JBC Papers in Press. Published on June 7, 2017 as Manuscript M117.788042
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M117.788042

Functional analyses yield detailed insight into the mechanism of thrombin inhibition by the antihemostatic salivary protein cE5 from *Anopheles gambiae*

Luciano Pirone[a], Jorge Ripoll-Rozada[b,c], Marilisa Leone[a], Raffaele Ronca[d], Fabrizio Lombardo[e], Gabriella Fiorentino[d], John F. Andersen[f], Pedro José Barbosa Pereira[b,c], Bruno Arcà[e*] and Emilia Pedone[a*]

a Institute of Biostructures and Bioimaging, National Research Council, Via Mezzocannone 16, 80134 Naples, Italy.

b IBMC-Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal

c Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen 208, 4200-135 Porto, Portugal

d Department of Biology, "Federico II" University, Via Cinthia, 80126 Naples, Italy.

e Department of Public Health and Infectious Diseases, Division of Parasitology, "Sapienza" University, P.le Aldo Moro 5, 00185 Rome, Italy.

f Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Twinbrook III, 12735 Twinbrook Parkway, National Institute of Health, Rockville, MD 20852, USA

Running title: *A. gambiae* cE5-thrombin interaction

*To whom correspondence should be addressed:
Emilia Pedone, Institute of Biostructures and Bioimaging, National Research Council, Via Mezzocannone 16, 80134 Naples, Italy. Tel. +39 081 2534521; e-mail: empedone@unina.it

Bruno Arcà, Department of Public Health and Infectious Diseases, Division of Parasitology, Sapienza University, Piazzale Aldo Moro 5, 00185 Rome, Italy. Tel. +39 06 4991 4413 (4647); e-mail: bruno.arca@uniroma1.it

Keywords: intrinsically disordered protein; protein-protein interaction; crystal structure; enzyme inhibitor; hemostasis; Anopheles; mosquito saliva.

ABSTRACT

Saliva of blood-feeding arthropods carries several antihemostatic compounds whose physiological role is to facilitate successful acquisition of blood. The identification of novel natural anticoagulants and the understanding of their mechanism of action may offer opportunities for designing new antithrombotics disrupting blood clotting. We report here an in depth structural and functional analysis of the anophelin family member cE5, a salivary protein from the major African malaria vector *Anopheles gambiae* that specifically tightly and quickly binds and inhibitss thrombin. Using calorimetry, functional assays and complementary structural techniques we show that the central region of the protein, encompassing amino acids Asp31-Arg62, is the region mainly responsible for α-thrombin binding and inhibition. As previously reported for the *Anopheles albimanus* orthologue anophelin, cE5 binds both thrombin exosite I with segment Glu35-Asp47 and the catalytic site with the region Pro49-Arg56, which includes the highly conserved DPGR tetrapeptide. Moreover, the N-terminal Ala1-Ser30 region of cE5 (which includes an RGD tripeptide) and the additional C-terminal serine-rich Asn63-Glu82 region (absent in orthologues from anophelines of the New World species *A. albimanus* and *.darlingi*) also played some functionally relevant role. Indeed, we observed decreased thrombin binding and inhibitory properties even when using the central cE5 fragment (Asp31-Arg62) alone. In summary, these results shed additional light on the mechanism of thrombin binding and inhibition by this family of salivary anticoagulants from anopheline mosquitoes.

The ability of hematophagous insects to use blood as a food source involves complex behavioral, morphological and physiological adaptations to find suitable hosts, pierce their skin and efficiently acquire and digest blood. As a result, the saliva of blood-feeding arthropods carries a complex cocktail of compounds playing crucial roles in counteracting the three powerful and highly redundant responses of vertebrates to tissue injury: hemostasis, inflammation and immunity (1). Blood feeding evolved independently several times in different taxa (2) and this resulted in a large variety of salivary anti-hemostatic factors targeting platelet aggregation, vasoconstriction and blood clotting (1,3). Salivary anticoagulants found in the mosquito family offer a good example of this convergent evolution: members of the culicine subfamily (i.e., *Aedes* and *Culex* species) inhibit factor Xa whereas species belonging to the anopheline subfamily adopted a thrombin-directed anticoagulant activity (4).

The anti-thrombin of anopheline mosquitoes was first identified in the South-American malaria vector *Anopheles albimanus* as a small cysteine-less polypeptide of 61 amino acids named anophelin (5). Synthetic anophelin was shown to be a highly specific, slow, tight-binding, reversible inhibitor of α-thrombin (6). It binds both the catalytic site and the anion binding exosite 1 (TABE1) of thrombin, a region that is known to be involved in the recognition of fibrinogen (7). A cDNA encoding the anophelin orthologue in the African malaria vector *Anopheles gambiae* had been previously identified by a selective cloning strategy and the encoded protein named cE5 (8). The two proteins, anophelin and cE5, are quite divergent (43% identity) with the slightly longer *A. gambiae* protein (82 amino acids) displaying an additional stretch of amino acids at the carboxy-terminal. Despite these differences, recombinant *A. gambiae* cE5 showed full preservation of anti-thrombin function, although the two proteins displayed slightly different binding kinetics. The *A. albimanus* anophelin is a slow-binding thrombin inhibitor (6), whereas the *A. gambiae* cE5 behaves as a fast-binding inhibitor (9), raising the possibility that the extended carboxy-terminal may participate in thrombin binding.

The crystal structure of the *A. albimanus* anophelin in complex with human thrombin revealed its unique inhibition mechanism in comparison to other known thrombin inhibitors (10). In particular, the C-terminal segment (^{A}E32-P61) appeared sufficient for thrombin inhibition, with residues ^{A}D33-F45 blocking the exosite I and the highly conserved DPGR tetrapeptide (^{A}D50-R53) occupying the active site cleft of the enzyme and disrupting the characteristic catalytic triad of the serine proteinase.

The release of the genomes of sixteen *Anopheles* mosquitoes (11) allowed identifying anophelin/cE5 orthologues from several additional species, offering further insights into the evolution of this unique family of thrombin inhibitors. Multiple alignment of cE5/anophelin family

members showed that, in comparison to the New World species *A. albimanus* and *A. darlingi*, orthologues from the Old World *Anopheles* species display an extended, serine-rich, carboxy-terminal region (10,12). In order to investigate the possible involvement of the C-terminal region (^{cE5}N63-E82) in thrombin-binding and inhibition, as well as the role of the conserved N-terminal portion (^{cE5}A1-S30), we performed a structural and functional analysis using both the cE5 protein and a set of cE5-derived peptides. Moreover, the three-dimensional structure of the human α-thrombin - *A. gambiae* cE5 complex unveiled the details of thrombin recognition and inhibition by the Old World anophelin orthologues.

RESULTS

Structural properties of cE5 as determined by Circular Dichroism and NMR spectroscopy- Prediction analysis carried out by DISOPRED (VL-XT predictor; http://www.pondr.com/) suggested that the cE5 protein, as previously reported for the *A. albimanus* anophelin (10), was intrinsically disordered in solution. However, in contrast to *A. albimanus* anophelin, cE5 displayed a slight propensity to be structured in the region encompassing amino acids ^{cE5}P54 to ^{cE5}S64 (Figure 1). Similar results were obtained using the database of intrinsically disordered proteins, MobiDB (not shown)(13). Furthermore, cE5 displayed a typical Circular Dichroism (CD) spectrum of an unstructured protein, with a minimum at 202 nm (Figure 2) . After the addition of 2,2,2-trifluoroethanol (TFE) as co-solvent, which may help exploring the intrinsic tendency of a polypeptide to assume secondary structure elements (14-16), we found a very low propensity of cE5 to adopt an alpha-helical conformation. Indeed, even with high concentrations of TFE (60%), an increase of only ~10% in the alpha helical content was observed (inset Figure 2).

The conformational properties of cE5 were also investigated using NMR spectroscopy. The [^{1}H-^{15}N] HSQC spectrum of ^{15}N-labelled cE5 showed a limited chemical shift range (Figure 3A), indicative of an unfolded molecule having few secondary structure elements, consistent with the random-coil state suggested by CD data. In addition, [^{1}H-^{15}N] HSQC spectra of uniformly ^{15}N-labelled cE5 (40 μM) in the absence and the presence of different amounts of TFE (10, 30 and 60% (v/v)) (Figure 3 B-D) were recorded in the same experimental conditions used for CD analyses. In complete agreement with the CD data, these NMR experiments indicated a continuous increase in signal dispersion (i.e., in structuration) with increasing TFE concentration. However, even at relatively high TFE concentrations, HSQC spectra still resemble those of canonical intrinsically disordered proteins (IDPs) (Figure 3 B-D).

Alignment of anophelin family members from different Anopheles species- The sequencing of the genomes of 16 anopheline species (11) spanning ~100 million years of evolution, allowed the identification of several additional members of the cE5/anophelin family (12). Alignment of the 18 full-length orthologues highlighted their considerable divergence: 18% to 77% amino acid identity among the different *Anopheles* species (excluding comparisons within the *A. gambiae* species complex). Overall, a block of sixteen amino acids, including the highly conserved tetrapeptide APQY, can be recognized at the N-terminal region and another highly conserved tetrapeptide, DPGR, is found towards the C-terminus (Figure 4). Between these two conserved blocks there is a more divergent central region enriched in acidic residues (D+E = 19-33%). A careful examination of the aligned proteins disclosed three features of potential functional relevance: (i) the N-terminal RGD tripeptide in *A. gambiae* and in a few other species of the complex; (ii) the conserved DPGR tetrapeptide, with proline to alanine replacement in *A. atroparvus* and *A. epiroticus*; (iii) the presence of a longer C-terminal in Old World anophelines.

The tripeptide RGD is known for its involvement in binding to integrins (17). RGD-containing proteins with anti-platelet activity (disintegrins) have been found in snake venoms and in the saliva of blood-feeding arthropods (ticks and horseflies), leeches and worms, but never in mosquitoes. Disintegrins act as antagonists of integrin αIIbβ3 and inhibit fibrinogen binding to platelets and subsequent platelet cross-linking (18). Typically, integrin-binding RGD motifs are positioned in the loop of peptide hairpins formed by disulfide bonds (19), which is not the case for the

cE5 RGD. We tested cE5 and the cE5-derived P1 peptide (Figure 4) in platelet aggregation assays. We found no or very little effect on platelet aggregation (Figure S1), which seems to rule out interaction of cE5 RGD with integrin αIIbβ3 (even though we cannot exclude binding to other integrins).

The DPGR tetrapeptide was previously found to play a crucial role in *in vitro* binding to α-thrombin (20) and shown to occupy the active site cleft of the enzyme in the anophelin/thrombin crystal structure (10). Substitutions of the catalytic triad-disrupting aspartate and of the S1-targeting arginine residues in the *A. albimanus* protein were detrimental for the inhibitory activity (10). We do not know the effect of replacing the proline in this tetrapeptide with an alanine. However, it is conceivable that given the compact size of alanine, it may replace the pivotal proline residue without much impact in inhibitor activity.

Intriguingly, Old World anophelines (subgenera *Cellia* and *Anopheles*) carry an additional serine-enriched (29-56%) C-terminal stretch of 7 to 21 amino acids which is absent in the New World species *A. albimanus* and *A. darlingi* (subgenus *Nyssorhynchus*). It has been previously shown that the anophelin C-terminal fragment ^{A}E32-P61 is mainly responsible for both binding to and inhibiting thrombin. Fragment ^{A}A1-D31 did not appear to play any functional role in thrombin targeting and inhibition (10). However, *A. gambiae* cE5 appeared to possess higher affinity for thrombin and more potent anticoagulant activity than *A. albimanus* anophelin (K_i = 3.5 pM versus 34 pM; prolongation of thrombin time from 16 to 221 seconds for cE5 and to 84 seconds for anophelin) (9,10). Moreover, anophelin and cE5 also displayed different kinetics of binding to thrombin, the former being a slow-binding thrombin inhibitor (6) and the latter behaving as a fast-binding inhibitor (9). To evaluate the possible role of the C-terminal serine-rich fragment of cE5 we designed the peptides P1 (^{cE5}A1-S30), P2 (^{cE5}D31-R62), P3 (^{cE5}N63-E82) and P5 (^{cE5}D31-E82) (Figure 4) and assessed their interaction with thrombin by isothermal titration calorimetry (ITC) and enzymatic assays.

Binding studies-As determined by ITC, cE5 formed a very stable complex with thrombin (Kd ≤ 1 nM), with an apparent affinity slightly higher than anophelin (Kd ≤ 2 nM; Table 1, Figure 5A-B). In addition, binding assays with the cE5-based peptides P1, P2, P3 and P5 suggested that the region involved in the interaction with thrombin resides almost completely on the fragment ^{cE5}D31-R62, corresponding to the P2 peptide (Kd ≤12 nM). This was underscored by the comparable ΔH values obtained for cE5 (-9.85 kcal/mol) and P2 (-9.66 kcal/mol) (Table 1, Figure 5C-E). Surprisingly, binding experiments with P5 yielded larger Kd (≤25 nM) and smaller ΔH (-4.89 kcal/mol) values than those obtained for P2, pointing to a negative contribution in thrombin binding of the region corresponding to P3 in the absence of the N-terminal portion encompassed by P1 (Table 1, Figure 5 F).

It is worth noting that the binding affinity (Ka) of cE5 to thrombin is beyond the limit of direct calorimetric determination. Under these conditions, ITC titration still provides an accurate measurement of the binding enthalpy (Table 1). However, ITC can be used to determine the complete binding thermodynamics of ligands with affinities down to the picomolar range using displacement titration (21), which was performed for this system in the presence of a weak thrombin inhibitor, the synthetic tetrapeptide D-Phe-Pro-D-Arg-Ala (22). Under these conditions, cE5 displayed a binding constant of 0.3 nM (Figure S2), in line with the value previously reported (9,22) and with that obtained from the thrombin amidolytic activity inhibition assays in this study (Table 2).

We also used NMR titration experiments and analysis of 2D [^{1}H, ^{15}N] HSQC spectra (23) to study protein-protein interactions. HSQC spectra of ^{15}N-labelled cE5 were recorded in the absence and the presence of increasing amounts of unlabeled thrombin. At cE5/thrombin ratios higher than 1/5 numerous spectral changes could be observed, pointing to interactions taking place between the two proteins (Figure S3). However, it was not possible to reach saturating conditions, which would have allowed gaining more detailed structural information. This is most likely due to the relatively large size of the complex, which caused severe line broadening and loss of sensitivity in the NMR spectra (Figure S3). In addition, 2D [^{1}H, ^{1}H] NMR spectra collected for the P2 peptide showed a similar lack of relevant chemical shift dispersion and the absence of a consistent number of Nuclear Overhauser Effect (NOE) contacts, confirming a disordered state for the P2 peptide in aqueous solution (Supplemental Figure S4, right panel).

Moreover, no significant change in either chemical shift and/or NOE intensity could be observed when comparing 2D [^{1}H, ^{1}H] NOESY experiments with the P2 peptide (470 μM) in the absence and the presence of a sub-stoichiometric amount of thrombin (25 μM) (Figure S4). These observations suggest that under our experimental conditions the peptide "dynamically" binds to thrombin by preserving disorder and flexibility.

Thrombin inhibition assays-The inhibitory effect of cE5 and of peptides P1, P2, P3 and P5 on the *in vitro* amidolytic activity of thrombin was measured following the hydrolysis of a fluorogenic substrate. A thrombin inhibition of approximately 80% was found for cE5, in line with previous measurements using a chromogenic substrate (9). The P2 peptide still retained inhibitory properties, even though it was significantly less effective than cE5 (~39% inhibition), whereas neither P1 nor P3 affected thrombin activity (Figure 6). The inhibitory effect of P5 was not significantly different from those observed using P2 alone or any combination including either P2 or P5.

Overall, thrombin inhibition results fit quite well with ITC data confirming that the region corresponding to the P2 peptide (^{cE5}D31-R62) is the main segment responsible for both binding to thrombin and for the inhibitory activity. The P1 peptide (^{cE5}A1-S30) did not bind thrombin and did not affect its activity, and these results are fully consistent with previous observations on *A. albimanus* anophelin (10). Moreover, our results seem to rule out any direct functional role of the serine-rich carboxy-terminal region in thrombin inhibition. In fact, the P3 peptide alone did not bind to or inhibit thrombin and the P5 peptide, which encompasses both P2 and P3, was not a better inhibitor or a more efficient binder than P2.

Overall structure of the A. gambiae cE5 / human α-thrombin complex-The recombinant P5 fragment (^{cE5}D31-E82) of *A. gambiae* cE5 was crystallized in complex with human α-thrombin and its structure determined at 1.45-Å resolution (Supplementary Table 1 and Figure 7). There are three copies (A, B, and C) of the complex in the asymmetric unit: complex A comprising thrombin residues ^{T}S1E to ^{T}R15 and ^{T}I16 to ^{T}G246, and inhibitor residues ^{cE5}E35 to ^{cE5}N63 (the chymotrypsinogen-based numbering will be used for thrombin); complex B encompasses thrombin residues ^{T}T1H to ^{T}R15 and ^{T}I16 to ^{T}G246, and inhibitor residues ^{cE5}E36 to ^{cE5}L42 and ^{cE5}A46 to ^{cE5}R62; finally, in complex C thrombin residues ^{T}G1D to ^{T}I14K, ^{T}I16 to ^{T}Q244, and inhibitor residues ^{cE5}R52 to ^{cE5}L61 could be modeled. In all complexes, thrombin's 149 insertion loop was disordered (from ^{T}W148 to ^{T}K149E, ^{T}W148 to ^{T}G149D, and ^{T}T147 to ^{T}G149D, for complexes A to C, respectively) and could not be modeled. Although the three complexes are structurally very similar (r.m.s.d. of 0.44 Å, 0.82 Å, and 0.93 Å for the superposition of the proteinase moieties of complexes A-B, A-C, and B-C, respectively), complex C displays an unusual flexibility of segments of the heavy chain of thrombin, encompassing residues ^{T}Q38-^{T}K70, ^{T}R101-^{T}E127, ^{T}M210-^{T}C220 and ^{T}F227-^{T}Q244. Therefore, only complex A will be described herewith (Figure 7). No significant density was present for the N-terminal tetrapeptide ^{cE5}D31 to ^{cE5}E34 or for the C-terminal 21 amino acids (^{cE5}S64 to ^{cE5}E82) of the P5 fragment, and these segments were therefore excluded from the model. Superposition of the structures of free thrombin (PDB entry 3U69 (22)) and the proteinase moiety of the A complex revealed that only minor rearrangements occur in the proteinase upon *A. gambiae* cE5 binding (r.m.s.d. of 0.36Å for 277 aligned Cα atoms).

Unsurprisingly, *A. gambiae* cE5 displays the same unique mode of thrombin inhibition observed for *A. albimanus* anophelin (10), further confirming it as a general mechanism for this class of inhibitors (MEROPS family I77 (24); Figure 7B). In agreement with the results of CD (Figure 2) and NMR measurements (Figure 3), the inhibitor adopts a mostly extended conformation, running in a reverse orientation to substrates on the surface of thrombin and interacting with both the exosite I and the active site region of the proteinase (Figure 7A, C, D). Further supporting the observed binding mode, both full-length *A. gambiae* cE5 and its truncated P5 fragment display three orders of magnitude larger inhibition constants towards the exosite I-disrupted γ-thrombin than towards α-thrombin (Table 2).

Bidentate binding of cE5 to thrombin-Similar to the *A. albimanus* anophelin-thrombin complex (10), the cE5 fragment spanning residues ^{cE5}E35 to ^{cE5}D47, blocks the exosite I of the proteinase (Figure 4 and Figure 7C). However, despite the overall

resemblance, the first half of this cE5 segment runs closer to the proteinase surface, with its main chain atoms deviating significantly from the path of the equivalent region of *A. albimanus* anophelin (Figure 7B). The inhibitor's tripeptide ^{cE5}E35-^{cE5}E36-^{cE5}F37 establishes water-mediated main chain-main chain contacts with thrombin's ^{T}I82-^{T}S83-^{T}M84 tripeptide. This region is also stabilized by an intramolecular hydrogen bond between the carbonyl oxygen of ^{cE5}D38 and the amide nitrogen of ^{cE5}L41, further reinforced by two interactions with the side chain of ^{T}Q38, involving the main chain nitrogen of ^{cE5}E43 and the carbonyl oxygen of ^{cE5}L41. This conformation results in the placement of the side chain of ^{cE5}F37 snuggly inside the hydrophobic depression bordered by the side chains of ^{T}L65, ^{T}I82 and ^{T}M84, to which it establishes Van der Walls interactions. Additionally, the aromatic ring of ^{cE5}F37 also contacts the side chain of the well-conserved ^{cE5}L42, in the same position of the homologous ^{A}L41 of *A. albimanus* anophelin, and in turn contacting ^{T}F34 and ^{T}Y76. The side chains of ^{cE5}P39 and ^{cE5}L41 pack on top of that of ^{cE5}F37, completing this strong hydrophobic cluster.

Further towards the active site cleft of thrombin, ^{cE5}D47 forms a salt bridge with ^{T}R73. Notably, the replacement of *A. albimanus* anophelin ^{A}F45 by the much shorter ^{cE5}A48 impairs the highly conserved stacking interaction with ^{T}F34 observed in numerous thrombin ligands, from protease-activated receptor 1 (PAR1) (25) to hirudin (26), and from boophilin (27) to anophelin (10). Instead, two polar interactions are observed between the carbonyl oxygen of ^{cE5}A48 and the side chains of ^{T}R73 and ^{T}Q151. From ^{cE5}P49 all the way to ^{cE5}R56, the interactions established between thrombin and the homologous ^{A}T47-R53 segment of *A. albimanus* anophelin are strictly conserved (10), in agreement with the remarkable amino acid sequence conservation observed in all members of the family (Figure 4). Therefore, ^{cE5}D53 interacts with the catalytic ^{T}H57 and ^{T}S195, while ^{cE5}R56 occupies the S1 specificity pocket of the proteinase (Figure 7D). The downstream ^{cE5}N57 interacts with the side chain of ^{T}E192, resembling the contact established by *A. albimanus* anophelin ^{A}R54 (10), but the main chain of the remaining ordered ^{cE5}P58-^{cE5}L61 tetrapeptide deviates considerably from the homologous segment of anophelin (Figure 7B). Despite that, the position of the side chain of ^{cE5}F60 closely resembles that of *A. albimanus* anophelin ^{A}L55, therefore occupying the aryl binding site of thrombin.

DISCUSSION

Overall, the cE5 central region represented by the P2 peptide appeared as the main responsible for both thrombin binding and inhibition. The fact that only the ^{cE5}E35-N63 segment of peptide P5 could be modeled in the crystal structure and no significant electron density was detectable for the serine-rich C-terminal portion reinforces this main role of the region corresponding to P2. Moreover, the C-terminal region was not constrained by crystal packing, explaining its observed disorder (Figure S5). The binding of cE5 to thrombin was to a large extent similar to that previously described for *A. albimanus* anophelin. Nonetheless, the cE5 N-terminal and C-terminal regions still appear to play some role in both binding and inhibition. Indeed, when these two disordered segments were missing we observed a decrease in binding affinity (≤1 nM for cE5 versus ≤12 nM for P2) as well as in thrombin inhibition (80% for cE5 versus 39% for P2). These more efficient binding and inhibitory properties of the full-length protein are in agreement with the flanking model, already proposed for interactions involving IDPs (28,29). Unexpectedly, ITC studies indicated a negative contribution of the C-terminal region encompassed by P3 (higher Kd and lower ΔH values for P5 than for either P2 or the entire cE5 protein). This conflicts with the conservation of the C-terminal serine-rich extension among Old World anophelines, suggesting a potential functional role for this region. However, specific experimental conditions and/or the *in vitro* nature of our study should be taken into account. The situation may be different *in vivo*, where the regions encompassed by peptides P1 and P3 may provide interaction sites for other binding partners (e.g the RGD tripeptide may be involved in binding to integrins) and/or undergo post-translational modifications (as suggested from the different mobility between recombinant and native cE5 protein in SDS-PAGE) (9). In conclusion, through structural, functional and binding studies we shed additional light on the unique mechanism of thrombin binding and inhibition by this family of salivary anticoagulants from anopheline mosquitoes. It is here anticipated that future proteomic and structural studies on salivary proteins

of hematophagous animals are expected to provide insights into novel antihemostatic molecules and novel mechanisms of disrupting blood clotting that may be of great help in the design of novel antithrombotics. This appears especially true considering (i) the fast evolutionary rate of salivary proteins (11,12), (ii) the convergent evolutionary nature of hematophagy and (iii) the fact that so far we acquired some knowledge on the salivary repertoires of just a few of the >14000 arthropod species estimated to feed on blood (30).

EXPERIMENTAL PROCEDURES

cE5 protein and peptides-Recombinant cE5 and ^{15}N-labelled cE5 proteins were expressed and purified as described (9). The ORF encoding P5 (^{cE5}D31-E82) was chemically synthesized (Life Technologies) and cloned into pETM11 (Novagen). Expression was performed as above. Protein integrity and purity was confirmed by mass spectrometry (AGILENT Q-TOF LC/MS). Peptides P1(^{cE5}A1-S30, APQYARGDVPTYDEEDFDEESLKPHSSSSS), P2 (^{cE5}D31-R62, DDGEEEFDPSLLEEHADAPTARDPGRNPEFLR) and P3 (^{cE5}N63-E82, NSNTDEQASAPAASSSESDE) were purchased from INBIOS S.r.l. (Naples, Italy). The *A. albimanus* anophelin was synthesized at the Peptide Synthesis Laboratory, Research Technologies Branch, NIAID (Rockville, MD) by Dr. Jan Lukszo.

Circular Dichroism analyses-Circular Dichroism (CD) spectra were recorded at 20°C in the far UV (190-260 nm) on a Jasco J-710 spectropolarimeter (31). Each spectrum was obtained averaging three scans, subtracting contributions from corresponding blanks and converting the signal to mean residue ellipticity in units of deg × cm^2 × dmol^{-1} × res^{-1}. Spectra were recorded using 10 μM cE5 in 10 mM phosphate buffer at pH 7.5 in the absence and in the presence of increasing concentrations (up to 60 % (v/v)) of 2,2,2-trifluoroethanol (TFE, 99.5% isotopic purity, Sigma-Aldrich). Prediction analysis of secondary structure content was performed with CDPro (31,32).

Isothermal titration calorimetry-Isothermal titration calorimetry (ITC) studies were performed at 22°C with an ITC200 calorimeter (MicroCal/GE Healthcare, Milan, Italy). Anophelin and cE5 (100 μM) or the peptides P1, P2, P3 and P5 (150 μM) were titrated into a solution of human α-thrombin (20 μM; Haematologic Technologies Inc., HCT-0020). Prior to all titration experiments, proteins and peptides were dialyzed against 20 mM HEPES pH 7.5, 50 mM NaCl. All data were analyzed and fitted using the Microcal Origin version 7.0 software package. Binding enthalpy, dissociation constants and stoichiometry were determined by fitting the data using a one-set-of-site binding model. ITC runs were repeated twice to evaluate the reproducibility of the results.

Nuclear magnetic resonance spectroscopy -Nuclear magnetic resonance (NMR) spectra were recorded at 298 K on a Varian Unity Inova 600 MHz spectrometer provided with a cold probe. 2D [^{1}H, ^{15}N] HSQC spectra of ^{15}N-labelled cE5 were acquired by implementing the following set of experimental conditions: 50 mM sodium phosphate buffer pH 7.8; 550 μl sample total volume; H_2O/D_2O (98% D, Armar Scientific, Switzerland) 90/10, 0.1%(w/v) NaN_3; 5-40 μM protein concentration range. Similar experiments were conducted in phosphate buffer containing increasing amounts (10%, 30% and 60% (v/v)) of TFE. Water suppression was achieved by Excitation Sculpting (33). NMR spectra were processed with Varian software (Vnmrj version 1.1D)(34) and analyzed with the software NEASY (http://www.nmr.ch/) (35). To monitor binding to thrombin, 2D [^{1}H, ^{15}N] HSQC spectra of ^{15}N-labelled cE5 (10 μM) were recorded for the protein in the free form and after addition of increasing amounts of unlabeled thrombin (cE5/thrombin molar ratios: 1/1.4, 1/2.7, 1/4.2, 1/7). Analysis of titration experiments and overlays of 2D spectra were performed with Sparky (36). In addition D [^{1}H, ^{1}H] NOESY 300 (Nuclear Overhauser Enhancement Spectroscopy) (37) and TOCSY 70 (Total Correlation Spectroscopy) (38) experiments were recorded for the P2 peptide (470 μM in 50 mM sodium phosphate pH 7.8) in its unbound form and complexed to thrombin (25 μM).

Thrombin inhibition assays

Inhibition of human α-thrombin by cE5 and peptides P1, P2, P3 or P5 was evaluated following

the hydrolysis of the fluorogenic substrate Boc-Val-Pro-Arg-7-amido-4-methylcoumarin hydrochloride (Sigma-Aldrich, B9385). Reactions were carried out in 100 µl total volume in Reaction Buffer (50 mM Tris-HCl pH 8.0, 50 mM $NaCl_2$, 20 mM $CaCl_2$, 0.01% (v/v) Triton X-100). Human α-thrombin (0.5 nM; Sigma-Aldrich, T7009) was pre-incubated for 10 min at 37°C with 100 nM recombinant cE5 or cE5-derived peptides. Reactions were initiated by addition of substrate (250 µM) and hydrolysis followed at 37°C for 60 min at 460 nm (excitation = 360 nm) on a Synergy HT microplate reader (BioTek). Data represent three independent experiments, each one in triplicate. Graphics and statistical analysis were performed using GraphPad Prism 6.0 (GraphPad Software, Inc. La Jolla, CA).

Thrombin amidolytic activity assay-The amidolytic activity of human α-thrombin or γ-thrombin (Haematologic Technologies) was followed using Tos-Gly-Pro-Arg-p-nitroanilide (Chromozym TH; Roche) as chromogenic substrate. Inhibition assays were performed in 50 mM Tris pH 8.0, 50 mM NaCl, 1 mg/ml bovine serum albumin with 0.2 nM human α-thrombin or γ-thrombin, 100 µM substrate and varying concentrations of inhibitor (0 to 80 nM for α-thrombin or 0 to 200 nM for γ-thrombin). Inhibition constants (Ki) were determined according to a tight-binding inhibitor model, using the Morrison equation (39) with GraphPad Prism 6.0. All reactions were initiated by addition of thrombin and were carried out at least in duplicate at 37°C in 96-well microtiter plates. Reaction progress was monitored at 405 nm for 30 minutes on a Synergy2 multi-mode microplate reader (BioTek).

Platelet assays-Inhibition of collagen-mediated platelet aggregation was evaluated using human platelet-rich plasma obtained from donors at the NIH blood bank and diluted to a density of 2×10^5 platelets /µL in Tyrode-BSA (40) (final volume 300 µL). cE5, P1 or an equal volume of Tyrode buffer was added to each sample which was then placed in an aggregometer (40) and stirred at 1200 rpm at 37 °C for 1 min prior to the addition of collagen (type-1 fibrils, Chrono-log) to a concentration of 1.6 µg/mL.

Crystallization of A. gambiae cE5 in complex with human α-thrombin-Human α-thrombin (Haematologic Technologies) was mixed in 20 mM HEPES pH 7.5, 125 mM NaCl with a four-fold molar excess of P5 peptide (^{cE5}D31-E82) and incubated on ice for one hour. The resulting complex was concentrated by ultrafiltration using a 2 kDa cutoff centrifugal device (Sartorius). An initial crystallization screen at 20 °C on sitting drop geometry was performed at the High Throughput Crystallization Laboratory of the European Molecular Biology Laboratory (Grenoble, France). Preliminary crystallization conditions were systematically optimized in-house until single monoclinic crystals belonging to space group C2 were obtained after 4-6 days at 20°C using the vapor diffusion sitting-drop method from drops consisting of equal volumes (1 µL) of protein complex (at 6.4 mg/mL) and precipitant solution (0.1 M PCTP pH 5.0, 25% (w/v) PEG 1500) equilibrated against a 300 µL reservoir.

Data collection and processing-Crystals were cryoprotected by brief immersion in 0.1 M PCTP pH 5.0, 20% (w/v) PEG1500, 20% (v/v) ethylene glycol or 0.1 M PCTP pH 5.0, 35% (w/v) PEG1500 and flash-cooled in liquid nitrogen. Diffraction data were collected from two isomorphous crystals (2100 and 2400 images in 0.05° oscillation steps and 0.037 s exposure) on a Pilatus 6M-F detector (Dectris) using a wavelength of 0.973 Å at beam line ID30B of the European Synchrotron Radiation Facility (Grenoble, France). Data were integrated with XDS (41), scaled with XSCALE (41) and reduced with utilities from the CCP4 program suite (42). Data collection statistics are summarized in Supplementary Table 1. The diffraction data (43,44) were deposited with the Structural Biology Data Grid (45).

Structure determination and refinement-The structure of the complex was solved by molecular replacement with PHASER (46) using the coordinates of free wild-type human α-thrombin (PDB entry 3U69 (22)) as search model. Alternating cycles of model building with COOT (47) and refinement with PHENIX (48) were performed until model completion (refinement statistics are summarized in Supplementary Table 1). All crystallographic software was supported by SBGrid (49). Refined coordinates and structure factors were deposited at the PDB with accession number 5NHU.

Acknowledgments

We wish to acknowledge JMC Ribeiro for kindly providing the anophelin peptide, the ESRF for provision of synchrotron radiation facilities and the ESRF staff for help with data collection. This work was funded by the EU FP7 Infravec project (228421) and Finanziamenti di Ateneo per la Ricerca Scientifica (C26A12SLP3) to B.A. and in part by Portuguese funds through FCT - Fundação para a Ciência e a Tecnologia in the form of postdoctoral fellowship SFRH/BPD/108004/2015 (to J.R.-R.) and of project "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274), co-funded by the European Regional Development Fund (FEDER) through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalization (POCI), PORTUGAL 2020. Support by the "Structured program on bioengineered therapies for infectious diseases and tissue regeneration" (Norte-01-0145-FEDER-000012), funded by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through FEDER is also acknowledged. Transnational Access to the High Throughput Crystallization Laboratory of the European Molecular Biology Laboratory Grenoble Outstation was supported by the European Community-Seventh Framework Program (FP7/2007-2013) Grant Protein Production Platform (PCUBE Agreement No. 227764).

Conflict of interest: The authors declare no conflict of interest

Author contributions: L.P., J.R.R., M.L., R.R, F.L. and J.A. performed research; L.P., M. L., G.F., P.J.B.P., B.A and E.P. designed research; L.P., J.R.R., M.L., J.A., P.J.B.P., B.A and E.P analyzed data; M.L., P.J.B.P., B.A and E.P. wrote the paper.

REFERENCES

1. Ribeiro, J. M. C., and Arcà, B. (2009) From Sialomes to the Sialoverse: An Insight into Salivary Potion of Blood-Feeding Insects. *Adv Insect Physiol* **37**, 59-118
2. Mans, B. J. (2011) Evolution of vertebrate hemostatic and inflammatory control mechanisms in blood-feeding arthropods. *J Innate Immun* **3**, 41-51
3. Ribeiro, J. M., Mans, B. J., and Arcà, B. (2010) An insight into the sialome of blood-feeding Nematocera. *Insect Biochem Mol Biol* **40**, 767-784
4. Stark, K. R., and James, A. A. (1996) Anticoagulants in vector arthropods. *Parasitol. Today* **12**, 430-437
5. Valenzuela, J. G., Francischetti, I. M., and Ribeiro, J. M. C. (1999) Purification, cloning, and synthesis of a novel salivary anti-thrombin from the mosquito Anopheles albimanus. *Biochemistry* **38**, 11209-11215.
6. Francischetti, I. M., Valenzuela, J. G., and Ribeiro, J. M. C. (1999) Anophelin: kinetics and mechanism of thrombin inhibition. *Biochemistry* **38**, 16678-16685.
7. De Cristofaro, R., and De Candia, E. (2003) Thrombin domains: structure, function and interaction with platelet receptors. *J Thromb Thrombolysis* **15**, 151-163
8. Arcà, B., Lombardo, F., de Lara Capurro, M., della Torre, A., Dimopoulos, G., James, A. A., and Coluzzi, M. (1999) Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci U S A* **96**, 1516-1521
9. Ronca, R., Kotsyfakis, M., Lombardo, F., Rizzo, C., Curra, C., Ponzi, M., Fiorentino, G., Ribeiro, J. M., and Arca, B. (2012) The Anopheles gambiae cE5, a tight- and fast-binding thrombin inhibitor with post-transcriptionally regulated salivary-restricted expression. *Insect Biochem Mol Biol* **42**, 610-620
10. Figueiredo, A. C., de Sanctis, D., Gutierrez-Gallego, R., Cereija, T. B., Macedo-Ribeiro, S., Fuentes-Prior, P., and Pereira, P. J. (2012) Unique thrombin inhibition mechanism by anophelin, an anticoagulant from the malaria vector. *Proc Natl Acad Sci U S A* **109**, E3649-3658
11. Neafsey, D. E., Waterhouse, R. M., Abai, M. R., Aganezov, S. S., Alekseyev, M. A., Allen, J. E., Amon, J., Arca, B., Arensburger, P., Artemov, G., Assour, L. A., Basseri, H., Berlin, A., Birren, B. W., Blandin, S. A., Brockman, A. I., Burkot, T. R., Burt, A., Chan, C. S., Chauve, C., Chiu, J. C., Christensen, M., Costantini, C., Davidson, V. L., Deligianni, E., Dottorini, T., Dritsou, V., Gabriel, S. B., Guelbeogo, W. M., Hall, A. B., Han, M. V., Hlaing, T., Hughes, D. S., Jenkins, A. M., Jiang, X., Jungreis, I., Kakani, E. G., Kamali, M., Kemppainen, P., Kennedy, R. C., Kirmitzoglou, I. K., Koekemoer, L. L., Laban, N., Langridge, N., Lawniczak, M. K., Lirakis, M., Lobo, N. F., Lowy, E., MacCallum, R. M., Mao, C., Maslen, G., Mbogo, C., McCarthy, J., Michel, K., Mitchell, S. N., Moore, W., Murphy, K. A., Naumenko, A. N., Nolan, T., Novoa, E. M., O'Loughlin, S., Oringanje, C., Oshaghi, M. A., Pakpour, N., Papathanos, P. A., Peery, A. N., Povelones, M., Prakash, A., Price, D. P., Rajaraman, A., Reimer, L. J., Rinker, D. C., Rokas, A., Russell, T. L., Sagnon, N., Sharakhova, M. V., Shea, T., Simao, F. A., Simard, F., Slotman, M. A., Somboon, P., Stegniy, V., Struchiner, C. J., Thomas, G. W., Tojo, M., Topalis, P., Tubio, J. M., Unger, M. F., Vontas, J., Walton, C., Wilding, C. S., Willis, J. H., Wu, Y. C., Yan, G., Zdobnov, E. M., Zhou, X., Catteruccia, F., Christophides, G. K., Collins, F. H., Cornman, R. S., Crisanti, A., Donnelly, M. J., Emrich, S. J., Fontaine, M. C., Gelbart, W., Hahn, M. W., Hansen, I. A., Howell, P. I., Kafatos, F. C., Kellis, M., Lawson, D., Louis, C., Luckhart, S., Muskavitch, M. A., Ribeiro, J. M., Riehle, M. A., Sharakhov, I. V., Tu, Z., Zwiebel, L. J., and Besansky, N. J. (2015) Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science* **347**, 1258522
12. Arca, B., Lombardo, F., Struchiner, C. J., and Ribeiro, J. M. (2017) Anopheline salivary protein genes and gene families: an evolutionary overview after the whole genome sequence of sixteen Anopheles species. *BMC Genomics* **18**, 153
13. Potenza, E., Di Domenico, T., Walsh, I., and Tosatto, S. C. (2015) MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res* **43**, D315-320

14. Anderson, V. L., Ramlall, T. F., Rospigliosi, C. C., Webb, W. W., and Eliezer, D. (2010) Identification of a helical intermediate in trifluoroethanol-induced alpha-synuclein aggregation. *P Natl Acad Sci USA* **107**, 18850-18855
15. Buck, M. (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* **31**, 297-355
16. Maestro, B., Galan, B., Alfonso, C., Rivas, G., Prieto, M. A., and Sanz, J. M. (2013) A New Family of Intrinsically Disordered Proteins: Structural Characterization of the Major Phasin PhaF from Pseudomonas putida KT2440. *Plos One* **8**
17. Zaccaro, L., del Gatto, A., Pedone, C., and Saviano, M. (2009) Peptides for Tumour Therapy and Diagnosis: Current Status and Future Directions. *Curr Med Chem* **16**, 780-795
18. Francischetti, I. M. (2010) Platelet aggregation inhibitors from hematophagous animals. *Toxicon* **56**, 1130-1144
19. Assumpcao, T. C., Ribeiro, J. M., and Francischetti, I. M. (2012) Disintegrins from hematophagous sources. *Toxins (Basel)* **4**, 296-322
20. Raffler, N. A., Schneider-Mergener, J., and Famulok, M. (2003) A novel class of small functional peptides that bind and inhibit human alpha-thrombin isolated by mRNA display. *Chem Biol* **10**, 69-79
21. Velazquez-Campoy, A., and Freire, E. (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat Protoc* **1**, 186-191
22. Figueiredo, A. C., Clement, C. C., Zakia, S., Gingold, J., Philipp, M., and Pereira, P. J. B. (2012) Rational design and characterization of D-Phe-Pro-D-Arg-derived direct thrombin inhibitors. *PLoS ONE* **7**, e34354
23. Pellecchia, M. (2005) Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. *Chem Biol* **12**, 961-971
24. Rawlings, N. D., Barrett, A. J., and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **44**, D343-350
25. Gandhi, P. S., Chen, Z., and Di Cera, E. (2010) Crystal structure of thrombin bound to the uncleaved extracellular fragment of PAR1. *J Biol Chem* **285**, 15393-15398
26. Liu, C. C., Brustad, E., Liu, W., and Schultz, P. G. (2007) Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. *J Am Chem Soc* **129**, 10648-10649
27. Macedo-Ribeiro, S., Almeida, C., Calisto, B. M., Friedrich, T., Mentele, R., Stürzebecher, J., Fuentes-Prior, P., and Pereira, P. J. B. (2008) Isolation, cloning and structural characterisation of boophilin, a multifunctional Kunitz-type proteinase inhibitor from the cattle tick. *PLoS ONE* **3**, e1624
28. Sharma, R., Raduly, Z., Miskei, M., and Fuxreiter, M. (2015) Fuzzy complexes: Specific binding without complete folding. *FEBS Lett* **589**, 2533-2542
29. Gruet, A., Dosnon, M., Blocquel, D., Brunel, J., Gerlier, D., Das, R. K., Bonetti, D., Gianni, S., Fuxreiter, M., Longhi, S., and Bignon, C. (2016) Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *Febs Journal* **283**, 576-594
30. Ribeiro, J. M. (1995) Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect Agents Dis* **4**, 143-152
31. de Paola, I., Pirone, L., Palmieri, M., Balasco, N., Esposito, L., Russo, L., Mazza, D., Di Marcotullio, L., Di Gaetano, S., Malgieri, G., Vitagliano, L., Pedone, E., and Zaccaro, L. (2015) Cullin3-BTB Interface: A Novel Target for Stapled Peptides. *PLoS One* **10**
32. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* **209**, 32-44
33. Hwang, T. L., and Shaka, A. J. (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J Magn Reson Ser A* **112**, 275-279
34. Contursi, P., Farina, B., Pirone, L., Fusco, S., Russo, L., Bartolucci, S., Fattorusso, R., and Pedone, E. (2014) Structural and functional studies of Stf76 from the Sulfolobus islandicus plasmid-virus pSSVx: a novel peculiar member of the winged helix-turn-helix transcription factor family. *Nucleic Acids Res* **42**, 5993-6011

35. Bartels, C., Xia, T. H., Billeter, M., Guntert, P., and Wuthrich, K. (1995) The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. *J Biomol Nmr* **6**, 1-10
36. Goddard, T. D., and Kneller, D. G. SPARKY 3, University of California, San Francisco.
37. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* **95**, 1-6
38. Griesinger, C., Otting, G., Wuethrich, K., and Ernst, R. R. (1988) Clean TOCSY for proton spin system identification in macromolecules. *J Am Chem Soc* **110**, 7870-7872
39. Williams, J. W., and Morrison, J. F. (1979) The kinetics of reversible tight-binding inhibition. *Methods Enzymol* **63**, 437-467
40. Lagrue, A. H., Francischetti, I. M. B., Guimaraes, J. A., and Jandrot-Perrus, M. (1999) Phosphatidylinositol 3 '-kinase and tyrosine-phosphatase activation positively modulate Convulxin-induced platelet activation. Comparison with collagen. *FEBS Lett* **448**, 95-100
41. Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
42. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
43. Ripoll-Rozada, J. P., P. J. B. . (2017) X-Ray Diffraction data for: Human thrombin-A. gambiae cE5 complex - dataset 1. *PDB Code 5NHU. Structural Biology Data Grid, V1*
44. Ripoll-Rozada, J. P., P. J. B. (2017) X-Ray Diffraction data for: Human thrombin-A. gambiae cE5 complex – dataset 2. *PDB Code 5NHU. Structural Biology Data Grid, V1*
45. Meyer, P. A., Socias, S., Key, J., Ransey, E., Tjon, E. C., Buschiazzo, A., Lei, M., Botka, C., Withrow, J., Neau, D., Rajashankar, K., Anderson, K. S., Baxter, R. H., Blacklow, S. C., Boggon, T. J., Bonvin, A. M., Borek, D., Brett, T. J., Caflisch, A., Chang, C. I., Chazin, W. J., Corbett, K. D., Cosgrove, M. S., Crosson, S., Dhe-Paganon, S., Di Cera, E., Drennan, C. L., Eck, M. J., Eichman, B. F., Fan, Q. R., Ferre-D'Amare, A. R., Christopher Fromme, J., Garcia, K. C., Gaudet, R., Gong, P., Harrison, S. C., Heldwein, E. E., Jia, Z., Keenan, R. J., Kruse, A. C., Kvansakul, M., McLellan, J. S., Modis, Y., Nam, Y., Otwinowski, Z., Pai, E. F., Pereira, P. J., Petosa, C., Raman, C. S., Rapoport, T. A., Roll-Mecak, A., Rosen, M. K., Rudenko, G., Schlessinger, J., Schwartz, T. U., Shamoo, Y., Sondermann, H., Tao, Y. J., Tolia, N. H., Tsodikov, O. V., Westover, K. D., Wu, H., Foster, I., Fraser, J. S., Maia, F. R., Gonen, T., Kirchhausen, T., Diederichs, K., Crosas, M., and Sliz, P. (2016) Data publication with the structural biology data grid supports live analysis. *Nature communications* **7**, 10882
46. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
47. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
48. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
49. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., and Sliz, P. (2013) Collaboration gets the most out of software. *eLife* **2**, e01456
50. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins* **42**, 38-48
51. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539

TABLE 1. Binding properties of cE5, anophelin and cE5-derived peptides to thrombin as determined by ITC studies.

Ligand	**Kd**	**ΔH (cal/mol)**	**n**
cE5	≤ 1 ± 0.2 nM	-9848 ± 60	1
cE5*	0.3 nM	-8272	1
P1 ($cE5^{1-30}$)	*nd*	*nd*	*nd*
P2 ($cE5^{31-62}$)	≤ 12 ± 1 nM	-9663 ± 34	1
P3 ($cE5^{63-82}$)	*nd*	*nd*	*nd*
P5 ($cE5^{31-82}$)	≤ 25 ± 1nM	-4887 ± 75	1
anophelin	≤ 2 ± 0.2 nM	-14120 ± 90	1

nd, not detected.

* From displacement titration experiments.

TABLE 2. Inhibition of human α- and γ-thrombin by *A. gambiae* cE5. The affinity for thrombin of the P5 fragment is comparable to that of full-length cE5. Disruption of thrombin's exosite I decreases the inhibition potency of cE5 (and P5) by three orders of magnitude. Ki values ± SEM given are representative of two independent experiments.

	α-thrombin	**γ-thrombin**
Inhibitor	K_i (pM)	K_i (nM)
cE5	5.50 ± 1.26	3.79 ± 0.16
P5 fragment	20.80 ± 0.99	10.99 ± 0.69

Legends to Figures

Figure 1. Disorder prediction of the *A. gambiae* cE5 protein. Prediction of naturally disordered regions in the primary sequence of cE5 (VL-XT predictor) (50). Residues exceeding a threshold value of 0.5 are considered disordered.

Figure 2. Far UV Circular Dichroism (CD) spectra of the cE5 protein. CD spectra of cE5 (10 µM in 10 mM phosphate buffer pH 7.5) alone (solid line) and after addition of 60% (v/v) TFE (dotted line). The insert shows the molar ellipticity at 220 nm in the presence of increasing concentrations of TFE.

Figure 3. NMR spectroscopy of the *A. gambiae* cE5 protein. A) [^{1}H-^{15}N] HSQC spectrum of uniformly ^{15}N-labelled cE5 (20 µM in 50 mM sodium phosphate pH 7.8; sample total volume: 550 µL with 10% D_2O). B-D) Monitoring the effect of TFE on the IDP cE5. Comparison of HSQC spectra of ^{15}N-labelled cE5 in the absence (red) and in the presence of TFE at increasing concentrations: B) 10% (green); C) 30% (cyan); D) 60% (blue). NaP stands for sodium phosphate buffer.

Figure 4. Multiple alignment of cE5/anophelin family members from 18 *Anopheles* species. The conserved N-terminal region and the DPGR tetrapeptide (highlighted in red) are boxed. Serine residues in the C-terminal region are shown in white lettering on a black background. Residues conserved in all or at least in 75% of the aligned sequences are highlighted in yellow and green, respectively. The peptides used in this study are represented above the aligned proteins. Regions of the cE5 protein shown by crystallographic analysis to interact with the exosite I and the catalytic site of thrombin are indicated below the alignment, as are their flanking residues. VectorBase accession numbers: *A. gambiae*, AGAP008004; *A. coluzzii*, ACOM030673; *A. arabiensis*, AARA005120; *A. quadriannulatus*, AQUA000660; *A. merus*, AMEM012461; *A. funestus*, AFUN004964; *A. minimus*, AMIN006263; *A. culicifacies*, ACUA026308; *A. stephensi*, ASTE006313; *A. dirus*, ADIR011024; *A. sinensis*, ASIS010361; *A. albimanus*, AALB014019. GenBank accession number: *A. darlingi*, GI:208657572. Sequences of the *A. melas*, *A. christyi*, *A. epiroticus*, *A. farauti* and *A. atroparvus* orthologues can be found in (12). The sequences were aligned with Clustal Omega (51).

Figure 5. ITC analysis of thrombin binding to *A. gambiae* cE5, *A. albimanus* anophelin, and cE5-derived peptides. Isothermal titration calorimetry of thrombin interaction with *A. gambiae* cE5 (A), *A. albimanus* anophelin (B), P1 (C), P2 (D), P3 (E), and P5 (F) is shown. Top and bottom panels report raw and integrated data, respectively.

Figure 6. Thrombin inhibition assays. Residual thrombin activity in the absence and the presence of *A. gambiae* cE5, or of the different peptides as indicated, was measured following the hydrolysis of the fluorogenic substrate Boc-Val-Pro-Arg-7-amido-4-methylcoumarin hydrochloride. Dots represent data from three independent experiments, each in triplicate. Thick lines show the mean values and bars indicate standard deviations.

Figure 7. The *A. gambiae* cE5 anticoagulant blocks both the active site and the exosite I regions of human *α*-thrombin. (A) The acidic ^{cE5}E35-D47 segment of cE5 (stick model with nitrogen atoms in blue, oxygen in red, and carbon in green) binds to the exosite I of thrombin (solid surface representation with positive surface electrostatic potential in blue and negative surface electrostatic potential in red), whereas the downstream segment, up to ^{cE5}R62, threads along the active site cleft of the proteinase, blocking the active site and the non-primed specificity subsites. The 2Fo-Fc electron density map (1.0-σ contour) for cE5 is represented as a black mesh. (B) Comparison between the *A. gambiae* cE5 P5 fragment (green) and *A. albimanus* anophelin (PDB entry 4E05 (10); blue) shown in cartoon representation. Thrombin, shown as a solid white surface, is in the same orientation as in panel A. (C) Close-up view of the interactions established between the N-terminal segment ^{cE5}E35-E43 of cE5 (colored as in panel A) and the exosite I of thrombin (gray cartoon with selected residues as sticks color-coded as cE5, except for carbon atoms, colored tan). Water molecules and hydrogen bonds are represented as red spheres and dotted black lines, respectively. (D) Close-up view of the interaction between the

strictly conserved ^{cE5}D53-R56 tetrapeptide and the active site region of thrombin (colors as in panel C). Figure prepared with PyMOL (http://www.pymol.org).

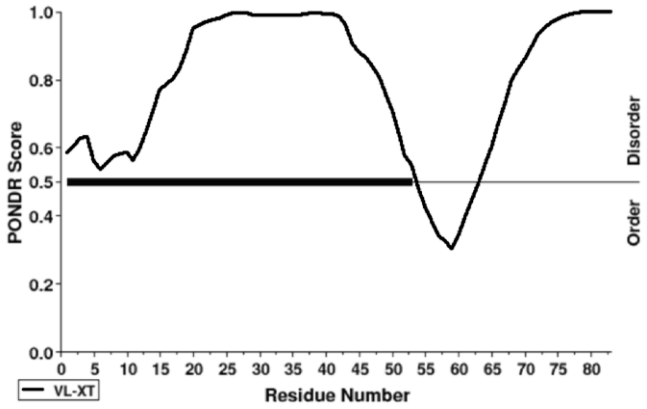

Figure 1

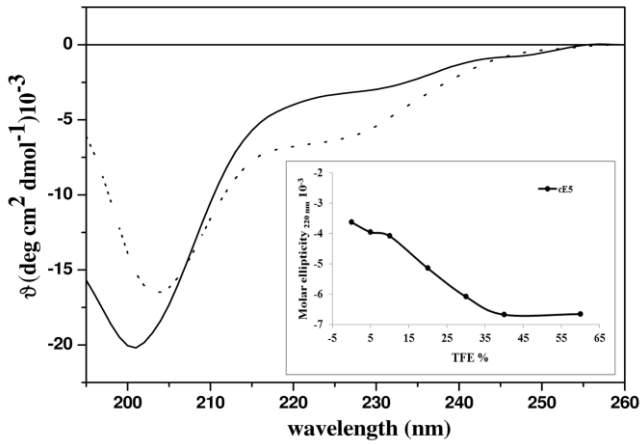

Figure 2

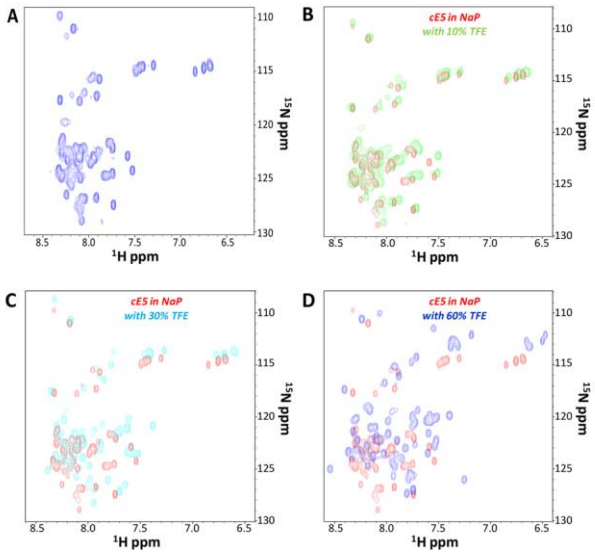

Figure 3

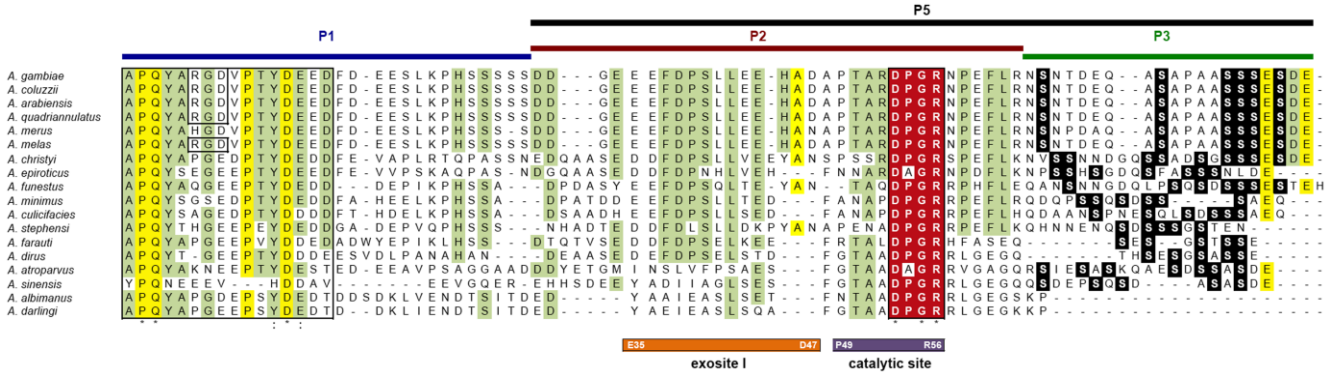

Figure 4

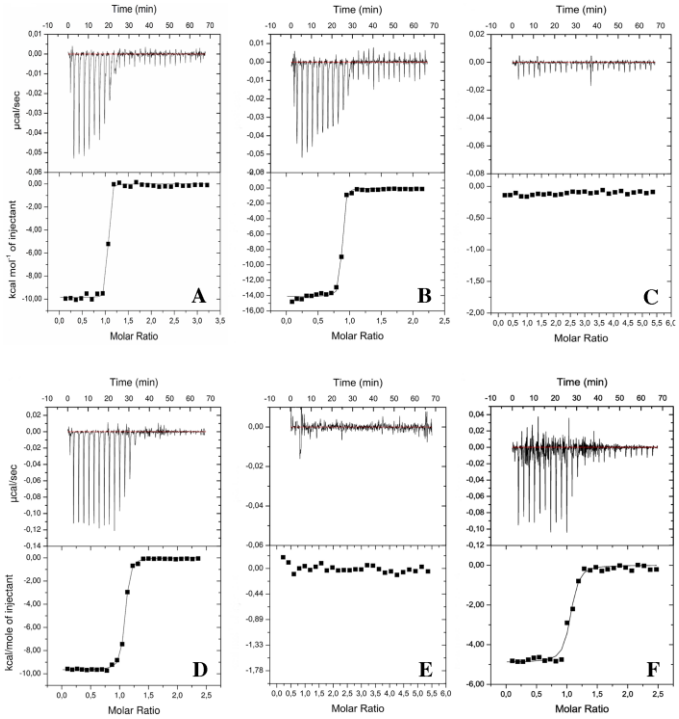

Figure 5

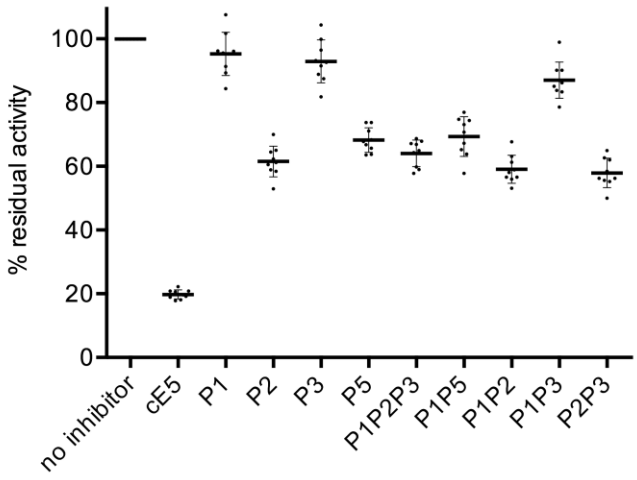

Figure 6

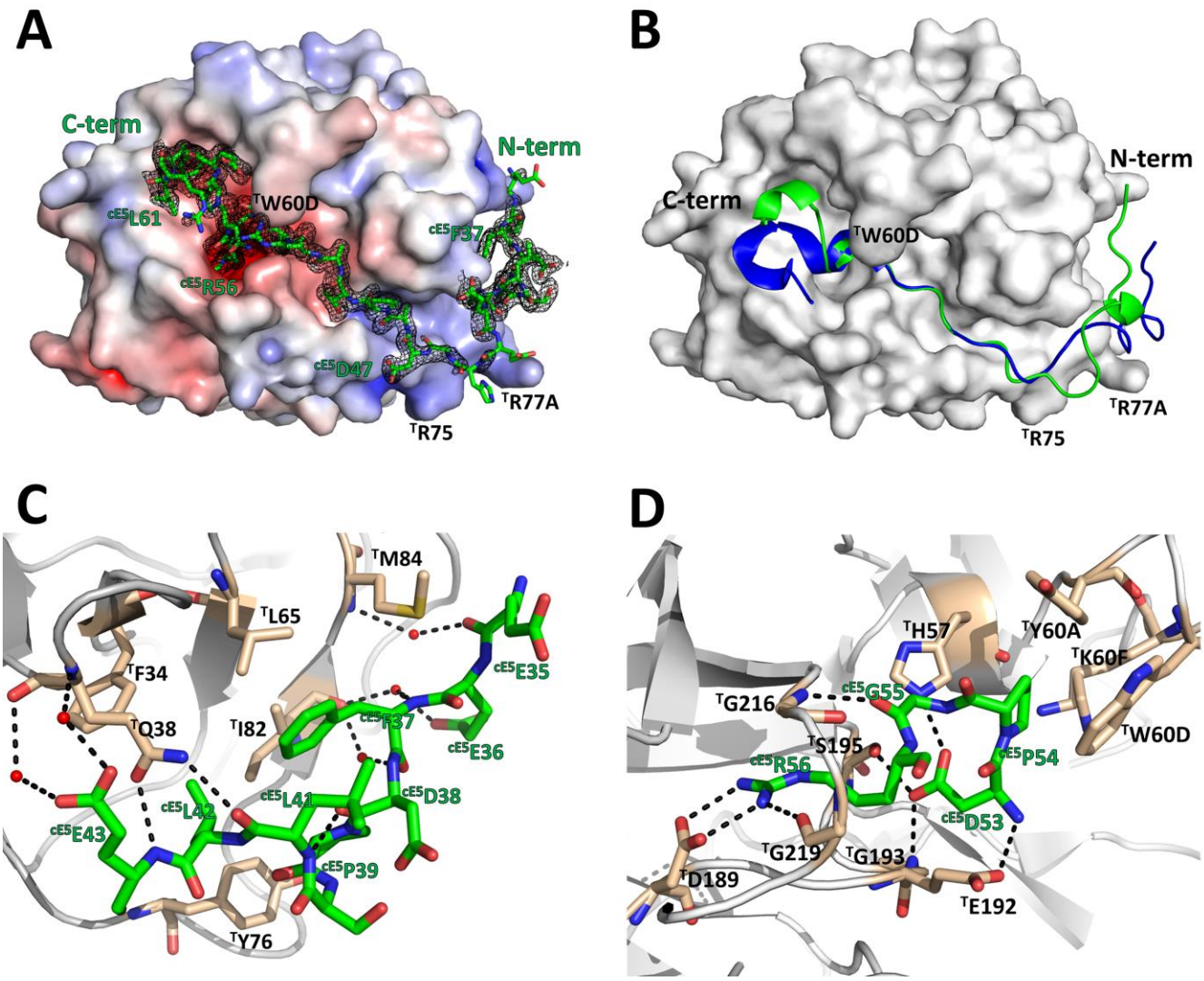

Figure 7

Functional analyses yield detailed insight into the mechanism of thrombin inhibition by the antihemostatic salivary protein cE5 from Anopheles gambiae

Luciano Pirone, Jorge Ripoll-Rozada, Marilisa Leone, Raffaele Ronca, Fabrizio Lombardo, Gabriella Fiorentino, John F. Andersen, Pedro José Barbosa Pereira, Bruno Arcà and Emilia Pedone

J. Biol. Chem. published online June 7, 2017
