Toward the discovery of inhibitors of babesipain-1, a *Babesia bigemina* cysteine protease: in vitro evaluation, homology modeling and molecular docking studies

Bianca Pérez · Sandra Antunes · Lídia M. Gonçalves ·
Ana Domingos · José R. B. Gomes ·
Paula Gomes · Cátia Teixeira

Abstract *Babesia bigemina* is a protozoan parasite that causes babesiosis, a disease with a world-wide distribution in mammals, principally affecting cattle and man. The unveiling of the genome of *B. bigemina* is a project in active progress that has already revealed a number of new targets with potential interest for the design of anti-babesiosis drugs. In this context, babesipain-1 has been identified as a proteolytically active enzyme whose three-dimensional structure has not been resolved yet, but which is known to be inhibited by cysteine proteases inhibitors such as E64, ALLN, leupeptin, and vinyl sulfones. In this work, we introduce (1) a homology model of babesipain-1; (2) a comparison between babesipain-1 and falcipain-2, a cysteine protease of the malaria parasite *Plasmodium falciparum*; (3) in vitro data for babesipain-1 inhibition by HEDICINs and HECINs, previously reported as modest inhibitors of falcipain-2; and (4) the docked binding conformations of HEDICINs and HECINs in the model of babesipain-1. HEDICINs presented similar preferred binding conformations for both babesipain-1 and falcipain-2. However, in vitro bioassay shows that HEDICINs and HECINs are better inhibitors of babesipain-1 than of falcipain-2, which could be explained by observed differences between the active pockets of these proteins in silico. Results presented herein provide a valuable contribution to future computer-aided molecular design of new babesipain-1 inhibitors.

Keywords Babesipain-1 · *Babesia bigemina* · Cysteine proteases · Falcipain-2 · Homology modelling · Molecular docking

B. Pérez · P. Gomes · C. Teixeira
Departamento de Química, Faculdade de Ciências, Centro de Investigação em Química da Universidade do Porto, Universidade do Porto, R. Campo Alegre, 687, 4169-007 Porto, Portugal
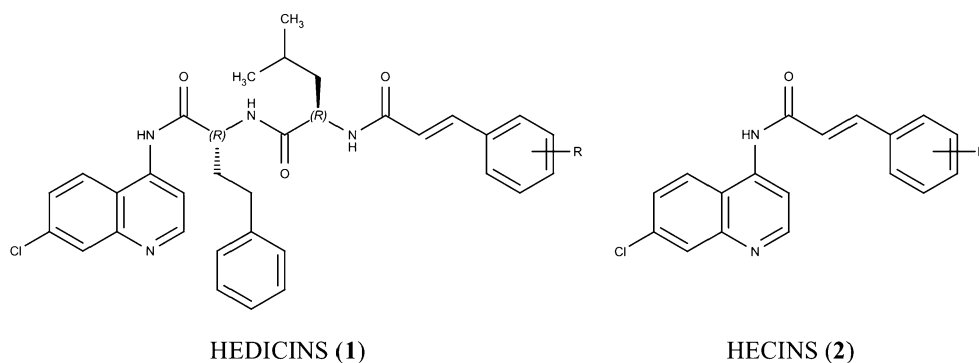
S. Antunes · A. Domingos
Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

L. M. Gonçalves
Faculdade de Farmácia, iMed-UL, Research Institute for Medicines and Pharmaceutical Sciences, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisbon, Portugal

A. Domingos
Instituto de Higiene e Medicina Tropical, Centro de Malária e outras Doenças Tropicais, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

J. R. B. Gomes · C. Teixeira (✉)
CICECO, Departamento de Química, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
e-mail: ca.teixeira@ua.pt

Introduction

Babesiosis is caused by intraerythrocytic protozoan parasites of the genus *Babesia*, which infect a wide range of domestic and wild animals, and occasionally man [1]. The major impact occurs in the cattle industry, where bovine babesiosis has had a huge economic effect [2]. One of the most important species that causes babesiosis in cattle is *Babesia bigemina*, which is distributed wherever

Springer

HEDICINS (**1**)   HECINS (**2**)

Rhipicephalus (*Boophilus*) sp. ticks are encountered, including North and South America, Southern Europe, Africa, Asia and Australia [1]. Although babesiosis can be controlled with vaccination and treated with antiparasitic drugs, the vaccines might not confer cross protection due to the existence of different local variants [3, 4]. Also, many effective anti-babesiosis drugs have been withdrawn from the market due to health or environmental safety concerns [5]. Hence, control of babesiosis currently requires more specific, fast acting, reliable and safer chemotherapeutic treatments; consequently, the identification and characterization of new drug targets for chemotherapy of bovine babesiosis is considered a pressing priority [6].

The sequencing of the genome of *B. bigemina* is in active finishing and the information retrieved so far has provided a better understanding of the biology of this parasite and has also unraveled potential targets, such as cysteine proteases, that may be of utility in prophylactic and therapeutic interventions [7]. Although the role of cysteine proteases in piroplasms is mostly unknown, the importance of these enzymes in the life cycle of *Babesia* sp. was demonstrated in inhibition studies. It was reported that the inhibition of cysteine proteases reduces in vitro invasion of erythrocytes and growth of *Babesia bovis* [8]. Three genes belonging to the C1 family of cysteine proteases from *B. bigemina*, BbiCPL1 to BbiCPL3, were previously identified [9], and found to share many features with papain, including: (1) a 20–23 amino acid putative transmembrane domain, (2) the presence of the ERFNIN and GNFD pro-sequence motifs typical of cathepsin L-like cysteine proteases [10], (3) conservation of catalytic residues, and (4) six cysteine residues predicted to be involved in disulfide bond formation in the mature protease sequence [8]. Recently, BbiCPL1 (or babesipain-1) was cloned and expressed as a recombinant enzyme which was active against typical peptide substrates of cysteine proteases, as well as inhibited by E-64, ALLN, cystatin and leupeptin [6].

Enzymes from the papain family are the most common proteases in protozoan parasites, and are essential to the life cycle and pathogenicity of these organisms [11]. Presently, protozoan cysteine proteases are recognized drug targets, and specific inhibitors are under validation for

chemotherapy of leishmaniasis, malaria, and trypanosomiasis [11, 12]. Consequently, it is very likely that cysteine proteases from *Babesia* sp. might become relevant targets for improving the control of bovine babesiosis. Moreover, babesipain-1 was recently found to be inhibited by artemisinin-vinyl sulfone hybrid molecules, which had been previously reported as inhibitors of *Plasmodium falciparum* cysteine proteases, falcipains, and to effectively inhibit the growth of *P. falciparum* cultured parasites [13, 14]. This finding paves the way for the rescuing of antimalarial compounds as potential anti-babesiosis agents, thus turning falcipain inhibitors as good starting points for the development of babesipain-1 inhibitors.

In view of the above, we have evaluated the inhibition of babesipain-1 by some compounds, known as HECINs **1** and HEDICINs **2** (Fig. 1), which were previously reported as micromolar falcipain inhibitors [15]. The interactions of these compounds with babesipain-1 were studied in silico, in order to rationalize in vitro data at the molecular level and guide future design of suitable inhibitors. However, since the three-dimensional (3D) structure for babesipain-1 is yet unavailable, enzyme-compound docking studies were preceded by the establishment of a rational 3D structure for the enzyme through homology modeling. This technique offers a rational alternative to predict protein structure based on sequence similarity among several proteins of the same class. The model obtained for babesipain-1 was validated with various structure/geometry verification tools. The docking studies with the test compounds also provided insight into the possible binding modes and interactions of ligands with the enzyme.

Materials and methods

Experimental

Synthesis of compounds 1–2

Compounds **1a–j** and **2a–j** were prepared as described in the literature. Analytical and structural information for

each compound was found to be in perfect agreement with reported data [15].

Activity assays

The full length babesipain-1 gene was amplified from *B. bigemina* genomic DNA by PCR, and cloned in the pGEX-6P-1 expression vector (GE Healthcare) as previously described [14]. Constructs were transformed into *Escherichia coli* BL21 cells (GE Healthcare), and liquid cultures were induced at an absorbance of 1.0 at 600 nm, for 3 h, with 1 mM IPTG. Insoluble inclusion bodies were washed with urea buffer, and the resulting denatured and reduced GST-babesipain-1 protein was refolded and later acidified to promote auto-activation to an active form. Babesipain-1 activity was assayed by a fluorimetric method as previously described [13, 14]. Briefly, assays were carried out in 200 μL assay buffer (10 mM PBS, pH 7.4, 5 mM DTT) containing 20 μL of babesipain-1 activated in assay buffer at 50 μg/mL, and 5 μL of each concentration of the tested inhibitors. Reactions were initiated by the addition of a fluorogenic substrate (Z-Leu-Leu-Arg-AMC, Bachem, Germany) and the activity was monitored (excitation 355 nm; emission 460 nm) for 30 min, at 37 °C on a Fluorescence Microplate Reader (FLUOstar Omega, BMGLABTECH GmbH, Germany). For all assays, saturated substrate concentrations were used in order to obtain linear fluorescence curves. Stock solutions of the inhibitors were prepared in DMSO and serial dilutions were made also in DMSO. Controls were performed using enzyme alone, substrate alone, enzyme with DMSO and a positive control (*trans*-epoxysucciny-L-leucyl-amido(4-guanidino)butane—E64 (Calbiochem, Germany). The IC_{50} values were determined using GraphPad PRISM software by non-linear regression analysis based on the log of the concentrations of the inhibitors versus the percentage of activity. All assays were performed in triplicate.

Computational

Homology modeling

The amino acid sequence of babesipain-1 was retrieved from the universal protein resource (UniProt) database (ID: C3VEH9) in fasta format [16]. The retrieved sequence consisted of 458 amino acids, including those from the prodomain present in papain-like cysteine proteases. As the interest of this study was to model the mature (functional state) cysteine protease, the amino acids from the prodomain were cleaved off, based on previously reported information [9]. Therefore, only the mature sequence, starting from residue Ser242 and having a total length of 217 amino acids, was used for deriving the homology model. The BLAST program against Protein Data Bank (PDB), available at National Center for Biotechnology Information (NCBI), was used to select template structures for homology modeling of babesipain-1. Among the homologous sequences, only hits matching the following criteria were selected: (1) E-value below 10^{-4}; (2) query coverage >90 % and sequence identity >35 % or query coverage >85 % and sequence identity >40 %; (3) pdb structure with a resolution <2.5 Å; (4) pdb structure without missing residues in the active site; (5) pdb structure with bound ligand.

The templates and the target sequences were then aligned using the PSI-Coffee mode of T-Coffee v.9.03 [17]. The babesipain-1 models were constructed based on different alignments, single and multiple. For each alignment, 100 models were generated using the standard "automodel" routine of Modeller v.9.11 [18]. The resulting modeled structures were ranked on the basis of an internal scoring function, and only those with the least internal scores were identified and used for model validation. In addition, the root-mean square deviation (RMSD) of the models was calculated by superimposing each model on the different template structures; the quality of the consistency between the templates and the babesipain-1 model was evaluated using ProSA-web [19, 20], during which energy criteria for the modeled structure were compared with the potential mean force obtained from a large set of known protein structures. The backbone conformation of the modeled structure was calculated by analyzing the phi (Φ) and psi (ψ) torsion angles using PROCHECK, as determined by Ramachandran plot statistics [21]. The quality of babesipain-1 models was estimated using the qualitative model energy analysis (QMEAN) and the protein quality predictor (ProQ) servers [22, 23].

Docking

The docking studies were performed as previously described [24]. Briefly, the proteins were protonated using the H++ server [25] assuming a pH of 5.5 and a salinity of 0.15 mol/L. The proteins were then minimized with the AMBER 11 program [26] by 500 steps of steepest descent, followed by 2,000 steps of conjugate gradient to remove bad contacts using a generalized-Born solvent model. The biomolecular force field ff03 [27] was used. Docking was performed with GOLD [28] version 5.0.1, allowing full flexibility for the ligand while keeping the protein fixed. The docking exploration consisted of 500 independent runs of the docking algorithm with each compound, using the default genetic algorithm (GA) search parameters and the GoldScore scoring function. The binding site was defined as 15 Å radius from the catalytic amino acid Cys25 of babesipain-1 (numbered according to enzyme mature domain) models.

Table 1 In vitro data on compounds **1a–j** and **2a–j** for babesipain-1 and falcipain-2 inhibition activity

Cpd	R	Babesipain-1 IC$_{50}$ (μM)	FP-2 IC$_{50}$ (μM)[a]	Cpd	R	Babesipain-1 IC$_{50}$ (μM)	FP-2 IC$_{50}$ (μM)[a]
1a	H	30.8	50	**2a**	H	20.3	>50
1b	*p*-Me	31.3	>50	**2b**	*p*-Me	13.4	28.5
1c	*p*-iPr	27.7	>50	–	–	–	–
1d	*p*-OMe	35.8	>50	**2d**	*p*-OMe	9.8	>50
1e	*m*-F	29.1	23.1	**2e**	*m*-F	20.9	>50
1f	*p*-F	28.1	50.0	**2f**	*p*-F	17.4	>50
1g	*p*-Cl	9.7	>50	**2g**	*p*-Cl	17.9	41.3
1h	*p*-Br	11.0	48.3	**2h**	*p*-Br	14.2	33.1
1i	o-NO$_2$	18.6	>50	–	–	–	–
1j	*m*-NO$_2$	10.2	28.1	**2j**	*m*-NO$_2$	48.3	14.2
E64[b]	–	13.7	24.6	–	–	13.7	24.6

[a] As reported in [15]

[b] In nM

Results and discussion

In vitro inhibition studies

Both HEDICINS **1** and HECINS **2** were evaluated in vitro for inhibition of babesipain-1, using a fluorimetric method [13, 14], as described under "Experimental". Close inspection of data obtained (Table 1, columns 3 and 7), and comparison with the falcipain inhibitory activities reported in [15] for the same compounds, shows the following:

1. HEDICINS **1** and HECINS **2** present mid-micromolar activities against babesipain-1 (IC$_{50}$ = 9.7–35.8 and IC$_{50}$ = 9.8–48.3 μM, respectively) which, in general, are lower than those formerly observed against falcipain-2 (IC$_{50}$ = 23.1 to >50 μM and IC$_{50}$ = 14.2 to >50 μM, respectively); noteworthy, the best babesipain-1 inhibitor in each series, **1g** (IC$_{50}$ = 9.7 μM) and **2d** (IC$_{50}$ = 9.8 μM), were found previously to be inactive (IC$_{50}$ > 50 μM) against falcipain-2 [15];

2. within the complete set of HEDICINS (**1**), the cinnamoyl substituent R has a slight effect on babesipain-1 inhibition activity; hence:

    *   the fact that the two best compounds of the series are **1 g** (R = *p*-Cl, IC$_{50}$ = 9.7 μM) and **1j** (R = *m*-NO$_2$, IC$_{50}$ = 10.2 μM), and the worst is **1d** (R = *p*-OMe; IC$_{50}$ = 35.8 μM), suggests that the inhibition activity is improved by electron-withdrawing substituents;
    *   comparison of **1i** (IC$_{50}$ = 18.6 μM) with **1j** (IC$_{50}$ = 10.2 μM) suggests that the inhibition activity benefits more from cinnamoyl substituents at position *meta* than at position *ortho*, while the difference between *meta* and *para* positions (**1e** vs.

1f; IC$_{50}$ = 29.1 and 28.1 μM, respectively) is negligible;

3. within the HECINS series (**2**), the influence of the cinnamoyl substituent R on babesipain-1 inhibition resulted as follows:

    *   in contrast to HEDICINS **1**, compounds **2** with electron-donating R groups seem to be more interesting inhibitors, as the best one is **2d** (R = *p*-OMe; IC$_{50}$ = 9.8 μM), whereas the chlorinated (**2g**), fluorinated (**2e**, **f**) and nitrated (**2j**) analogues are the worst (IC$_{50}$ **2g** ≈ **2f** < **2e** ≪ **2j**);
    *   as for HEDICINS **1**, no preference between *meta* and *para* position was observed for the cinnamoyl substituent R in HECINS **2** (**2e** vs. **2f**; IC$_{50}$ values of 20.9 and 17.4 μM, respectively).

4. the influence of both the lipophilicity of the test compounds (clogP values) [29] and the bulkiness of their cinnamoyl ring substituents R (Charton's steric parameter, ν) [30] on inhibition activity was also assessed, but no clear correlation was found (supporting information; Table S1).

Template selection and sequence alignment

Since the accuracy of a homology model is related to the degree of sequence identity and to the similarity between template and target, template search and sequence alignment are crucial steps in homology modeling. Among the homologous sequences identified with the BLAST program against PDB, only those matching the following criteria were selected as templates: (1) E-value below 10^{-4}; (2)

Table 2 Babesipain-1 homologs identified with BLAST according to the established search criteria (see text)

PDB code (chain used)	Uniprot code	Organism	Identity (%)	Coverage (%)	PDB resolution (Å)	E-value
1EWM (A)	P25779	*Trypanosoma cruzi*	43	91	2.00	1×10^{-41}
1S4 V (A)	O65039	*Ricinus communis*	41	87	2.00	2×10^{-43}
2BDZ (C)	P84346	*Jacaratia mexicana*	38	95	2.10	4×10^{-43}
2OUL (A)	Q9N6S8	*Plasmodium falciparum*	41	94	2.20	1×10^{-46}
2P7U (A)	Q95PM0	*Trypanosoma brucei rhodesiense*	43	90	1.65	2×10^{-42}
3BWK (B)	Q9NAW4	*Plasmodium falciparum*	41	94	2.42	3×10^{-45}

query coverage >90 % and sequence identity >35 % or query coverage >85 % and sequence identity >40 %; (3) pdb structure with resolution <2.5 Å; (4) pdb structure without missing residues in the active site; (5) pdb structure with bound ligand.

Table 2 shows PDB codes, chain, UniProt accession numbers, scientific organism that each template sequence belongs to, sequence identity, query coverage, resolution of the structures and E-value for all sequences chosen as templates.

A multiple alignment of the sequences of the template and of the target is presented in Fig. 2. For simplicity, the numbering was started at 1 for all the alignments performed. The average sequence homology of babesipain-1 with the six homologs was 41 %, ranging from 38 to 43 %. Based on their sequences, babesipain-1 and selected templates are classifiable as cysteine peptidases belonging to Clan CA, subfamily C1A. This peptidase subfamily utilizes catalytic glutamine (Gln19), cysteine (Cys25), histidine (His155) and asparagine (Asn177) residues, always keeping this ordering [31]. These four amino acids are present in three separate, well conserved regions of the primary sequence of the mature protease, known as the cysteine, histidine, and asparagine active site regions of cysteine proteases (Fig. 2). Alignment of babesipain-1 with the selected templates showed strict conservation of the catalytic residues, and low polymorphism in their surrounding areas. Notably, the tryptophan (Trp179) that forms the "oxyanion hole" together with Gln19 is also preserved [32], and other additional structurally conserved regions are observed, thus making the selected hits suitable templates.

Model evaluation

In general, the Modeller code appears to perform best when using two or three templates compared to a single one [33]. Therefore, homology models were built based on different alignments, performed with PSI-Coffee mode of T-Coffee, using a single 3D selected template and all the possible combinations of two to three selected homolog proteins,

giving a total of 41 different alignments. For each alignment, 100 models were obtained using Modeller [18], totalizing 4,100 models for babesipain-1.

The outputs from Modeller were evaluated using a statistical evaluation method, Z-DOPE, a normalized atomic distance-dependent statistical potential based on known protein structures, where a score of <-1 indicates a "reliable" model (i.e., it indicates that 80 % of its Cα atoms are within 3.5 Å of their correct positions) [34]. Therefore, only those models with a Z-DOPE score equal to or lower than -1 were taken for further validation analysis (Table 3). To select the final model of babesipain-1, among the nine listed models, additional validation tools were employed, such as PROCHECK, Prosa, ProQ, QMean, and the following criteria were set: (1) >90 % of amino acids in the most favorable region as determined by PROCHECK; (2) Prosa Z-score in accordance with those obtained for the pdb structures of the templates used to derive each of the models; (3) a LG score >5 and a MaxSub >0.5 as determined by ProQ; and (4) a QMEAN score >0.75 [19–21, 23, 35].

By comparing the different parameters shown in Table 3, only model 7 (1S4V-2BDZ_51), derived from the multiple sequence alignment of babesipain-1 with templates 1S4V and 2BDZ, was found to match the four criteria established. When compared to the original structures, the alignment used to obtain this model showed one of the highest PSI-coffee alignment scores and one of the lowest RMSDs, which reinforces the good quality of the model. In more detail, the validation results for model 7 of babesipain-1, determined by Ramachandran plot statistics (Fig. 3) performed with PROCHECK [21], revealed that 90.9, 7.0, 1.1 and 1.1 % of the residues were located in the most favorable, additionally allowed, generously allowed and disallowed regions, respectively. Although this model babesipain-1 presented two amino acids (Arg102, Asp199) in disallowed regions, they were out of the binding cavity; the Cα of the closest amino acid, Asp199, was distant from approximately 20 Å of Cα of the catalytic Cys25. Moreover, the PROCHECK G-factor value of -0.07 for the final model also indicated the good quality of the constructed model.

Fig. 2 Multiple sequence alignment performed with PSI-Coffee mode of T-Coffee [17] of the amino acid sequence of babesipain-1 (bbspn1) and of the selected templates. The *boxes* indicate the cysteine, histidine and asparagine catalytic regions, respectively. The *arrows* show the catalytic residues glutamine (Gln19), cysteine (Cys25), histidine (His155) and asparagine (Asn177), with residue *numbering* according to the sequence of mature babesipain-1

```
bbspn1   --------------------SFENVNWRTPGAVSPVKDQGHCGSCWAFAAI  31
1ewm_A   --------------------APAAVDWRARGAVTAVKDQGQCGSCWAFSAI  31
1s4v_A   T-------------------VPASVDWRKKGAVTSVKDQGQCGSCWAFSTI  32
2bdz_C   --------------------YPESIDWREKGAVTPVKNQNPCGSCWAFSTV  31
2oul_A   QMN-YEEVIKKYRGE-ENFDHAAYDWRLHSGVTPVKDQKNCGSCWAFSSI  48
2p7u_A   --------------------APAAVDWREKGAVTPVKDQGQCGSCWAFSTI  31
3bwk_B   YEANYEDVIKKYKPADAKLDRIAYDWRLHGGVTPVKDQALCGSCWAFSSV  50

bbspn1   GAVESLFRMEKHQYLTFSEQELVSCDL-QSSGCGGGFSDFALEYIKK--N  78
1EWM_A   GNVECQWFLAGHPLTNLSEQMLVSCDK-TDSGCSGGLMNNAFEWIVQENN  80
1S4V_A   VAVEGINQIKTNKLVSLSEQELVDCDTDQNQGCNGGLMDYAFEFIKQ--R  80
2BDZ_C   ATIEGINKIITGQLISLSEQELLDCER-RSHGCDGGYQTTSLQYVVD--N  78
2OUL_A   GSVESQYAIRKNKLITLSEQELVDCSF-KNYGCNGGLINNAFEDMIE--L  95
2P7U_A   GNIEGQWQVAGNPLVSLSEQMLVSCDT-IDFGCGGGLMDNAFNWIVNSNG  80
3BWK_B   GSVESQYAIRKKALFLFSEQELVDCSV-KNNGCYGGYITNAFDDMID--L  97

bbspn1   -GVTSSEDWNYEASNGE---CSAH-HGIRYYIKDYVSARGANV--ASTLL  121
1EWM_A   GAVYTEDSYPYASGEGISPPCTTSGHTVGATITGHVELPQDEAQIAAWLA  130
1S4V_A   GGITTEANYPYEAYDGT--CDVSKENAPAVSIDGHENVPENDENALLKAV  128
2BDZ_C   -GVHTEREYPYEKKQGR--CRAKDKKGPKVYITGYKYVPANDEISLIQAI  125
2OUL_A   GGICPDGDYPYVSDAPN--LCNIDRCTEKYGIKNYLSVPDNKL--KEALR  141
2P7U_A   GNVFTEASYPYVSGNGEQPQCQMNGHEIGAAITDHVDLPQDEDAIAAYLA  130
3BWK_B   GGLCSQDDYPYVSNLPE--TCNLKRCNERYTIKSYVSIPDDKF--KEALR  143

bbspn1   VRAPTVVYVAFT-EDLFHYSGGVYNGECPEEQLNHAVLLVGEGYDAT---  167
1EWM_A   VNGPVAVAVDAS--SWMTYTGGVMTSCVS-EQLDHGVLLVGYNDSA----  173
1S4V_A   ANQPVSVAIDAGGSDFQFYSEGVFTGSCG-TELDHGVAIVGYGTTI----  173
2BDZ_C   ANQPVSVVTDSRGRGFQFYKGGIYEGPCG-TNTDHAVTAVGYGK------  168
2OUL_A   FLGPISISVAVS-DDFAFYKEGIFDGECG-DQLNHAVMLVGFGMKEIVNP  189
2P7U_A   ENGPLAIAVDAT--SFMDYNGGILTSCTS-EQLDHGVLLVGYNDAS----  173
3BWK_B   YLGPISISIAAS-DDFAFYRGGFYDGECG-AAPNHAVILVGYGMKDIYNE  191

bbspn1   -----VKKRYWLIKNSWGSDWGEDGFIRLERTDK-GYDKCGVLSFGFIPT  211
1EWM_A   ------AVPYWIIKNSWTTQWGEEGYIRIAKGS----NQCLVKEEASSAV  213
1S4V_A   -----DGTKYWTVKNSWGPEWGEKGYIRMERGISDKEGLCGIAMEASYPI  218
2BDZ_C   --------TYLLLKNSWGPNWGEKGYIRIKRASGRSKGTCGVYTSSFFPI  210
2OUL_A   LTKKGEKHYYYIIKNSWGQQWGERGFINIETDESGLMRKCGLGTDAFIPL  239
2P7U_A   ------NPPYWIIKNSWSNMWGEDGYIRIEKGT----NQCLMNQAVSSAV  213
3BWK_B   DTGRMEKFYYYIIKNSWGSDWGEGGYINLETDENGYKKTCSIGTEAYVPL  241

bbspn1   GAVLAH  217
1EWM_A   ----VG  215
1S4V_A   KKSSNN  224
2BDZ_C   ----KG  212
2OUL_A   ----IE  241
2P7U_A   ----VG  215
3BWK_B   -----L  242
```

Model 7 was also evaluated with the ProSA-web program by examining whether or not the interactions of each residue with the rest of the protein structure are favorable [19, 20]. The Z-score, provided by ProSA-web from the calculation of the knowledge-based mean fields, is used to judge the quality of protein folds, thus indicating the overall quality of the model. The value of the Z-score is displayed in a plot that contains the Z-scores of all experimentally determined protein chains in current PDB. Figure 4 shows the Z-score plots of 1S4V (A), model 7 of babesipain-1 (B) and 2BDZ (C), which are −7.98, −6.48 and −6.85, respectively.

Although the Z-score of model 7 is slightly lower than that of template 1S4V, it is in the same range of template 2BDZ and it is a perfect fit within the structures in PDB.

ProQ is a neural network based method developed to predict the quality of protein models by recognizing folds that are not compatible with a protein sequence [23]. The quality of the model is quantified by two indices: LG score (i.e., the −log of a *p* value) [36] and MaxSub (ranging 0–1) [37]. Depending on the specific values of these indices, the model can be qualified as: correct if LG score >1.5 and MaxSub >0.1, as good if LG score >3 and MaxSub >0.5,

Table 3 Validation scores from Modeller, PROCHECK, ProQ, ProSA and Qmean for the constructed models

Model	Alignment (PSI-coffee score)	Modeller (Z-DOPE)	PROCHECK (%)[a]	ProQ LG score	ProQ MaxSub	ProSA (Z-score)	Qmean score (Z-score)[b]	RMSD (Å)[c]
1	1EWM-1S4V-2BDZ_42 (91)	−1.06	88.7 10.2 1.1 0.0	5.068	0.536	−6.43	0.773 (0.02)	1.14 0.80 0.64
2	1S4V-2BDZ_18 (93)	−1.06	91.4 7.0 0.0 1.6	5.118	0.555	−6.31	0.742 (−0.32)	0.91 0.65
3	1S4V-2BDZ_54 (93)	−1.05	91.4 7.5 0.0 1.1	5.318	0.560	−6.20	0.706 (−0.70)	0.94 0.67
4	1EWM-1S4V-2BDZ_51 (91)	−1.04	87.6 11.3 0.5 0.5	5.091	0.508	−6.58	0.720 (−0.55)	1.11 0.74 0.61
5	1EWM-1S4V-2BDZ_76 (91)	−1.04	89.8 8.1 2.2 0.0	5.358	0.547	−6.25	0.718 (−0.57)	1.15 0.80 0.67
6	2BDZ-2P7U_77 (89)	−1.03	82.3 14.5 1.6 1.6	5.394	0.575	−6.60	0.703 (−0.73)	1.08 0.87
7	1S4V-2BDZ_51 (93)	−1.01	90.8 7.0 1.1 1.1	5.384	0.565	−6.48	0.751 (−0.22)	0.79 0.71
8	1S4V-2P7U_13 (89)	−1.01	88.7 10.2 1.1 0.0	5.391	0.543	−6.39	0.677 (−1.01)	0.82 1.10
9	1S4V-2P7U_68 (89)	−1.00	90.3 8.1 1.6 0.0	5.297	0.538	−6.64	0.684 (−0.93)	0.61 1.10

[a] Values correspond to the percentages of residues that are located in the most favorable, additionally allowed, generously allowed and disallowed regions, respectively

[b] The first value corresponds to the QMEAN global score, while the value between parenthesis refers to the QMEAN Z-score

[c] RMSD values of bebesipain-1 model compared to the 3D structures of the templates used for the alignment (listed in the same order as column 2)

and as very good if LG score >5 and MaxSub >0.8). Thus, model 7 of babesipain-1 was evaluated as "very good" according to the LG score (5.384) and "good" according to the MaxSub index (0.565).
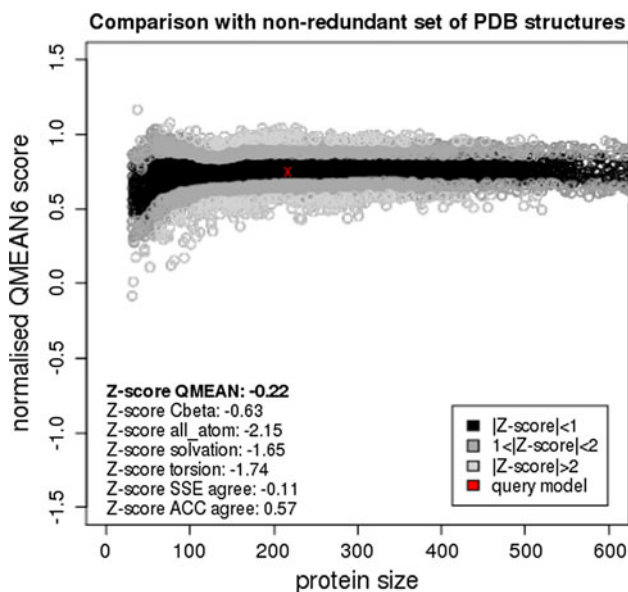
The QMEAN score corresponds to the global score of the whole model, on the basis of a linear combination of six structural descriptors, reflecting the predicted model reliability ranging from 0 to 1 with higher scores for reliable models. Accordingly, the global score of 0.751 reflects the reliability of the babesipain-1 model. Furthermore, the quality of the model can be compared to reference structures of high resolution obtained from X-ray
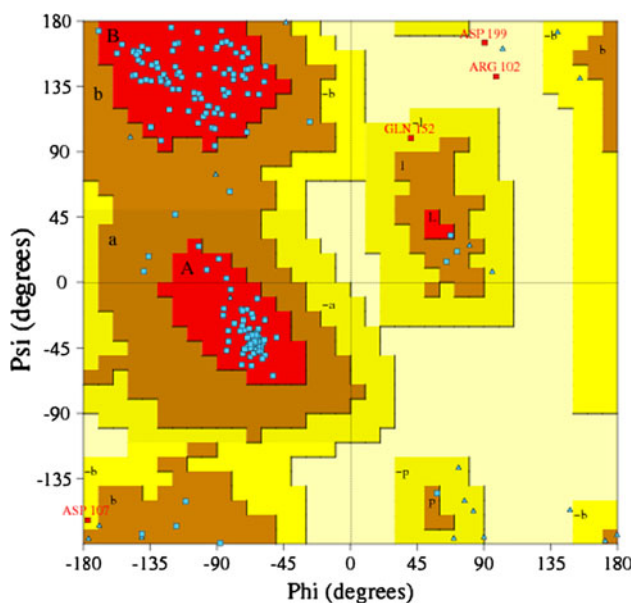
Fig. 3 Ramachandran plot of the homology modeled structure of babesipain-1. The different *colored areas* indicate "disallowed" (*beige*), "generously allowed" (*yellow*), "additional allowed" (*brown*), and "most favored" (*red*) regions


Fig. 5 Absolute quality of babesipain-1 model as assessed by QMEAN Z-score. Good models are generally located in the *dark zone*. The *red marker* indicates the positioning of babesipain-1 model 7

crystallography analysis through QMEAN Z-score, where a value of 0 is the average value for a good model [35]. According to Benkert et al. [35], QMEAN Z-score provides an estimation of the "degrees of nativeness" of the structural features observed in a model and indicates if the model has a quality comparable to experimental structures. In the present analysis, QMEAN Z-score for babesipain-1 model is −0.22 (Fig. 5) that, together with the other validation analyses presented above, reinforces the good quality of the derived model for babesipain-1. The model coordinates are supplied as Supporting Information.
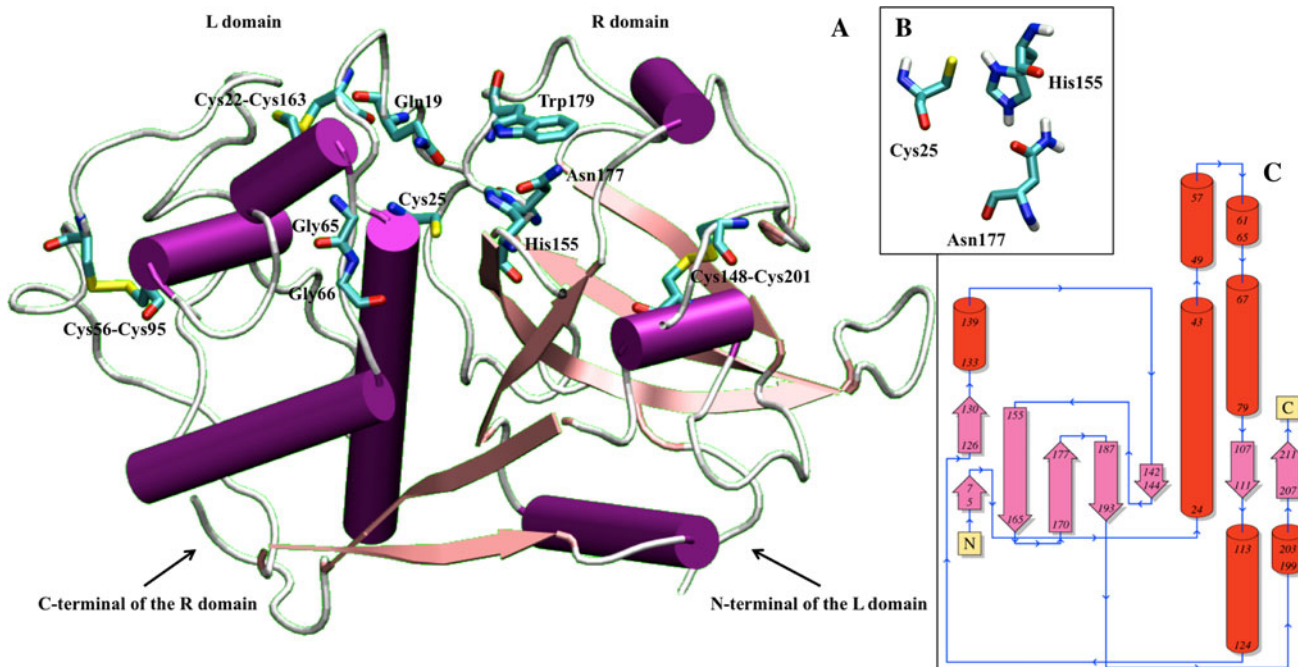
Babesipain-1 structure

Typical papain-like cysteine protease features were observed for the modeled structure of babesipain-1, composed of two domains, an α-helix-rich (L) domain and a β-sheet-rich (R) domain, separated by a groove containing the active site (Fig. 6a). The L domain is composed of four helices and the R domain is formed by six β-sheets and three small helices at the surface (Fig. 6a, c), which are typical features of the C1 papain-like fold [38]. The C- and N-termini of the R and L domains, respectively, bind to the L and R domains to stabilize the binding region.
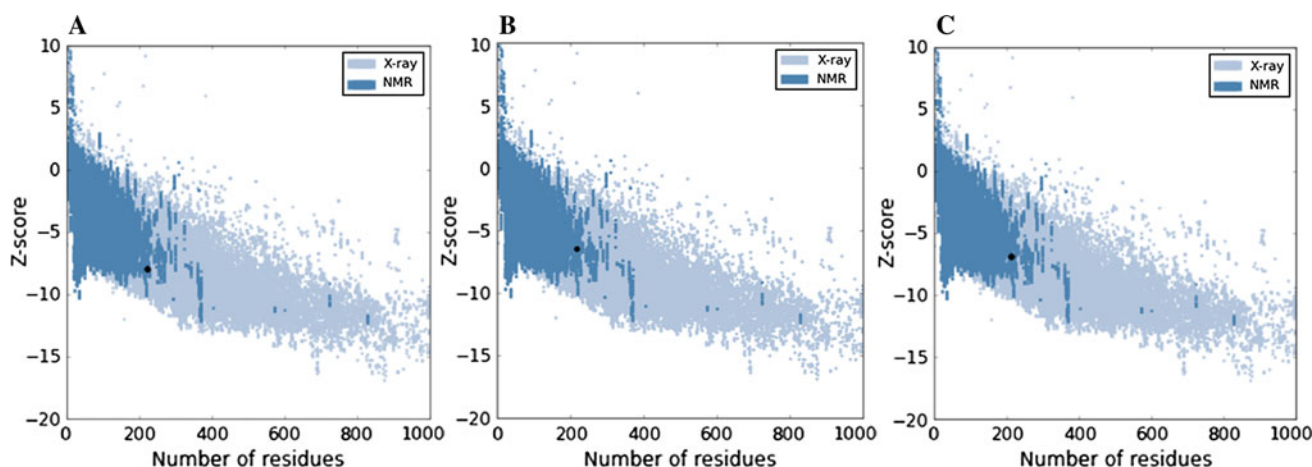
Fig. 4 ProSA analysis for the model structure of babesipain-1 (**b**) the template structures, 1S4V (**a**) and 2BDZ (**c**) with Z-scores values of −6.48, −7.98 and −6.85, respectively

**Fig. 6 a** Modeled structure of babesipain-1, where some conserved catalytic residues and the disulfide bonds are also represented. The model contains two typical L (N-terminal) and R (C-terminal) domains of papain-like cysteine proteases sharing "V" shaped active site. **b** The active site residues (Cys25, His155, Asn177) of babesipain-1. **c** Predicted secondary structure of babesiapin-1 using PDBSum server showing two domains of the protein. The N-terminal domain is dominated by α-helices (*red*) while the C-terminal domain is dominated by anti-parallel β-sheets (*pink*)

Babesipain-1 contains the seven cysteine residues common to the papain family, six of them involved in disulfide bonds (Cys22–Cys63, Cys56–Cys95 and Cys148–Cys201); the seventh residue, Cys25, is the active catalytic residue. Residues that constitute the binding pocket surround the catalytic Cys25, located in the L domain at the N-terminus of helix-1, as shown in Fig. 6a. In the vicinity of Cys25 is His155, placed at the R domain at the N-terminus of sheet-5, and Asn177, which may facilitate the appropriate orientation for the formation of the thiolate/imidazolium ion pair (Fig. 6b). In babesipain-1, as in other enzymes of the family, Gln19 and Trp179, whose side chains form the "oxyanion hole", are in a similar orientation (Fig. 6a); this is an important feature for the enzyme's proteolytic activity, as the "oxyanion hole" stabilizes the tetrahedral adduct during the nucleophilic attack of the thiolate anion to the appropriate electron deficient carbonyl of the substrate [32]. Additionally, we observed the typical glycine-rich region, comprising mainly Gly65 and Gly66, that in other papain-like cysteine proteases was found to provide additional stability to the complex by forming a constellation of hydrogen bonds with the substrate [39].

Usually, the active site of papain-like cysteine proteases is constituted by four pockets [32], S1, S1′, S2 and S3, as shown in Fig. 7. Residues within 6 Å of the active site Cys51 contour the binding pocket and are listed in Table 4.
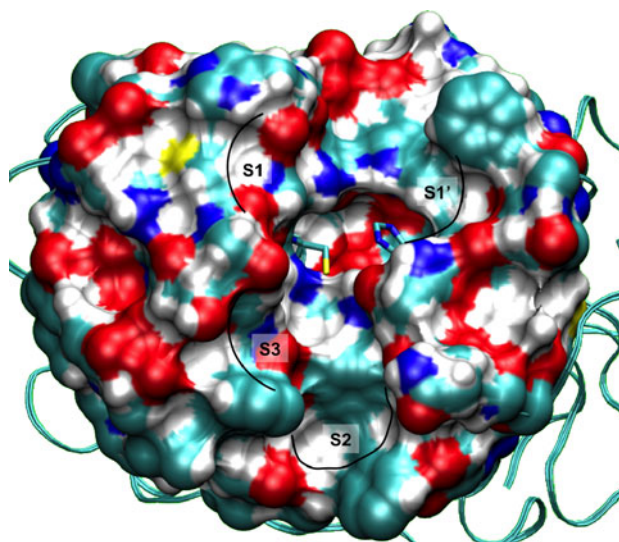
Fig. 7 Representation of the babesipain-1 catalytic site surface showing the S1, S1′, S2 and S3 pockets

The S1 pocket is the least defined pocket in cysteine proteases, which comprises Gln19 of the "oxyanion hole." The most well-defined pocket governing ligand specificity is the S2 pocket. One of the highly conserved residues in the S1′ pocket is Trp179, the other amino acid that forms the "oxyanion hole". The glycine-rich region of the binding site represents the S3 pocket. Although some

Table 4 Important residues lining the binding pockets of babesipain-1 and falcipain-2

Cysteine protease	S1′	S1	S2	S3
Babesipain-1	Val130-***Phe132-Thr133-Phe137***-His155-Trp179	Gln19-Gly23-Cys63-***Gly64***	***Phe67-Ser68-Tyr***129-Leu-153-***Phe206***	***Gln59-Ser60-Ser61***-Gly65-Gly66
Falcipain-2	Val150-***Val152-Ser153-Ala157***-His174-Trp206	Gln36-Gly40-Cys80-***Asn81***	***Leu84-Ile85-Ser149***-Leu172-***Asp234***	***Lys76-Asn77-Tyr78***-Gly82-Gly83

The residues are numbered according to enzyme mature domain and those differing in the two enzymes are shown in bold italic

differences are observed, the overall topology of the active site of babesipain-1 is similar to that of other family members as the majority of the binding site residues are conserved.

Comparing the sequences of babesipain-1 and falcipain-2, significant differences are found in the active sites of the corresponding proteases (Table 4). As previously reported, these differences are observed mainly in the S2 site [6], while cavity S1 is more conserved. The nature of the S2 pocket, and in particular of the residue present in the hollow end of the pocket, is thought to be essential to the substrate specificity of clan CA enzymes [32]. Hydrophobic residues usually constitute the S2 pocket, but the key residue (residue 205 in papain) present at the bottom of the pocket is not conserved [32]. In this critical position of falcipain-2 and of babesipain-1, we find the polar residue Asp234 [40] and the bulky hydrophobic Phe206 residue, respectively. This difference was suggested to be responsible for a narrower S2 pocket in babesipain-1 compared to that in falcipain-2 and, hence, responsible for the P2 rank ordering observed for babesipain-1, Val > Leu > Phe, in contrast with that in falcipain-2 following Leu > Phe > Val ordering [6]. However, the comparison of babesipain-1 and falcipain-2 structures shows that the difference in the rank ordering is more likely due to the presence of the bulky hydrophobic Tyr129 in babesipain-1, whereas falcipain-2 exhibits the considerably smaller Ser149. Indeed, we noticed that the S2 pocket in babesipain-1 is not narrower than that in falcipain-2 in terms of their heights and widths, but it is definitively shallower due to the presence of Tyr129 (Fig. 8). Moreover, S2 subsite of babesipain-1 is lined by three bulky hydrophobic residues Phe67, Tyr129 and Phe206, allowing that small hydrophobic residues, as valine, are better accommodated than bulky hydrophobic ones, as phenylalanine, which is in agreement with the P2 preference rank ordering obtained for babesipain-1. In summary, the above-mentioned analyses indicate that the model structure is consistent with the current understanding of the protein structure.

Docking results

Docking calculations were performed to predict the structures of complexes between babesipain-1 and the two families of compounds **1** and **2** in order to understand their inhibitory activities in vitro. While docking algorithms have been reasonably successful in predicting binding modes, scoring the poses to predict the binding affinity has proved to be more challenging. Thus, when accurate calculations of binding energies are required, more precise methods should be used, such as MM-PBSA. However, as our main objective was to analyze the interactions established between the ligands and the protein, we didn't engage into more computationally demanding techniques.

Computational results suggested a preferred binding mode for HEDICINS, **1**, into babesipain-1 binding site by placing the 7-chloroquinolyl, *homo*-phenylalanyl, leucyl and cinnamoyl groups into S2′, S1′, S1 and S2 subsites, respectively (Fig. 9a). A similar conformation was previously obtained when docking the same compounds against falcipain-2 [15]. However, we noticed that the 7-chloroquinoline ring of compounds **1** fails to form π–π interactions with the conserved Trp179 due to steric hindrance by residue Phe137 in babesipain-1. Instead, in falcipain-2, close to Trp206 is the small Ala157 residue, which does not cause such a steric effect, allowing the aforementioned π–π interactions to occur. We also observed that the first amide bond following the 7-chloroquinoline ring establishes a hydrogen bond with the NH of Trp179 side chain. Other relevant interactions observed between this family of compounds and babesipain-1 were (1) a hydrogen bond between the second carbonyl of the ligand and catalytic His155, and (2) π–π interactions between the aromatic ring of *homo*-Phe and Phe137.

The vinyl bond of compounds **1** was placed within 3–5.5 Å of the catalytic Cys thiolate. Although 3 Å seems to be a reasonable distance for a covalent bond between the enzyme and compounds **1**, the placement of the rather rigid cinnamoyl group in S2 severely hinders an attack by the catalytic Cys, which may account for the modest inhibitory activity shown by HEDICINS against babesipain-1. Noteworthy, docking of **1c** shows that this compound is able to fit into the babesipain-1 binding site, which corroborates its in vitro activity (IC$_{50}$ = 27.7 μM), but contrasts with the compound's behavior as a falcipain-2 inhibitor [15]: **1c** was previously found to be inactive against falcipain-2 due to the bulky *para*-isopropyl group blocking its fitting into the

Fig. 8 Amino acids that constitute the S2 cavity of babesipain-1 (*purple*) and falcipain-2 (*pink*) are represented in *licorice*, *ball* and *sticks*, respectively
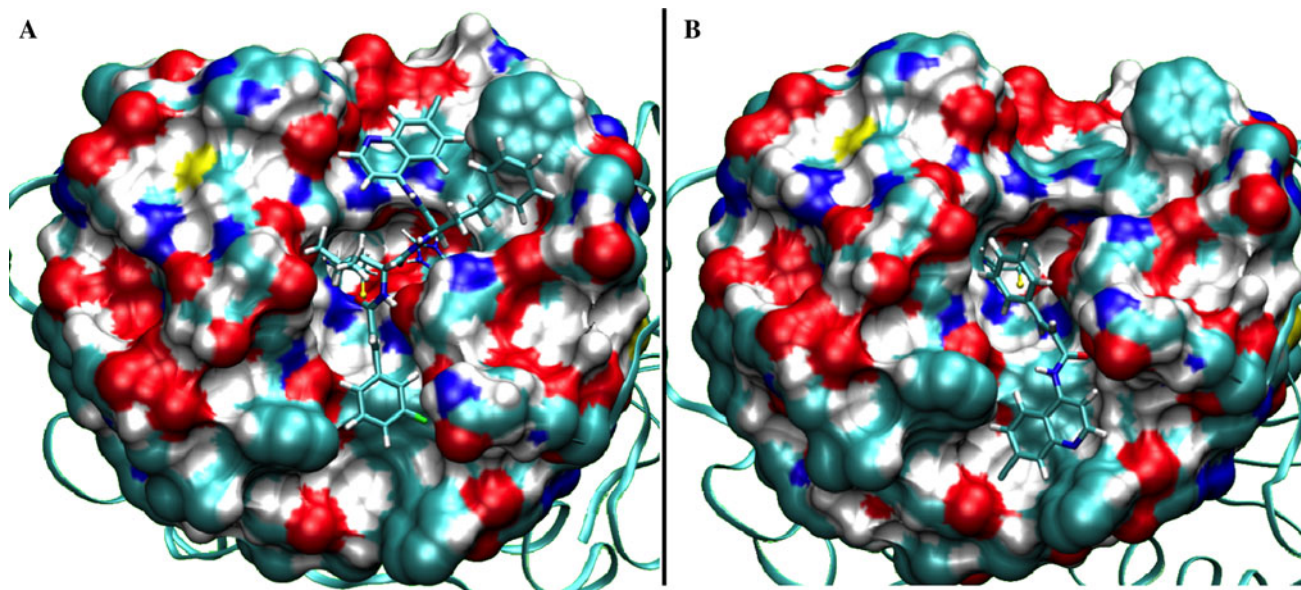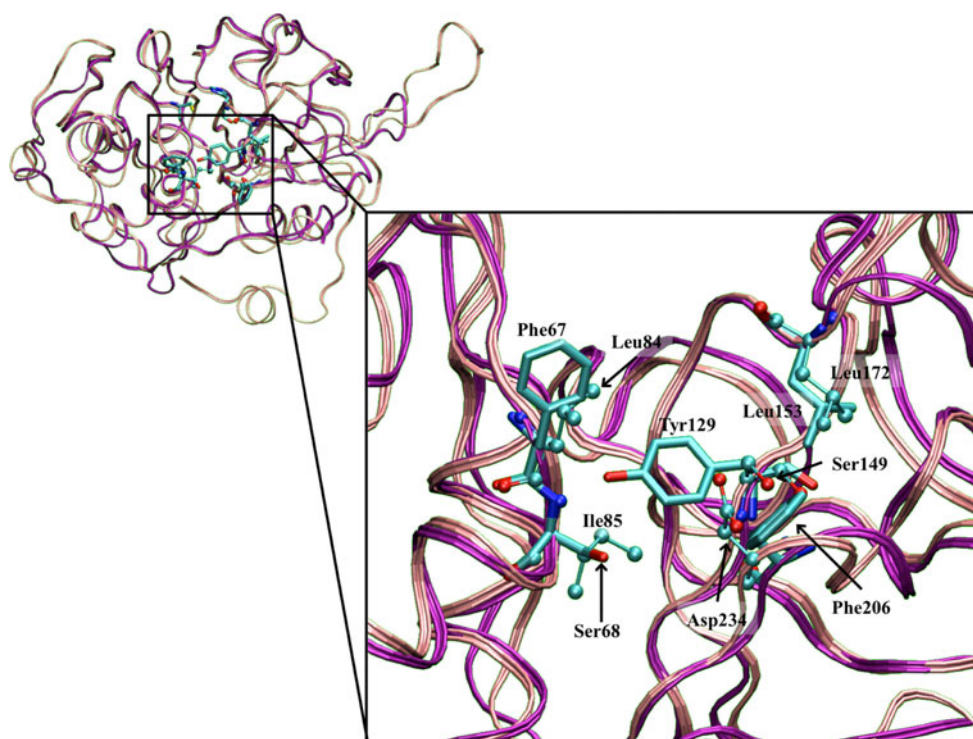
Fig. 9 Preferred binding modes of **a** HEDICINS **1** and **b** HECINS **2** into babesipain-1 binding site

falcipain's binding site. Hence, it seems that, while shallower, babesipain-1 S2 pocket is wider than its falcipain-2 equivalent, allowing the accommodation of the rigid cinnamoyl ring presenting a bulky *para*-isopropyl substituent. Still, we believe that, although S2 subsite of babesipain-1 can accommodate the cinnamoyl group better than falcipain-2, a smaller group at this position would be preferable.

Further analyses of the structure of babesipain-1 suggested that a polar group in the *para* position of the cinnamoyl ring is not favored, as the hydrophobic side chains of

the residues Phe67, Leu153 and Phe206 are placed at the bottom of babesipain-1 S2 cavity. This observation is in agreement with in vitro data, where compound **1d**, bearing the polar *p*-OMe cinnamoyl substituent, was the worst inhibitor of the series, with an IC_{50} value of 35.8 μM. Nevertheless, the positioning of polar groups at *ortho* or *meta* positions of the cinnamoyl ring is not unfavorable, because the substituents will be oriented either to the left or to the right side, being able to establish electrostatic contacts with the backbone of Gly66 and the backbone of Leu153,

respectively. Again, this is in agreement with in vitro results where compounds **1i** and **1j** presented better activities (IC$_{50}$ values of 18.6 and 10.2 μM, respectively) than compound **1d**. The higher activity of compound **1j** compared to that of **1i** is most likely due to the closer proximity of the carbon prone to suffer the nucleophilic attack. For HEDICINS bearing a nitro substituent, this happens in the case of the α-carbon of the α, β-unsaturated carbonyl moiety [15] that, in the case of compounds **1j** and **1i**, is at a distance of 3.0 and 4.2 Å from the catalytic Cys thiolate, respectively.

By analyzing the docked binding modes of HECINS, **2**, we observed a preference for the positioning of the 7-chloroquinoline group at the S2 cavity, with the cinnamoyl group pointing toward the catalytic cysteine (Fig. 9b) and placing the vinyl bond within ∼4.5 Å of this residue. The only exception to this binding mode was found for compound **2j**, which presents an upside-down orientation with the vinyl bond located far away from the catalytic site, hence excluding a possible reaction with the catalytic thiolate (supporting information, Figure S1). This result is in agreement with in vitro data, where compound **2j** (IC$_{50}$ = 48.3 μM) was the least active of the HECINS series. The 7-chloroquinoline ring of the remaining HECINS establishes several hydrophobic contacts with the hydrophobic residues of the S2 subsite. Additionally, the slight preference observed for cinnamoyl substituents with a higher electron-donating character could be explained with the previously reported atomic Fukui indices for HECINS [15], which were used as a measure of the activation, i.e., electron density imbalance of the vinyl double bond. Except for **2j**, the β-carbon of the α, β-unsaturated carbonyl moiety, which is the carbon of the vinyl bond closer to the babesipain-1 catalytic cysteine, was the preferred site of nucleophilic attack for all HECINS **2**. Thus, substituents with a higher electron-donating character favor electron delocalization toward the carbonyl group, which will allow a higher activation of the double bond, and, consequently, favor the interaction of the β-carbon with catalytic Cys. Again, these results are in agreement with in vitro data, as the two compounds showing the higher activation of the double bond in the β-carbon [15], **2b** and **2d**, were also the best inhibitors of the series, with IC$_{50}$ values of 13.4 and 9.8 μM. Finally, as outlined for HEDICINS **1**, although the S2 cavity is wide enough to accommodate bulky hydrophobic groups, such as the 7-chloroquine moiety, changing those groups by smaller ones might favor the inhibitory activity against babesipain-1.

Conclusions

The best babesipain-1 3D model structure was obtained through homology modeling by combining templates 1S4V and 2BDZ. The model structure was well validated by PROCHECK, ProQ, ProSA and QMEAN, and presented all typical features of papain-like cysteine proteases. Comparison of falcipain-2 with babesipain-1 demonstrated that the active cavity of the latter is globally wider, shallower and more hydrophobic. In silico docking studies showed that all HEDICINS **1** are placed approximately in the same conformation inside the binding cavity; moreover, differences between IC$_{50}$ values of compounds **1** against babesipain-1 were perfectly explained by stereoelectronic aspects of the interactions between the distinct ligands and the enzyme. Similar observations were made in the docking of HECINS **2** to babesipain-1: all but one of compounds **2** were docked in approximately the same conformation, where slight differences were in agreement with results from in vitro experiments; the outlier compound of this series (**2j**), whose vinyl bond was farther apart from the enzyme's catalytic Cys, was also the worst babesipain-1 inhibitor in vitro. Altogether, these results undeniably demonstrate the validity of the babesipain-1 3D model constructed in the present work, which represents a new doorway toward design and discovery of novel anti-babesia drugs. Further in vitro studies need to be conducted in order to analyze whether the activity of compounds herein reported against babesipain-1 correlates with their ability to impair growth of *B. bigemina* parasites. Still, babesipain-1 has similar characteristics to falcipain-2, and inhibition of the latter is known to strongly affect normal development of intraerythrocytic malaria parasites [6, 9, 14]. Therefore, we are strongly inclined to believe that babesipain-1 will become a relevant therapeutic target against babesiosis.

References

1. Schnittger L, Rodriguez AE, Florin-Christensen M, Morrison DA (2012) Babesia: a world emerging. Infect Genet Evol 12(8):1788
2. Mosqueda J, Olvera-Ramirez A, Aguilar-Tipacamu G, Canto GJ (2012) Current advances in detection and treatment of babesiosis. Curr Med Chem 19(10):1504
3. de Waal DT, Combrink MP (2006) Live vaccines against bovine babesiosis. Vet Parasitol 138(1–2):88
4. Fish L, Leibovich B, Krigel Y, McElwain T, Shkap V (2008) Vaccination of cattle against *B. bovis* infection with live attenuated parasites and non-viable immunogens. Vaccine 26(Suppl 6):G29

5. Vial HJ, Gorenflot A (2006) Chemotherapy against babesiosis. Vet Parasitol 138(1–2):147
6. Martins TM, do Rosario VE, Domingos A (2012) Expression and characterization of the *Babesia bigemina* cysteine protease BbiCPL1. Acta Trop 121(1):1
7. http://www.sanger.ac.uk/resources/downloads/protozoa/babesia-bigemina.html. Accessed 08 May 2013
8. Okubo K, Yokoyama N, Govind Y, Alhassan A, Igarashi I (2007) *Babesia bovis*: effects of cysteine protease inhibitors on in vitro growth. Exp Parasitol 117(2):214
9. Martins TM, do Rosario VE, Domingos A (2011) Identification of papain-like cysteine proteases from the bovine piroplasm *Babesia bigemina* and evolutionary relationship of piroplasms C1 family of cysteine proteases. Exp Parasitol 127(1):184
10. Sijwali PS, Shenai BR, Gut J, Singh A, Rosenthal PJ (2001) Expression and characterization of the *Plasmodium falciparum* haemoglobinase falcipain-3. Biochem J 360(Pt 2):481
11. McKerrow JH, Rosenthal PJ, Swenerton R, Doyle P (2008) Development of protease inhibitors for protozoan infections. Curr Opin Infect Dis 21(6):668
12. Teixeira C, Gomes JR, Gomes P (2011) Falcipains, *Plasmodium falciparum* cysteine proteases as key drug targets against malaria. Curr Med Chem 18(10):1555
13. Capela R, Oliveira R, Goncalves LM, Domingos A, Gut J, Rosenthal PJ, Lopes F, Moreira R (2009) Artemisinin-dipeptidyl vinyl sulfone hybrid molecules: design, synthesis and preliminary SAR for antiplasmodial activity and falcipain-2 inhibition. Bioorg Med Chem Lett 19(12):3229
14. Martins TM, Goncalves LM, Capela R, Moreira R, do Rosario VE, Domingos A (2010) Effect of synthesized inhibitors on babesipain-1, a new cysteine protease from the bovine piroplasm *Babesia bigemina*. Transbound Emerg Dis 57(1–2):68
15. Perez BC, Teixeira C, Figueiras M, Gut J, Rosenthal PJ, Gomes JR, Gomes P (2012) Novel cinnamic acid/4-aminoquinoline conjugates bearing non-proteinogenic amino acids: towards the development of potential dual action antimalarials. Eur J Med Chem 54:887
16. Apweiler R, O'onovan C, Magrane M, Alam-Faruque Y, Antunes R, Bely B, Bingley M et al (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 40(Database issue):D71
17. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF, Notredame C (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39(Web Server issue):W13
18. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779
19. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17(4):355
20. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35(Web Server issue): W407
21. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283
22. Benkert P, Schwede T, Tosatto SC (2009) QMEANclust: estimation of protein model quality by combining a composite

23. scoring function with structural density information. BMC Struct Biol 9:35
24. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12(5):1073
25. Teixeira C, Gomes JR, Couesnon T, Gomes P (2011) Molecular docking and 3D-quantitative structure activity relationship analyses of peptidyl vinyl sulfones: *Plasmodium falciparum* cysteine proteases inhibitors. J Comput Aided Mol Des 25(8):763
26. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res 33(Web Server issue):W368
27. Case DA, Darden TA, Cheatham TE, Simmerling CLI, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry KF, Wong KF, Paesani F, Vanicek F, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco
28. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24(16):1999
29. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267(3):727
30. MarvinSketch 5.12.2 (2013) ChemAxon. http://www.chemaxon.com
31. Charton M (1975) Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. J Am Chem Soc 97(6):1552
32. Rawlings ND, Barrett AJ, Bateman A (2010) MEROPS: the peptidase database. Nucleic Acids Res 38(Database issue):D227
33. Sajid M, McKerrow JH (2002) Cysteine proteases of parasitic organisms. Mol Biochem Parasitol 120(1):1
34. Larsson P, Wallner B, Lindahl E, Elofsson A (2008) Using multiple templates to improve quality of homology models in automated homology modeling. Protein Sci 17(6):990
35. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15(11):2507
36. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27(3):343
37. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A (2001) A study of quality measures for protein threading models. BMC Bioinform 2:5
38. Siew N, Elofsson A, Rychlewski L, Fischer D (2000) MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics 16(9):776
39. Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG (1968) Structure of papain. Nature 218(5145):929
40. Brinen LS, Hansell E, Cheng J, Roush WR, McKerrow JH, Fletterick RJ (2000) A target within the target: probing cruzain's P1′ site to define structural determinants for the Chagas' disease protease. Structure 8(8):831
41. Kerr ID, Lee JH, Pandey KC, Harrison A, Sajid M, Rosenthal PJ, Brinen LS (2009) Structures of falcipain-2 and falcipain-3 bound to small molecule inhibitors: implications for substrate specificity. J Med Chem 52(3):852