. | 0.04  | 29   | 30 | 1  | -   | K.KEPSLVPQLEQTLR.H                         |              |
| 9642  | 9569  | 9148  | 21    | 4    | 7  |    |     |                                            |              |
| 1642. | 1641. | 1641. | 0.10  | 11   | 12 | 1  | -   | K.TEESGQNGKYSPAFK.<br>V                    |              |
| 8794  | 8721  | 7634  | 87    | 96   | 10 |    |     |                                            |              |
| 1784. | 1783. | 1783. | 0.06  | 21   | 21 | 0  | -   | R.YTINEIYQAADQISR.W                        |              |
| 9470  | 9397  | 8740  | 57    | 35   | 49 |    |     |                                            |              |
| 1788. | 1787. | 1787. | 0.06  | 12   | 13 | 0  | -   | R.VCSHPDDDTWAINMK.<br>K                    |              |
| 8324  | 8251  | 7607  | 44    | 3    | 7  |    |     |                                            |              |
| 1837. | 1836. | 1836. | 0.05  | 21   | 21 | 0  | -   | R.EAFYQTLADMEHVQR.<br>E                    |              |
| 9136  | 9063  | 8464  | 99    | 02   | 16 |    |     |                                            |              |
| 1942. | 1941. | 1940. | 0.06  | 48   | 50 | 0  | 137 | K.SLGNMGVMTHDVAE<br>VLIR.Y                 |              |
| 0580  | 0507  | 9811  | 96    | 5    | 2  |    |     |                                            |              |
| 1942. | 1941. | 1940. | 0.06  | 48   | 50 | 0  | -   | K.SLGNMGVMTHDVAE<br>VLIR.Y                 |              |
| 0580  | 0507  | 9811  | 96    | 5    | 2  |    |     |                                            |              |
| 1958. | 1957. | 1956. | 0.06  | 48   | 50 | 0  | -   | K.SLGNMGVMTHDVAE<br>VLIR.Y + Oxidation (M) |              |
| 0520  | 0447  | 9760  | 87    | 5    | 2  |    |     |                                            |              |
| 2284. | 2283. | 2283. | 0.08  | 30   | 32 | 1  | -   | R.HLCDITKDEIQADAAV<br>ELDK.A               |              |
| 1926  | 1853  | 1052  | 01    | 8    | 7  |    |     |                                            |              |
| 2380. | 2379. | 2379. | 0.07  | 36   | 38 | 0  | -   | K.SLYFDAIAFVHEPESVI<br>IMAK.E              |              |
| 2988  | 2915  | 2184  | 32    | 1    | 1  |    |     |                                            |              |
| 2     | 802.5 | 801.5 | 801.4 | 0.03 | 33 | 34 | 0   | -                                          | R.IPLQGFK.E  |
|       | 182   | 109   | 748   | 61   | 5  | 1  |     |                                            |              |
|       | 958.6 | 957.6 | 957.5 | 0.04 | 33 | 34 | 1   | 20                                         | R.RIPLQGFK.E |
|       | 259   | 186   | 759   | 27   | 4  | 1  |     |                                            |              |

|       |       |       |      |    |    |   |     |                    |   |
|-------|-------|-------|------|----|----|---|-----|--------------------|---|
| 958.6 | 957.6 | 957.5 | 0.04 | 33 | 34 | 1 | -   | R.RIPLQGFK.E       |   |
| 259   | 186   | 759   | 27   | 4  | 1  |   |     |                    |   |
| 978.6 | 977.6 | 977.5 | 0.04 | 27 | 28 | 0 | -   | R.ISQGHLVPK.R      |   |
| 133   | 060   | 658   | 02   | 5  | 3  |   |     |                    |   |
| 991.6 | 990.6 | 990.5 | 0.04 | 54 | 55 | 0 | -   | R.IAAYLTALR.C      |   |
| 354   | 281   | 862   | 19   | 9  | 7  |   |     |                    |   |
| 1002. | 1001. | 1001. | -0.0 | 13 | 13 | 0 | -   | K.LQGPVTFIK.T      |   |
| 5464  | 5391  | 5909  | 518  | 34 | 42 |   |     |                    |   |
| 1016. | 1015. | 1015. | 0.04 | 61 | 70 | 0 | -   | K.GVEGGLVETR.W     |   |
| 5804  | 5731  | 5298  | 34   |    |    |   |     |                    |   |
| 1101. | 1100. | 1100. | 0.04 | 53 | 54 | 0 | -   | K.LVDFAVRPGK.N     |   |
| 6819  | 6746  | 6342  | 05   | 4  | 3  |   |     |                    |   |
| 1141. | 1140. | 1140. | 0.04 | 28 | 29 | 0 | -   | R.LLYDYETPK.K      |   |
| 6245  | 6172  | 5703  | 70   | 5  | 3  |   |     |                    |   |
| 1142. | 1141. | 1141. | 0.09 | 16 | 16 | 0 | -   | K.TWDIAFDMK.A      | + |
| 6163  | 6090  | 5114  | 76   | 69 | 77 |   |     | Oxidation (M)      |   |
| 1153. | 1152. | 1152. | 0.04 | 24 | 24 | 0 | 19  | K.VVRPAYGTYK.F     |   |
| 6858  | 6785  | 6291  | 94   | 0  | 9  |   |     |                    |   |
| 1153. | 1152. | 1152. | 0.04 | 24 | 24 | 0 | -   | K.VVRPAYGTYK.F     |   |
| 6858  | 6785  | 6291  | 94   | 0  | 9  |   |     |                    |   |
| 1173. | 1172. | 1172. | 0.15 | 13 | 13 | 1 | -   | K.DSCFMELKK.Q      | + |
| 6781  | 6708  | 5206  | 03   | 72 | 80 |   |     | Oxidation (M)      |   |
| 1221. | 1220. | 1220. | 0.05 | 10 | 11 | 0 | 102 | K.YPLVVAMTDGR.V    |   |
| 6807  | 6734  | 6223  | 11   | 9  | 9  |   |     |                    |   |
| 1221. | 1220. | 1220. | 0.05 | 10 | 11 | 0 | -   | K.YPLVVAMTDGR.V    |   |
| 6807  | 6734  | 6223  | 11   | 9  | 9  |   |     |                    |   |
| 1237. | 1236. | 1236. | 0.04 | 10 | 11 | 0 | -   | K.YPLVVAMTDGR.V    | + |
| 6729  | 6656  | 6172  | 84   | 9  | 9  |   |     | Oxidation (M)      |   |
| 1297. | 1296. | 1296. | 0.05 | 28 | 29 | 1 | 20  | K.RLLYDYETPK.K     |   |
| 7323  | 7250  | 6714  | 37   | 4  | 3  |   |     |                    |   |
| 1297. | 1296. | 1296. | 0.05 | 28 | 29 | 1 | -   | K.RLLYDYETPK.K     |   |
| 7323  | 7250  | 6714  | 37   | 4  | 3  |   |     |                    |   |
| 1308. | 1307. | 1307. | 0.05 | 25 | 26 | 0 | 46  | K.FVEAEQUESTLR.L   |   |
| 6952  | 6879  | 6357  | 22   | 0  | 0  |   |     |                    |   |
| 1308. | 1307. | 1307. | 0.05 | 25 | 26 | 0 | -   | K.FVEAEQUESTLR.L   |   |
| 6952  | 6879  | 6357  | 22   | 0  | 0  |   |     |                    |   |
| 1425. | 1424. | 1424. | 0.09 | 20 | 20 | 0 | -   | K.IDFEYKPEQTR.A    |   |
| 7997  | 7924  | 6936  | 89   | 04 | 14 |   |     |                    |   |
| 1441. | 1440. | 1440. | 0.15 | 20 | 20 | 1 | -   | R.AMKDFLMSEVDR.V   |   |
| 8372  | 8299  | 6741  | 59   | 15 | 26 |   |     |                    |   |
| 1457. | 1456. | 1456. | 0.04 | 86 | 88 | 0 | -   | K.AHSFIGYDAYVSK.S  |   |
| 7511  | 7438  | 6986  | 52   | 8  | 0  |   |     |                    |   |
| 1486. | 1485. | 1485. | 0.05 | 26 | 27 | 0 | 42  | R.LTSSDVPTPDIVSR.I |   |
| 8329  | 8256  | 7675  | 82   | 1  | 4  |   |     |                    |   |

|       |       |       |       |      |    |    |      |                        |             |   |
|-------|-------|-------|-------|------|----|----|------|------------------------|-------------|---|
| 1486. | 1485. | 1485. | 0.04  | 72   | 73 | 1  | -    | K.IQEELDV              | SIRER.R     |   |
| 8329  | 8256  | 7787  | 70    | 0    | 1  |    |      |                        |             |   |
| 1509. | 1508. | 1508. | 0.05  | 29   | 30 | 0  | -    | K.EPSLVPQLEQTLR.H      |             |   |
| 8833  | 8760  | 8198  | 62    | 5    | 7  |    |      |                        |             |   |
| 1637. | 1636. | 1636. | 0.05  | 29   | 30 | 1  | 13   | K.KEPSLVPQLEQTLR.H     |             |   |
| 9744  | 9671  | 9148  | 23    | 4    | 7  |    |      |                        |             |   |
| 1637. | 1636. | 1636. | 0.05  | 29   | 30 | 1  | -    | K.KEPSLVPQLEQTLR.H     |             |   |
| 9744  | 9671  | 9148  | 23    | 4    | 7  |    |      |                        |             |   |
| 1644. | 1643. | 1643. | 0.05  | 60   | 61 | 0  | -    | R.SLLTNILIPDNFER.D     |             |   |
| 9456  | 9383  | 8882  | 01    | 3    | 6  |    |      |                        |             |   |
| 1680. | 1679. | 1679. | 0.00  | 12   | 12 | 1  | -    | R.GRPAVD               | FEAHIAAKAK  |   |
| 9198  | 9125  | 9107  | 19    | 29   | 44 |    |      | .F                     |             |   |
| 1788. | 1787. | 1787. | 0.06  | 12   | 13 | 0  | -    | R.VCSHPDDDDTWAINMK.    |             |   |
| 8378  | 8305  | 7607  | 98    | 3    | 7  |    |      | K                      |             |   |
| 1804. | 1803. | 1803. | 0.06  | 12   | 13 | 0  | -    | R.VCSHPDDDDTWAINMK.    |             |   |
| 8302  | 8229  | 7556  | 73    | 3    | 7  |    |      | K + Oxidation (M)      |             |   |
| 1942. | 1941. | 1940. | 0.06  | 48   | 50 | 0  | -    | K.SLGNGM               | GVMTDVAE    |   |
| 0562  | 0489  | 9811  | 78    | 5    | 2  |    |      | VLIR.Y                 |             |   |
| 2284. | 2283. | 2283. | 0.08  | 30   | 32 | 1  | -    | R.HLCDITKDEIQADAAV     |             |   |
| 2004  | 1931  | 1052  | 79    | 8    | 7  |    |      | ELDK.A                 |             |   |
| 2332. | 2331. | 2331. | 0.12  | 14   | 14 | 1  | -    | K.IAVDATIMPESGSQTRS    |             |   |
| 3062  | 2989  | 1740  | 50    | 72   | 93 |    |      | NIDVK.V                |             |   |
| 2380. | 2379. | 2379. | 0.08  | 36   | 38 | 0  | -    | K.SLYFDAIAFVHEPESVI    |             |   |
| 3083  | 3010  | 2184  | 27    | 1    | 1  |    |      | IMAK.E                 |             |   |
| 2396. | 2395. | 2395. | 0.08  | 36   | 38 | 0  | -    | K.SLYFDAIAFVHEPESVI    |             |   |
| 3040  | 2967  | 2133  | 34    | 1    | 1  |    |      | IMAK.E + Oxidation (M) |             |   |
| 802.5 | 801.5 | 801.4 | 0.03  | 33   | 34 | 0  | -    | R.IPLQGFK.E            |             |   |
| 182   | 109   | 748   | 61    | 5    | 1  |    |      |                        |             |   |
| 958.6 | 957.6 | 957.5 | 0.04  | 33   | 34 | 1  | 20   | R.RIPLQGFK.E           |             |   |
| 259   | 186   | 759   | 27    | 4    | 1  |    |      |                        |             |   |
| 958.6 | 957.6 | 957.5 | 0.04  | 33   | 34 | 1  | -    | R.RIPLQGFK.E           |             |   |
| 259   | 186   | 759   | 27    | 4    | 1  |    |      |                        |             |   |
| 3     | 958.3 | 957.3 | 957.4 | -0.0 | 42 | 42 | 0    | 34                     | K.YMDNIFR.E |   |
| 692   | 619   | 378   | 759   | 1    | 7  |    |      |                        |             |   |
| 958.3 | 957.3 | 957.4 | -0.0  | 42   | 42 | 0  | -    | K.YMDNIFR.E            |             |   |
| 692   | 619   | 378   | 759   | 1    | 7  |    |      |                        |             |   |
| 974.3 | 973.3 | 973.4 | -0.0  | 42   | 42 | 0  | (11) | K.YMDNIFR.E            |             | + |
| 630   | 557   | 327   | 770   | 1    | 7  |    |      | Oxidation (M)          |             |   |
| 974.3 | 973.3 | 973.4 | -0.0  | 42   | 42 | 0  | -    | K.YMDNIFR.E            |             | + |
| 630   | 557   | 327   | 770   | 1    | 7  |    |      | Oxidation (M)          |             |   |
| 976.4 | 975.4 | 975.5 | -0.0  | 32   | 33 | 0  | -    | R.DMLILESR.I           |             |   |
| 139   | 066   | 059   | 993   | 3    | 0  |    |      |                        |             |   |
| 992.4 | 991.4 | 991.5 | -0.0  | 32   | 33 | 0  | -    | R.DMLILESR.I           |             | + |
| 600   | 527   | 008   | 481   | 3    | 0  |    |      | Oxidation (M)          |             |   |

|       |       |       |      |    |    |   |    |                     |   |
|-------|-------|-------|------|----|----|---|----|---------------------|---|
| 1029. | 1028. | 1028. | -0.0 | 54 | 55 | 0 | -  | K.LSPGANTITR.S      |   |
| 4873  | 4800  | 5614  | 814  | 7  | 6  |   |    |                     |   |
| 1078. | 1077. | 1077. | -0.0 | 10 | 11 | 0 | -  | K.HWFSLFNK.R        |   |
| 4626  | 4553  | 5396  | 842  | 4  | 1  |   |    |                     |   |
| 1113. | 1112. | 1112. | -0.0 | 59 | 60 | 0 | 3  | K.YHSGLGLPNR.F      |   |
| 4926  | 4853  | 5726  | 873  | 3  | 2  |   |    |                     |   |
| 1113. | 1112. | 1112. | -0.0 | 59 | 60 | 0 | -  | K.YHSGLGLPNR.F      |   |
| 4926  | 4853  | 5726  | 873  | 3  | 2  |   |    |                     |   |
| 1137. | 1136. | 1136. | -0.0 | 90 | 98 | 0 | 19 | R.LLQELNDHR.L       |   |
| 5056  | 4983  | 5938  | 955  |    |    |   |    |                     |   |
| 1137. | 1136. | 1136. | -0.0 | 90 | 98 | 0 | -  | R.LLQELNDHR.L       |   |
| 5056  | 4983  | 5938  | 955  |    |    |   |    |                     |   |
| 1159. | 1158. | 1158. | -0.1 | 33 | 34 | 0 | -  | R.DAIAHGYITAK.D     |   |
| 5079  | 5006  | 6033  | 027  | 3  | 3  |   |    |                     |   |
| 1231. | 1230. | 1230. | -0.0 | 32 | 33 | 1 | -  | R.VRDMLILES.R       |   |
| 5859  | 5786  | 6754  | 968  | 1  | 0  |   |    |                     |   |
| 1247. | 1246. | 246.6 | -0.0 | 32 | 33 | 1 | -  | R.VRDMLILES.R       | + |
| 5791  | 5718  | 703   | 985  | 1  | 0  |   |    | Oxidation (M)       |   |
| 1293. | 1292. | 1292. | -0.1 | 89 | 98 | 1 | -  | K.RLLQELNDHR.L      |   |
| 6008  | 5935  | 6949  | 014  |    |    |   |    |                     |   |
| 1315. | 1314. | 1314. | -0.1 | 47 | 56 | 0 | -  | K.QHDVNYLLWK.V      |   |
| 5726  | 5653  | 6720  | 067  |    |    |   |    |                     |   |
| 1361. | 1360. | 1360. | -0.1 | 56 | 57 | 0 | -  | K.DSAVTVPDVPSFK.D   |   |
| 5828  | 5755  | 6874  | 119  | 0  | 2  |   |    |                     |   |
| 1428. | 1427. | 1427. | -0.1 | 33 | 34 | 1 | -  | R.IRDAIAHGYITAK.D   |   |
| 6854  | 6781  | 7885  | 103  | 1  | 3  |   |    |                     |   |
| 1548. | 1547. | 1547. | -0.1 | 53 | 54 | 1 | -  | R.WNAIELDKFWVK.L    |   |
| 6967  | 6894  | 8136  | 242  | 5  | 6  |   |    |                     |   |
| 1599. | 1598. | 1598. | -0.1 | 64 | 66 | 0 | 39 | K.YPDNRPHGYPLDR.R   |   |
| 6436  | 6363  | 7590  | 226  | 9  | 1  |   |    |                     |   |
| 1599. | 1598. | 1598. | -0.1 | 64 | 66 | 0 | -  | K.YPDNRPHGYPLDR.R   |   |
| 6436  | 6363  | 7590  | 226  | 9  | 1  |   |    |                     |   |
| 1633. | 1632. | 1632. | -0.2 | 55 | 57 | 1 | -  | R.SGKDSAVTVPDVPSFK. |   |
| 5582  | 5509  | 8358  | 849  | 7  | 2  |   |    | D                   |   |
| 1649. | 1648. | 1648. | -0.1 | 52 | 53 | 0 | 41 | K.YDNNGVEYSFNDGR.   |   |
| 5533  | 5460  | 6753  | 293  | 1  | 4  |   |    | W                   |   |
| 1649. | 1648. | 1648. | -0.1 | 52 | 53 | 0 | -  | K.YDNNGVEYSFNDGR.   |   |
| 5533  | 5460  | 6753  | 293  | 1  | 4  |   |    | W                   |   |
| 1748. | 1747. | 1747. | -0.2 | 90 | 10 | 1 | -  | R.LLQELNDHRLLEQK.H  |   |
| 7150  | 7077  | 9580  | 503  |    | 3  |   |    |                     |   |
| 1769. | 1768. | 1768. | -0.1 | 25 | 26 | 0 | 46 | K.GENFFWVHHQLTVR.F  |   |
| 7501  | 7428  | 8797  | 369  | 2  | 5  |   |    |                     |   |
| 1769. | 1768. | 1768. | -0.1 | 25 | 26 | 0 | -  | K.GENFFWVHHQLTVR.F  |   |
| 7501  | 7428  | 8797  | 369  | 2  | 5  |   |    |                     |   |

|       |         |       |      |     |    |   |    |                     |
|-------|---------|-------|------|-----|----|---|----|---------------------|
| 1828. | 1827.   | 1827. | -0.1 | 39  | 41 | 0 | 37 | K.FDLPPGVLEHFETATR. |
| 7789  | 7716    | 9155  | 439  | 6   | 1  |   |    | D                   |
| 1828. | 1827.   | 1827. | -0.1 | 39  | 41 | 0 | -  | K.FDLPPGVLEHFETATR. |
| 7789  | 7716    | 9155  | 439  | 6   | 1  |   |    | D                   |
| 1889. | 1888.   | 1888. | -0.1 | 49  | 50 | 0 | -  | R.LTHSDFSLSLGVTTNNN |
| 7838  | 7765    | 9278  | 513  | 0   | 7  |   |    | GK.E                |
| 1897. | 1896.   | 1896. | -0.1 | 25  | 26 | 1 | -  | R.KGENFFWVHHQLTVR.  |
| 8322  | 8249    | 9747  | 497  | 1   | 5  |   |    | F                   |
| 2111. | 2110.   | 2111. | -0.1 | 57  | 59 | 0 | -  | K.DLIEQTEAALSSGSAL  |
| 9187  | 9114    | 0746  | 631  | 3   | 2  |   |    | HLEK.Y              |
| 2519. | 2518.   | 2519. | -0.1 | 29  | 32 | 0 | 19 | K.YGGYFSPRPDHSVHFE  |
| 9846  | 9773    | 1617  | 844  | 9   | 0  |   |    | DVDGVAR.V           |
| 2519. | 2518.   | 2519. | -0.1 | 29  | 32 | 0 | -  | K.YGGYFSPRPDHSVHFE  |
| 9846  | 9773    | 1617  | 844  | 9   | 0  |   |    | DVDGVAR.V           |
| 2548. | 2547.   | 2547. | -0.2 | 38  | 41 | 1 | -  | R.QGDPHGKFDLPPGVLE  |
| 0542  | 0469    | 2506  | 036  | 9   | 1  |   |    | HFETATR.D           |
| 2853. | 2852.   | 2852. | -0.2 | 43  | 45 | 0 | -  | K.DSLTPYTTQELEFPGV  |
| 2126  | 2053    | 4079  | 026  | 1   | 6  |   |    | AIDSVTLNLR.L        |
| 3479. | 3478.   | 3478. | -0.2 | 57  | 88 | 1 | -  | K.VYEDIRDPALQGLSQS  |
| 4192  | 4119    | 6276  | 157  |     |    |   |    | FNPSGGHYDDDGAV      |
| 4     | 883.3   | 882.3 | -0.0 | 40  | 46 | 0 | 48 | R.QQFFASR.Q         |
|       | 816     | 743   | 348  | 604 |    |   |    |                     |
|       | 1041.   | 1040. | -0.0 | 24  | 25 | 0 | -  | K.AAVCVAHWQ.-       |
|       | 4210    | 4137  | 4862 | 724 | 4  | 2 |    |                     |
|       | 1075.48 | 1074. | -0.0 | 15  | 16 | 0 | 10 | R.GLGEVTVMNR.R      |
|       | 4       | 4771  | 5491 | 720 | 4  | 3 |    |                     |
|       | 1075.   | 1074. | -0.0 | 15  | 16 | 0 | -  | R.GLGEVTVMNR.R      |
|       | 4844    | 4771  | 5491 | 720 | 4  | 3 |    |                     |
|       | 1091.   | 1090. | -0.0 | 15  | 16 | 0 | -  | R.GLGEVTVMNR.R +    |
|       | 4832    | 4759  | 5440 | 681 | 4  | 3 |    | Oxidation (M)       |
|       | 1158.   | 1157. | -0.0 | 16  | 17 | 0 | 26 | R.GSFDLYSVDR.T      |
|       | 4673    | 4600  | 5353 | 752 | 5  | 4 |    |                     |
|       | 1158.   | 1157. | -0.0 | 16  | 17 | 0 | -  | R.GSFDLYSVDR.T      |
|       | 4673    | 4600  | 5353 | 752 | 5  | 4 |    |                     |
|       | 1169.   | 1168. | -0.0 | 24  | 25 | 1 | -  | R.KAAVCVAHWQ.-      |
|       | 5063    | 4990  | 5811 | 821 | 3  | 2 |    |                     |
|       | 1235.   | 1234. | -0.0 | 19  | 20 | 0 | 88 | R.DLASTAPYEIR.G     |
|       | 5422    | 5349  | 6193 | 844 | 7  | 7 |    |                     |
|       | 1235.   | 1234. | -0.0 | 19  | 20 | 0 | -  | R.DLASTAPYEIR.G     |
|       | 5422    | 5349  | 6193 | 844 | 7  | 7 |    |                     |
|       | 1314.   | 1313. | -0.0 | 16  | 17 | 1 | 24 | R.RGSFDLYSVDR.T     |
|       | 5559    | 5486  | 6364 | 877 | 4  | 4 |    |                     |
|       | 1314.   | 1313. | -0.0 | 16  | 17 | 1 | -  | R.RGSFDLYSVDR.T     |
|       | 5559    | 5486  | 6364 | 877 | 4  | 4 |    |                     |

|         |       |       |       |      |    |    |    |                    |             |
|---------|-------|-------|-------|------|----|----|----|--------------------|-------------|
| 2320.   | 2319. | 2320. | -0.1  | 13   | 14 | 0  | 15 | R.VYHFSWCNLPGHGFS  |             |
| 9124    | 9051  | 0595  | 544   | 0    | 9  |    |    | QSAAR.D            |             |
| 2320.   | 2319. | 2320. | -0.1  | 13   | 14 | 0  | -  | R.VYHFSWCNLPGHGFS  |             |
| 9124    | 9051  | 0595  | 544   | 0    | 9  |    |    | QSAAR.D            |             |
| 1314.   | 2755. | 2755. | -0.1  | 17   | 19 | 0  | -  | R.TTDFDYFFELLQHQIP |             |
| 5559    | 2268  | 3646  | 377   | 5    | 6  |    |    | SVWTR.D            |             |
| 883.3   | 882.3 | 882.4 | -0.0  | 40   | 46 | 0  | 48 | R.QQFFASR.Q        |             |
| 816     | 743   | 348   | 604   |      |    |    |    |                    |             |
| 1041.   | 1040. | 1040. | -0.0  | 24   | 25 | 0  | -  | K.AAVCVAHWQ.-      |             |
| 4210    | 4137  | 4862  | 724   | 4    | 2  |    |    |                    |             |
| 1075.48 | 1074. | 1074. | -0.0  | 15   | 16 | 0  | 10 | R.GLGEVTVMNR.R     |             |
| 4       | 4771  | 5491  | 720   | 4    | 3  |    |    |                    |             |
| 1075.   | 1074. | 1074. | -0.0  | 15   | 16 | 0  | -  | R.GLGEVTVMNR.R     |             |
| 4844    | 4771  | 5491  | 720   | 4    | 3  |    |    |                    |             |
| 1091.   | 1090. | 1090. | -0.0  | 15   | 16 | 0  | -  | R.GLGEVTVMNR.R     | +           |
| 4832    | 4759  | 5440  | 681   | 4    | 3  |    |    | Oxidation (M)      |             |
| 1158.   | 1157. | 1157. | -0.0  | 16   | 17 | 0  | 26 | R.GSFDLYSVDR.T     |             |
| 4673    | 4600  | 5353  | 752   | 5    | 4  |    |    |                    |             |
| 1158.   | 1157. | 1157. | -0.0  | 16   | 17 | 0  | -  | R.GSFDLYSVDR.T     |             |
| 4673    | 4600  | 5353  | 752   | 5    | 4  |    |    |                    |             |
| 1169.   | 1168. | 1168. | -0.0  | 24   | 25 | 1  | -  | R.KAAVCVAHWQ.-     |             |
| 5063    | 4990  | 5811  | 821   | 3    | 2  |    |    |                    |             |
| 1235.   | 1234. | 1234. | -0.0  | 19   | 20 | 0  | 88 | R.DLASTAPYEIR.G    |             |
| 5422    | 5349  | 6193  | 844   | 7    | 7  |    |    |                    |             |
| 1235.   | 1234. | 1234. | -0.0  | 19   | 20 | 0  | -  | R.DLASTAPYEIR.G    |             |
| 5422    | 5349  | 6193  | 844   | 7    | 7  |    |    |                    |             |
| 1314.   | 1313. | 1313. | -0.0  | 16   | 17 | 1  | 24 | R.RGSFDLYSVDR.T    |             |
| 5559    | 5486  | 6364  | 877   | 4    | 4  |    |    |                    |             |
| 1314.   | 1313. | 1313. | -0.0  | 16   | 17 | 1  | -  | R.RGSFDLYSVDR.T    |             |
| 5559    | 5486  | 6364  | 877   | 4    | 4  |    |    |                    |             |
| 2320.   | 2319. | 2320. | -0.1  | 13   | 14 | 0  | 15 | R.VYHFSWCNLPGHGFS  |             |
| 9124    | 9051  | 0595  | 544   | 0    | 9  |    |    | QSAAR.D            |             |
| 2320.   | 2319. | 2320. | -0.1  | 13   | 14 | 0  | -  | R.VYHFSWCNLPGHGFS  |             |
| 9124    | 9051  | 0595  | 544   | 0    | 9  |    |    | QSAAR.D            |             |
| 1314.   | 2755. | 2755. | -0.1  | 17   | 19 | 0  | -  | R.TTDFDYFFELLQHQIP |             |
| 5559    | 2268  | 3646  | 377   | 5    | 6  |    |    | SVWTR.D            |             |
| 5       | 883.3 | 882.3 | 882.4 | -0.0 | 40 | 46 | 0  | 30                 | R.QQFFASR.Q |
|         | 753   | 680   | 348   | 667  |    |    |    |                    |             |
| 1041.   | 1040. | 1040. | -0.0  | 24   | 25 | 0  | -  | K.AAVCVAHWQ.-      |             |
| 4141    | 4068  | 4862  | 793   | 4    | 2  |    |    |                    |             |
| 1075.   | 1074. | 1074. | -0.0  | 15   | 16 | 0  | 15 | R.GLGEVTVMNR.R     |             |
| 4751    | 4678  | 5491  | 813   | 4    | 3  |    |    |                    |             |
| 1075.   | 1074. | 1074. | -0.0  | 15   | 16 | 0  | -  | R.GLGEVTVMNR.R     |             |
| 4751    | 4678  | 5491  | 813   | 4    | 3  |    |    |                    |             |

|       |       |       |      |    |    |   |     |                    |   |
|-------|-------|-------|------|----|----|---|-----|--------------------|---|
| 1091. | 1090. | 1090. | -0.0 | 15 | 16 | 0 | -   | R.GLGEVTVMNR.R     | + |
| 4680  | 4607  | 5440  | 833  | 4  | 3  |   |     | Oxidation (M)      |   |
| 1158. | 1157. | 1157. | -0.0 | 16 | 17 | 0 | 14  | R.GSFDLYSVDR.T     |   |
| 4574  | 4501  | 5353  | 851  | 5  | 4  |   |     |                    |   |
| 1158. | 1157. | 1157. | -0.0 | 16 | 17 | 0 | -   | R.GSFDLYSVDR.T     |   |
| 4574  | 4501  | 5353  | 851  | 5  | 4  |   |     |                    |   |
| 1169. | 1168. | 1168. | -0.0 | 24 | 25 | 1 | -   | R.KAAVCVAHWQ.-     |   |
| 5093  | 5020  | 5811  | 791  | 3  | 2  |   |     |                    |   |
| 1235. | 1234. | 1234. | -0.0 | 19 | 20 | 0 | 72  | R.DLASTAPYEIR.G    |   |
| 5321  | 5248  | 6193  | 945  | 7  | 7  |   |     |                    |   |
| 1235. | 1234. | 1234. | -0.0 | 19 | 20 | 0 | -   | R.DLASTAPYEIR.G    |   |
| 5321  | 5248  | 6193  | 945  | 7  | 7  |   |     |                    |   |
| 1314. | 1313. | 1313. | -0.0 | 16 | 17 | 1 | 19  | R.RGSFDLYSVDR.T    |   |
| 5442  | 5369  | 6364  | 994  | 4  | 4  |   |     |                    |   |
| 1314. | 1313. | 1313. | -0.0 | 16 | 17 | 1 | -   | R.RGSFDLYSVDR.T    |   |
| 5442  | 5369  | 6364  | 994  | 4  | 4  |   |     |                    |   |
| 1982. | 1981. | 1981. | -0.1 | 21 | 39 | 0 | 158 | R.FSYDGTPLADDVVVG  |   |
| 7893  | 7820  | 9381  | 560  |    |    |   |     | SGQR.Q             |   |
| 1982. | 1981. | 1981. | -0.1 | 21 | 39 | 0 | -   | R.FSYDGTPLADDVVVG  |   |
| 7893  | 7820  | 9381  | 560  |    |    |   |     | SGQR.Q             |   |
| 3205. | 3204. | 3204. | -0.2 | 21 | 24 | 0 | -   | R.SGDSTAGQDCLSLEAS |   |
| 1477  | 1404  | 3823  | 419  | 3  | 2  |   |     | TDGLFLLQDSCSDR     |   |

---