

Bioinformatic Comparison of Kunitz Protease Inhibitors in *Echinococcus Granulosus* and *E. Multilocularis* and the Genes Expressed in Different Developmental Stages of *E. Granulosus*

Hui Zhang

Basic Medicine College, Xinjiang Medical University

Mengxiao Tian

Basic Medicine College, Xinjiang Medical University

Wenjing Qi

Xinjiang Medical University, The First Affiliated Hospital of Xinjiang Medical University

Juan Wu

Basic Medicine College, Xinjiang Medical University

Huajun Zheng

Chinese National Human Genome Center at Shanghai

Gang Guo

Xinjiang Medical University, The First Affiliated Hospital of Xinjiang Medical University

Liang Zhang

Chinese National Human Genome Center at Shanghai

Shiwanthi L. Ranasinghe

QIMR Berghofer Medical Research Institute

Donald P. McManus

QIMR Berghofer Medical Research Institute

Jun Li

Xinjiang Medical University, The First Affiliated Hospital of Xinjiang Medical University

Wenbao Zhang (✉ wenbaozhang2013@163.com)

Xinjiang Medical University, The First Affiliated Hospital of Xinjiang Medical University

Research Article

Keywords: Kunitz protease inhibitors, *E. multilocularis*, *E. granulosus*, Bioinformatic comparison

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-318603/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Cystic and alveolar echinococcosis caused by the tapeworms *Echinococcus granulosus* and *E. multilocularis*, respectively, are important zoonotic diseases. Protease inhibitors are crucial for the survival of both *Echinococcus spp.* Kunitz-type inhibitors play a regulatory role in the control of protease activity. In this study we identified all the Kunitz-type protease inhibitors present in the genomes of these two tapeworms and analyzed the gene sequences using computational, structural bioinformatics and phylogenetic approaches to evaluate the evolutionary relationships of these genes.

Results: A total of 19 genes from *E. multilocularis* and 23 genes from *E. granulosus* contained single or multiple Kunitz-domains. A neighbor-joining phylogenetic tree indicated that the *E. granulosus* and *E. multilocularis* Kunitz domain peptides were divided into three branches containing 9 clusters. Based on available transcriptome data, we analyzed the expression of these Kunitz-domain protease inhibitors in four major developmental stages of *E. granulosus* and found they were differentially expressed.

Conclusion: We identified 19 and 23 Kunitz protease inhibitors in *E. multilocularis* and *E. granulosus* respectively; the majority of these genes were expressed in one or four stages of *E. granulosus* with some being highly expressed in adult worms indicating that these genes likely play different roles in the different developmental stages.

Background

Cyst echinococcosis (CE) and alveolar echinococcosis (AE) are both medically and economically important diseases caused by the metacestode stages of *Echinococcus granulosus* and *E. multilocularis* respectively. The diseases impact on hundreds of millions of people in Asia, Europe, America and Africa [1]. The control and treatment of echinococcosis are difficult. High frequency of dosing dogs with the drug praziquantel has played a key role in the control of the disease [2, 3], but undertaking control measures is challenging in remote areas. A vaccine against adult worms in dogs is urgently needed [4].

The life-cycle of these two tapeworms involves four major developmental stages present in their definitive and intermediate hosts. The survival of these tapeworms relies on evading host immune responses and avoiding attack by proteases; this is especially important for the adult parasites which reside in the gastrointestinal duct, a location where high concentration of proteases are present which are harmful and toxic for the worms.

Eukaryote proteases including serine (trypsin/chymotrypsin-like), cysteine (thiol) and aspartic (pepsin/cathepsin/rennin) proteases play a fundamental role in the regulation of protein function. Their functions are controlled largely by protease inhibitors which play crucial roles in the regulation of proteases involved in a range of biological processes system including cell proliferation, inflammation, immune mechanisms and cell homeostasis [5–7]; protease inhibitors act mainly through the control of potentially disadvantageous, excess or inopportune proteolytic activity. Protease inhibitors including aspartic, cysteine, metallo, serine, and threonine inhibitors are super-families based on their similarities at

the amino acid sequence level and tertiary structure [8]. Similarities in primary structure and tertiary structure support the common ancestry of many protease inhibitor families.

Kunitz-type domain protease inhibitors (KDPIs) are an important type of protease inhibitor and belong to the I2 family of protease inhibitors [8, 9]. These inhibitors contain at least a cysteine-rich peptide chain (Kunitz-type domain) with α and β sheets. The Kunitz domain consists of around 60 amino acids including six conserved cysteine residues forming three disulphide bonds in a characteristic pattern (C1-C6, C2-C4, and C3-C5) [9]. The protein is stabilized by three conservative disulfide bonds. These inhibitors have been characterized from animals and plants [9, 10] including helminths [11–13]. A previous study described eight genes (EgKU1-EgKU8) isolated from *E. granulosus* protoscoleces treated with pepsin/H (+)[14]. We previously cloned and characterized two *E. granulosus* KDPIs, EgKI-1 (EG_08721 (GenBank: EUB56407.1)) and EgKI-2 (EG_07242 (GenBank: EUB57880.1))[13]. EgKI-1 is highly expressed in the oncosphere (egg) stage and is a potent chymotrypsin and neutrophil elastase inhibitor that binds calcium and reduced neutrophil infiltration in a local inflammation model. EgKI-2 is highly expressed in adult worms, it is a potent inhibitor of trypsin and is a potential vaccine candidate against echinococcosis in dogs [13]. Beyond these other *E. granulosus* and *E. multilocularis* KDPIs have received little attention.

In the present study, we identified all KDPI sequences predicted in the *E. granulosus* and *E. multilocularis* genomes and used computerized programs to characterize these Kunitz domain protease inhibitors. We show that the majority of the *E. granulosus* KDPIs are expressed and are differentially expressed in different life cycle stages and some have a range of GO numbers indicating these inhibitors likely function in different ways in the tapeworm's development.

Results

General characterizations of Kunitz domain protease inhibitors

InterproScan and Motif scan identified 19 and 23 genes encoding KDPIs from the *E. multilocularis* and *E. granulosus* genomes, respectively (Table 1). The KDPI family has a typical Kunitz domain of about 50 amino acids in size (Fig. 1) with a special secondary structure formed by 3 disulfide bonds or bridges (Additional file 3: Fig. S1). The echinococcal Kunitz domains contain an average of 52.85 aa (range 47–55 aa) with the majority comprising 53 aa (Fig. 1).

Table 1

Summary of physiological and biological characteristics of Kunitz protease inhibitors in *E. multilocularis* and *E. granulosus*

Species	<i>E. multilocularis</i>	<i>E. granulosus</i>
KDPIs/single KDPIs	19/16	23/21
Number of amino acids	333.47	269.96
Molecular weight (Da)	37153.2	30117.02
Isoelectric points	7.44	7.84
No. of trans-domain (%)	4(21.05)	5(21.74)
No. of cysteine in Kunitz domain	5.63	5.65
No. of cysteine in the protein	24.89	18.13
Instability index	45.17	47.36
Stable protein (No/yes)	12/7	15/8
Aliphatic index	72.35	71.37
Grand average of hydropathicity	-0.22	-0.22
Signal peptides (%)	17(89.47) *	14(60.87)
No. of Kunitz Motifs	1.79	1.17
En-t-In(T/C)	9/2	9/4
Note: KDPIs, Kunitz domain protease inhibitors; No. of aa, number of amino acids; No. of tran-domain, percentage of containing transmembrane domains; Aver of cysteine, average of cysteine per sequence; Aliphatic indexes; GRSVY, hydropathic index; En-t-In (T/C), enzyme targeting inhibitors, trypsin inhibitors(T) or chymotrypsin inhibitors(C).		
There are significant difference of signal peptides between <i>E. multilocularis</i> and <i>E. granulosus</i> KDPIs ($P < 0.05$).		

Among these KDPIs, *E. multilocularis* has 16 KDPIs containing a single Kunitz domain with the complete proteins having sizes of 75–534 aa. There are 3 proteins containing multi-domains with a maximum 8 Kunitz domains (EmuJ_001181950) in size ranging from 610 to 2425 aa. *E. granulosus* has 21 single Kunitz domain KDPIs of 75–976 aa in size, and 2 multiple domain KDPIs sized from 878 to 1540 aa. The molecular weights of these KDPIs range from 8.34 kDa to 266.77 kDa with isoelectric points from 4.52 to 10.52 (Additional file 1: Table S1). The majority of single Kunitz domain proteins comprise less than 100 amino acid(Additional file 1: Table S1).

We used the instability index to estimate the stability of the KDPIs. An instability value > 40 is an unstable protein. The index value representing rigidity/flexibility of each peptide varied (26.51–94.41). The average

value was 45.17 for *E. multilocularis* and 47.37 for *E. granulosus*, suggesting the peptides may be flexible. The analysis showed that *E. multilocularis* has 12 unstable and 7 stable KDPIs and *E. granulosus* has 15 unstable and 8 stable KDPIs (Additional file 1: Table S1).

Sequence analysis showed that these echinococcal KDPI sequences contain a high percentage of hydrophobic residues including alanine (A), valine (V), leucine (L) and isoleucine (I). The hydrophobic index (Grand Average of Hydropathy: GRAVY) for *E. granulosus* and *E. multilocularis* are -0.223 ± 0.333 and -0.222 ± 0.312 , respectively. Aliphatic indexes (AI) are 71.37 ± 14.08 and 72.36 ± 13.75 for *E. granulosus* and *E. multilocularis*, respectively (Table 1 and Additional file 1: Table S1).

The average numbers of negatively charged residues (Asp + Glu) are 38.42 and 28.91, accounting for 8.84% and 9.13% of *E. multilocularis* and *E. granulosus* KDPIs, respectively. There are 35.16 and 30.70 positively charged residues (Arg + Lys) in the *E. multilocularis* and *E. granulosus* KDPIs accounting for 12.73% and 12.23% of the total amino acids. Neutral amino acid residues are 259.89 and 210.35 aa on average and account for 78.42% and 78.63% of the KDPIs in *E. multilocularis* and *E. granulosus*, respectively (Table 1 and Additional file 1: Table S1).

The average aliphatic indexes are 72.36 (51.54–89.78) and 71.37 (49.44–100.73) for the *E. multilocularis* and *E. granulosus* KDPIs, respectively. The hydrophobicity indexes of *E. multilocularis* and *E. granulosus* KDPIs are -0.222 (ranging from -0.978 to 0.340) and -0.223 (ranging from -0.996 to 0.371) for respectively. The results indicate that the KDPIs in both parasites are likely hydrophilic proteins (Table 1 and Additional file 1: Table S1).

E. multilocularis and *E. granulosus* have 4 and 5 KDPIs containing transmembrane regions, respectively, and 78.94% and 78.26% of the *E. multilocularis* and *E. granulosus* KDPIs are extracellular (Table 1), which matches the GO analysis (Table 2 and Additional file 2: Table S2), indicating that the most KDPIs may involve host and parasite interface responses. The TopPred program indicated that 4 *E. multilocularis* and 5 *E. granulosus* KDPIs are located in the cytoplasm (Additional file 1: Table S1) with others, including 15 Em-KDPI sequences and 18 Eg-KDPIs, being extracellular.

Table 2

The expression of Kunitz-type domain protease inhibitors in four stages of *E. granulosus* showing Hi-seq reads

Gene ID	Adt	Onc	PSC	CM	SEq. Description	Seq. Lth	#GOs	GOs
EG_07244	217	0	1	10	serine protease inhibitor	106	1	F: serine-type endopeptidase inhibitor activity
EG_08716	155	0	0	0	kunitz-type protease inhibitor 3-like	84	3	F: peptidase inhibitor activity; F: protein binding; P: transforming growth factor beta receptor signaling pathway
EG_10096	70	0	3	4	kunitz domain-containing	98	42	C: cytoplasmic vesicle; P: apoptosis; P: neuromuscular process controlling balance; P: ionotropic glutamate receptor signaling pathway; P: regulation of epidermal growth factor receptor activity; F: receptor binding; P: regulation o
EG_07242	46	0	1	5	serine protease inhibitor	83	7	F: peptidase inhibitor activity; C: nematocyst; F: ion channel inhibitor activity; F: potassium channel inhibitor activity; F: serine-type endopeptidase inhibitor activity; C: extracellular region; P: pathogenesis
EG_07243	35	0	2	5	wap four-disulfide core domain 6b	75	2	C: cytoplasm; F: serine-type endopeptidase inhibitor activity
EG_03480	24	0	1	5	four-domain proteases inhibitor	88	1	F: peptidase inhibitor activity

Note: Adt, adult worms; Onc, oncospheres; PSC, protoscolecetes; CM, cyst membrane.

Gene ID	Adt	Onc	PSC	CM	SEq. Description	Seq. Lth	#GOs	GOs
EG_09490	17	2	14	51	spon-1 protein	976	3	F: serine-type endopeptidase inhibitor activity; C: proteinaceous extracellular matrix; C: extracellular region
EG_03481	16	0	8	2	trypsin inhibitor	242	4	F: peptidase inhibitor activity; P: multicellular organismal process; F: extracellular matrix structural constituent; C: proteinaceous extracellular matrix
EG_08720	10	0	0	0	kunitz bovine pancreatic trypsin inhibitor domain protein	84	4	F: serine-type endopeptidase inhibitor activity; F: peptidase activity; F: peptidase inhibitor activity; C: extracellular region
EG_07944	9	0	22	14	kunitz bovine pancreatic trypsin inhibitor domain protein	539	2	C: extracellular region; F: hydrolase activity
EG_09269	8	0	0	1	tissue factor pathway inhibitor 2-like	92	2	F: extracellular matrix structural constituent; C: proteinaceous extracellular matrix
EG_05317	7	0	19	16	Kunitz-like protease inhibitor precur	1540		
EG_08721	6	108	0	0	serine protease inhibitor- with kunitz and wap domains 1	79	1	C: acrosomal vesicle
EG_04958	5	0	1	2	elegans protein partially confirmed by transcript evidence	135	2	F: serine-type endopeptidase inhibitor activity; P: epidermis development

Note: Adt, adult worms; Onc, oncospheres; PSC, protoscolecocytes; CM, cyst membrane.

Gene ID	Adt	Onc	PSC	CM	Seq. Description	Seq. Lth	#GOs	GOs
EG_05483	5	12	0	2	secreted protein with kunitz	191	2	C: extracellular region; F: hydrolase activity
EG_09007	4	0	0	0	mechanosensory abnormality family member (mec-1)	86	2	P: extracellular structure organization; P: mechanosensory behavior
EG_09008	3	0	0	0	acp24a4	102	2	C: extracellular region; F: peptidase activity
EG_01779	1	0	13	3	isoform g	878	1	F: hydrolase activity
EG_05482	1	2	0	0	kunitz bovine pancreatic trypsin inhibitor domain containing protein	130	1	C: extracellular region
EG_07266	1	0	0	3	serine protease inhibitor	129	3	F: binding; F: serine-type endopeptidase inhibitor activity; C: extracellular region
EG_05316	0	0	3	2	kunitz-type protease inhibitor 3-like	239	1	F: peptidase inhibitor activity
EG_08718	0	2	0	0	kunitz domain-containing	144	4	F: peptidase activity; F: serine-type endopeptidase inhibitor activity; F: peptidase inhibitor activity; C: extracellular region
EG_09006	0	0	0	0	single kunitz protease inhibitor	89	5	F: peptidase activity; F: serine-type endopeptidase inhibitor activity; F: peptidase inhibitor activity; C: extracellular region; P: multicellular organismal development

Note: Adt, adult worms; Onc, oncospheres; PSC, protoscolecies; CM, cyst membrane.

Signal peptide analysis showed that there are 17/19(89.47%) *E. multilocularis* KDPIs having an 18–26 amino acid (aa) signal peptide and 2/19(10.53%) KDPIs without. In contrast, *E. granulosus* has 14/23 (60.87%) KDPIs containing signal peptide sequences and 9/23 (39.13%) KDPIs without (Table 1). There is significant difference of signal peptides between *E. multilocularis* and *E. granulosus* KDPIs ($P < 0.05$; Table 1).

Cluster and phylogenetic analysis of Kunitz protease inhibitors

Multiple sequence alignment and phylogenetic analysis of the amino acid sequences were used to infer the evolutionary relationships between the *E. multilocularis* and *E. granulosus* KDPIs and to make a comparison with other species. Figure 2 shows the different evolutionary distances of the genes of a single Kunitz domain of the KDPIs using the neighbor-joining method. The analysis indicated that the *E. granulosus* and *E. multilocularis* Kunitz domain peptides were divided into three branches containing 9 clusters.

Comparison of KDPI genes predicated from the *E. granulosus* and *E. multilocularis* genomes

We compared the KDPI genes predicted from the genomes of *E. granulosus* and *E. multilocularis* and found that some genes are species-specific. *E. multilocularis* does not have homologues of *E. granulosus* sequences EG_07242, EG_07266, EG_07243, EG_09006 and EG_09008; whereas EmuJ_001136700.1 and EmuJ_001137100.1 are specific to *E. multilocularis*.

The specificity of a protease inhibitor against a protease is mainly determined by the nature of the amino acid residue at position P1 of its active site. It has been shown that Lys(K) and Arg(R) mutants of bovine pancreatic trypsin inhibitor (BPTI) bind to bovine trypsin about 10^5 -fold stronger than BPTI with P1 Tyr(T) [15]. In addition, it has been shown that typical trypsin inhibitors have Arg(R) or Lys(K) at P1, and chymotrypsin inhibitors have Leu (L) or Met (M) at the P1 position [16]. Therefore, the sequence analysis shows that the Em-KDPIs have 8 sequences containing R and 1 sequence containing K whereas the Eg-KDPIs have 8 sequences containing R at P1, which belong to typical trypsin inhibitors. Furthermore, the two tapeworms have 3 or 4 sequences containing L at P1, which are chymotrypsin inhibitors (Fig. 1 and Table 1).

Two D and three D of Kunitz domain protease inhibitors

The majority of the single *E. multilocularis* and *E. granulosus* KDPIs are small proteins sized 16-kDa and contain a relatively high percentage of Lys and Arg residues at the C-terminus. Like most Kunitz domain protease inhibitors, the Em- and Eg-KDPIs contain a conserved Kunitz type sequence with 6 cysteine residues forming 3 disulfide bridges (C1-C6, C2-C4 and C3-C5) (Additional file 3: Fig. S1) and these play a key role in the formation of the 2D and 3D structure of these KDPIs. For the single Kunitz domain sequences, the secondary structure prediction revealed 19.01–52.71% and 18.6-60.35% of α -helix and random coil structures in Eg-KDPIs, followed by extended strands and β -turn structure, accounting for 13.1-26.67 and 1.89–10.84%, respectively. Em-KDPIs α -helix and random coil structures account for 19-

40.45% and 32.5-55.99% of the protein sequence respectively, followed by extended strands and β -turn, accounting for 8-36.25% and 0-10.71% (Table 3).

Table 3
The secondary structure prediction of the single Eg-KDPIs and Em-KDPIs

Accession number	Alpha helix(aa/%)	Extended strand(aa/%)	Beta turn(aa/%)	Random coil(aa/%)
EG_03480	32/36.36	19/21.59	3/3.41	34/38.64
EG_03481	46/19.01	57/23.55	19/7.85	120/49.59
EG_04958	33/24.44	28/20.74	11/8.15	63/46.67
EG_05316	92/38.17	54/22.41	17/7.05	78/32.37
EG_05482	54/41.54	18/13.85	8/6.15	50/38.46
EG_05483	66/34.55	41/21.47	10/5.24	74/38.74
EG_07242	25/30.12	13/15.66	9/10.84	36/43.37
EG_07243	20/26.67	20/26.67	5/6.67	30/40
EG_07244	41/38.68	23/21.7	2/1.89	40/37.74
EG_07266	68/52.71	29/22.48	8/6.2	24/18.6
EG_07944.1	138/25.6	91/16.88	34/6.31	276/51.21
EG_08716	31/36.9	11/13.1	3/3.57	39/46.43
EG_08718	41/28.47	34/23.61	8/5.56	61/42.36
EG_08720	25/29.76	18/21.43	4/4.76	37/44.05
EG_08721	35/44.3	13/16.46	3/3.8	28/35.44
EG_09006	28/31.46	22/24.72	9/10.11	30/33.71
EG_09007	23/26.74	13/15.12	7/8.14	43/50
EG_09008	43/42.16	22/21.57	7/6.86	30/29.41
EG_09269	37/40.22	15/16.3	4/4.35	36/39.13
EG_09490	203/20.8	145/14.86	39/4.0	589/60.35
EG_10096	20/20.41	23/23.47	3/3.06	52/53.06
EmuJ_000077700.1	65/30.95	42/20	17/8.1	86/40.95
EmuJ_000077800.1	33/24.26	39/28.68	10/7.35	54/39.71
EmuJ_000302900.1	128/23.97	76/14.23	31/5.81	299/55.99
EmuJ_000419200.1	37/40.22	15/16.30	0/0.00	40/43.48

Note: aa/%, number/percentage of of amino acids in each secondary structure

Accession number	Alpha helix(aa/%)	Extended strand(aa/%)	Beta turn(aa/%)	Random coil(aa/%)
EmuJ_000534800.1	24/32.00	6/8.00	5/6.67	40/53.33
EmuJ_000548800.1	23/23.23	21/21.21	4/4.04	51/51.52
EmuJ_000549400.1	19/19.00	27/27.00	8/8.00	46/46.00
EmuJ_001136500.1	36/40.45	16/17.98	2/2.25	35/39.33
EmuJ_001136600.1	31/36.90	13/15.48	6/7.14	34/40.48
EmuJ_001136700.1	20/25.64	18/23.08	8/10.26	32/41.03
EmuJ_001136800.1	36/40.45	16/17.98	2/2.25	35/39.33
EmuJ_001136900.1	21/23.33	17/18.89	4/4.44	48/53.33
EmuJ_001137000.1	26/30.95	16/19.05	4/4.76	38/45.24
EmuJ_001137100.1	22/26.19	20/23.81	9/10.71	33/39.29
EmuJ_001137300.1	20/25	29/36.25	5/6.25	26/32.5
EmuJ_001137400.1	26/30.59	16/18.82	4/4.71	39/45.88
Note: aa/%, number/percentage of of amino acids in each secondary structure				

It is accepted that there is a close relationship between the structure and function of a protein. Therefore, we used SWISS-MODEL to predict 3D structures based on the homology modeling of KDPI templates from PDB (protein database) including single and multiple Kunitz domain proteins.

Three D structure analysis showed that a single Kunitz domain sequence with 3 disulfide bonds has a similar structure containing a α -helix and random coils with similar structures (Fig. 3). Some single Kunitz domain sequences losing the second cysteine (C2) the structure is different from 3 disulfide bonds (Fig. 1 and Table 1).

Expression of *E. granulosus* KDPIs in different developmental stages

To estimate expression of the KDPIs, Hi-seq techniques were employed to obtain the transcript reads of these genes from total RNA from each of 4 developmental stages of *E. granulosus*. The transcript read information was published in a previous paper of ours [17].

The transcriptome analysis showed that these Kunitz peptides were differentially expressed in the different developmental stages of *E. granulosus* (Table 2). All the inhibitors, except EG_09006, were expressed in one or 4 stages of *E. granulosus* with some being highly and differentially expressed in one or two stages. EG_03480 (extra), EG_03481 (intra), EG_07242 (extra), EG_07243 (intra), EG_07244 (intra), EG_08716 (extra), EG_08720 (extra), EG_09490 (extra) and EG_10096 (extra) were significantly highly

expressed in the adult worm stage (Table 2 and Additional file 1: Table S1). EG_08716 is an extracellular protease inhibitor and has 42 predicted GOs, including cytoplasmic vesicle for neuromuscular process controlling balance, ionotropic glutamate receptor signaling pathway, regulation of the activity of epidermal growth factor receptor and synapse, regulation of mitotic cell cycle and translation and cellular copper and calcium ion homeostasis (Additional file 1: Table S1). The expression analysis indicated that this gene may play an important role in adult worm development and against host protease attack. EG_07244 is also an endopeptidase, indicating that the protein has two functions, as a peptidase and as a protease inhibitor in adult worms.

EG_08721 is an extracellular inhibitor and was differentially highly expressed in the oncosphere compared with the other stages, indicating this protease inhibitor plays an important role in oncosphere biology, the only stage for primarily infection and EG_08721 may play an important role in oncosphere against host protease attack which may be a candidate for vaccine development.

Although we activated PSC with pepsin, only three KDPIs (EG_01779 EG_05317 and EG_07944) were slightly elevated in this stage. Importantly, we found that EG_09490, EG_09268 and EG_09490 were highly expressed in the cyst membrane and the proteins expressed by these genes may be potential targets for drug development.

Discussion

KDPIs occur in almost all living organisms from bacteria to plants and animals. Kunitz peptides show diverse biological activities including inhibition of proteases and/or blocking or modulating ion channels. Gastrointestinal helminths survive in an environment containing proteases and these parasites must have mechanisms to control protease activation. Therefore, Kunitz domain inhibitors are important for parasite survival, especially intestinal dwelling helminth parasites, to counteract protease attack.

A remarkable difference between the larval stages of *E. multilocularis* and *E. granulosus* is the difference in the lesion pathology in the intermediate hosts. The metacestode of *E. multilocularis* is a tumor-like, infiltrating structure consisting of many small vesicles embedded in the stroma of connective tissue. The continual growth of parasite vesicles in a proliferative style causes damage of liver tissues, which results in high mortality of AE. In contrast, *E. granulosus* cysts develop in internal organs (mainly liver and lungs) of humans and other intermediate hosts as unilocular fluid-filled bladders with clear edge between cyst and host tissue. CE causes mortality in very few patients and there is a relatively good prognosis after surgical removal of the cystic lesion. Contrastingly, AE causes severe damage to the liver and patients require extensive treatment with albendazole to prevent relapse. However, little is known about the molecular mechanisms underpinning biological differences between the two parasites and the diseases they cause.

In this study, based on the genomic information available for *E. granulosus* and *E. multilocularis* we identified 23 and 19 KDPIs, respectively. The differential expression of these KDPI genes between *E. granulosus* and *E. multilocularis* may be associated with the differences in pathology caused by the

metacestodes of the two species. It would be informative to determine whether these genes play a role in determining the different pathologies resulting from infection by the two cestodes in their intermediate hosts.

Signal peptide analysis showed that 89.47% of *E. multilocularis* KDPIs having signal peptide compared to only 60.87% of *E. granulosus* KDPIs containing signal peptide sequences. It is not known whether the differential KDPIs of *E. multilocularis* with signal peptides and being extracellular are associated with a more virulent pathology of AE lesion.

E. granulosus has 5 genes EG_07242, EG_07266, EG_07243, EG_09006 and EG_09008, that *E. multilocularis* does not have. Whereas, these two genes, EmuJ_001136700.1 and EmuJ_001137100.1 are only existed in *E. multilocularis* genome. These differential presented genes may play a role in the difference of pathology between the two parasites.

Two *Echinococcus* stages, the oncosphere and adult worm, are found in the gastrointestinal duct. The oncosphere is activated in the stomach and penetrates through the intestinal wall before being passed into the internal organs, whereas the adult worm spends its whole life in the gastrointestinal duct which contains high concentrations of proteases such as pepsin, trypsin and chymotrypsin. We previously showed that two KDPIs, EgKI-1(EG_08721) and EgKI-2 (EG_7242) function as protease inhibitors. EgKI-1 (also has accession number EUB56407.1) is highly expressed in the oncosphere and EgKI-2 (GenBank: EUB57880.1) is highly expressed in the adult worm [13]. These KDPIs are differentially expressed and stage-specifically protect *E. granulosus* from protease attack [12]. In this study, we showed that 11 out 25 Eg-KDPIs were highly expressed in adult worms. These Eg-KDPIs likely protect against protease attacks in the gut during adult worm development. EG_05483 and EG_08721 were relatively highly expressed in oncospheres, suggesting their expressed products might be potential vaccine candidates for use in dogs against adult worm of *E. granulosus*.

In this study, we did not find any KDPIs that were differentially and highly expressed in protoscoleces, although a previous study described a multigene family of eight (EgKU1-EgKU8) secreted Kunitz proteins from *E. granulosus* protoscoleces preferentially expressed by pepsin/H (+)-treated worms [14].

The secondary structures of proteins, especially the α -helix and β -strands play key roles in molecular function, cell stability, mechanical signaling, and tissue constitution as? random coils are easily folded and exposed to the protein surface [18]. The basic structure of a Kunitz peptide domain contains a typical sequence with six highly conserved cysteine residues connecting 3 disulfide bridges (C1-C6, C2-C4 and C3-C5) which stabilizes the protein structure. Among the disulphide bonds, the C1-C6 and C3-C5 bridges are required for the maintenance of native confirmation [19],whereas the C2-C4 bond stabilizes the folded structure [20]. We found 10 sequences had lost the #2 cysteine, including 5 from *E. granulosus*, indicating no C2-C4 bridge in these proteins. It is not known whether these 5 proteins formed different bridges impacting on the function of these KDPIs, indicating that these genes may have a different functional role.

Hydrophilicity analysis showed that the Em- and Eg-KDPIs have high hydrophobicity, which is a typical characteristic of membrane proteins. The transmembrane regions consist of 20 hydrophobic amino acids, which could have an anchoring effect on cell membranes.

We previously showed that EgKI-1 is highly expressed in the oncosphere, indicating this protein helps protect this stage from digestion by trypsin, chymotrypsin and pancreatic elastase before it penetrates the intestinal wall.

Conclusion

In conclusion, based on whole genome analysis, 19 and 23 Kunitz domain protease inhibitors were identified in the two *Echinococcus* species and these included single and multi-domain inhibitors. The differential expression of these KDPIs in different developmental stages of *E. granulosus* suggests they may have different in regulation of host immune responses, but further investigations will be required to determine precisely what roles they play in echinococcal development as such information may provide new insights for the prevention and treatment of both cyst and alveolar echinococcosis.

Materials And Methods

Identification of *E. granulosus* and *E. multilocularis* Kunitz domain sequences

The *E. granulosus* and *E. multilocularis* genomes were previously completed by the Chinese National Human Genome Centre in Shanghai (CHGC) and the Wellcome Sanger Institute, United Kingdom in 2013 [17, 21]. Based on the DNA genomic sequences, 11325 and 10429 genes were predicted for *E. granulosus*, and *E. multilocularis*, respectively. The InterProScan program (<https://www.ebi.ac.uk/interpro/result/InterProScan/>) and Motif scan (https://myhits.sib.swiss/cgi-bin/motif_scan) were used to identify the Kunitz domain protease inhibitor (KDPI) sequences.

Physiological/biochemical characters

The physiological/biochemical characters of KDPIs including molecular weight, isoelectric point and instability index were predicted using the ProtParam online software (<http://web.expasy.org/protparam/>). Signal peptides were predicted with the SignalP 5.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>). Post-translational modification sites were identified by MotifScan (http://hits.isb-sib.ch/cgi-bin/motif_scan/).

The conservative structural domain of each KDPI was predicted using the Conserved Domain program (<http://www.ncbi.nlm.nih.gov/cdd/>); their subcellular localization was predicted using ProtCompv. 9.0 (<http://linux1.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc/>). Transmembrane regions were predicted by TMPred (http://embnet.vital-it.ch/cgi-bin/TMPRED_form_parser) and TopPred 1.10 (<http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::toppred>). The hydrophilicity plot was predicted by ProtScale (<http://web.expasy.org/protscale/>).

The secondary structures of KDPIs were predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma/). These genes were annotated to the Gene Ontology (GO) database for biological process (BP), molecular function (MF), and cellular component (CC) using Blast2GO PRO (<https://www.blast2go.com/>).

Three-dimensional (3D) structures of KDPIs were constructed using the automated modeling program within the online service SWISS-MODEL. The 3D models of KDPIs were assessed by Verify_3D (http://services.mbi.ucla.edu/Verify_3D/).

The multiple sequence alignment was analyzed and ordered by Clustalomega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). The phylogenetic tree was constructed by MEGA version 6.06 (<http://www.megasoftware.net/>). The complete KDPI protein sequences of single domain KDPIs were used for phylogenetic tree analysis.

Expression of Kunitz domain inhibitors in *E. granulosus*

Transcript reads were obtained for each of the KDPI genes expressed in the adult worm, oncosphere, protoscolex and cyst (cyst membrane) of *E. granulosus* using Hiseq techniques as described [17].

Statistical analysis

Data are presented as means or median. Two-tailed Student's t test and Mann-Whitney U test was used for comparisons between two groups. Chi square test followed by Fisher's Exact Test was used to compare the sample rate (or constituent ratio) between the two groups. $P < 0.05$ was considered significant in statistical analysis.

Declarations

Acknowledgments

The authors thank Chuanchuan Wu for his help in producing the figures.

Authors' contributions

Hui Zhang and Mengxiao Tian participated in biological and physiological characteristic and phylogenetic analysis. Mengxiao Tian, Wenjing Qi and Juan Wu participated in KDPI sequence confirmation and 3 D analysis. Jun Hua, Gang Guo and Liang Zhang participated in InterProScan and Motif scan, expression and GO analysis. Jun Li and Wenbao Zhang planned the experiments and wrote the article. Shiwanthi L. Ranasinghe and Donald P. McManus were involved in the discussion the article. All authors read and approved the final manuscript.

Funding

This study was funded by the National Natural Science Foundation of China (grant numbers 81830066 and U1803282).

Availability of data and materials

All data generated or analyzed during this study are included in this published article and the additional data file.

Ethics approval and consent to participate

not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Wen H, Vuitton L, Tuxun T, Li J, Vuitton DA, Zhang W, et al. Echinococcosis: Advances in the 21st Century. *Clin Microbiol Rev.* 2019; 32(2).
2. Larrieu E, Gavidia CM, Lightowlers MW. Control of cystic echinococcosis: Background and prospects. *Zoonoses Public Health.* 2019;66(8):889-99.
3. Craig PS, Hegglin D, Lightowlers MW, Torgerson PR, Wang Q. Echinococcosis: Control and Prevention. *Adv Parasitol.* 2017;96:55-158.
4. Wa ZC, Du T, Hu HT, Lu MD. Microwave ablation combining surgery for the treatment of multiorgan cystic echinococcosis: A case report. *Parasitol Int.* 2020;74:101921.
5. Lopez-Otin C, Bond JS. Proteases: multifunctional enzymes in life and disease. *J Biol Chem.* 2008;283(45):30433-7.
6. Ranasinghe S, McManus DP. Structure and function of invertebrate Kunitz serine protease inhibitors. *Dev Comp Immunol.* 2013;39(3):219-27.
7. Ascenzi P, Bocedi A, Bolognesi M, Spallarossa A, Coletta M, De Cristofaro R, et al. The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein. *Curr Protein Pept Sci.* 2003;4(3):231-51.
8. Rawlings ND, Tolle DP, Barrett AJ. Evolutionary families of peptidase inhibitors. *Biochem J.* 2004;378(Pt 3):705-16.
9. Mishra M. Evolutionary Aspects of the Structural Convergence and Functional Diversification of Kunitz-Domain Inhibitors. *J Mol Evol.* 2020;88(7):537-48.

10. Rustgi S, Boex-Fontvieille E, Reinbothe C, von Wettstein D, Reinbothe S. The complex world of plant protease inhibitors: Insights into a Kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Commun Integr Biol*. 2018;11(1):e1368599.
11. Ranasinghe SL, Fischer K, Gobert GN, McManus DP. Functional expression of a novel Kunitz type protease inhibitor from the human blood fluke *Schistosoma mansoni*. *Parasit Vectors*. 2015;8:408.
12. Ranasinghe SL, Duke M, Harvie M, McManus DP. Kunitz-type protease inhibitor as a vaccine candidate against schistosomiasis mansoni. *Int J Infect Dis*. 2018;66:26-32.
13. Ranasinghe SL, Fischer K, Zhang W, Gobert GN, McManus DP. Cloning and Characterization of Two Potent Kunitz Type Protease Inhibitors from *Echinococcus granulosus*. *PLoS Negl Trop Dis*. 2015;9(12):e0004268.
14. Gonzalez S, Flo M, Margenat M, Duran R, Gonzalez-Sapienza G, Grana M, et al. A family of diverse Kunitz inhibitors from *Echinococcus granulosus* potentially involved in host-parasite cross-talk. *PLoS One*. 2009;4(9):e7009.
15. Krowarsch D, Dadlez M, Buczek O, Krokoszynska I, Smalas AO, Otlewski J. Interscaffolding additivity: binding of P1 variants of bovine pancreatic trypsin inhibitor to four serine proteases. *J Mol Biol*. 1999;289(1):175-86.
16. Grzesiak A, Helland R, Smalas AO, Krowarsch D, Dadlez M, Otlewski J. Substitutions at the P(1) position in BPTI strongly affect the association energy with serine proteinases. *J Mol Biol*. 2000;301(1):205-17.
17. Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, et al. The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat Genet*. 2013;45(10):1168-75.
18. Khrustalev VV. Random Coils of Proteins Situated Between a Beta Strand and an Alpha Helix Demonstrate Decreased Solvent Accessibility. *Protein J*. 2020;39(4):308-17.
19. Creighton TE. The two-disulphide intermediates and the folding pathway of reduced pancreatic trypsin inhibitor. *J Mol Biol*. 1975;95(2):167-99.
20. Laskowski M, Jr., Kato I. Protein inhibitors of proteinases. *Annu Rev Biochem*. 1980;49:593-626.
21. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496(7443):57-63.

Figures

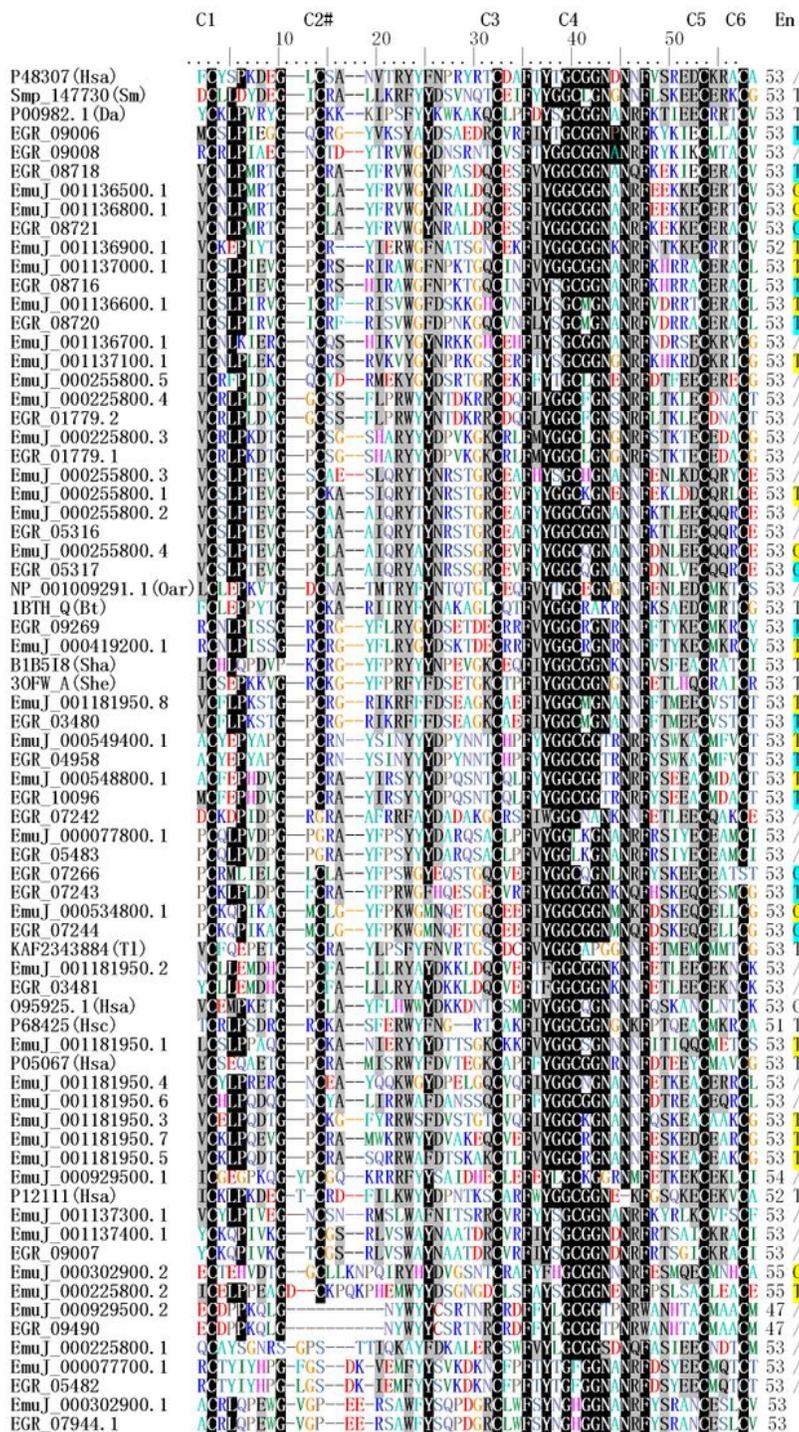


Figure 1

Alignment and clustering of *E. multilocularis* and *E. granulosus* Kunitz-type domain protease inhibitors. These inhibitor protein sequences are compared with the homologue sequences (blocked in gray background) from human including P48307(Hsa), P05067(Hsa), P12111(Hsa) and 095925(Hsa), *Ovis aries* (NP_001009291.1(Oar)), cattle (1BTH_Q(Bt)), *Schistosoma mansoni* (Smp_147730 (Sm)), *Dendroaspis angusticeps* (P00982.1(Da)), *Stichodactyla haddoni*(B1B5I8(Sha)), *Stichodactyla helianthus*

(30FW_A(She)), *Trinorchestia longiramus*(KAF2343884(TI)), and *Haplopelma schmidt* (P68425(Hsc)). "En", "T", "C" and "/" represent enzyme and inhibitor of trypsin, chymotrypsin, and nonprediction, respectively.

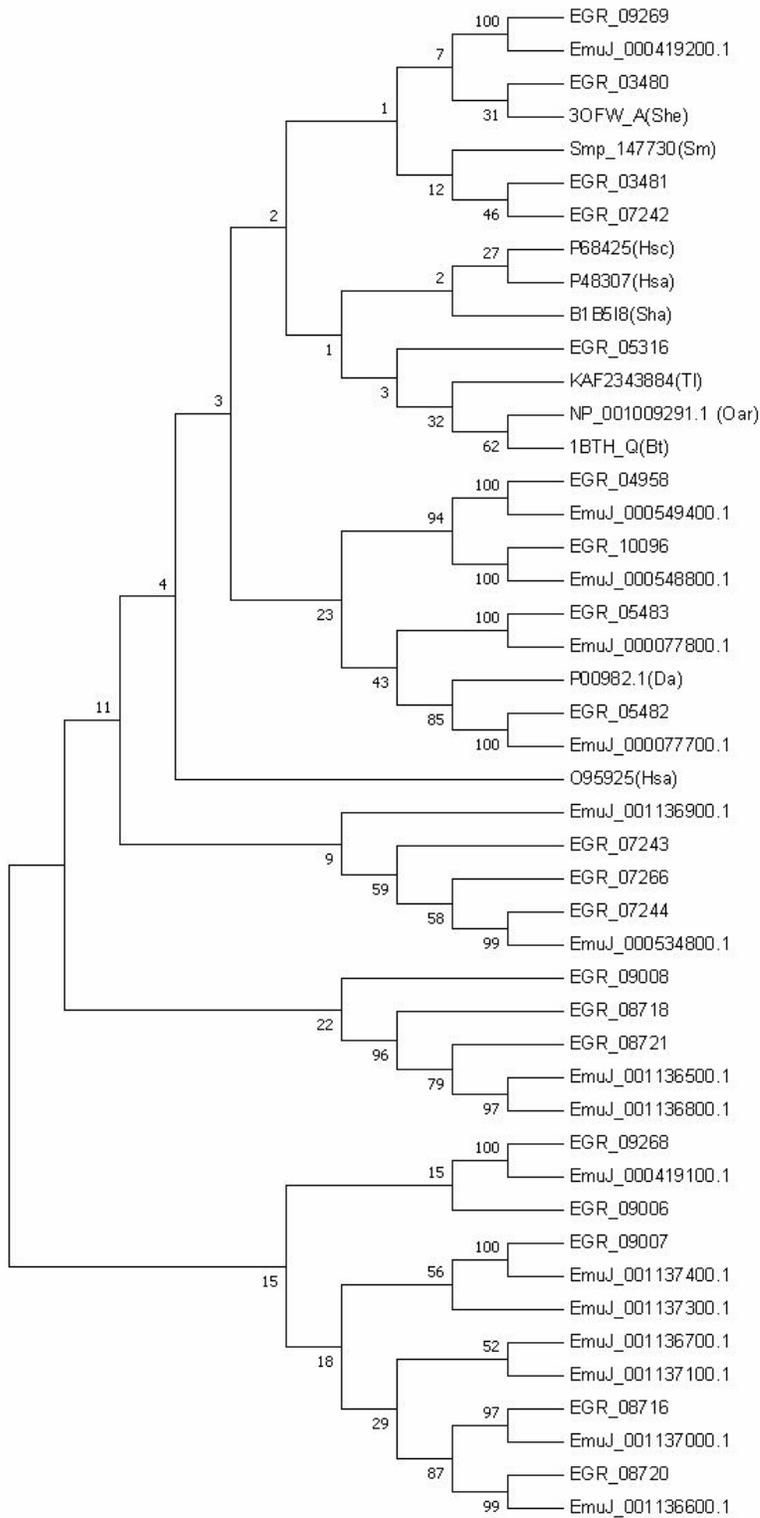


Figure 2

Phylogram constructed using the neighbor joining method to compare the sequences of *E. granulosus*- and *E. multilocularis*-KDPIs with KDPIs from bovine and humans and other species. The accession

numbers for sequences included in the phylogenetic analysis are shown after the branches. The bootstrap distance values are shown after the accession numbers. Differentially expressed genes of *E. granulosus* and *E. multilocularis* are blocked in grey.

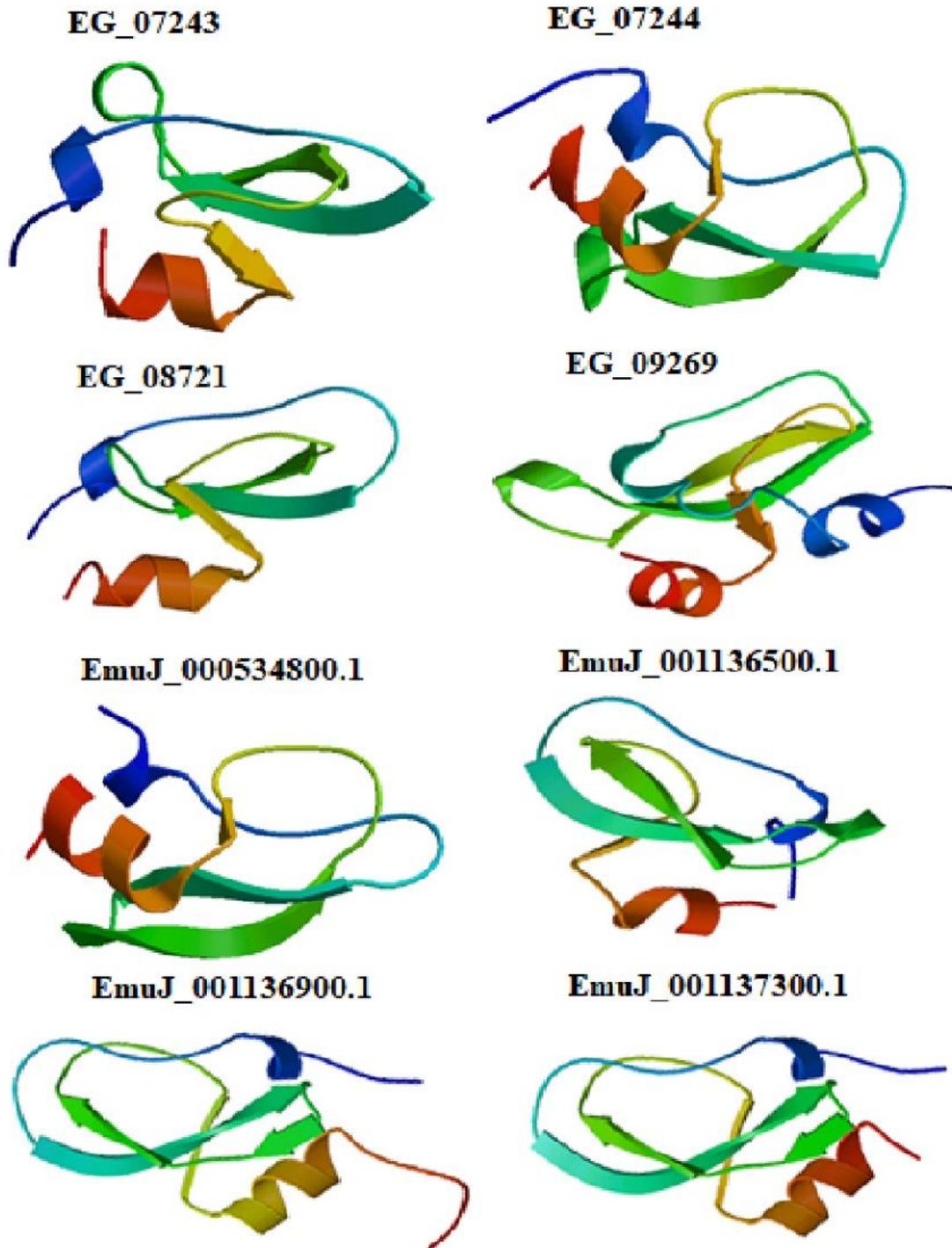


Figure 3

Three dimensional structures of single Kunitz domain protease inhibitors in *E. granulosus* and *E. multilocularis* using SWISS-MODEL.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.tif](#)