Lopap, a prothrombin activator from *Lonomia obliqua* belonging to the lipocalin family: recombinant production, biochemical characterization and structure–function insights

Cleyson Valença REIS*, Sonia Aparecida ANDRADE†, Oscar Henrique Pereira RAMOS*, Celso Raul Romero RAMOS‡, Paulo Lee HO‡, Isabel de Fátima Correia BATISTA* and Ana Marisa CHUDZINSKI-TAVASSI*¹

*Laboratório de Bioquímica e Biofísica, Instituto Butantan, 1500 Av. Vital Brazil, CEP 05503-900, São Paulo, SP, Brazil, †Laboratório de Hemostasia, Instituto de Ensino e Pesquisa Hospital Sírio Libanês, 69 Rua Cel. Nicolau dos Santos, CEP 01308-050, São Paulo, SP, Brazil, and ‡Centro de Biotecnologia, Instituto Butantan, 1500 Av. Vital Brazil, CEP 05503-900, São Paulo, SP, Brazil

Using a cDNA library made from *Lonomia obliqua* caterpillar bristles, we identified a transcript with a 603 bp open reading frame. The deduced protein corresponds to Lopap, a prothrombin activator previously isolated by our group from the bristles of this species. The mature protein is composed by 185 amino acids and shares similarity with members of the lipocalin family. The cDNA encoding the mature form was amplified by PCR, subcloned into pAE vector and used to transform *Escherichia coli* BL21(DE3) cells. As for the native Lopap, the recombinant fusion protein shows enzymatic activity, promotes prothrombin hydrolysis, generates fragments similar to prethrombin-2 and fragment 1.2 as intermediates, and generates thrombin as the final product. In ad-

dition, structural bioinformatics studies indicated several interesting molecular features, including the residues that could be responsible for Lopap's serine protease-like activity and the role of calcium binding in this context. Such catalytic activity has never been found in other members of the lipocalin family. This is the first report describing the recombinant production and biochemical characterization of a *Lonomia obliqua* lipocalin, as well as the structural features that could be responsible for its serine protease-like catalytic activity.

Key words: catalytic lipocalin, lipocalin, *Lonomia obliqua*, Lopap, prothrombin activator, structure–function relationship.

INTRODUCTION

The contact of human skin with *Lonomia obliqua* caterpillar bristles leads to an envenoming characterized by consumption coagulopathy and secondary fibrinolysis [1,2]. Some studies showed that the crude bristle extract from this caterpillar induces the clot formation by triggering activation of both prothrombin and FX (Factor X) [3,4]. A prothrombin activator named Lopap (*Lonomia obliqua* prothrombin activator protein) was purified from *L. obliqua* bristle extract. The native Lopap is a 69 kDa homotetrameric protein that is able to generate thrombin and prethrombin-2, by hydrolysing the Arg²⁸⁴-Thr²⁸⁵ peptide bond of the human prothrombin molecule. Its proteolysis is improved in the presence of calcium [5], and serine protease inhibitors, such as PMSF, inhibit it. Furthermore, it is able to induce activation, expression of adhesion molecules and to exert an anti-apoptotic effect on HUVECs (human umbilical vein endothelial cells) [6].

The sequenced N-terminus of the native Lopap (first 46 residues) presents 55 % identity with bombyrin from *Bombyx mori* [7] and 51 % identity with the insecticynin from the haemolymph of *Manduca sexta* [5,8], both members of the lipocalin protein family [9]. The lipocalins are highly diverse and found in a variety of species, playing important roles in retinol transport, pheromone transport, regulation of the immune response and cell homeostatic mediation [9–11]. All of these functions are related to lipocalin-binding activity. There are, however, at least three

previous reports describing lipocalins with enzymatic activity: (i) prostaglandin D synthase, a lipocalin responsible for the biosynthesis of PGD (prostaglandin D) [12]; (ii) violaxanthin and zeaxanthin epoxidase, lipocalins that catalyse carotenoid interconversions [13]; and (iii) tear lipocalin and β -lactoglobulin, lipocalins with nuclease activity [14].

In the present paper, we describe the cloning, sequence analysis, bacterial expression and structural insights regarding Lopap. The results reveal that the recombinant protein exhibits similar characteristics when compared with native Lopap; therefore it should be considered as a lipocalin family member with a new catalytic activity.

MATERIALS AND METHODS

cDNA library construction

Total RNA from *L. obliqua* bristles was extracted using TRIzol® (Life Technologies) [15], and mRNAs were subsequently purified with an oligo(dT)–cellulose column (Amersham Biosciences) according to the manufacturer's instructions. The cDNAs were synthesized with Superscript plasmid system (Life Technologies) using a synthetic oligonucleotide dT18-NotI primer-adaptor (Amersham Biosciences). Reverse transcription was performed with SuperScript reverse transcriptase (Life Technologies). The

Abbreviations used: FX, etc., Factor X, etc.; IPTG, isopropyl β -D-thiogalactoside; Lopap, *Lonomia obliqua* prothrombin activator protein; rLopap, recombinant Lopap; PGD, prostaglandin D; L-PGDS, lipocalin-like PGD synthase; RMSD, root mean square deviation; TFA, trifluoroacetic acid.

¹ To whom correspondence should be addressed (email amchudzinski@butantan.gov.br).

The nucleotide sequence data reported for *Lonomia obliqua* prothrombin activator have been deposited in the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession number AY908986.1.

The amino acid sequence data reported for *Lonomia obliqua* prothrombin activator have been deposited in the NCBI Protein Sequence Database under the accession number AAW88441.1.

resulting cDNAs were ligated to EcoRI adapters (Amersham Biosciences) and the products were digested with both restriction enzymes (NotI and EcoRI) and unidirectionally cloned into the plasmid pGEM-11Zf(+) (Promega). Finally, *Escherichia coli* DH5 α cells were transformed by the recombinant plasmids.

Lopap cDNA cloning

In order to obtain the cDNA that encodes Lopap, a degenerate sense primer was designed according to the N-terminal sequence of the mature protein (Asp¹⁷–Ala²³), and PCRs were performed using the constructed cDNA library as template. The amplification product was sequenced using the BigDye dideoxy method using an ABI 377 Automated DNA Sequencer (Applied Biosystems). Based on the obtained sequence, a reverse primer was designed (5'-CTCAACACTTCCACTATCACCGTTCGC-3'). Sense and reverse primers respectively carrying BamHI and EcoRI restriction sites were used in PCRs aiming to amplify the cDNA that encodes the mature Lopap. The amplification product was unidirectionally cloned into the pAE expression vector [16], after BamHI and EcoRI restriction of both molecules. The construction was analysed and sequenced. The resulting plasmid was called pAE-Lopap.

rLopap (recombinant Lopap) production

E. coli BL21(DE3) cells were transformed with pAE-Lopap. This system was designed for the expression of rLopap fused to a minimal N-terminal His₆-tag. Transformed cells were inoculated in LB (Luria–Bertani) medium with 100 mg/ml ampicillin and cultured at 37 °C. When the *D*₆₀₀ reached 0.5, the rLopap expression was induced with 0.5 mM IPTG (isopropyl β -D-thiogalactoside). After 4 h, the cells were harvested by centrifugation at 3200 *g* for 12 min at 4 °C and disrupted in a French Press. Inclusion bodies were washed in 50 mM Tris/HCl, pH 8.0, 1 % Triton X-100 and 1 M urea and dissolved in solubilization buffer (50 mM Tris/HCl, pH 8.0, 500 mM NaCl, 8 M urea and 5 mM 2-mercaptoethanol). The solution was submitted to refolding by a 200-fold dilution (drop-by-drop) into refolding buffer (50 mM Tris/HCl, pH 8.0, 500 mM NaCl, 50 mM CaCl₂, 5 mM imidazole and 5 mM 2-mercaptoethanol).

rLopap purification

The refolded protein was loaded on to a Ni²⁺–Sephacrose column (Amersham Biosciences) at flow rate of 0.5 ml/min, and non-specifically bonded molecules were washed with 10 column vol. of 50 mM Tris/HCl, pH 6.8, 500 mM NaCl and 20 mM imidazole. Then rLopap was eluted with 4 column vol. of 50 mM Tris/HCl, pH 6.8, 100 mM NaCl and 1 M imidazole, and fractions of 1 ml were collected, analysed by SDS/PAGE (15 % gels) and dialysed against 0.15 M NaCl. Aiming to select rLopap's refolding forms with serine protease activity, fractions containing the partially purified rLopap were pooled and applied (at 0.5 ml/min) on to a benzamidine–Sephacrose column (Amersham Biosciences), equilibrated previously with 50 mM Tris/HCl, pH 7.4, containing 500 mM NaCl. rLopap was eluted from the column by pH shift using 0.05 M glycine, pH 4.0. The eluted fractions (1.0 ml each) were collected in 50 μ l of Tris/HCl, pH 9.0, and dialysed against 150 mM NaCl for further structural and kinetic experiments.

Purified rLopap was analysed by SDS/PAGE (15 % gels) and HPLC using a J. T. Baker C₄ column (4.6 mm $\times$ 250 mm). The HPLC procedure was conducted using an acetonitrile gradient (20–80 %) in TFA (trifluoroacetic acid) (0.1 %) at room temperature (20–25 °C) at 1 ml/min for 20 min.

Sequence alignment and structure analysis

Selected sequences were aligned using ClustalX (gap opening: 10; gap extension: 0.2; Gonnet series matrix). The sequence similarities were analysed according to the RISLER matrix [17] using ESPript [18].

Aiming to find molecular features of Lopap structure that could be related to its serine protease-like activity, we made use of structural modelling by satisfaction of spatial restraints to investigate its monomeric and tetrameric forms. With regard to the monomeric form, the structures of insecticynin from the tobacco hornworm *Manduca sexta* (PDB code 1Z24) and engineered bilin-binding protein variants from the cabbage white butterfly *Pieris brassicae* (PDB codes 1KXO and 1NOS) were taken as templates for the model construction using Modeller 8v1 [19,20]. From 20 initial models, the one with the best variable target function score was chosen for the subsequent steps. The model was then checked for the presence of serine protease catalytic residues using Catalytic Site Atlas (CSA; <http://www.ebi.ac.uk/thornton-srv/databases/CSA/>).

The residues indicated by CSA were refined by rotamer searching and the whole model was submitted to energy minimization (200 steepest descent steps followed by 200 conjugate gradient steps and 200 steepest descent steps) using GROMOS96 implemented in Swiss PDB Viewer. The normality of the stereochemistry, the chemical environment and the atomic contacts of the protein were evaluated using, respectively, Procheck at 2.0 Å (1 Å = 0.1 nm) resolution [21], Verify3D [22] and the module Quality of WHAT IF [23]. The structure of the Lopap monomer can be seen at <http://www.BiochemJ.org/bj/398/bj3980295add.htm>. The modelling of the tetrameric Lopap was conducted with the same methodological approach adopting the structure of bilin-binding protein from *Pieris brassicae* (PDB code 1BBP) as a template.

Finally, since *in vitro* observations indicate that the catalytic activity of Lopap is positively affected by the presence of calcium, the GG software was employed for calcium-binding site prediction [24].

CD measurements

The CD spectra of native Lopap and rLopap treated with urea (0, 3 or 6 M) diluted in 20 mM Tris/HCl, pH 7.4, were recorded using a JASCO J-810 spectropolarimeter with a Peltier system to control the cell temperature. Each spectrum represented the average of eight records read between wavelengths of 190 and 260 nm, with 0.2 nm resolution, 0.5 nm bandwidth, 4 s response time, 100 mdeg sensitivity and 20 nm/min scan speed in cells of 0.2 and 1 mm path length. All sample spectra were adjusted for background removal by subtracting buffer spectra (blank). The CD intensities were expressed as molar ellipticity (deg $\cdot$ cm² $\cdot$ dmol⁻¹). The percentages of the different secondary structures (α -helix, β -sheet, β -turn and random coil) were estimated with 5 % prediction errors on the range 190–260 nm using the CDNN program, version 2.1, ACGT Progenomics.

rLopap activity on human plasma

To evaluate procoagulant activity on human plasma (re-calcification clotting time), rLopap (0.5–5.0 μ M) was incubated at 37 °C with normal human plasma (100 μ l) in the presence of CaCl₂ (6.25 mM) in a final volume of 400 μ l. The same reaction in the absence of rLopap was taken as control.

Activation of prothrombin by rLopap

Prothrombin activation was determined by measuring the generation of thrombin amidolytic activity towards chromogenic

substrate S-2238 (*H*-D-phenylalanyl-L-picolyl-L-arginine-*p*-nitroanilide dihydrochloride) (Chromogenix). rLopap (10 nM) was added to 50 mM Tris/HCl and 100 mM NaCl, pH 8.3 (37°C), containing prothrombin (25 µM) (Enzyme Research Labs) and CaCl₂ (5 mM). After various time intervals (0, 1, 5, 10 and 20 min) samples were withdrawn from the reaction mixture and transferred to 50 mM Tris/HCl buffer, 100 mM NaCl, pH 8.3, 100 mM EDTA and 40 µM S-2238. The amidolytic activity was quantified as described previously [5].

FX activation

FX activation by purified native Lopap and rLopap (15 nM) was determined indirectly by testing FXa formation from 30 nM FX (Sigma Chemical Co.) using 20 µM chromogenic substrate S-2765 (*N*-α-benzoyloxycarbonyl-D-arginyl-L-glycyl-L-arginine-*p*-nitroanilide) (Chromogenix). FX activation by rLopap was carried out in appropriate buffer (25 mM Tris/HCl, 200 mM NaCl and 10 mM CaCl₂, pH 8.3, to a final volume of 200 µl). After 20 min of pre-incubation at 37°C, S-2765 was added, and FXa formation was followed spectrophotometrically at a wavelength of 405 nm. Direct amidolytic activity was measured using the same conditions without FX.

SDS/PAGE analysis of the prothrombin activation

Prothrombin (10 µM) was activated by rLopap (2 µM) or FXa (10 nM) (Sigma Chemical Co.) in the presence or absence of 50 µM phospholipids (phosphatidylserine/phosphatidylcholine, 7:3, v/v) and 100 nM FVa (Haematologic Technologies) in reaction buffer (0.02 M Tris/HCl, 0.15 M NaCl, pH 8.0, and 15 mM CaCl₂). After a 1 h incubation, aliquots (10 µl) were removed to perform SDS/PAGE (10% gels) analysis.

rLopap (2 µM), incubated previously at 37°C for 30 min with 4 µM PMSF (Acros Organics) or saline (used as a control) was submitted to the same protocol, and samples (10 µl) were collected for subsequent SDS/PAGE (10% gels) analysis.

RESULTS

Deduced protein primary structure

Lopap's cDNA sequence was obtained by PCR using primer corresponding to the N-terminal sequence of the native protein, T7 promoter primer and the *L. obliqua* bristle cDNA library as template. A DNA fragment of approx. 600 bp was amplified, cloned and sequenced (Figure 1). The conceptual translation of the amplified sequence is compatible with the N-terminal sequencing [5] of the native Lopap (shaded grey in Figure 1), suggesting that the cloned cDNA encodes the target protein. Based on the sequence obtained, an antisense primer was designed to amplify the 5' end of the Lopap cDNA. With SP6 promoter primer, PCRs were carried out using the *L. obliqua* cDNA library as template. The analysis of the amplified sequence showed an additional amino acid sequence corresponding to a signal peptide, as suggested by SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP>). The cleavage site was predicted between positions 16 and 17 (TAA↓DV). This prediction and the comparison between the protein sequence of native Lopap and the full-length cDNA indicate that Lopap is expressed as a 22.5 kDa protein with a 16-amino-acid N-terminal signal peptide that is cleaved during secretion of this protein to the extracellular medium, resulting in a mature protein of 20.8 kDa. Figure 1 shows the complete cDNA (717 bp) that includes an open reading frame (603 bp) and a 3'-untranslated region (111 bp).

Figure 1 cDNA nucleotide sequence and the predicted amino acid sequence of Lopap

The sequences of amino acids shaded grey were obtained by proteolysis followed by peptide sequencing of the native Lopap protein. The arrow indicates the sequence that inspired the reverse primer used for the obtention of the 5' region of the cDNA that codes for Lopap. The signal peptide-cleavage site and N-terminus of the mature protein are indicated by the arrow. The black background indicates the residues possibly involved in the serine protease-like activity. ATG (underlined) and TAA (italic) represent the start codon and the stop codon respectively. The nucleotide sequence inside the box represents the polyadenylation signal.

Sequence and structural analysis

We searched non-redundant sequences of GenBank® CDS (coding sequence) translations, PDB, SwissProt, PIR and PRF databases with the full-length Lopap protein sequence using BLAST (National Center for Biotechnology Information). The results showed that Lopap has identity with several members of the lipocalin protein family (up to 50% in amino acid sequence identity when comparing with biliverdin-binding protein-1 from *Samia cynthia ricini* caterpillar; GenBank® accession number BAB85482.1). Figure 2 shows the sequence alignment of mature Lopap protein with selected lipocalins. Moreover, the three lipocalin motifs present in the Lopap sequence (GXWY, TDYXXY and IXSR), as well as the positions of four conserved cysteine residues in lipocalins, are indicated.

The similarity of Lopap's amino acid sequence to other lipocalins is also reflected in the structure modelling. The identity observed between the model and its templates ranges from 32.7 to 38.5% and the RMSD (root mean square deviation) of backbone atoms was between 1.61 and 2.31 Å. Figure 3(A) presents the Lopap monomer model showing the characteristic lipocalin basket-like β-barrel formed by eight β-strands (β2–β9) and a conserved α-helix (α1). The search for serine protease active-site residues using CSA predicted that His¹⁶⁸, Glu¹⁷¹ and Ser¹¹⁹ could be related to such activity.

Figure 2 Sequence alignment of Lopap with other members of the lipocalin protein family

Sequence similarities according to the RISLER matrix are shown in boxes using ESPrnt [18]. Secondary structures of bilin-binding protein (PDB code 1BBP) are drawn above the alignment: α -helices are represented as helices, β -strands as arrows, β -turns are identified by 'TT' and 3_{10} -helices by η . The three motifs that characterize the lipocalin family are marked. The cysteine residues that form disulfide bonds are numbered in grey and italic below the corresponding positions. The aligned proteins are PH2IP (prostaglandin H_2 D-isomerase precursor; accession number Q8WNM1), Lopap (accession number AAW88441), BBP (bilin-binding protein precursor; accession number P09464), APOD (apolipoprotein D; accession number NP_001638), INS (insecticynin A; accession number P00305) and bombyrin (accession number BAB47155).

The tetrameric Lopap model compared with its template (bilin-binding protein) exhibited 36.3% identity and an RMSD of 0.42 Å for the backbone atoms. The model obtained suggests that each monomer interacts with a second monomer by its C-terminal portion (including $\alpha 1$ helix) and with a third monomer by β -strand 1 ($\beta 1$) and the loops between $\beta 2$ and $\beta 3$, $\beta 4$ and $\beta 5$, and $\beta 8$ and $\beta 9$.

rLopap production

The pAE-Lopap plasmid encodes the mature Lopap protein with an N-terminus (MHHHHHHLEGS), containing the His₆-tag. The predicted molecular mass for the recombinant protein was 22.1 kDa. As expected, *E. coli* BL21(DE3) cells carrying the pAE-Lopap plasmid produced a recombinant protein of approx. 20 kDa clearly visible by SDS/PAGE in the bacterial cell extracts. The intensity of this band increased by further incubation of the bacterial culture with IPTG, whereas no band was observed for cells transformed with the empty vector (results not shown). rLopap production in this system typically yields 3.5 mg/l of culture. Figure 4 shows the final product of rLopap's purification, analysed by SDS/PAGE (Figure 4A) and by HPLC (Figure 4B), which indicates the high level of purity of the recombinant protein.

Secondary structure analysis of Lopap

Native and rLopap showed very similar secondary-structure compositions (Table 1). The secondary-structure element patterns of both proteins are in good agreement with those present in the crystallographic structures of lipocalins [9,10]. The spectral decon-

volution of the rLopap recorded in solution with increasing amounts of urea (1 and 3 M) revealed that the negative intensity of the spectrum around 222 nm increased with higher urea concentrations, thus suggesting a slow shift from native to unfolded conformations with an concomitant increase of non-native α -helix content (Figure 4C).

rLopap activity on normal human plasma

rLopap (0.5–5.0 μ M) significantly reduced the normal citrated human plasma recalcification time (control, 290 s; 0.5 μ M rLopap, 260 s; 2.5 μ M rLopap, 150 s; and 5.0 μ M rLopap, 70 s).

Prothrombin hydrolyses

rLopap (10 nM) recognizes prothrombin (25 μ M) as a substrate. The time courses of prothrombin activation were linear with time and proportional to the amount of enzyme present in the reaction mixture (results not shown).

FX activation

Lopap and rLopap were not able to activate FX. None of them exhibited direct amidolytic activity on the S-2765 chromogenic substrate, demonstrating an absence of serine protease contaminants such as FXa in the protein preparations.

SDS/PAGE analysis

Prothrombin (72 kDa) was hydrolysed by rLopap producing three bands: a 52 kDa band which probably corresponds to fragment

Figure 3 Schematic representation (cartoons) of Lopap in the monomeric and tetrameric forms

(A) Structural model for Lopap monomer. β -strands ($\beta 1$ – $\beta 9$) and the conserved α -helix ($\alpha 1$) are identified in accordance with the structure features presented in the Figure 2. Residues that were predicted to be involved in the serine protease-like catalytic activity are labelled, in agreement with Figure 1. Bonded calcium ions (Ca) are represented by spheres. To indicate the hydrophobic pocket location, the biliverdin hydrophobic ligand was superposed on the Lopap model. The spatial folding of Lopap resembles a β -barrel structure characteristic of the lipocalins. (B) Structural model for Lopap in the tetrameric form. The monomers that are near to the observer are represented in black, whereas the ones that are further away are represented in white. Arrows indicate the hydrophobic pocket entrance in two subunits, and the broken line ellipses indicate the active site in two subunits. Both molecular regions remain accessible for ligands in the tetramer.

1.2, a 36 kDa fragment, corresponding to thrombin or prothrombin 2, and a 24 kDa band corresponding to fragment 1 [26,27]. The same hydrolysis pattern was obtained in the presence or absence of prothrombinase complex components (Figure 5A).

Under reducing conditions, FXa in the presence of prothrombinase complex produces a fragment of 56 kDa corresponding to fragments 1.2 and A, indicating the meizothrombin formation. When the prothrombin hydrolysis induced by rLopap

Figure 4 SDS/PAGE analysis of purified rLopap

The recombinant protein, produced in *E. coli*, was purified with a Chelating Sepharose Fast Flow column (Amersham Biosciences). (A) SDS/PAGE (12.5 % gels) under reducing condition. Lane 1, molecular mass standards (sizes given in kDa); lane 2, purified rLopap. (B) HPLC analysis observed when using a 20–80 % acetonitrile (Acn) gradient in 0.1 % TFA for 20 min (1 ml/min flow rate) at room temperature. (C) Far-UV (190–260 nm) CD spectra of rLopap. The recombinant protein was submitted to CD analysis under appropriate conditions (20 mM Tris/HCl buffer, pH 7.4, at 25 °C) with different concentrations of urea (0, 3 and 6 M).

Table 1 Comparison of the percentage of secondary structures estimated from far-UV CD spectra of rLopap and of β -lactoglobulin

Analyses were conducted using the CDNN program. The CD spectrum of β -lactoglobulin was published by Qi et al. [40].

Protein	Temperature (°C)	pH	α -Helix (%)	β -Sheet (%)	β -Turn (%)	Random coil (%)
Native Lopap	25	7.4	6.7	55.9	16.5	27.5
rLopap	25	7.4	6.8	53.2	17.0	28.1
β -Lactoglobulin*	20–25	6.7	10	50.0	8.0	35.0

was compared with the hydrolysis by FXa, in the presence and absence of the prothrombinase complex components, we observed that rLopap produced fragments similar to those produced by the FXa in the absence of prothrombinase complex (Figure 5A).

Experiments performed in the presence of the inhibitors showed that prothrombin activator activity was abolished by PMSF (Figure 5B).

Figure 5 Prothrombin hydrolysis by rLopap

(A) Comparison between prothrombin hydrolysis by rLopap and FXa. Prothrombin (10 μ M) was incubated with rLopap and FXa (2 μ M) in the presence or absence of prothrombinase complex components for 1 h and analysed by SDS/PAGE (10% gels) under non-reduced (lanes 2–5) and reduced conditions (lanes 6–9). Lane *, molecular mass standards (sizes indicated in kDa); lane 1, purified prothrombin (PT); lane 2, PT + rLopap; lane 3, PT + rLopap + prothrombinase complex; lane 4, PT + FXa; lane 5, PT + FXa + prothrombinase complex; lane 6, PT + rLopap; lane 7, PT + rLopap + prothrombinase complex; lane 8, PT + FXa; lane 9, PT + FXa + prothrombinase complex. (B) Inhibition of rLopap prothrombin activation by 4 μ M PMSF (non-reducing SDS/PAGE analysis). Prothrombin (10 μ M) was incubated for 18 h at 37 °C with rLopap (2 μ M) pre-treated or not with PMSF and analysed by SDS/PAGE (10% gels). Lane *, molecular mass standards (sizes indicated in kDa); lane 1, prothrombin (PT); lane 2, PT + rLopap treated with PMSF; lane 3, PT + rLopap.

DISCUSSION

Lonomia obliqua is a caterpillar found mainly in the southern region of Brazil. The contact of its bristles with human skin causes a severe haemorrhagic syndrome due to a consumption coagulopathy triggered by procoagulating agents, including a prothrombin activator and an FX activator. Our findings strongly suggest that Lopap contributes to this syndrome through prothrombin activation, resulting in a consumption coagulopathy [28].

In the present study, we constructed a cDNA library from mRNA of *L. obliqua* bristles, and cloned the Lopap cDNA. Analysis of the Lopap sequence revealed that the protein is synthesized as 201 amino acid residues (Figure 1). The presence of an N-terminal signal peptide (16 amino acids) was established by comparison of the cloned sequence with the N-terminal sequence of the native protein. Furthermore, the cleavage site of the signal peptide agrees with the predictions made using the SignalP 3.0 Server.

Sequence analysis also indicated that Lopap is a member of the lipocalin family of proteins, since it presents identity of between 20 and 59% with other lipocalins, and contains three sequence signatures that define this protein family (Figure 2) [9]. The first motif found in the mature Lopap is the most highly conserved among lipocalin proteins: an invariant glycine (Gly²⁴), an invariant tryptophan (Trp²⁶) and an aromatic group (Tyr²⁷). Motif II presents the conserved triad Thr¹⁰⁴-Asp¹⁰⁵-Tyr¹⁰⁶. Motif III shows high similarity to other lipocalins, especially with regard to its first four amino acids, consisting of a hydrophobic residue (Ile¹³⁰), a polar hydrophilic residue (Ser¹³²) and a basic residue (Arg¹³³).

The secondary-structure analysis by CD measurements shows that, under physiological conditions, the contents of such structures in Lopap are comparable with those of β -lactoglobulin (Table 1).

Lipocalins unusually display low levels of overall sequence conservation, with pairwise sequence identities often falling below 20%, which is a threshold for reliable alignment. However, all lipocalins share sufficient similarities for a certain definition of family membership. In contrast with their low conservation at the

sequence level, analysis of available lipocalin crystal structures [29–34] shows that the overall folding pattern is highly conserved. The nature of this common structure is well-described [35,37,38], allowing us to obtain a structural model for the Lopap protein (Figure 3). The model presents the conserved C-terminal α -helix and a highly symmetrical β -barrel structure composed of an eight-stranded antiparallel β -sheet, closed back on itself, forming a continuous hydrogen-bonding network and resulting in a basket-like structure [9].

Mature rLopap protein fused to a His₆-tag was obtained, and the analysis of CD spectra indicated that the native protein, as well as the recombinant protein, have a predominantly β -structure, as expected by the structure modelling (Figure 4C).

In addition, stability studies showed that, after unfolding induced by urea, there was a remarkable increase in the α -helix content, and a decrease in the β -sheet content of Lopap, as can be seen in the CD spectra shown in Figure 4(C), and the deconvolution data presented in Table 1.

Three lipocalins have been described presenting enzymatic activity. The lipocalin-type PGD synthase (L-PGDS) is responsible for the biosynthesis of PGD₂, a potent endogenous sleep-inducing substance found in the central nervous system. The overall architecture of L-PGDS shows an eight-stranded β -barrel with a hydrophobic cavity, in which Cys⁶⁵ is part of the active site [12]. The other examples are the violaxanthin de-epoxidase and the zeaxanthin epoxidase which catalyse the interconversions between the carotenoids violaxanthin, antheraxanthin and zeaxanthin in plants. In these lipocalin structures, there is only one cysteine residue, which is predicted to be located in the loop structure between the first and second antiparallel β -strands of the barrel [13]. Nuclease activity of lipocalin members was also reported in [14].

So far, Lopap is the only member of the lipocalin family to be described that presents protease activity. Lipocalins that show enzymatic activity exhibit their active sites inside their barrel structure [12,13]. Using the CSA, a putative serine protease-like catalytic site was found outside the β -barrel of Lopap. The residues suggested to be responsible for this enzymatic activity lie in

the C-terminal α -helix (Glu¹⁶⁷, His¹⁶⁸ and Glu¹⁷¹) and in β 7 (Ser¹¹⁹). The glutamate residues that could be involved in the catalytic mechanism do not perfectly fit the serine protease model of the classic catalytic triad, aspartate, histidine and serine. However, further thought allows us to conclude that the glutamate residue (or residues, since two glutamate residues are in close proximity to His¹⁶⁸) that is replaced by an aspartate residue implies a conservative change. Furthermore, the inferred catalytic residues can reach sterical conformation that allows interactions between the acidic residues (Glu¹⁶⁷ and Glu¹⁷¹) and His¹⁶⁸, as well as the interactions of the latter with Ser¹¹⁹. In this case, the catalytic residues would be positioned at the bottom of a shallow and superficial cleft located in the opposite side of the hydrophobic pocket entrance. Interestingly, a similar serine protease-like catalytic site was also found in insecticynin in an alternative site of its tertiary structure. These results seem to indicate a new case of convergent evolution giving rise to serine protease-like enzymes belonging to the lipocalin family.

Moreover, two putative calcium-binding sites were predicted (using GG software) in opposite sides of Lopap's molecular structure. The first is located laterally to the β -barrel and the second is located in the loop between β 9 and the C-terminal α 1. Especially, the second calcium binding-site could influence the enzymatic activity by stabilizing the active site. Nevertheless, the experimental probing of catalytically relevant residues remains to be conducted using techniques such as site-directed mutagenesis. X-ray crystallography could also contribute valuable information. Our group has already initiated both approaches.

The Lopap tetrameric form presents an interaction between each monomer in such a way that the predicted catalytic site remains exposed to solvent and the hydrophobic pocket entrance remains accessible (Figure 3B), to ensure the molecule's functionality.

Under physiological conditions, prothrombin activation is catalysed by the prothrombinase complex, composed of FXa, FVa, phospholipids and calcium ions. Initially, the activated intermediate meizothrombin is formed due to cleavage of the Arg³²⁰-Ile³²¹ bond, followed by a second cleavage at the Arg²⁷¹-Thr²⁷² bond, that releases fragment 1.2 and leads to thrombin formation. Prothrombin can be activated by physiological concentrations of FXa in the absence of phospholipids, generating fragments of 52, 36, 24 and 17 kDa that correspond to fragment 1.2, prethrombin 2 or thrombin, fragment 1 and fragment 2. However, in the presence of the prothrombinase complex, this activation is more effective and the fragments generated are meizothrombin (72 kDa), thrombin (36 kDa), fragment 1 (24 kDa) and fragment 2 (17 kDa) [26,27].

Native Lopap is a tetrameric protein, of 69 kDa [5], while rLopap is expressed as an active 21 kDa protein. When the prothrombin activation of both proteins was compared, the native protein showed higher activity. Despite the fact that the chromatographic (HPLC) profile of rLopap results in a major peak, the lower activity of the recombinant protein could be explained by the diversity of folding intermediates as evidenced by light-scattering experiments (results not shown). In fact, alternative folding forms were already found in a single HPLC peak [36].

rLopap recognizes and hydrolyses the prothrombin (in concentrations similar to the human plasma), leading to an active thrombin generation which follows linear kinetics. It is also able to decrease the recalcification time (in normal plasma). Given these two findings, we conclude that its proteolytic activity is similar to that of the native Lopap [5,8].

The comparison of the hydrolysis products of prothrombin generated by the native Lopap with those generated by other pro-

thrombin activators [8] suggested the following mechanism of action: thrombin generation would start by prethrombin 2 formation, followed by thrombin generation by two consecutive hydrolyses. Preliminary studies indicated that rLopap is able to hydrolyse two fluorogenic substrates derived from the human prothrombin sequence (Phe²⁸⁰-Gly²⁸⁹ and Tyr³¹⁶-Ser³²⁵), leading to thrombin formation. This assumption is based on the fact that, apparently, meizothrombin is not formed from prothrombin by Lopap and that products with molecular masses similar to prethrombin 2 are generated.

The prothrombin fragments obtained by rLopap hydrolysis in the presence or absence of prothrombinase components (phospholipids, FVa and calcium) were of similar size: 52, 36 and 27 kDa. The migration patterns were the same in every experiment, suggesting that there is no difference in the prothrombin activation process by Lopap and rLopap acting as FXa in absence of prothrombinase components. Recently, Lilla et al. [39] described a protein with FXa-like activity with a molecular mass of approx. 20 kDa, but, unfortunately, they did not observe that the this protein sequence corresponds to Lopap, identified in GenBank® with accession number AY908986.

Figure 5 demonstrates the differences between prothrombin hydrolysis induced by FXa and rLopap in the presence of prothrombinase complex components. FXa generates meizothrombin as an intermediate fragment observed by a 56 kDa band (fragments 1.2 and A) under reducing conditions, differently from what was observed for rLopap.

When rLopap was pre-incubated with PMSF, prothrombin hydrolysis did not occur (Figure 5B), which demonstrates its protease nature. Taken together, these results corroborate the view that rLopap belongs to the lipocalin family, has a serine protease-like activity and acts on prothrombin, similar to FXa in the absence of prothrombinase components.

We thank Dr Beatriz L. Fernandes from the Center for Applied Toxinology (CAT/CEPID) for help with manuscript revision before resubmission. This work was supported by the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), CAT/CEPID and by COINFAR (Consórcio das Indústrias Farmacêuticas).

REFERENCES

- 1 Zannin, M., Lourenco, D. M., Motta, G., Dalla Costa, L. R., Grando, M., Gamborgi, G. P., Noguti, M. A. and Chudzinski-Tavassi, A. M. (2003) Blood coagulation and fibrinolytic factors in 105 patients with hemorrhagic syndrome caused by accidental contact with *Lonomia obliqua* caterpillar in Santa Catarina, southern Brazil. *Thromb. Haemostasis* **89**, 355–364.
- 2 Fritzen, M., Schattner, M., Ribeiro, A. L., Batista, I. F., Ventura, J., Prezoto, B. C. and Chudzinski-Tavassi, A. M. (2003) *Lonomia obliqua* venom action on fibrinolytic system. *Thromb. Res.* **112**, 105–110.
- 3 Donato, J. L., Moreno, R. A., Hyslop, S., Duarte, A., Antunes, E., Le Bonniec, B. F., Rendu, F. and de Nucci, G. (1998) *Lonomia obliqua* caterpillar spicules trigger human blood coagulation via activation of factor X and prothrombin. *Thromb. Haemostasis* **79**, 539–542.
- 4 Kelen, E. M. A., Picarelli, Z. P. and Duarte, A. C. (1995) Hemorrhagic syndrome induced by contact with caterpillars of the genus *Lonomia* (Saturniidae, Hemoleucinae). *J. Toxinol. Toxin* **14**, 283–308.
- 5 Reis, C. V., Portaro, F. C., Andrade, S. A., Fritzen, M., Fernandes, B. L., Sampaio, C. A., Camargo, A. C. and Chudzinski-Tavassi, A. M. (2001) A prothrombin activator serine protease from the *Lonomia obliqua* caterpillar venom (Lopap): biochemical characterization. *Thromb. Res.* **102**, 427–436.
- 6 Chudzinski-Tavassi, A. M., Schattner, M., Fritzen, M., Pozner, R. G., Reis, C. V., Lourenco, D. and Lazzari, M. A. (2001) Effects of Lopap on human endothelial cells and platelets. *Haemostasis* **31**, 257–265.
- 7 Sakai, M., Wu, C. and Suzuki, K. (2001) Nucleotide and deduced amino acid sequences of a cDNA encoding a lipocalin protein in the central nervous system of *Bombyx mori*. *J. Insect Biotechnol. Sericol.* **70**, 105–111.

- 8 Riley, C. T., Barbeau, B. K., Keim, P. S., Kezdy, F. J., Heinrikson, R. L. and Law, J. H. (1984) The covalent protein structure of insecticynin, a blue biliprotein from the hemolymph of the tobacco hornworm, *Manduca sexta* L. *J. Biol. Chem.* **259**, 13159–13165
- 9 Flower, D. R. (1996) The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14
- 10 Flower, D. R., North, A. C. and Sansom, C. E. (2000) The lipocalin protein family: structural and sequence overview. *Biochim. Biophys. Acta* **1482**, 9–24
- 11 Gutierrez, G., Ganfornina, M. D. and Sanchez, D. (2000) Evolution of the lipocalin family as inferred from a protein sequence phylogeny. *Biochim. Biophys. Acta* **1482**, 35–45
- 12 Irikura, D., Kumasaka, T., Yamamoto, M., Ago, H., Miyano, M., Kubata, K. B., Sakai, H., Hayaishi, O. and Urade, Y. (2003) Cloning, expression, crystallization, and preliminary X-ray analysis of recombinant mouse lipocalin-type prostaglandin D synthase, a somnogen-producing enzyme. *J. Biochem. (Tokyo)* **133**, 29–32
- 13 Hieber, A. D., Bugos, R. C. and Yamamoto, H. Y. (2000) Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim. Biophys. Acta* **1482**, 84–91
- 14 Yusifov, T. N., Abduragimov, A. R., Gasyimov, O. K. and Glasgow, B. J. (2000) Endonuclease activity in lipocalins. *Biochem. J.* **347**, 815–819
- 15 Junqueira de Azevedo, I. L., Farsky, S. H., Oliveira, M. L. and Ho, P. L. (2001) Molecular cloning and expression of a functional snake venom vascular endothelium growth factor (VEGF) from the *Bothrops insularis* pit viper: a new member of the VEGF family of proteins. *J. Biol. Chem.* **276**, 39836–39842
- 16 Ramos, C. R., Abreu, P. A., Nascimento, A. L. and Ho, P. L. (2004) A high-copy T7 *Escherichia coli* expression vector for the production of recombinant proteins with a minimal N-terminal His-tagged fusion peptide. *Braz. J. Med. Biol. Res.* **37**, 1103–1109
- 17 Risler, J. L., Delorme, M. O., Delacroix, H. and Henaut, A. (1988) Amino acid substitutions in structurally related proteins: a pattern recognition approach: determination of a new and efficient scoring matrix. *J. Mol. Biol.* **204**, 1019–1029
- 18 Gouet, P., Courcelle, E., Stuart, D. I. and Metz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308
- 19 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815
- 20 Sali, A. (1995) Comparative protein modeling by satisfaction of spatial restraints. *Mol. Med. Today* **1**, 270–277
- 21 Laskowski, R. A. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
- 22 Bowie, J. U., Luthy, R. and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170
- 23 Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* **8**, 52–56
- 24 Deng, H., Chen, G., Yang, W. and Yang, J. J. (2006) Predicting calcium-binding sites in proteins: a graph theory and geometry approach. *Proteins* **64**, 34–42
- 25 Reference deleted
- 26 Fischer, B. E., Schlokot, U., Mitterer, A., Grillberger, L., Reiter, M., Mundt, W., Dorner, F. and Eibl, J. (1996) Differentiation between proteolytic activation and autocatalytic conversion of human prothrombin: activation of recombinant human prothrombin and recombinant D419N-prothrombin by snake venoms from *Echis carinatus* and *Oxyuranus scutellatus*. *Protein Eng.* **9**, 921–926
- 27 Brufatto, N. and Nesheim, M. E. (2003) Analysis of the kinetics of prothrombin activation and evidence that two equilibrating forms of prothrombinase are involved in the process. *J. Biol. Chem.* **278**, 6755–6764
- 28 Reis, C. V., Farsky, S. H., Fernandes, B. L., Santoro, M. L., Oliva, M. L., Mariano, M. and Chudzinski-Tavassi, A. M. (2001) *In vivo* characterization of Lopap, a prothrombin activator serine protease from the *Lonomia obliqua* caterpillar venom. *Thromb. Res.* **102**, 437–443
- 29 Bocskai, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavaggioni, A., Findlay, J. B. and North, A. C. (1992) Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature (London)* **360**, 186–188
- 30 Cowan, S. W., Newcomer, M. E. and Jones, T. A. (1990) Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins* **8**, 44–61
- 31 Holden, H. M., Rypniewski, W. R., Law, J. H. and Rayment, I. (1987) The molecular structure of insecticynin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J.* **6**, 1565–1570
- 32 Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H. (1987) Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol.* **198**, 499–513
- 33 Monaco, H. L. and Zanotti, G. (1992) Three-dimensional structure and active site of three hydrophobic molecule-binding proteins with significant amino acid sequence similarity. *Biopolymers* **32**, 457–465
- 34 Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C., Findlay, J. B., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. and Kraulis, P. J. (1986) The structure of β -lactoglobulin and its similarity to plasma retinol-binding protein. *Nature (London)* **324**, 383–385
- 35 Flower, D. R., North, A. C. and Attwood, T. K. (1993) Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* **2**, 753–761
- 36 Rosenfeld, R. D., Noone, N. M., Lauren, S. L., Rohde, M. F., Narhi, L. O. and Arakawa, T. (1993) Mutation of Arg^{55/56} to Leu⁵⁵/Ala⁵⁶ in insulin-like growth factor-I results in two forms different in disulfide structure and native conformation but similar under reverse-phase conditions. *J. Protein Chem.* **12**, 247–254
- 37 Chikenji, G. and Kikuchi, M. (2000) What is the role of non-native intermediates of β -lactoglobulin in protein folding? *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14273–14277
- 38 Kuwata, K., Hoshino, M., Era, S., Batt, C. A. and Goto, Y. (1998) $\alpha \rightarrow \beta$ transition of β -lactoglobulin as evidenced by heteronuclear NMR. *J. Mol. Biol.* **283**, 731–739
- 39 Lilla, S., Pereira, R., Hyslop, S., Donato, J. L., Le Bonniec, B. F. and de Nucci, G. (2005) Purification and initial characterization of a novel protein with factor Xa activity from *Lonomia obliqua* caterpillar spicules. *J. Mass Spectrom.* **40**, 405–412
- 40 Qi, X. L., Holt, C., McNulty, D., Clarke, D. T., Brownlow, S. and Jones, G. R. (1997) Effect of temperature on the secondary structure of β -lactoglobulin at pH 6.7, as determined by CD and IR spectroscopy: a test of the molten globule hypothesis. *Biochem. J.* **324**, 341–346

Received 25 February 2005/18 May 2006; accepted 31 May 2006

Published as BJ Immediate Publication 31 May 2006, doi:10.1042/BJ20060325