

Table S1. The *E. granulosus* Kunitz protein family.

Name	cDNA sequence data ^a			Predicted protein parameters ^b					Similar Kunitz domain ^e	Close platyhelminth proteins ^e
	Cluster ID (reads)	cDNA source	Accession numbers	Consensus (nt)	SP/MP (aa)	MW (Da) ^c	pI	charge		
EgKU-1	EGC000603 (6)	PS/peps/H ⁺	CN650596, CN650753, CN650941, CN652374, CN652433, BI244287	359	20/58 ^d	6599.5	9.2	+6	sp P00985 IVB2_HEMHA - [23-75] 1e ⁻¹⁰ Venom Basic Inhibitor 2 <i>H. haemacathus</i>	_9
EgKU-2	EGC000598 (5)	PS/peps/H ⁺	CN650229, CN650436, CN650784, CN652059, BI244281	626 (long 3'UTR)	20/101	12258.8	8.8	+4	tr Q8NT7L9 BOMMO - [28-81] 0.0009 Serine Protease Inhibitor 1 <i>Bombyx mori</i>	EL746785 ^h - <i>T. solium</i> larva 85% identity (98/115 aa)
EgKU-3	EGC001164 (1)	PS/peps/H ⁺	BQ173695	354	18/57	6405.4	5.1	-1	sp P31713 ISH1_STOHE - [23-73] 6e ⁻¹³ Kunitz Inhibitor 1 <i>Stoichactis helianthus</i>	XvEMa04137 - <i>E. multilocularis</i> larva 98% identity (74/75 aa)
EgKU-4	EGC004311 (1)	PS/peps/H ⁺	CN652174	535 (long 5'UTR)	20/64	7288.4	10.1	+14	sp P00978 AMBP_BOVIN - [25-75] 7e ⁻¹² Alpha-1-microglobulin inter-alpha-trypsin inhibitor light chain <i>Bos taurus</i>	XvEMa03312 - <i>E. multilocularis</i> larva 92% identity (78/84 aa) EL763407 - <i>T. solium</i> adult 63% identity (53/84 aa) ⁱ
EgKU-5	EGC002977 (2)	PS PS/peps/H ⁺	CN649551 CN650837	317	29/60	7239.3	9.6	+9	tr O54819 TFPI_MOUSE - [37-87] 2e ⁻¹² Tissue factor pathway inhibitor 1 <i>Mus musculus</i>	BW635664 - <i>D. ryukyuvensis</i> adult 60% identity (31/51 aa, KU domain) DN300487 ^j and DN307650 ^j - <i>S. mediterranea</i> mature hermaphrodites 56% identity (29/51 aa, KU domain)
EgKU-6	EGC003000 (2)	PS	CN649590, CN649996	416	18/82	9778.9	6.9	0	tr O45881 CAEEL - [24-74] 7e ⁻¹² Protein W01F3.3, partially confirmed <i>Caenorhabditis elegans</i>	EL743839 - <i>T. solium</i> larva 75% ident. (75/100 aa)
EgKU-7	EGC002672 (1)	PS/peps/H ⁺	CN648947	327 (no 3'UTR)	20/79	9108.0	5.1	-3	sp Q868Z9 PPN_DROME - [25-75] 7e ⁻¹² Papilin (precursor) <i>D. melanogaster</i>	_9
EgKU-8	-	PS/peps/H ⁺	FJ031017	228	18/57 ^d	6519.4	7.8	+1	tr Q8WPI2 BOOMI - [23-73] 1e ⁻¹² <i>Boophilus microplus</i> boophilin	XvEMa16368 - <i>E. multilocularis</i> larva 98% identity (74/75 aa)

^aFor *EgKU-1-EgKU-7*, identified while surveying the protoscolex transcriptome, data include: i) the EST cluster identification in PartiGeneDB (<http://www.compsysbio.org/partigene/>), EGC00XXXX and the number of reads in the cluster; ii) the libraries from which ESTs were derived, designed as: PS, untreated protoscolexes; PS/peps/H⁺, protoscolexes treated with pepsin at pH 2; iii) the NCBI accession number of ESTs in the cluster; iv) the size and relevant features of the consensus sequence (except for *EgKU-7*, this accounts for the full-length cDNA). For *EgKU-8*, data include: i) the source of RNA used for RT-PCR; ii) the NCBI accession number of the cDNA cloned on the basis of the putative *E. multilocularis* orthologue, as detailed in the text; iii) the size of the full coding sequence thus isolated.

^bAnalyses were carried out with the tools available at the ExPASy Proteomics Server (<http://expasy.org.ch/>): "Translate", for prediction of protein sequences; "SignalP", for prediction of signal peptides (SP), signal peptidase cleavage sites and, correspondingly, mature proteins (MP); and "ProtParam", for calculation of the theoretical MW (average) and pI of predicted mature proteins; the charge was estimated as the difference between positively and negatively charged amino acids. The consensus sequence of clustered ESTs was used for *EgKU-1-EgKU-7*.

^cThe theoretical MW of predicted mature proteins (estimated as detailed in b) minus 6 Da is indicated, under the assumption that the six conserved Cys residues of the Kunitz domains form disulphide bridges in native proteins.

^dThe experimentally determined N-terminus of the native protein directly isolated from protoscolexes confirmed the SignalP prediction.

^eThe predicted full-length *EgKU-1-EgKU-8* were analyzed by InterProScan at ExPASy and found to hit InterPro 002223: Proteinase inhibitor I2, Kunitz metazoa. The Kunitz domain showing the highest score in PD000222 was recorded for every protein. Data include: the accession number in Swiss-Prot (sp) or TrEMBL (tr) and name of the protein bearing the domain, the score and the similarity region in the *E. granulosus* protein.

^fRelated platyhelminth proteins were identified blasting the predicted full-length *EgKU-1-EgKU-8* against 'non-human, non-rodent ESTs' in dbEST (option 'est_others' at the NCBI server), and *E. multilocularis* ESTs available at 'Full-Echinococcus' (<http://fullmal.hgc.jp/em/docs/blast.html/>). For each putative protein, data include: i) an identifier of the corresponding EST - the NCBI accession number for *T. solium*, *D. ryukyuvensis* and *S. mediterranea*; and the clone name in Full-Echinococcus for *E. multilocularis*; ii) the cDNA library from which each EST was derived; iii) the level of identity with the corresponding *E. granulosus* protein and the length of the alignment. In the case of *EgKU-5*, identity is over the 51 aa of the Kunitz domain (between conserved Cys). Only the predicted protein showing the highest similarity with every *E. granulosus* molecule is indicated for each platyhelminth.

^gProteins close to *EgKU-1* and *EgKU-7* were the same as for *EgKU-4* and *EgKU-6*, respectively, but the identity level was lower than for the latter.

^hEST does not contain the full coding sequence: as shown in Fig 1b (*Ts-2*), the sequence translated from EL746785 has no starting Met.

ⁱTwo additional *T. solium* ESTs with about 60% identity with *EgKU-4* were identified: EL763445 (adult) and EL740737 (larva).

^jThese ESTs were assigned to transcript mk4.029059.00.01 in the *S. mediterranea* Genome Database (SmedGD: <http://smedgd.neuro.utah.edu>).