


# Identification and characterization of a cathepsin L-like cysteine protease from *Rhipicephalus (Boophilus) annulatus*

Shahin Saidi<sup>1</sup> · Sedighe Nabian<sup>1,2</sup> · Elahe Ebrahimzade<sup>1</sup> ·  
Ali Najafi<sup>3</sup> · Mehrdad Moosazadeh Moghaddam<sup>4</sup> ·  
Alireza Sazmand<sup>5</sup>  · Masoud Torkzadeh-Mahani<sup>6</sup> ·  
Saeed Sattari Tabrizi<sup>1</sup>

Received: 10 August 2015 / Accepted: 5 November 2015 / Published online: 23 November 2015  
© Springer International Publishing Switzerland 2015

**Abstract** The tick *Rhipicephalus (Boophilus) annulatus* is one of the most important ectoparasites of bovines and is responsible for the transmission of different pathogens such as *Babesia* and *Anaplasma*. Cysteine proteases are involved in several host-tick interactions including invasion of host tissues, immune evasion, pathogen transmission,

---

✉ Sedighe Nabian  
nabian@ut.ac.ir

Shahin Saidi  
shahin.saydi1366@yahoo.com

Elahe Ebrahimzade  
eebrahimzade@ut.ac.ir

Ali Najafi  
najafi74@yahoo.com

Mehrdad Moosazadeh Moghaddam  
mm.genetics@gmail.com

Alireza Sazmand  
alireza\_sazmand@yahoo.com

Masoud Torkzadeh-Mahani  
mtmahani@gmail.com

Saeed Sattari Tabrizi  
saeed\_satari@ut.ac.ir

<sup>1</sup> Department of Parasitology, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran

<sup>2</sup> Iranian Research Center for Tick and Tick-borne Diseases, Tehran, Iran

<sup>3</sup> Molecular Biology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

<sup>4</sup> Applied Biotechnology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

<sup>5</sup> Department of Agriculture, Payame Noor University, Tehran, Iran

<sup>6</sup> Department of Biotechnology, Institute of Science, High Technology and Environmental Science, Graduate University of Advanced Technology, Kerman, Iran

embryogenesis and blood digestion. In this study, the gene encoding *R. annulatus* cathepsin L-like enzyme (RaCL1) was cloned into pTZ57R/T vector, sequenced and analyzed using bioinformatics approaches. The nucleotide length of RaCL1 was 999 bp. Bioinformatics analysis showed 332 amino acids with an approximate molecular weight of 36.3 kDa which contained a signal peptide sequence (18 amino acids), pro-region (97 amino acids) and mature enzyme (217 amino acids). Multiple sequence alignment of the RaCL1 revealed high similarity to cathepsin L-like cysteine proteases from other tick species such as *Rhipicephalus (Boophilus) microplus* and *Amblyomma variegatum*. Based on bioinformatics analyses, results of this work suggest that RaCL1 can be a suitable candidate for the development of vaccine against *R. annulatus*.

**Keywords** *Rhipicephalus (Boophilus) annulatus* · Cathepsin L · Cysteine protease · Gene expression · cDNA · Bioinformatics

## Introduction

The tick *Rhipicephalus annulatus* (formerly *Boophilus annulatus*) is a circum-Mediterranean species with distribution in sites with adequate moisture content. Bovidae are the preferred hosts but this tick species can also be found on several mammalia species including humans. *R. annulatus* is the principal vector of *Babesia bovis* and *Babesia bigemina* in the middle East and has been recognized or considered so far as potential vector for several rickettsial agents such as *Anaplasma marginale*, *Rickettsia aeschlimannii*, *Rickettsia africae* and *Rickettsia sibiricamongolitimonae* (Estrada-Peña et al. 2004; Parola et al. 2013). More than pathogen transmission heavy burdens of ticks on livestock can cause economic losses due to effects on leather, meat and milk production. Currently using acaricides is the most common route for tick control. However, due to different degrees of resistance of ticks to chemicals, residues in livestock products and environmental contamination, alternative control strategies including tick vaccines have been developed recently. The goal of ongoing research in the field of tick vaccine development is to strengthen the effect of reduction of tick infestations and tick-borne pathogen transmissions.

Different proteinases have significant roles in tick biology and physiology. Cysteine proteases with involvement in several remarkable tick-host reactions including hemoglobin digestion in the gut (Renard et al. 2002), vitellin degradation (Seixas et al. 2003; Estrela et al. 2007, 2010) and pathogen transmission are considered as candidates for chemotherapy and immunoprophylaxis (Willadsen and Kemp 1988; Sajid and McKerrow 2002; Renard et al. 2002). Cathepsin L which is synthesized as a proenzyme is one of the cysteine proteases belonging to the family C1 (cathepsin L- and cathepsin B-like). The mature form of cathepsin L is shown to be involved in tissue remodeling, immune system as well as modulation and alteration of cell function (Dickinson 2002), and its functional activities have been described in several tick species such as *R. annulatus* (Taheri et al. 2014a), *Rhipicephalus (Boophilus) microplus* (Renard et al. 2000, 2002; Seixas et al. 2003; Estrela et al. 2007), *Ixodes ricinus* (Franta et al. 2011), *Haemaphysalis longicornis* (Tsuji et al. 2008; Yamaji et al. 2009) and *Ornithodoros moubata* (Fagotto 1990). Several other pathogenic proteins and enzymes of *R. annulatus* ticks such as serine protease, vitellogenin and tropomyosin have been studied (Nikpay et al. 2012; Nabian et al. 2013, 2014; Ranjbar et al. 2013, 2014, 2015; Taheri et al. 2014b).

Genetic analyses of genes that encode essential tick proteins are helpful in developing novel control strategies. Accordingly, in this study for the first time we have done sequence analysis and characterization of the *R. annulatus* cathepsin L-like (RaCL1) gene from two

geographical regions of Iran. The structures of the extracted sequences were analyzed, and prediction of the structure, function and antigenic epitopes of the enzyme were done by bioinformatics approaches.

## Materials and methods

### Parasite specimens

The first engorged female *R. annulatus* ticks were collected from infested cattle in Guilan and Mazandaran provinces in Northern Iran as described before and reared in Hereford cattle (Brown et al. 1984). This tick species is distributed only in the Caspian Sea region of Iran (Rahbari et al. 2007). After fully feeding and detachment they were kept in glass tubes in incubator at 28 °C and 85 % relative humidity for oviposition. The eggs were collected and placed in the same condition until hatching. Then larvae were separated at 20th day after hatching and stored at –70 °C until further use.

### RNA isolation and RT-PCR

Total RNA was obtained from two grams of grounded *R. annulatus* larvae by Tripure Isolation Reagent<sup>®</sup> (Roche, Basel, Switzerland) according to the manufacturer instructions. One microgram of total RNA was used for synthesis of cDNA by means of Power cDNA Synthesis Kit<sup>®</sup> (Intron Biotechnology, Kyungki-Do, Korea). For cDNA synthesis 0.5 µl AMV-RT enzyme (10 U/µl), 1 µl oligo (dT) 15 primer (0.2 mM), 2 µl RNase free dNTP (10 mM), 1 µl RNase inhibitor (10 U/µl), 2 µl DTT (0.1 M) and 4 µl 5× RT buffer in an end volume of 20 µl were used. The cDNA synthesis was performed at 42 °C for 1 h followed by 70 °C for 5 min. The amplification of cDNA was performed using primers derived from the nucleotide sequence of *R. microplus* cathepsin L-like proteinase precursor (BmCL1) mRNA (GenBank<sup>®</sup> accession number AF227957.1). Nucleotide sequences of primers were F: 5'ATGCTTAGATTAAGCGTACTTTG3' and R: 5'TTAGAC GAGGGGGTAGCTGGCC3'. Then 1 µl of cDNA was amplified using specific primers under the following conditions: initial denaturing at 95 °C for 5 min, followed by 35 cycles of 95 °C for 45 s and 67 °C for 1 min, 72 °C for 45 s and final extension step at 72 °C for 10 min. *R. annulatus* COI sequences derived primers were used as positive control.

### Cloning into pTZ57R/T vector

One hundred µl of RT-PCR product was purified with PCR Product Purification Kit (MBST, Tehran, Iran) according to the manufacturer protocol. Three µl of the purified product was added to the solution containing 1 µl of plasmid vector pTZ57R/T DNA (0.18 pmol ends), 2 µl of 5× ligation buffer, 0.33 µl T4 DNA ligase and 3.7 µl deionized water, then incubated at 4 °C for 15 h. Preparation of competent cells of *Escherichia coli* DH5α strain was performed and the recombinant plasmid was transferred to the competent *E. coli* DH5α. The *E. coli* clones were separately grown (37 °C, under shaking condition) and the plasmids were isolated using Plasmid Isolation Kit<sup>®</sup> (MBST).

## Comparative nucleotide and amino acid sequence analysis

The recombinant plasmid containing cathepsin L-like was extracted using the Plasmid Extraction Kit<sup>®</sup> (MBST) as described by the manufacturer and sequenced by Sanger Cycle Sequencing (Takapouzist, Tehran, Iran). The nucleotide sequences of cathepsin L-like from geographical isolate of *R. annulatus* (Guilan and Mazandaran isolates) were analyzed by Emboss needle tool ([http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)). The nucleotide sequences of *R. annulatus* cathepsin L-like mRNA from two geographical regions (Mazandaran and Guilan) were submitted to the ExPASy translate tool, deduced amino acid sequences of cathepsin L-like were directed to NCBI BlastP program and similarity to other sequence records in GenBank<sup>®</sup> were evaluated.

## Phylogenetic analysis of *Rhipicephalus annulatus* cathepsin L

Available cathepsin L-like amino acid sequences from several ticks in the NCBI database (<http://ncbi.nlm.nih.gov>) were aligned with the resulted sequences in this study using Clustal W algorithm. Phylogenetic analysis was performed with MEGA5<sup>®</sup> software, applying the neighbor-joining method with bootstrap analysis (1000 replicates).

## Cathepsin protein structure modeling

To our knowledge, no three dimensional structure has been recorded for *R. annulatus* cathepsin L-like protein, therefore for 3D structure prediction, homology modeling was separately applied to construct the 3D proteins using I-Tasser server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>).

## Predicting antigenic propensity and B cell epitope

Several methods based on various physiochemical properties of amino acid sequences and protein structure such as flexibility, hydrophobicity and accessibility have been developed for the prediction of antigenic regions and epitopes (Kyte and Doolittle 1982; Gershoni et al. 1997). Antigenic activity of cathepsin L-like coding region were determined using hydrophobicity plot and Kolaskar–Tongankar method for prediction of hydrophilic regions and linear epitopes from protein sequence (Kyte and Doolittle 1982; Kolaskar and Tongankar 1990). SoEMBOS Pepwindow<sup>®</sup> ([http://www.ebi.ac.uk/Tools/seqstats/emboss\\_pepwindow/](http://www.ebi.ac.uk/Tools/seqstats/emboss_pepwindow/)) and EMBOS Pepinfo<sup>®</sup> ([http://www.ebi.ac.uk/Tools/seqstats/emboss\\_pepinfo/](http://www.ebi.ac.uk/Tools/seqstats/emboss_pepinfo/)) tools were used. Also, for the prediction of hydrophobicity and B cell epitopes, IEDB web-based tools were used. Then, the Swiss-PdbViewer<sup>®</sup> (<http://spdbv.vital-it.ch/>) was used for rendering and mapping of the predicted epitopes on 3D structure (Zhang et al. 2008).

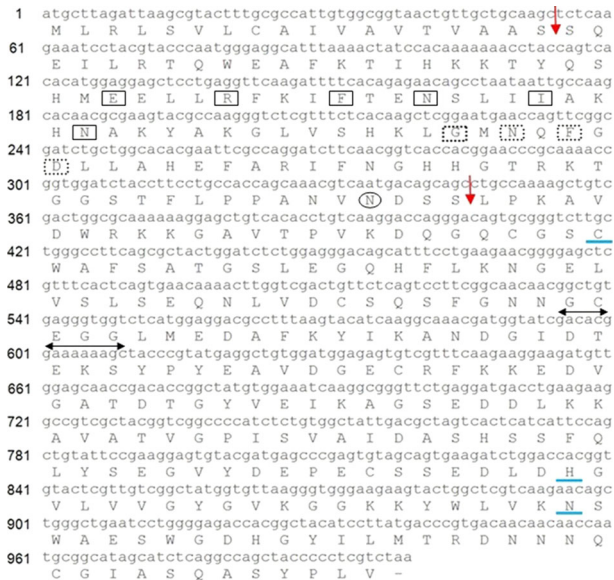
## Results

The nucleotide sequences of *R. annulatus* cathepsin L-like from two geographical regions (Guilan; GenBank<sup>®</sup> accession number: KM272201.1 and Mazandaran; GenBank<sup>®</sup> accession number: KM272202.1) showed 99.7 % identity in pairwise sequence alignment analyzed by Emboss-needles tools. There were three mismatch nucleotides between Guilan

and Mazandaran isolates at positions of 171 (Guilan: A, Mazandaran: C), 615 (Guilan: G, Mazandaran: A) and 633 (Guilan: A, Mazandaran: C). Nucleotide sequences of *R. annulatus* cathepsin L-like from two geographical regions of Iran showed 98 % identity with *R. microplus*: AFQ98385.1 (NCBI), KC707946 (UniProt) and AAF61565.1 (NCBI). They also had 90 % identity with *Rhipicephalus appendiculatus* (AY208824), 88 % with *Rhipicephalus haemaphysaloides* (AY336797), 83 % with *Dermacentor variabilis* (EU025855) and 83 % with *Hyalomma anatolicum anatolicum* (KC707937).

A long open reading frame which encodes 332 amino acids was obtained after the nucleotide sequences were submitted to the NCBI ORF Finder®. The deduced amino acid sequences showed 100 % identity. The nucleotide sequences of *R. annulatus* cathepsin L-like and its amino acid sequences are shown in Fig. 1.

Alignment of the *R. annulatus* cathepsin L-like sequences showed the similarity to the peptidase C1 family with available sequences in the Uniprot® database. It consists of a pre-region, pro-region and mature enzyme containing 18, 97 and 217 amino acids respectively. Bioinformatics analysis of RaCL1 showed 332 amino acids with an approximate molecular weight of 36.33 kDa which contained a signal peptide sequence (1.8 kDa), pro-region (11.06 kDa) and mature enzyme (23.47 kDa). The potential cleavage sites for releasing of the pre-region and pro-region are placed at Ser<sup>18</sup>–Ser<sup>19</sup> and Ser<sup>115</sup>–Leu<sup>116</sup> respectively. The conserved amino acid residues Cys<sup>25</sup>, His<sup>164</sup> and Asn<sup>184</sup> that are involved in the catalysis of *R. annulatus* cathepsin L-like are present in our sequences. The sequences around these residues (Cys<sup>25</sup>, His<sup>164</sup> and Asn<sup>184</sup>) are also conserved in cysteine proteinases. Like in peptidase C1 family six cysteine residues involved in disulfide bond formation (Cys<sup>22</sup>, Cys<sup>56</sup>, Cys<sup>65</sup>, Cys<sup>98</sup>, Cys<sup>157</sup> and Cys<sup>206</sup>) were observed in the present protein. Similar to



**Fig. 1** Nucleotide and the deduced amino acid sequences of *Rhipicephalus annulatus* cathepsin L-like gene (RaCL1). The arrows indicate the cleavage sites for the pre and pro enzyme, respectively. The underlined amino acids represent the conserved residues which are involved for active site information. ERFNIN and GNFD motifs are indicated by the boxes and dotted boxes respectively. The ellipse presents a potential glycosylation site. The GCEGG motif was showed by bidirectional arrow. (Color figure online)

the most of the members of cysteine protease family our sequence also contains a proline residue at position 2 in the mature enzyme (Lee et al. 2012). Based on multiple alignment amino acid sequences of *R. annulatus* cathepsin L-like from Iran showed 98 % identity with *R. microplus* (AFQ98385.1 and AAF61565), 88 % with *R. haemaphysaloides* (AAQ16117), 82 % with *D. variabilis* (ABS70713), 78 % with *R. appendiculatus* (AAQ60048.1), 76 % with *H. longicornis* (BAH86062.1) and 66 % with *A. variegatum* (DAA34687.1) with query cover 99 % in BlastX. Moreover, it showed 87 % identity with *H. anatolicum* (AFQ98384.1) and 68 % with *I. ricinus* (ABO26562.1) with query cover less than 99 % in BlastP. Amino acid changes are shown in Fig. 2. Phylogenetic analysis suggested a closer genetic relationship between RaCL1 and *R. microplus* cathepsin compared with others (Fig. 3).

The prediction of antigenic sequences was performed by determination of high hydrophilic regions in protein sequence as accessible regions and analysis of the B cell epitopes and the results are shown in Figs. 4 and 5. Based on hydrophobicity a total of four regions in protein sequence were identified as areas that are exposed to the protein surface which therefore can be considered as potential antigenic sequences. The B cell epitope analysis for RaCL1 identified several regions as epitopes but according to max\_scores, 4–23 and 278–287 regions have higher score in comparison with other regions (1.244 and 1.187) (Fig. 5a, b). 3D structure prediction showed that these regions have helix and coil structures and are accessible on the surface of protein (Fig. 5c).

## Discussion

Cysteine proteases in parasites have important roles in several vital phenomena including host entry, feeding, host immune responses, parasite development and homeostasis (Li et al. 2006; Dou and Carruthers 2011; Kissoon-Singh et al. 2011). In ticks most of the cysteine proteases belong to the papain-like super family and are essential for the development (Sojka et al. 2011) and hemoglobin digestion (Franta et al. 2011; Sonenshine and Roe 2014). Hence characterization of these enzymes and studying their role in physiological and biological procedures is crucial for chemotherapy and control strategies. In this study the gene encoding RaCL1 was identified and analyzed using bioinformatics approaches. The identity between tick isolates of two localities was 99.7 % at nucleotide level. Despite slight differences presented in cathepsin L-like nucleotide sequences, the deduced amino acid sequence showed 100 % identity which indicates no polymorphism is present between RaCL1 in ticks of different regions. So cathepsin L-like could be suggested potentially as a target for chemotherapy or vaccination. The calculated molecular weight for RaCL1 is 36.3 kDa which is approximately similar with *R. microplus* and *H. longicornis* (Renard et al. 2000; Yamaji et al. 2009). Also 98 % identity was found in amino acids sequence of RaCL1 and cathepsin-L like cysteine proteases of *R. microplus* (AFQ98385.1 and AFQ98389.1). It should be mentioned that there was no records of *R. annulatus* cathepsin protease in the sequence databases. Based on this topic a similar study has been done by (Li et al. 2006), for identification and characterization of cathepsin L-like cysteine protease from *Taenia solium* metacestode. Alignments of *T. solium* cathepsin L showed low similarity to other helminths at nucleotide sequences.

In contrast of these results, RaCL1 amino acid sequences are similar to other cathepsin-L like cysteine proteases but its identity is higher to ticks cathepsin L-like proteases especially *R. microplus* (AFQ98385.1 and AFQ98389.1) amino acid sequence with 98 %

CLUSTAL O(1.2.1) multiple sequence alignment

```

ABO26562      mlrclvlclllvaa--aaavsyqevlgaewsafkakhgksyvseteevfrlkiymenrhki
DAA34687      mrgfvvlcfcicaamtaaaithqelvgawsafkalhgkeyqsseteeyrklkiymenrmmi
AAO60048      mlrlsllcaivav--tvaansqeilrtqweafkthhkksyqsnmeellrfkiftenslii
BAH86062      mlrlaflcgcvaav--aiaassqeilrtqweafkshqhkaysshveellrfkiftentllv
ABS70713      mlrlsllcafvvv--ttaassheilrtqweafkathhkksyqsnmeellrfkiftensllv
Bacl-Iran     MLRLSLVLCAlVAV--TVAASSQEILRTQWEAFKTHHKTYQSHMEELLRFKIFTENSLII
AFQ98385      mlrlsvlcaiaav--tvaassqeilrtqweafkthhkktvqshmeellrfkiftenslii
AAF61565      mlrlsvlcaivav--tvaassqeilrtqweafkthhkksyqshmeellrfkiftenslii
AAQ16117     mlrlsllcaivav--tvaansheilrtqweafkthhkksyqshmeellrfkiftenslii
AFQ98384     -----eilrtqweafkthhkksyqskmeellrykiftenslii
                *::: :*:.*:*: * . * * . * * * * * * * * * * * * * * * * * * * * * *
ABO26562      akhnekyargevpsymamnefgdmlhhefvstrngfkrnykdqpregstylepeniedfs
DAA34687      arhnekyannkvsyklamneygdmlhhefvstrngfrdryrskprqgsfyiepegiedkh
AAO60048      akhnakyakglvsyklgmngqfgdlllahefakifngyhgka--hnr-ridlpatsnindss
BAH86062      akhnakyakglvsyklamnkfgdllphefakmvnyrgkq--nkeqrptfippanlndss
ABS70713      arhnekyarglvsyklgmngqfgdllphefarmfnyrgar--tagrgstflppanvnvss
Bacl-Iran     AKHNAYAKGLVSHKLGMMNQFGDLLAHEFARIFNGHHGTR--K-TGGSTFLPPANVNDS
AFQ98385      akhnakyakglvsyklgmngqfgdlllahefari fnghhgtr--k-tggs flppanvnvss
AAF61565      akhnakyakglvsyklgmngqfgdlllahefari fnghhgtr--k-tggs flppanvnvss
AAQ16117     akhnakyakglvsyklgmngqfgdlllahefakifngyrgqr--t-srgstfmppanvnvss
AFQ98384     akhnakyakglvsyklgmngqfgdllphefakmfngygher--k-grgstflppanvnvss
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
ABO26562      lpktvdwrkkgavtpvknqggcgscwafsatgslsegghfrksgsmvslseqnlvgcstdf
DAA34687      lpktvdwrkkgavtpvknqggcgscwafsttgslegqghfrksgdvmvslseqnlvdcstaf
AAO60048      cqvc-rlaerschrtdqggcgscwafsatgslsegqghllkdgelvsseqnlvdcsqsf
BAH86062      lpttdwrkkgavtpvknqggcgscwafsttgslegqghfrktgklvslseqnlvdcsddf
ABS70713      lpktvdwrkkgavtpvknqggcgscwafsttgslegqghflkngelvsseqnlvdcsdtdf
Bacl-Iran     LPKAVDWRKKGAVTPVKDQGGCGSCWAFSATGSLSEGQHFLKNGELVLSLSEQNLVDCSQSF
AFQ98385      lpkavdwrkkgavtpvknqggcgscwafsatgslsegghflkngelvsseqnlvdcsqsf
AAF61565      lpkvdwrkkgavtpvkdqggcgscwafsatgslsegghflkngelvsseqnlvdcsqsf
AAQ16117     lpstvdwrkkgavtpvkdqggcgscwafsatgslsegghflkngelvsseqnlvdcsqsf
AFQ98384     lpktvdwrkkgavtpvkdqggcgscwafsatgslsegghflksgklvslseqnlvdcsqsf
                : : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
ABO26562      gnngeggllmdafkyirankgidteksypngtdgtchfkkstvgatdsgfvdi kegse
DAA34687      gnngeggllmdnafkyikanggidteksypngtdgtchfkkstvgatdsgfvdipegne
AAO60048      gnngeggllmdnafnyikandgidteegyypaevdgecrfkkeedvgatdtgfvdiqggie
BAH86062      gnngeggllmdngfkyikanggidteeshpytaqgdgckfkkadvgatdagfvdiqggse
ABS70713      gnngeggllmdnafkyikanggidteksypyeavdgecrfkkeedvgatdtgfvdiqggse
Bacl-Iran     GNNCEGGLMEDAFKYIKANDGIDTEKSYPYEAVDGECECRFKKEDVGTDTGYVEIKAGSE
AFQ98385      gnngeggllmedafkyikandgidteksypyeavdgecrfkkeedvgatdtgyveikagse
AAF61565      gnngeggllmedafkyikandgidteksypyeavdgecrfkkeedvgatdtgyveikagse
AAQ16117     gnngeggllmdnafkyikandgidaeesyypaeamddkcrfkkeedvgatdtgfvdiqggse
AFQ98384     gnngeggllmdnafkyikandgidteesyypaeamddkcrfkkeedvgatdtgfvdiqggse
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
ABO26562      tqlkkavatvgrisvaidashesfqfysdgydepedcsesldhgvlvvgygtlngtdyw
DAA34687      dlkkavatvgrisvaidashsqsfqfysgvydepedcsesldhgvlvvgygtkddqdyw
AAO60048      ddllkkaasfcwppppv--lwrspssfqlysegyydepedcsesldhgvlvvgyvkngkkyw
BAH86062      ddllkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
ABS70713      ddllkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
Bacl-Iran     DDLKKA VATVGRISVAIDASHSSFQLYSEGYDEPEDCSSELDHGVLVVGYVKNGKKYW
AFQ98385      ddllkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
AAF61565      vdlkkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
AAQ16117     ddllkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
AFQ98384     ddllkavatvgrisvaidashsqsfqlysegyydepedcsesldhgvlvvgyvkngkkyw
                * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
ABO26562      fvknswggttwgdegyirmsrnknkcgiassaasyplv
DAA34687      lvknswggttwgdgyiyimtrnkdnqcgiasaasyplv
AAO60048      lvknswaeswgdqgyilmsrdknqcgiasaasyplv
BAH86062      lvknswggdwgdngyilmsrdknqcgiasaasyplv
ABS70713      lvknswaeswgdngyilmsrdknqcgiasaasyplv
Bacl-Iran     LVKNSWAESWGDHGYILMTRDNNNQCIGIASASASYPLV
AFQ98385      lvknswaeswgdqgyilmsrdnnnqcgiasqasyplv
AAF61565      lvknswaeswgdqgyilmsrdnnnqcgiasqasyplv
AAQ16117     lvknswgsgwgdngyilmsrdknqcgiasaasyplv
AFQ98384     lvknswaetwgdngyilmsrdknqcgiasaasyplv
                : * * * * . * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

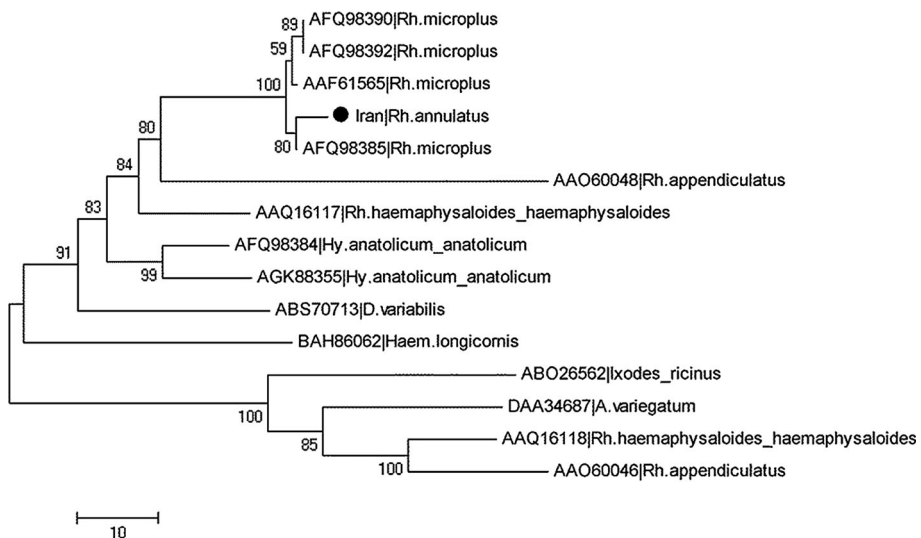
Fig. 2 Multiple sequence alignment of the deduced amino acid sequence of RaCL1 (named: Bacl-Iran) with other tick cathepsin L cysteine proteases. *Rhipicephalus microplus* (AFQ98385, AAF61565), *Haemaphysalis longicornis* (BAH86062), *Rhipicephalus appendiculatus* (AAO60048), *Rhipicephalus haemaphysaloides* (AAQ16117), *Ixodes ricinus* (ABO26562), *Hyalomma anatolicum* (AFQ98384), *Dermacentor variabilis* (ABS70713) and *Amblyomma variegatum* (DAA34687) were aligned together

identity. But in comparison with these sequences, in RaCL1 amino acid sequence alanine, threonine, tyrosine and glutamine amino acids have been replaced with valine, isoleucine, histidine and histidine, at positions 12, 33, 72 and 308 respectively. According to

hydrophobicity index, (Monera et al. 1995), valine (+76) and isoleucine (+99) are strongly hydrophobic while histidine (+8) is a neutral amino acid. On the other hand, alanine (+41) and tyrosine (+63) are hydrophobic amino acid but threonine (+13) and glutamine (−10) are neutral, so amino acid substitution in RaCL1 has led to the increase hydrophobicity in these positions. Also similar to our findings, in reported cathepsin-L sequences of *R. microplus* valine is dominant amino acid at position 11.

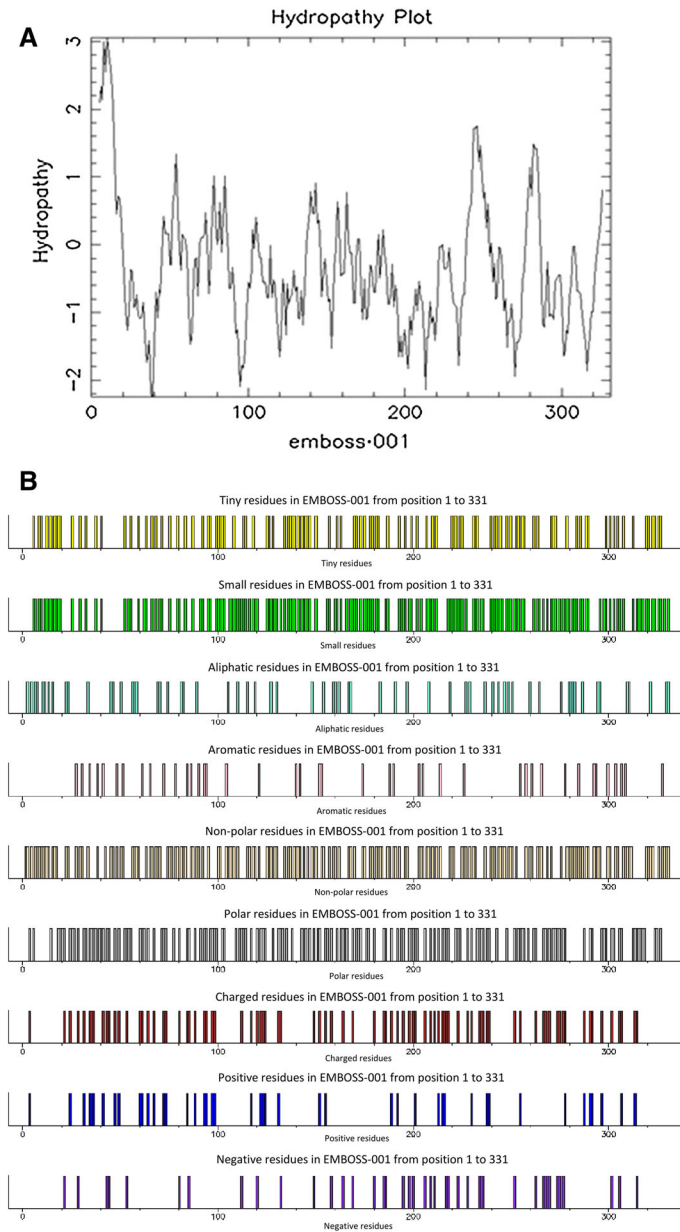
Referring to Uniprot database the RaCL1 consisted of a signal peptide, pro-region and mature enzyme which has a potential cleavage site for releasing of the signal peptide and pro-region placed at Ser<sup>18</sup>–Ser<sup>19</sup> and Ser<sup>115</sup>–Leu<sup>116</sup> respectively. It has been shown that for prevention of deleterious effects of peptidase C1 family enzymes on living cells they are synthesized as zymogens (Carmona et al. 1996; Khan and James 1998). Also, this protein contained a proline residue at position 2 in the mature enzyme that might have a role in preventing unwanted N-terminal proteolysis (Rawlings and Barrett 1994). Moreover, similar to *R. microplus* a single N-glycosylation motif (Asn<sup>112</sup>) was identified in the pro-region sequence of RaCL1 (Renard et al. 2000). It has been shown that glycosylation with mannose 6-phosphate in this position serves as an important sorting signal for routing immature enzyme into lysosomes and cleavage of the pro-region (Reiser et al. 2010).

In this study cysteine, histidine and asparagine were found to be conserved in RaCL1 sequences as the catalytic triad amino acid residues in ticks' cathepsin L-like proteases (Figs. 6, 7). Glutamine to histidine substitution at position 308 is nearest mutation to main active site amino acids which is neighboring with asparagine. As histidine is an amino acid with electrically charged side chain while glutamine has polar neutral side chain similar to asparagine, this substitution can be effective on Asp interaction and active site conformation however, further investigation is needed. Similar to cathepsin-L like cysteine



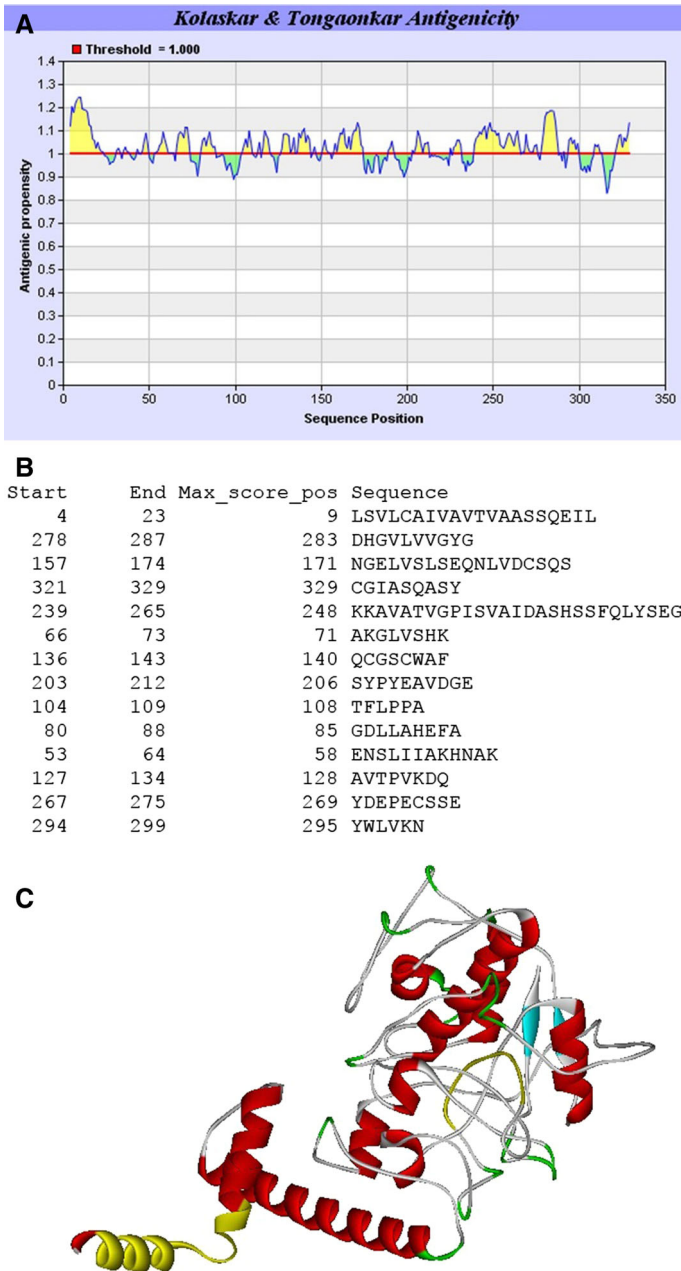
**Fig. 3** Phylogenetic analysis of some genus of hard ticks cathepsin L-like. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the number of differences method and are in the units of the number of amino acid differences per sequence. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5





**Fig. 4** Hydrophobicity analysis and amino acid property of RaCL1 protein sequence. **a** Hydropathy plot: the sequences above zero are hydrophobic regions but the sequences below zero are hydrophilic regions. **b** Amino acid properties of RaCL1 protein. (Color figure online)

protease sin *R. microplus*, *D. variabilis*, *R. appendiculatus* and *I. ricinus* RaCL1 sequence contains GCNGG conserved motif and the central asparagine residue as the single variant is exchanged with glutamic acid (Fig. 6). The conservation of this motif suggests that it has



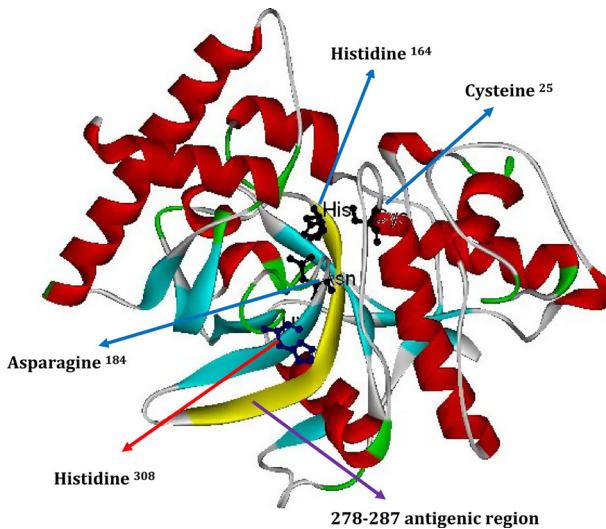
**Fig. 5** **a** Prediction of the linear B cell epitopes: potential regions are shown in yellow, the sequences of 4–23 and 278–287 are regions with high antigenic propensity. **b** Sequence of antigenic regions that are sorted from highest to lowest score. **c** Structure of RaCL1: two major B cell epitopes are shown in yellow. These regions are located in helical and coil conformation. (Color figure online)

```

MLRLSYDCAK VAVTVAASSQ ELPRTQWEAF KTIHKKTYQS HMEELLRFKI
FTENSLIIAK HNAYAKGLV SHKLGMNQFG DLAHEFARI FNGHHGTRKT
GGSTFLPPAN VNDSSLPKAV DWRKKGAVTP VKDQQQCGS WAFSATGSLE
GQHFLKNGEL VSLSEQNLVD CSQSFGNNGC EGGLMEDAFK YIKANDGIDT
EKSYPYEAVD GECRFKEDV GATDTGYVEI KAGSEDLKK AVATVGPISV
AIDASHSSFQ LYSEGVYDEP ECSSELDHG VLVVGYGVKG GKKYWLVKNS
WAESWGDHGY ILMTRDNNNQ CGIASQASYP LV*
    
```

- The linear B cell epitopes
- Conserve motif
- The catalytic triad amino acid residues
- X Substitution amino acids
- XXX Four hydrophilic regions
- Gray region: Signal sequence
- Green region: Pro-region sequence
- Black region: Mature enzyme sequence

**Fig. 6** A summary of the analysis done on the RaCL1 amino acid sequence. (Color figure online)



**Fig. 7** RaCL1 protein structure containing conserved amino acid residues: Cys<sup>25</sup>, His<sup>164</sup> and Asn<sup>184</sup>, that are involved in the catalysis of RaCL1; 278–287 antigenic region as major epitope and glutamine to histidine substitution at position 308 as nearest mutation to main active site amino acids. (Color figure online)

an important structural role, for example in papain the cysteine residue is involved in a disulfide bridge (Renard et al. 2000). Also the ERFNIN-like and GNFD conserved motifs, characterizing L-like cathepsin and mammalian cathepsin H and L, were found in pro-region sequence that are presented by Glu<sup>43</sup>, Arg<sup>47</sup>, Phe<sup>51</sup>, Asn<sup>54</sup>, Ile<sup>58</sup> and Asn<sup>62</sup> and

Gly<sup>75</sup>, Asn<sup>77</sup>, Phe<sup>79</sup> and Asp<sup>81</sup> respectively. The ERFNIN motif inhibits proteinase activity and converts the protein into its enzymatically active form after the removal of the pro-region (Karrer et al. 1993). The GNFD motif is responsible for correct folding and stability of enzyme in several papain family proteases (Kumar et al. 2004).

According to previous studies cathepsins are considered as potential therapeutic targets based on the results of enzyme inhibitor studies such as inhibition of cathepsin gene expression by antisense or treatment with chemical cathepsin inhibitors (Lustigman et al. 2004; Teo et al. 2007; Zhao et al. 2013). To our knowledge, since no studies have described the analysis and investigation of RaCL1 antigenic properties for using in antibody-based immunity against *R. annulatus*. Recently cathepsin B-like and cathepsin L-like proteins of *Toxoplasma gondii* have been introduced as a strong candidate for development of a DNA vaccine (Zhao et al. 2013). For that aim, they used bioinformatics approaches to identify antigenic epitopes on TgCBP and TgCPL. Their experimental results were consistent with the bioinformatics prediction of antigenic epitopes. It is remarkable that TgCPB and TgCPL are mainly expressed in the vacuolar compartment, but a tiny amount of TgCPL has been identified in the late endosome and this amount was elicited strong humoral and cellular immune responses in mice. Their results showed that TgCPB and TgCPL make good vaccine antigens, thus highlighting the reliability of the bioinformatics approaches that were used herein.

So, to address RaCL1 antigenic properties, analysis of the amino acids sequence were done based on two parameters including prediction of hydrophilic regions and the linear B cell epitopes by Kolaskar and Tongaonkar method. This method shows if the hydrophobic residues—Cys, Val and Leu—occur on the surface of a protein and therefore are more likely to be a part of antigenic sites with about 75 % accuracy. According to hydrophobicity plot approximately the sequences 18–44, 90–100, 190–218, and 260–280 are significantly hydrophilic that can be considered as the areas with the antigenic property.

Also analysis of amino acid properties in these regions was studied based on polar and side charged residues by Pepinfo tool and the results are summarized in Table 1. The linear B Cell epitope analysis of RaCL1 showed several regions but two regions of 4–23 and 278–287 which are located in helix and coiled structural conformation, were found to have significant antigenic propensity. In RaCL1, histidine is located in 278–287 sequence so this epitope can be one of the regions that are involved in forming the active site and antibody interaction (Fig. 7). Peptides that are located in non-helix structures are more antigenic than those in helical regions because this will increase the odds of peptide sequence recognition by antibody. Generally, antigenic sequences are located in solvent accessible regions and contain both hydrophobic and hydrophilic residues hence, epitopes in the N- and C-terminal regions of the protein are suitable epitopes because these regions are usually solvent accessible and unstructured which can increase their identification by antibodies. The sequence 320–328 which is a C-terminal epitope with optimum max\_score (1.136) and

**Table 1** Analysis of amino acids in four hydrophilic regions which predicted as the areas with the antigenic property by hydropathy plot (Fig. 4)

Regions (Seq)	Number of polar residues	Number of charged residues
18–44	17	10
90–100	7	4
190–218	16	13
260–280	10	8

located in mature enzyme sequence can also be a suitable antigen candidate. Some regions like 4–23, 18–44 and 90–100, as parts of the signal peptide and pro-region sequences, are not present in the mature enzyme structure therefore these regions are not suitable epitopes.

It is noteworthy that the localization of the enzyme for eliciting immune response is important. Generally cathepsin L enzymes have been localized to endosomes or lysosomes compartments with acidic pH. At such low pH antibodies may not function optimally which would negate its neutralizing effects and its potential as vaccine. But maturation of cathepsin enzymes occurs via two different trafficking pathways of traditional and non-traditional. In traditional type after disulfide bond formation and enzyme glycosylation with high-mannose glycans in endoplasmic reticulum mannose residues are phosphorylated to form mannose 6-phosphate (m6p) in the Golgi apparatus. M6p routes the protein into the endosomal/lysosomal compartment via the m6p receptor that in continues acidic condition leads to cleavage of the pro-region and enzyme activation. But in non-traditional type such as protein overexpression, a part of the cathepsin is not converted to m6p and as a result, immature enzyme is released into the extracellular matrix via exocytosis pathway (Reiser et al. 2010). Therefore, according to Zhao et al. report this type of enzyme can be used as a strong target for development of antibody-based immunity against *R. amulatus*. Also based on this topic, for increasing therapeutic potential, antigenic regions in pro-region sequence can be considered in parallel with these regions in mature enzyme.

In conclusion, cathepsin L-like would be suggested as a potential target for chemotherapy or vaccination against tick because of its role in physiological and biological procedures and good antigenic properties based on bioinformatics approaches. Assessment of RaCL1 gene expression patterns in larvae, nymphs and adult tissues (salivary gland, guts, ovaries) will shed light on functions of this enzyme in the tick.

**Acknowledgments** Authors wish to acknowledge the financial support of the research council of Tehran University and Veterinary Medicine Faculty of Tehran University.

**Compliance with ethical standards**

**Conflict of interest** To the authors' knowledge there are no conflicts of interest.

## References

- Brown SJ, Shapiro SZ, Askenase PW (1984) Characterization of tick antigens inducing host immune resistance. I. Immunization of guinea pigs with *Amblyomma americanum*-derived salivary gland extracts and identification of an important salivary gland protein antigen with guinea pig anti-tick antibodies. *J Immunol* 133:3319–3325
- Carmona E, Dufour E, Plouffe C, Takebe S, Mason P, Mort JS, Ménard R (1996) Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* 35:8149–8157
- Dickinson DP (2002) Cysteine peptidases of mammals: their biological roles and potential effects in the oral cavity and other tissues in health and disease. *Crit Rev Oral Biol Med* 13:238–275
- Dou Z, Carruthers VB (2011) Cathepsin proteases in *Toxoplasma gondii*. *Adv Exp Med Biol* 712:49–61
- Estrada-Peña A, Bouattour A, Camicas J-L, Walker AR (2004) Ticks of domestic animals in the Mediterranean region: a guide to identification of species. University of Zaragoza, Zaragoza
- Estrela A, Seixas A, Termignoni C (2007) A cysteine endopeptidase from tick (*Rhipicephalus (Boophilus) microplus*) larvae with vitellin digestion activity. *Comp Biochem Physiol B Biochem Mol Biol* 148:410–416
- Estrela AB, Seixas A, Teixeira VDON, Pinto AFM, Termignoni C (2010) Vitellin- and hemoglobin-digesting enzymes in *Rhipicephalus (Boophilus) microplus* larvae and females. *Comp Biochem Physiol B Biochem Mol Biol* 157:326–335
- Fagotto F (1990) Yolk degradation in tick eggs: I. Occurrence of a cathepsin L-like acid proteinase in yolk spheres. *Arch Insect Biochem Physiol* 14:217–235

- Franta Z, Sojka D, Frantova H, Dvorak J, Horn M, Srba J, Talacko P, Mares M, Schneider E, Craik CS, McKerrow JH, Caffrey CR, Kopacek P (2011) IrCL1—the haemoglobinolytic cathepsin L of the hard tick, *Ixodes ricinus*. *Int J Parasitol* 41:1253–1262
- Gershoni JM, Stern B, Denisova G (1997) Combinatorial libraries, epitope structure and the prediction of protein conformations. *Immunol Today* 18:108–110
- Karrer KM, Peiffer SL, DiTomas ME (1993) Two distinct gene subfamilies within the family of cysteine protease genes. *Proc Natl Acad Sci USA* 90:3063–3067
- Khan AR, James MN (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci* 7:815–836
- Kissoon-Singh V, Mortimer L, Chadee K (2011) *Entamoeba histolytica* cathepsin-like enzymes: interactions with the host gut. *Adv Exp Med Biol* 712:62–83
- Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276:172–174
- Kumar A, Dasaradhi PVN, Chauhan VS, Malhotra P (2004) Exploring the role of putative active site amino acids and pro-region motif of recombinant falcipain-2: a principal hemoglobinase of *Plasmodium falciparum*. *Biochem Biophys Res Commun* 317:38–45
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157:105–132
- Lee AR, Bak HJ, Kim NY et al (2012) Cloning, heterologous expression, and enzymatic characterization of cathepsin L from starfish (*Asterina pectinifera*). *Biosci Biotechnol Biochem* 76:2342–2346
- Li AH, Moon S-U, Park Y-K et al (2006) Identification and characterization of a cathepsin L-like cysteine protease from *Taenia solium* metacystode. *Vet Parasitol* 141:251–259
- Lustigman S, Zhang J, Liu J, Oksov Y, Hashmi S (2004) RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. *Mol Biochem Parasitol* 138:165–170
- Monera OD, Sereda TJ, Zhou NE et al (1995) Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci* 1:319–329
- Nabian S, Taheri M, Mazaheri Nezhad Fard R, Aramoon M (2013) Identification of tropomyosin and its immunological properties from larvae of cattle tick, *Boophilus annulatus*. *Iran J Parasitol* 8:242–248
- Nabian S, Taheri M, Ranjbar MM et al (2014) Assessment and partial purification of serine protease inhibitors from *Rhipicephalus (Boophilus) annulatus* larvae. *Rev Bras Parasitol Vet* 23:187–193
- Nikpay A, Nabian S, Taheri M (2012) Analysis of immunogenic relevant proteins in *Rhipicephalus (Boophilus) annulatus* tick. *J Arthropod Borne Dis* 6:36–44
- Parola P, Paddock CD, Socolovschi C, Labruna MB, Mediannikov O, Kernif T, Abdad MY, Stenos J, Bitam I, Fournier PE, Raoult D (2013) Update on tick-borne rickettsioses around the world: a geographic approach. *Clin Microbiol Rev* 26:657–702
- Rahbari S, Nabian S, Shayan P (2007) Primary report on distribution of tick fauna in Iran. *Parasitol Res* 101(Suppl):S175–S177
- Ranjbar MM, Nabian S, Nasab SDM, Taheri M, Ghorban K, Dadmanesh M, Malekan M, Ahmadi NA, Imanzad M (2013) Comprehensive phylogenetic, similarity and allergenicity analysis of *Boophilus* genus tick tropomyosin protein. *J Paramed Sci* 4:70–82
- Ranjbar MM, Nabian S, Taheri M, Nikbakht GR, Nikpay A (2014) Immunoinformatic survey of *Boophilus* tick tropomyosin protein. *J Vet Res* 4:335–345 (in Persian with English abstract)
- Ranjbar MM, Gupta SK, Ghorban K, Nabian S, Sazmand A, Taheri M, Esfandyari S, Taheri M (2015) Designing and modeling of complex DNA vaccine based on tropomyosin protein of *Boophilus* genus tick. *App Biochem Biotechnol* 175:323–339
- Rawlings ND, Barrett AJ (1994) Families of cysteine peptidases. *Methods Enzymol* 244:461–486
- Reiser J, Adair B, Reinheckel T (2010) Specialized roles for cysteine cathepsins in health and disease. *J Clin Invest* 120:3421–3431
- Renard G, Garcia JF, Cardoso FC et al (2000) Cloning and functional expression of a *Boophilus annulatus* cathepsin L-like enzyme. *Insect Biochem Mol Biol* 30:1017–1026
- Renard G, Lara FA, de Cardoso FC et al (2002) Expression and immunolocalization of a *Boophilus microplus* cathepsin L-like enzyme. *Insect Mol Biol* 11:325–328
- Sajid M, McKerrow JH (2002) Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol* 120:1–21
- Seixas A, Dos Santos PC, Velloso FF, Da Silva Vaz I, Masuda A, Horn F, Termignoni C (2003) A *Boophilus microplus* vitellin-degrading cysteine endopeptidase. *Parasitology* 126:155–163
- Sojka D, Francischetti IMB, Calvo E, Kotsyfakis M (2011) Cysteine proteases from bloodfeeding arthropod ectoparasites. In: Robinson MW, Dalton JP (eds) *Cysteine proteases of pathogenic organisms*. Springer, Boston, pp 177–191
- Sonenshine DE, Roe RM (2014) *Biology of ticks*, vol 2, 2nd edn. Oxford University Press, New York

- Taheri M, Nabian S, Nikbakht GR, Yousefi P (2014a) Study of cathepsins involved in haemoglobin and vitellin digestion in *Rhipicephalus (Boophilus) annulatus* larvae by one- and two-dimensional zymography. *J Vet Res* 69:25–31 (in Persian with English abstract)
- Taheri M, Nabian S, Ranjbar MM, Mazaheri Nezhad R, Gerami Sadeghian A, Sazmand A (2014b) Study of vitellogenin in *Boophilus annulatus* tick larvae and its immunological aspects. *Trop Biomed* 31:398–405
- Teo CF, Zhou XW, Bogyo M, Carruthers VB (2007) Cysteine protease inhibitors block *Toxoplasma gondii* microneme secretion and cell invasion. *Antimicrob Agents Chemother* 51:679–688
- Tsuji N, Miyoshi T, Battsetseg B, Matsuo T, Xuan X, Fujisaki K (2008) A cysteine protease is critical for *Babesia* spp. transmission in *Haemaphysalis* ticks. *PLoS Pathog* 4:e1000062
- Willadsen P, Kemp DH (1988) Vaccination with “concealed” antigens for tick control. *Parasitol Today* 4:196–198
- Yamaji K, Tsuji N, Miyoshi T, Islam MK, Hatta T, Alim MA, Anisuzzaman, Takenaka A, Fujisaki K (2009) Hemoglobinase activity of a cysteine protease from the ixodid tick *Haemaphysalis longicornis*. *Parasitol Int* 58:232–237
- Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, Bui H, Buus S, Frankild S, Greenbaum J, Lund O, Lundegaard C, Nielsen M, Ponomarenko J, Sette A, Zhu Z, Peters B (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36:W513–W518
- Zhao G, Zhou A, Lv G, Meng M, Sun M, Bai Y, Han Y, Wang L, Zhou H, Cong H, Zhao Q, Zhu XQ, He S (2013) *Toxoplasma gondii* cathepsin proteases are undeveloped prominent vaccine antigens against toxoplasmosis. *BMC Infect Dis* 13:207