The Kazal-type inhibitors infestins 1 and 4 differ in specificity but are similar in three-dimensional structure

I. T. N. Campos,[a,b]*‡ T. A. C. B. Souza,[c]‡ R. J. S. Torquato,[a] R. De Marco,[a] A. M. Tanaka-Azevedo,[d] A. S. Tanaka[a]* and J. A. R. G. Barbosa[c]§

[a]Departamento de Bioquímica, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Três de Maio 100, 04044-020 São Paulo-SP, Brazil, [b]Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Goiás, ICB, Campus Samambaia, CP 131, 74001-970 Goiânia-GO, Brazil, [c]Laboratório Nacional de Biociências, Centro Nacional de Pesquisas em Energia e Materiais, Campinas, Brazil, and [d]Instituto Butantan, Avenida Vital Brasil 1500, 05503-900 São Paulo-SP, Brazil

‡ These authors contributed equally to this work.
§ Present address: Pós-Graduação em Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, SGAN 916, Módulo C, Avenida W5 Norte, Asa Norte, 70790-160 Brasília-DF, Brazil.

Correspondence e-mail: itncampos@gmail.com, tanaka.bioq@epm.br

Blood coagulation is an important process in haemostasis, and disorders of blood coagulation can lead to an increased risk of haemorrhage and thrombosis. Coagulation is highly conserved in mammals and has been comprehensively studied in humans in the investigation of bleeding or thrombotic diseases. Some substances can act as inhibitors of blood coagulation and may affect one or multiple enzymes throughout the process. A specific thrombin inhibitor called infestin has been isolated from the midgut of the haematophagous insect *Triatoma infestans*. Infestin is a member of the nonclassical Kazal-type serine protease inhibitors and is composed of four domains, all of which have a short central α-helix and a small antiparallel β-sheet. Domains 1 and 4 of infestin (infestins 1 and 4) possess specific inhibitory activities. Infestin 1 inhibits thrombin, while infestin 4 is an inhibitor of factor XIIa, plasmin and factor Xa. Here, the structure determination and structural analysis of infestin 1 complexed with trypsin and of infestin 4 alone are reported. Through molecular modelling and docking, it is suggested that the protein–protein binding site is conserved in the infestin 1–thrombin complex compared with other Kazal-type inhibitors. Infestin 4 is able to bind factor XIIa, and the F9N and N11R mutants selected by phage display were shown to be more selective for factor XIIa in comparison to the wild type.

1. Introduction

Two main mechanisms for the control of blood loss form part of the haemostatic system of vertebrates. Initially, activated platelets adhere and aggregate at the injury site, forming a platelet plug that reduces or blocks blood loss. Concomitantly, the blood-coagulation cascade is a process that is activated by two distinct pathways, the intrinsic and the extrinsic pathways, which are both initiated by limited proteolysis of several inactive serine proteases circulating in the blood. Factor XII and plasma kallikrein are enzymes of the intrinsic pathway, while activated factor VII and tissue factor are associated with the extrinsic pathway. Both pathways converge to a common pathway resulting in the generation of factor Xa, which in turn activates prothrombin to thrombin. Finally, thrombin converts fibrinogen to fibrin, which then polymerizes, with subsequent stabilization of the fibrin polymer factor XIIIa (Davie *et al.*, 1991).

Haematophagous animals have an essential need to overcome the haemostatic barrier of their host and maintain blood fluidity. To meet this requirement, they have developed efficient mechanisms of control. The first report of the purification of substances from haematophagous animals that were able to block or delay vertebrate blood coagulation dates from the 19th century (Moser *et al.*, 1998), since when

doi:10.1107/S0907444912009067 **695**

numerous substances from haematophagous animals with anti-clotting, anti-platelet-aggregation, vasodilatory and fibrinolytic activities have been described (Arocha-Piñango *et al.*, 1999; Crab *et al.*, 2002; Iwanaga *et al.*, 2003; Francischetti *et al.*, 2002; Lee & Vlasuk, 2003). Among these substances are blood-coagulation inhibitors, which are mainly serine protease inhibitors that are selective for factor Xa and thrombin.

Three potent thrombin inhibitors of the Kazal-type family have previously been described in kissing bugs (Triatominae; Friedrich *et al.*, 1993; Mende *et al.*, 1999; Campos *et al.*, 2002). These protease inhibitors are composed of two Kazal domains in tandem, but their genes contain up to eight Kazal domains which are transcribed by a unique mRNA molecule (Araujo *et al.*, 2007). The inhibitors contain one or two Kazal domains which are released by an unknown post-translational modification (Lovato *et al.*, 2006) and their inhibitory specificity depends on the reactive-site amino-acid sequence.

With very few exceptions, Kazal-type family members have several common structural features: a canonical binding loop (Bode & Huber, 1992), a conserved cysteine distribution pattern, a typical VCG*X*D sequence (Schlott et al., 2002), two to three helices connected by turns and a short three-stranded antiparallel sheet (Mühlhahn *et al.*, 1994).

The Kazal-type inhibitors are mainly grouped into classical and nonclassical inhibitors (Hemmi *et al.*, 2005). Both classical and nonclassical Kazal-type inhibitors are composed of one or more domains linked in tandem by loops. The nonclassical Kazal-type inhibitors (Fink *et al.*, 1986) include Kazal-type inhibitors from invertebrates, such as leech-derived tryptase inhibitor (LDTI), rhodniin, bdellin B-3, *Aedes aegypti* trypsin inhibitor (AaTI) and infestin (Fink *et al.*, 1986; Friedrich *et al.*, 1993; Sommerhoff *et al.*, 1994; Watanabe *et al.*, 2010; Campos *et al.*, 2002). The nonclassical Kazal inhibitors can be divided into groups I and II. The characteristic feature of the group I inhibitors is that the disulfide bond between the first and fifth cysteine residues is shifted towards the C-terminus when compared with the corresponding residues of a classical Kazal-type inhibitor such as turkey ovomucoid third domain (OMTKY3). Inhibitors belonging to group II have a cysteine-stabilized α-helical (CSH) motif in their sequences and there is only a single-amino-acid spacer between the first and second cysteine, whereas the classical Kazal-type inhibitors have a seven- to eight-amino-acid spacer between the first and second cysteines (Hemmi *et al.*, 2005; Shenoy *et al.*, 2011).

In previous studies (Campos *et al.*, 2002; Campos, Tanaka-Azevedo *et al.*, 2004; Lovato *et al.*, 2006, 2011), we have described the identification, expression and characterization of different Kazal domains, named infestins, that are present in a unique transcript composed of seven Kazal domains expressed in *Triatoma infestans* midgut. The first domain, INF1R, inhibits subtilisin and neutrophil elastase, while the second and third domains as well the fourth and fifth domains code for two thrombin inhibitors (infestins 1 and 2) with two Kazal domains each. The last two domains, infestins 3 and 4, code for a plasmin, factor Xa and factor XIIa inhibitor which is also composed of two domains. Analyzing the last domain, infestin 4 (Campos, Tanaka-Azevedo *et al.*, 2004), we observed that this Kazal domain was responsible for the inhibition of plasmin, factor Xa and factor XIIa. Although several Kazal-type serine protease inhibitors have been described and their tertiary structures have been solved, only infestin 4 presents a high inhibitory activity for factor XIIa ($K_i = 100$ pM).

Factor XII (FXII) is one of the enzymes involved at the beginning of the intrinsic pathway of the blood-coagulation cascade. Initially, it was considered that FXII was not relevant to blood clotting, since fibrin formation is mainly mediated by the extrinsic pathway. It was thus proposed that hereditary deficiency of FXII is not associated with bleeding (Rosen *et al.*, 1997; Bugge *et al.*, 1996). However, this idea is being revised owing to studies that have shown that FXII-deficient mice are protected against arterial thrombosis and ischaemic stroke and, like FXII-deficient humans, do not display any detectable alteration in haemostasis, suggesting that FXII inhibitors may be good candidates as antithrombotic agents that avoid an increased bleeding risk (Renné *et al.*, 2005; Kleinschnitz *et al.*, 2006). Under normal conditions, FXII is found as a weakly activated zymogen in blood in the vascular injury site when exposed to negative surfaces (*e.g.* sub-endothelial collagen). Plasma kallikrein activates FXII to factor XIIa (FXIIa) and also cleaves kininogen to release bradykinin, bridging the coagulation process to inflammation response. Plasma kallikrein plays a major role in inflammation (Maeda *et al.*, 1999), releasing kinins and indirectly producing chemotactic factors (Muldbjerg *et al.*, 1993) and other mediator molecules. Thus, FXIIa not only participates in the blood-coagulation process, but also plays a role in other pathophysiological processes such as inflammation, fibrinolysis and the complement system (Fuhrer *et al.*, 1990). Additionally, FXIIa activates kallikrein, its own activator. Kallikrein activates urokinase plasminogen activator and also induces tissue plasminogen activator *in vivo*, considerably increasing the transformation of plasminogen to plasmin, which is responsible for the degradation of fibrin polymers (Braat & Rijken, 1999). In the complement system, FXIIa directly activates C1 to C1r and indirectly activates C3 to C3a through thrombin and plasmin activation, respectively (Fuhrer *et al.*, 1990). Serine proteases of the blood-coagulation cascade, mainly of the intrinsic pathway (contact system), have been associated with pathological conditions such as recurrent thrombosis in patients treated with fibrinolytic agents (Ewald & Eisenberg, 1995). Latacha *et al.* (2004) showed that thrombin produced in response to administration of fibrinolytic agents may be mediated by factor XIIa activated in the contact system. Alternatively, heparin has been used, but it also produces some undesirable side effects such as haemorrhagic episodes (Szabó *et al.*, 2002). To minimize these side effects, protease inhibitors for enzymes of the contact system have been proposed (Szabó *et al.*, 2002) and some clinical protocols have been tested (Hadjistavropoulos *et al.*, 2005). Another pathological situation involving the contact system is the cardiopulmonary bypass, in which high doses of heparin are injected into the patient to reduce the risk of thrombosis. However, this procedure leads to side effects (Wagner *et al.*, 1994). An alternative that may overcome these problems may be found in the specific inhibitors for enzymes involved in the

blood-coagulation contact system that have been developed. Recombinant human albumin fused with infestin 4 has been shown to have a powerful protective action against ischaemic cardiovascular and cerebrovascular events, inhibiting straight FXIIa (Hagedorn *et al.*, 2010). In this paper, we present for the first time the three-dimensional structures of two Kazal-type inhibitors derived from the kissing bug *T. infestans*, infestins 1 and 4, which are thrombin- and factor XIIa-specific inhibitors, respectively. We also report the evaluation of infestin 4 mutants selected for factor XIIa using a phage display system to obtain more selective and powerful factor XIIa inhibitors.

2. Materials and methods

2.1. Materials

Bovine trypsin was purchased from ICN Biochemicals (Costa Mesa, USA). Human plasma kallikrein was prepared according to Sampaio *et al.* (1974). Human plasmin and human factor Xa were purchased from Boehringer Mannheim (Mannheim, Germany). Bovine thrombin was purchased from Sigma (St Louis, USA). The chromogenic substrates H-D-Phe-Pip-Arg-*p*-nitroanilide (S2238), H-D-Val-Leu-Lys-*p*-nitroanilide (S2251), H-D-Pro-Phe-Arg-*p*-nitroanilide (S2302) and Bz-Ile-Glu-Gly-Arg-*p*-nitroanilide (S2222) were purchased from Chromogenix (Mölndal, Sweden). DEAE-Sephadex (A-50) and CNBr-activated Sepharose were obtained from Amersham Pharmacia (Uppsala, Sweden) to prepare trypsin-Sepharose according to the manufacturer's instructions.

2.2. Expression and purification of recombinant infestin 1-2

Infestin 1-2 was produced as described previously (Campos *et al.*, 2002). The culture medium containing infestin 1-2 was applied onto DEAE-Sephadex A-50 (batch form), the resin was washed with 50 mM Tris–HCl pH 8.0 containing 350 mM NaCl and the inhibitor was eluted with the same buffer containing 0.6 M NaCl. Briefly, medium containing infestin 1-2 was applied onto a trypsin-Sepharose column previously equilibrated with 0.1 M Tris–HCl pH 8.0. Proteins that were weakly bound to the column were washed out using 0.1 M Tris–HCl pH 8.0 containing 0.15 M NaCl. Finally, the inhibitor was eluted with 0.5 M KCl pH 2.0 solution and immediately neutralized with 1 M Tris–HCl pH 8.0. Infestin 1-2 was analyzed by SDS–PAGE, reverse-phase HPLC in a C$_8$ column (Vydac) and protein sequencing (model PPSQ-23; Shimadzu, Japan).

2.3. Crystallization and crystallographic data collection of the infestin 1-2–trypsin complex

Purified infestin 1-2 was incubated with trypsin (2:1) for 10 min at room temperature. The enzyme–inhibitor complex was separated from the inhibitor by gel-filtration chromatography on Sephadex G-50. The material corresponding to the complex was pooled and concentrated to 10 mg ml^{-1} in 10 mM Tris–HCl buffer pH 8.0 using a Centriprep YM-10 centrifugal filter.

Crystallization was carried out by vapour diffusion in Cryschem sitting-drop plates using Crystal Screen (Hampton Research), mixing equal volumes (2 µl) of protein solution and reservoir solution. Crystals from condition No. 17 [0.1 M Tris–HCl pH 8.5, 0.2 M lithium sulfate, 30%(w/v) polyethylene glycol 4000] were used in data collection.

X-ray diffraction data were collected on the DB03-MX1 beamline at Laboratório Nacional de Luz Síncrotron (Polikarpov *et al.*, 1998). The crystal was soaked for 5 s in the growth solution containing 10%(w/v) glycerol to prevent ice formation. The crystal was flash-cooled and maintained in a gaseous nitrogen stream at 100 K. The wavelength of the radiation source was set to 1.427 Å and a MAR CCD detector was used to record the X-ray diffraction data. The data were indexed, integrated, reduced and scaled using the *HKL*-2000 package (Otwinowski & Minor, 1997). Molecular replacement was performed with the *MOLREP* program (Vagin & Teplyakov, 2010) using an edited molecule obtained from the structure of PDB entry 1tbr (rhodniin; van de Locht *et al.*, 1995) as a model. Amino-acid residues 1–5, 34–39 and 49–103 were missing in the rhodniin search model.

2.4. Crystallization and crystallographic data collection of infestin 4

Infestin 4 was produced as described previously (Campos, Tanaka-Azevedo *et al.*, 2004). A crystal was obtained in 0.1 M sodium cacodylate, 0.1 M ammonium acetate pH 5.6, 30% PEG 4000. X-ray diffraction data collection and processing were performed as described in Campos, Guimarães *et al.* (2004).

2.5. Construction of models of the protein–protein complex

The thrombin structure (PDB entry 1tbr; van de Locht *et al.*, 1995) was superposed onto that of trypsin in the structure of trypsin–infestin 1 using the program *Coot* (Emsley & Cowtan, 2004). The atomic position of thrombin was saved and merged with infestin 1 to produce the thrombin–infestin 1 complex. The complex was minimized using the steepest-descent minimization algorithm as implemented in the *UCSF Chimera* software (Pettersen *et al.*, 2004). *PyMOL* (DeLano, 2002) was used to generate images of protein structures.

2.6. Infestin 4 DNA-fragment cloning into pCANTAB 5E

The infestin 4 gene fragment was cloned into pCANTAB 5E according to the method described by Tanaka *et al.* (1999). Infestin 4 DNA was obtained by polymerase chain reaction (PCR) amplification using the construct pIC 11.1.1 (Campos, Tanaka-Azevedo *et al.*, 2004), with forward primer MUT4-PCAN (5′-GCGGCCCAGCCGGCCGAGGTCAGAAACC-CTTGCGCCTGTNNBNNBNNBNNBGTTNNBGTGTGTG-GTTCAGATGGG-3′) and reverse primer INF4PCANR (5′-GGGAAGAAGTCCAAGCGGCCGCAAACTGTTCAAC-3′) from Invitrogen (São Paulo, Brazil).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outermost resolution shell.

	Infestin 1–trypsin	Infestin 4
Data collection		
Beamline	D03B-MX1, LNLS	D03B-MX1, LNLS
Wavelength (Å)	1.427	1.427
Space group	$P2_12_12_1$	$P2_12_12$
Unit-cell parameters (Å, °)	$a = 57.993$,	$a = 25.765$,
	$b = 62.692$,	$b = 45.377$,
	$c = 67.837$,	$c = 56.771$,
	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$
Resolution range (Å)	46.1–2.5	40.0–1.4
No. of unique reflections	8955	13618
Multiplicity	4.5 (3.3)	7.1 (5.3)
Completeness (%)	99.7 (98.1)	99.0 (91.0)
$\langle I/\sigma(I) \rangle$	11.7 (2.6)	52.5 (5.7)
R_{merge}† (%)	12.4 (44.3)	3.7 (36.7)
Matthews coefficient (Å^3 Da^{-1})	2.10	2.68
Solvent content (%)	41.36	54.03
Model refinement		
PDB code	2f3c	2erw
No. of protein chains	2	1
No. of water molecules	91	74
R_{work}‡ (%)	17.0	18.1
R_{free} (%)	25.2	19.0
R.m.s.d.		
From ideal bond lengths (Å)	0.014	0.014
From ideal angles (°)	1.540	1.486
Average B factor (Å^2)	26.61	24.05
Ramachandran plot (%)		
Most favoured regions	96.93	98.04
Additional allowed regions	2.69	1.96
Disallowed regions	0.38	0.00

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl) \rangle$ is the average over all observations of reflection hkl. ‡ $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. R_{free} is R_{work} calculated using 5% of the data that were omitted from refinement.

2.7. Construction of infestin 4 mutant library (P2–P2′ and P4′)

The correct DNA sequence of the infestin 4 gene fragment was confirmed using S1 and S6 primers for pCANTAB 5E. The new vector was named pIC 13.1.4. The degenerate oligonucleotide (MUT4PCAN) was synthesized with several mutations. It contained the nucleotide sequence *NNSDBB* (*N* = A/T/C/G, *S* = G/C, *B* = C/G/T and *D* = A/G/T), which codes for a restricted pool of amino acids at the P2–P2′ and P4′ positions of infestin 4 (Campos, Tanaka-Azevedo *et al.*, 2004). The library was constructed as described by Tanaka *et al.* (1999).

2.8. Selection of infestin 4 mutants binding factor XIIa

Transformed *Escherichia coli* TG1 cells harbouring the phagemids of the infestin 4 mutant library were grown to an optical density A_{550} of 0.5–0.7 at 303 K in 2×YT medium containing 200 µg ml^{-1} ampicillin and 2%(w/v) glucose; helper phages (M13KO7) were added (multiplicity of infection = 100) in order to produce fusion phages and growth was continued for 1 h at 310 K. The bacterial culture medium was centrifuged; the bacteria were collected and resuspended in fresh 2×YT medium (200 µg ml^{-1} ampicillin and 50 µg ml^{-1}

kanamycin) and grown for 12–15 h at 310 K. The bacterial culture was centrifuged, the pellet was discarded and the fusion phage particles (supernatant) were used for repetitive selection by affinity binding to factor XIIa as described by Tanaka *et al.* (1999). After the third round of selection, bound fusion phages were eluted and used for infection of *E. coli* cells. 80 colonies were randomly selected and plasmidial DNAs were sequenced.

2.9. Cloning, expression and purification of infestin 4 mutants

Infestin 4 mutant DNAs were amplified by PCR using pIC 14.1.3 and pIC 14.1.15 as template and the primers INF4DREV (5′-CCCAAGCTTTCAAAACTGTTCAAC-ATTGCTGCGCTGACAACGTCC-3′) and INF4DFWD (5′-ATCTCTAGATAAAAGAGAGGTCAGAAACCCTTGC-GCC-3′). The PCR product was digested with *Xho*I and *Not*I restriction enzymes and was ligated into the plasmid pPIC9 (Invitrogen). The resulting plasmids were linearized with *Sal*I restriction enzyme to transform competent *Pichia pastoris* GS115 yeast strain prepared according to the manufacturer's instructions. Yeast cells were incubated for 5 d in BMMY medium. After fermentation, yeast cells were harvested by centrifugation (4000g, 20 min, 277 K). Infestin 4 mutants were purified by affinity chromatography on a trypsin-Sepharose column.

2.10. Kinetics assay and determination of equilibrium dissociation constants (K_i)

The equilibrium dissociation constants for complexes of purified mutated infestins with bovine trypsin (EC 3.4.21.4), human factor Xa (EC 3.4.21.6), human factor XIIa (EC 3.4.21.38) and human plasmin (EC 3.4.21.7) were determined according to Bieth (1980) using the chromogenic substrates (Chromogenix) S2222 (Bz-Ile-Gly-Arg-pNa) for trypsin and factor Xa, S2302 (H-D-Pro-Phe-Arg-pNa) for factor XIIa and S2251 (H-D-Val-Leu-Lys-pNa) for plasmin. Apparent K_i values were calculated by fitting the steady-state velocities to the equation describing a tight-binding inhibitor model using nonlinear regression analysis (Morrison, 1969).

3. Results and discussion

3.1. Overall structures of infestins 1 and 4

In the crystal of the complex of infestin 1-2 and trypsin, only electron density for infestin 1 and trypsin was found (Supplementary Fig. S1[1]). Double-headed protease inhibitors present a flexible inter-domain loop that can be cleaved by proteases present in crystallization experiments. Ibrahim & Pattabhi (2004) showed that an ovomucoid inhibitor with three Kazal domains was processed to single domains in the solved structures. After confirming that the infestin 1-2

inter-domain did not contain any potential cleavage site for trypsin, we hypothesized that a small amount of chymotrypsin was contaminating the trypsin used for complex formation and therefore could be responsible for the hydrolysis between amino-acid residues Phe55 and Glu56 in the inter-domain during the crystallization assay. This hypothesis was subsequently reinforced when we confirmed the presence of chymotrypsin activity in the commercial trypsin that was used using a chromogenic substrate.

The crystals of infestin 1 in complex with trypsin presented high-quality electron density, allowing modelling of residues 16–238 and 5–50 of trypsin and infestin 1, respectively. In the structure of infestin 4 the high quality of the electron-density map allowed the construction of residues 4–56. Almost all of the exposed side chains could be defined in the electron-density maps. The data-collection and refinement statistics are described in Table 1.

Infestins consist of a short central α-helix (infestin 1, Asn25–Lys37; infestin 4, Asn25–Lys36) and a small anti-parallel β-sheet (infestin 1, Ala7–Pro9, Val15–Gly17, Thr22–Tyr23 and Gln42–Glu45; infestin 4, Val15–Gly17, Thr22–Tyr23 and Leu42–Glu45). This topology and the presence of three conservative disulfide bonds are characteristic of Kazal-type inhibitors (Fig. 1a; Engel et al., 1987; Esch et al., 1983; Bolander et al., 1988; Bode & Huber, 1991; Friedrich et al., 1993; van de Locht et al., 1995; Hohenester et al., 1997).

The Kazal-type inhibitor family contains proteins consisting of 46–61 amino acids containing six conserved cysteines that can be divided into classical and nonclassical clades. In group I nonclassical Kazal-type inhibitors, such as rhodiniin, the first and the second cysteine residues are separated by only a single amino-acid residue. In contrast, the classical Kazal-type inhibitors have a spacer of seven to eight amino acids between the first and second cysteine residues (Supplementary Fig. S2; van de Locht et al., 1995).

Structural alignment with other known Kazal-type inhibitors shows that the major structural differences between infestins and other Kazal-type inhibitors can be visualized on superimposing their structures (Fig. 1a). We found that the most different regions were loops Ala7–Asn11 and Lys36–Leu40, the first of which corresponds to the reactive site. At the reactive site the largest distance between infestin 1 and infestin 4 is 2.15 Å between the P1 residues of the inhibitors (Fig. 1b).

3.2. Complex between infestin 1 and trypsin

The structure of the complex between infestin 1 and trypsin was solved (Fig. 2a) and shows that infestin 1 interacts with trypsin in a substrate-like manner (Fig. 2b). The P1 and P1′ residues confirmed in the reactive site of infestin domain 1 are Arg10 and Val11, respectively. The arginine extends into the trypsin active site and its side chain interacts with six residues of the trypsin structure (Figs. 2c and 2d). Interestingly, one β-strand of infestin 1 (residues Ala7–Pro9) interacts with an antiparallel β-strand of trypsin (Fig. 2a).

Sequence alignment of infestin 1 and other Kazal-type protease inhibitors shows that thrombin inhibitors have Arg or His residues at P1 in the reactive site (Supplementary Fig. S2) and that these residues are not conserved in other types of serine protease inhibitors that have the specificity of thrombin inhibitors.

Previous studies have shown the capability of infestin 1-2 to inhibit thrombin activity (Campos et al., 2002). α-Thrombin (EC 3.4.21.5) is a serine protease that participates in the control of blood-coagulation processes. Thrombin has trypsin-like specificity for Arg-Xaa and Lys-Xaa peptide bonds, but with a higher selectivity than trypsin (DiBella & Scheraga, 1996). The surface of the thrombin structure displays three well known sites: the active site, which is responsible for the
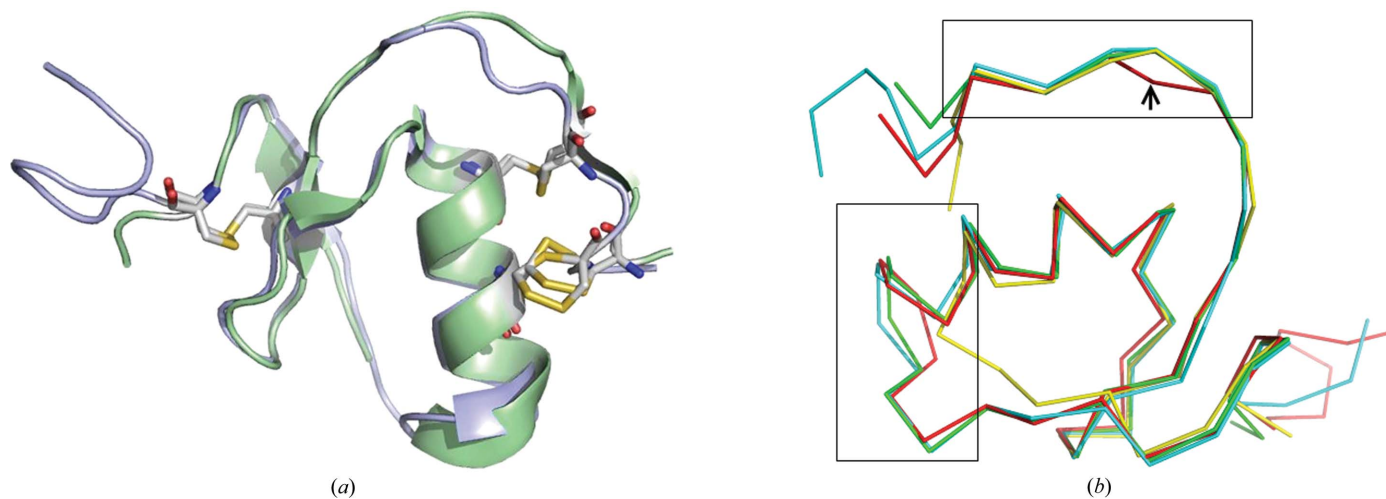

(a)

(b)

Figure 1
Overall structures of infestins 1 and 4. (a) Superposed three-dimensional structures of infestin 1 (PDB entry 2f3c; green) and infestin 4 (PDB entry 2erw; blue). (b) Structural alignment of infestin 1 (green), infestin 4 (red), rhodniin (PDB entry 1tbr; blue) and leech-derived tryptase inhibitor (PDB entry 1an1; yellow). The major differing regions are indicated and comprised the loops Ala7–Asn11 and Lys36–Leu40; the first loop corresponds to the reactive site. The arrow indicates the location of the P1 residues.

Table 2
Analysis of protein–protein interactions between serine protease inhibitors and the enzymes they inhibit.

Interaction	Buried surface area (Å^2)
Infestin 1–thrombin	719.0
Infestin 1–trypsin	703.1
Rhodniin–thrombin	1703.2

proteolytic cleavage of fibrinogen, the fibrinogen-binding site, also called the anion-binding exosite or exosite I, and the heparin-binding site or exosite II (Fenton *et al.*, 1988).

A theoretical structural model of thrombin and infestin 1 was obtained by molecular docking. Interface analysis of thrombin and trypsin complexed with infestin 1 indicated that the interacting residues are conserved and the interface areas of both complexes are of the same magnitude (Table 2). Rhodniin is composed of two Kazal-type domains, like infestin 1-2, and plays a role in thrombin inhibition. A superposition of rhodniin–thrombin (PDB entry 1tbr) and infestin 1–thrombin complex models shows that with the exception of the P1 and P1′ residues the crucial interacting residues (Cys6, Cys8,

Leu12 and Ser24; infestin 1 numbering) are structurally conserved in the first domain of rhodniin (Fig. 3). However, more interactions between rhodniin and thrombin were found owing to the presence of the second domain.

3.3. Phage display library

Owing to the importance of our finding the first specific factor XIIa inhibitor in haematophagous animals, a phage display library using infestin 4 as a template was constructed and selected for factor XIIa. In an attempt to find structural features, selected infestin 4 mutants were analyzed and a consensus Arg-Arg was found at P1–P1′ (Supplementary Fig. S3). Two mutants named inf4mut15 and inf4mut3 were expressed, purified and characterized. Inf4mut15 was shown to be more selective for factor XIIa than wild-type infestin 4. The F9T and N10R mutations in inf4mut15 increased the affinity for factor XIIa to picomolar values (K_i = 3.6 pM), making this inhibitor more specific for factor XIIa than wild-type infestin 4 and suggesting that it is a better candidate
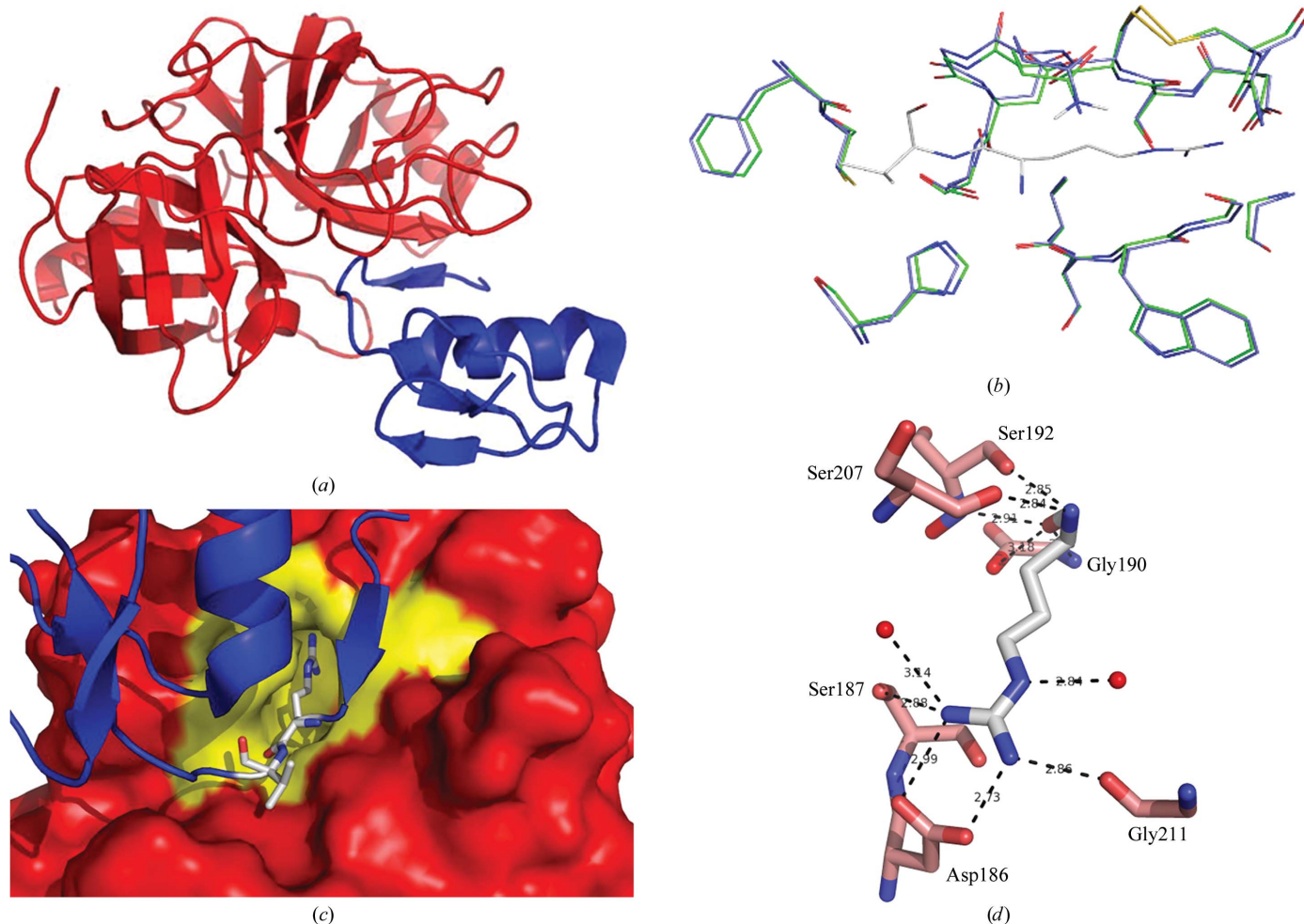

Figure 2
Complex between infestin 1 and trypsin. (*a*) Three-dimensional structures of infestin 1 (blue) and trypsin (red). (*b*) Conserved amino-acid residues of thrombin (C atoms in green) and trypsin (C atoms in blue); these residues are less than 4 Å from the P1 and P1′ residues of infestin 1. (*c*, *d*) The P1 and P1′ residues of the infestin 1 active site extend into the trypsin active site. The P1 side chain interacts with six residues of trypsin.
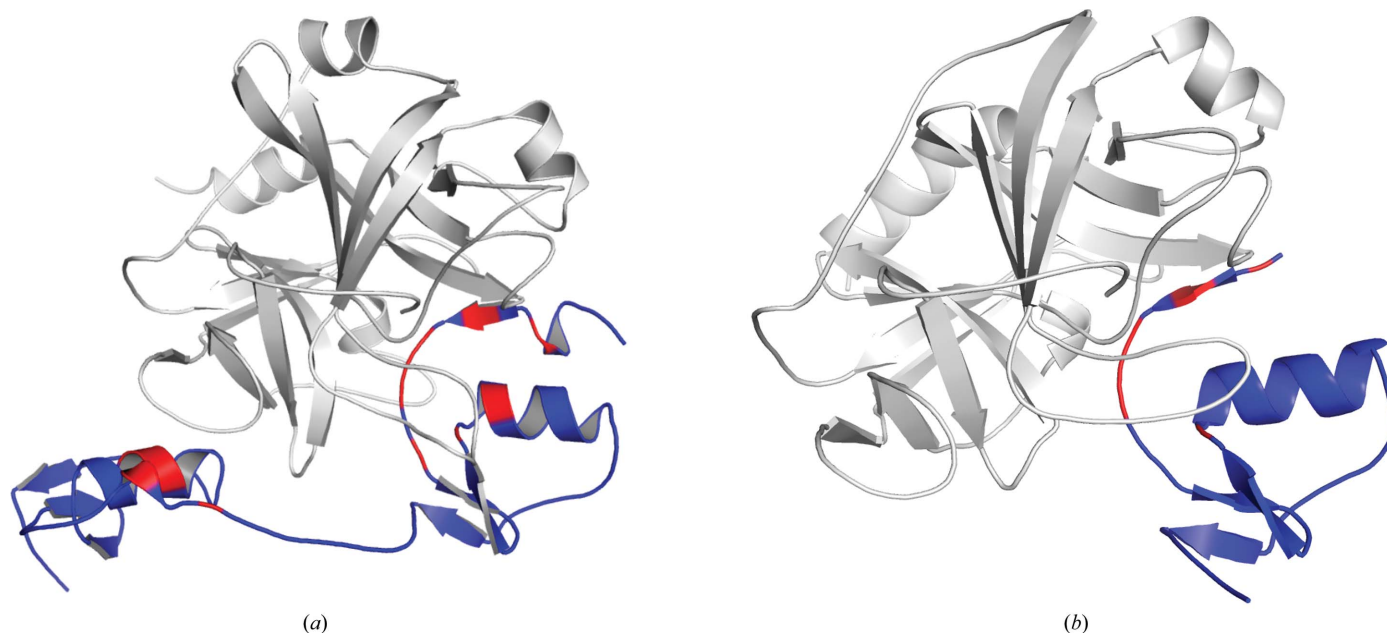
Figure 3
Analysis of rhodniin and infestin 1 with thrombin. (*a*) Interacting amino-acid residues (red) found in infestin 1 (blue) docked to thrombin (grey). (*b*) Interacting residues (red) found in the thrombin (grey)–rhodniin (blue) interface (PDB entry 1tbr).

Table 3
Dissociation constants of infestin 4, inf4mut15 and inf4mut3 for different serine proteases.

Inf4mut15 and inf4mut3 are infestin 4 mutants selected for factor XIIa by phage display. ni, not inhibited.

	Sequence, P2–P2′–P4′	K_i (n*M*)			
		Factor XIIa	Trypsin	Plasmin	Factor Xa
Infestin 4	FRNYvP	0.1	11.3	2.1	53.0
Inf4mut15	TRRFvA	0.0036	2.3	ni	ni
Inf4mut3	SRRLvH	4.6	9.5	ni	ni

molecule for antithrombotic therapy (Table 3; Campos, Tanaka-Azevedo *et al.*, 2004).

Models of wild-type infestin 4 and inf4mut15 were docked with factor XIIa. Analyzing their interactions, it was observed that the side chain of Arg11 at P1′ of inf4mut15 is longer than that of Asn11 found in native infestin 4 and that it may interact with Ser191 and Thr37 of thrombin. Those interactions are not observed in the native infestin 4 structure and may be among the structural features that are responsible for the high affinity and selectivity for factor XIIa of inf4mut15 (Supplementary Fig. S4).

Analyzing the inf4mut3 and inf4mut15 reactive sites, it was observed that in addition to the longer side chain of Arg11, the presence of a residue with an aromatic ring at position 12 may also contribute to factor XIIa inhibitory activity.

In conclusion, the specificity of infestin 4 for factor XIIa is in agreement with the highest gain in solvation free energy upon complex formation. Using a phage display system and infestin 4, a mutant (inf4mut15) with strong and specific binding of factor XIIa was selected by making only four amino-acid replacements in the entire molecule, and the presence of an Arg residue at the P1′ position was shown to be extremely important for selectivity for factor XIIa.

References

Araujo, R. N., Campos, I. T. N., Tanaka, A. S., Santos, A., Gontijo, N. F., Lehane, M. J. & Pereira, M. H. (2007). *Int. J. Parasitol.* **37**, 1351–1358.

Arocha-Piñango, C. L., Marchi, R., Carvajal, Z. & Guerrero, B. (1999). *Blood Coagul. Fibrinolysis*, **10**, 43–68.

Bieth, J. G. (1980). *Bull. Eur. Physiopathol. Respir.* **16**, Suppl., 183–197.

Bode, W. & Huber, R. (1991). *Biomed. Biochim. Acta*, **50**, 437–446.

Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.

Bolander, M. E., Young, M. F., Fisher, L. W., Yamada, Y. & Termine, J. D. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 2919–2923.

Braat, E. A. & Rijken, D. C. (1999). *Thromb. Haemost.* **81**, 657.

Bugge, T. H., Xiao, Q., Kombrinck, K. W., Flick, M. J., Holmbäck, K., Danton, M. J., Colbert, M. C., Witte, D. P., Fujikawa, K., Davie, E. W. & Degen, J. L. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 6258–6263.

Campos, I. T. N., Amino, R., Sampaio, C. A. M., Auerswald, E. A., Friedrich, T., Lemaire, H.-G., Schenkman, S. & Tanaka, A. S. (2002). *Insect Biochem. Mol. Biol.* **32**, 991–997.

Campos, I. T. N., Guimarães, B. G., Medrano, F. J., Tanaka, A. S. & Barbosa, J. A. R. G. (2004). *Acta Cryst.* D**60**, 2051–2053.

Campos, I. T. N., Tanaka-Azevedo, A. M. & Tanaka, A. S. (2004). *FEBS Lett.* **577**, 512–516.

Crab, A., Noppe, W., Pelicaen, C., Van Hoorelbeke, K. & Deckmyn, H. (2002). *Thromb. Haemost.* **87**, 899–904.

research papers

Davie, E. W., Fujikawa, K. & Kisiel, W. (1991). *Biochemistry*, **30**, 10363–10370.

DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.

DiBella, E. E. & Scheraga, H. A. (1996). *Biochemistry*, **35**, 4427–4433.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Engel, J., Taylor, W., Paulsson, M., Sage, H. & Hogan, B. (1987). *Biochemistry*, **26**, 6958–6965.

Esch, F. S., Böhlen, P., Ling, N. C., Brazeau, P. E., Wehrenberg, W. B. & Guillemin, R. (1983). *J. Biol. Chem.* **258**, 1806–1812.

Ewald, G. A. & Eisenberg, P. R. (1995). *Circulation*, **91**, 28–36.

Fenton, J. W. II, Olson, T. A., Zabinski, M. P. & Wilner, G. D. (1988). *Biochemistry*, **27**, 7106–7112.

Fink, E., Rehm, H., Gippner, C., Bode, W., Eulitz, M., Machleidt, W. & Fritz, H. (1986). *Biol. Chem. Hoppe Seyler*, **367**, 1235–1242.

Francischetti, I. M., Valenzuela, J. G., Andersen, J. F., Mather, T. N. & Ribeiro, J. M. (2002). *Blood*, **99**, 3602–3612.

Friedrich, T., Kröger, B., Bialojan, S., Lemaire, H. G., Höffken, H. W., Reuschenbach, P., Otte, M. & Dodt, J. (1993). *J. Biol. Chem.* **268**, 16216–16222.

Fuhrer, G., Gallimore, M. J., Heller, W. & Hoffmeister, H. E. (1990). *Blut*, **61**, 258–266.

Hadjistavropoulos, H., Bierlein, C., Neville, S., Tuttosi-White, T., Quine, A. & Sagan, M. (2005). *Healthc. Manage. Forum*, **18**, 22–26.

Hagedorn, I., Schmidbauer, S., Pleines, I., Kleinschnitz, C., Kronthaler, U., Stoll, G., Dickneite, G. & Nieswandt, B. (2010). *Circulation*, **121**, 1510–1517.

Hemmi, H., Kumazaki, T., Yoshizawa-Kumagaye, K., Nishiuchi, Y., Yoshida, T., Ohkubo, T. & Kobayashi, Y. (2005). *Biochemistry*, **44**, 9626–9636.

Hohenester, E., Maurer, P. & Timpl, R. (1997). *EMBO J.* **16**, 3778–3786.

Ibrahim, B. S. & Pattabhi, V. (2004). *Biochem. Biophys. Res. Commun.* **313**, 8–16.

Iwanaga, S., Okada, M., Isawa, H., Morita, A., Yuda, M. & Chinzei, Y. (2003). *Eur. J. Biochem.* **270**, 1926–1934.

Kleinschnitz, C., Stoll, G., Bendszus, M., Schuh, K., Pauer, H. U., Burfeind, P., Renné, C., Gailani, D., Nieswandt, B. & Renné, T. (2006). *J. Exp. Med.* **203**, 513–518.

Latacha, M. P., Schaiff, W. T., Eisenberg, P. R. & Abendschein, D. R. (2004). *J. Thromb. Haemost.* **2**, 128–134.

Lee, A. Y. & Vlasuk, G. P. (2003). *J. Intern. Med.* **254**, 313–321.

Locht, A. van de, Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., Höffken, W. & Bode, W. (1995). *EMBO J.* **14**, 5149–5157.

Lovato, D. V., Amino, R., González, Y., Miranda, A., Schenkman, S. & Tanaka, A. S. (2011). *Exp. Parasitol.* **129**, 362–367.

Lovato, D. V., Nicolau de Campos, I. T., Amino, R. & Tanaka, A. S. (2006). *Biochimie*, **88**, 673–681.

Maeda, H., Wu, J., Okamoto, T., Maruo, K. & Akaike, T. (1999). *Immunopharmacology*, **43**, 115–128.

Mende, K., Petoukhova, O., Koulitchkova, V., Schaub, G. A., Lange, U., Kaufmann, R. & Nowak, G. (1999). *Eur. J. Biochem.* **266**, 583–590.

Morrison, J. F. (1969). *Biochim. Biophys. Acta*, **185**, 269–286.

Moser, M., Auerswald, E., Mentele, R., Eckerskorn, C., Fritz, H. & Fink, E. (1998). *Eur. J. Biochem.* **253**, 212–220.

Mühlhahn, P., Czisch, M., Morenweiser, R., Habermann, B., Engh, R. A., Sommerhoff, C. P., Auerswald, E. A. & Holak, T. A. (1994). *FEBS Lett.* **355**, 290–296.

Muldbjerg, M., Markussen, S., Magnusson, S. & Halkier, T. (1993). *Blood Coagul. Fibrinolysis*, **4**, 47–54.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). *J. Synchrotron Rad.* **5**, 72–76.

Renné, T., Pozgajová, M., Grüner, S., Schuh, K., Pauer, H. U., Burfeind, P., Gailani, D. & Nieswandt, B. (2005). *J. Exp. Med.* **202**, 271–281.

Rosen, E. D., Chan, J. C. Y., Idusogie, E., Clotman, F., Vlasuk, G., Luther, T., Jalbert, L. R., Albrecht, S., Zhong, L., Lissens, A., Schoonjans, L., Moons, L., Collen, D., Castellino, F. J. & Carmeliet, P. (1997). *Nature (London)*, **390**, 290–294.

Sampaio, C., Wong, S.-C. & Shaw, E. (1974). *Arch. Biochem. Biophys.* **165**, 133–139.

Schlott, B., Wöhnert, J., Icke, C., Hartmann, M., Ramachandran, R., Gührs, K. H., Glusa, E., Flemming, J., Görlach, M., Grosse, F. & Ohlenschläger, O. (2002). *J. Mol. Biol.* **318**, 533–546.

Shenoy, R. T., Thangamani, S., Velazquez-Campoy, A., Ho, B., Ding, J. L. & Sivaraman, J. (2011). *PLoS One*, **6**, e18838.

Sommerhoff, C. P., Söllner, C., Mentele, R., Piechottka, G. P., Auerswald, E. A. & Fritz, H. (1994). *Biol. Chem. Hoppe Seyler*, **375**, 685–694.

Szabó, G., Barabás, E., Kedves, R., Csomor, K. & Kuszmann, J. (2002). *Thromb. Res.* **107**, 357–363.

Tanaka, A. S., Silva, M. M., Torquato, R. J. S., Noguti, M. A. E., Sampaio, C. A. M., Fritz, H. & Auerswald, E. A. (1999). *FEBS Lett.* **458**, 11–16.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Wagner, W. R., Johnson, P. C., Thompson, K. A. & Marrone, G. C. (1994). *Ann. Thorac. Surg.* **58**, 734–740.

Watanabe, R. M. O., Soares, T. S., Morais-Zani, K., Tanaka-Azevedo, A. M., Maciel, C., Capurro, M. L., Torquato, R. J. S. & Tanaka, A. S. (2010). *Biochimie*, **92**, 933–939.