Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# The structure of leech anti-platelet protein, an inhibitor of haemostasis

**Eric G. Huizinga,[a,b] Arie Schouten,[b] Thomas M. Connolly,[c] Jan Kroon,[b] Jan J. Sixma[a] and Piet Gros[b]***

[a]Thrombosis and Haemostasis Laboratory, Department of Haematology, Institute of Biomembranes, University Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands, [b]Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, and [c]Department of Biological Chemistry, Merck Sharp and Dohme Research Laboratories, West Point, Pennsylvania 19486, USA

Correspondence e-mail: p.gros@chem.uu.nl

Received 13 January 2001
Accepted 3 May 2001

**PDB Reference:** leech anti-platelet protein, 1i8n.

Leech anti-platelet protein (LAPP) from the leech *Haementeria officinalis* is a collagen-binding protein that inhibits the collagen-mediated adhesion of blood platelets. The crystal structure of recombinant LAPP has been determined using single isomorphous replacement with anomalous scattering combined with solvent flattening and threefold molecular averaging. The model of LAPP has been refined to 2.2 Å resolution ($R$ factor 21.5%; free $R$ factor 24.0%). LAPP contains an 89-residue C-terminal domain consisting of a central six-stranded antiparallel $\beta$-sheet flanked on one side by an $\alpha$-helix and on the other side by two extended loops with little secondary structure. A 36-residue N-terminal region is not visible in the electron-density map. This region is rich in glycine and lacks hydrophobic residues. It probably does not have a compact globular fold, but instead has an extended conformation and is flexible. The crystal packing suggests that LAPP may form tightly interacting dimers. The fold of the C-terminal domain of LAPP closely resembles that of the N-domain of hepatocyte growth factor (HGF), which classifies LAPP as a PAN domain. However, no significant sequence homology exists between LAPP and other PAN domains. Common structural features between LAPP and the HGF N-domain include two disulfide bonds that link the $\alpha$-helix to the central region of the protein and five residues with a conserved hydrophobic nature that are located in the core of the domain. These conserved structural features may be an important determinant of the PAN-domain type of fold.

## 1. Introduction

The saliva of leeches contains many factors that block haemostasis and help to maintain a fluid blood meal. One of these factors is leech anti-platelet protein (LAPP) isolated from the leech *H. officinalis* (Connolly *et al.*, 1992). LAPP is a 13.6 kDa protein with no detectable sequence homology to any other known protein sequence (Keller *et al.*, 1992). LAPP interferes with haemostasis by blocking the adhesion of platelets to fibrillar collagen, an important first step in the arrest of bleeding (Van Zanten *et al.*, 1995). LAPP inhibits platelet adhesion on collagen types I, III and IV with $IC_{50}$ values of 70, 600 and 90 n$M$, respectively. Calin, a 65 kDa protein isolated from the leech *Hirudo medicinalis*, has a similar activity to LAPP, but the difference in molecular weight suggests that these proteins are not homologous (Munro *et al.*, 1991; Depraetere *et al.*, 1999). Compounds with activities similar to LAPP and calin are potentially useful as antithrombotic agents, as inhibition of platelet adhesion to collagen may prevent thrombotic complications resulting from atherosclerosis (Deckmyn *et al.*, 1995).

The adhesion of platelets to collagen is mediated by receptors located on the platelet surface (for a review, see Barnes *et al.*, 1998). Important adhesive receptors are integrin $\alpha_2\beta_1$ and glycoprotein Ib. Integrin $\alpha_2\beta_1$ binds to collagen *via* its I-domain, which is located in the $\alpha$-chain. The I-domain recognizes the collagen sequence GFOGER, where O represents hydroxyproline (Knight *et al.*, 1998). Glycoprotein Ib interacts with collagen *via* a multimeric plasma glycoprotein called von Willebrand Factor (vWF). The collagen-binding site of vWF is located in its A3 domain. The collagen sequence recognized by A3 has not yet been established. Both the vWF-A3 domain and the $\alpha_2$-I domain belong to the vWF A-type domain family. The crystal structures of the vWF-A3 domain (Huizinga *et al.*, 1997; Bienkowska *et al.*, 1997) and of the $\alpha_2$-I domain (Emsley *et al.*, 1997) reveal a dinucleotide-binding fold. Although sharing the same fold, the $\alpha_2$-I domain and the vWF-A3 domain have their collagen-binding sites located on different faces of the domain. The $\alpha_2$-I domain contains a metal-ion dependent collagen-binding site in its top face (Emsley *et al.*, 2000), while the collagen-binding site of vWF-A3 is located close to the bottom face of the domain (Romijn *et al.*, 2001).

Recent data from our laboratory indicate that a G*X*R motif in collagen is involved in the primary binding site of LAPP (Verkleij *et al.*, 1999). *X* can be a number of different amino acids, but not glutamate, as LAPP does not bind a peptide containing the $\alpha_2\beta_1$ recognition sequence GFOGER (Verkleij, 1999). Therefore, LAPP and $\alpha_2\beta_1$ (and perhaps also von Willebrand factor) bind to different collagen sequences. How LAPP blocks collagen binding of $\alpha_2\beta_1$ without binding to its primary recognition sequence remains to be elucidated.

Here, we present the 2.2 Å structure of recombinant LAPP. The structure of LAPP together with the structures of the vWF-A3 domain and the $\alpha_2$-I domain may provide a basis for understanding the mechanism underlying the inhibition of platelet adhesion by LAPP.

## 2. Experimental procedures

### 2.1. Crystallization

Recombinant LAPP was expressed in the yeast *Saccharomyces cerevisae* as previously described (Keller *et al.*, 1992). The protein was purified according to Connolly *et al.* (1992).

For crystallization trials, the freeze-dried protein was dissolved in water at a concentration of approximately 10 mg ml$^{-1}$ and dialysed against 10 m*M* sodium phosphate pH 7.0. Crystallization experiments were performed by the hanging-drop method at 277 K. Crystals were first observed in drops equilibrated against 20%(*v*/*v*) 2-propanol, 0.1 *M* sodium acetate pH 4.5. These crystals were extremely fragile and temperature sensitive. Stabler crystals were obtained by equilibration against 40%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD), 0.1 *M* sodium citrate pH 4.3. Preliminary diffraction studies indicated that the crystals belong to the space group $P4_122$ or $P4_322$. The unit-cell parameters were $a = b = 73.8$, $c = 200.6$ Å. Assuming a specific protein volume of 0.74 cm$^3$ g$^{-1}$, the asymmetric unit is likely to contain three or four protein molecules. This would correspond to a $V_M$ of 3.36 or 2.53 Å$^3$ Da$^{-1}$ (Matthews, 1968) and a solvent content of 63.4 or 51.4%, respectively.

**Table 1**
Diffraction data.

Values in parentheses are for data in high-resolution shells: native 1, 2.95–2.90 Å; $K_2PtCl_4$ derivative, 3.16–3.10 Å; native 2, 2.23–2.20 Å resolution.

| Data set | Native 1 | $K_2PtCl_4$ derivative | Native 2 |
|---|---|---|---|
| Resolution (Å) | 20.0–2.9 | 20.0–3.1 | 20.0–2.2 |
| Unit-cell parameters (Å) | | | |
| *a*, *b* | 73.79 | 73.68 | 73.45 |
| *c* | 200.61 | 199.25 | 199.49 |
| $R_{sym}$ | 0.085 (0.287) | 0.110 (0.305) | 0.037 (0.253) |
| $I > 3\sigma(I)$ (%) | 78.6 (52.0) | 85.7 (67.9) | 81.4 (53.5) |
| Multiplicity | 4.1 | 5.9 | 4.4 |
| No. of unique reflections | 12549 | 10659 | 27864 |
| Completeness (%) | 96.6 (96.2) | 95.6 (98.9) | 97.1 (96.5) |
| Mosaicity (°) | 0.39 | 0.61 | 0.37 |

For heavy-atom derivative searches crystals of LAPP were transferred to synthetic mother liquor containing 45%(*v*/*v*) MPD and 50 m*M* sodium succinate pH 4.3. One isomorphous heavy-atom derivative was obtained by soaking a crystal for 20 h in 1 m*M* $K_2PtCl_4$.

### 2.2. X-ray data collection and processing

For X-ray data collection, crystals were cooled to a temperature of 120 K using a Cryostream cooler (Oxford Cryosystems, Oxford, England). Native crystals were cryocooled directly from the drop. Heavy-atom-soaked crystals were cryocooled in the soaking solution. One crystal was used for each data set. A native data set to 2.9 Å resolution and a derivative data set to 3.1 Å resolution were collected in house using Cu $K\alpha$ radiation from a FR571 rotating-anode generator (Nonius, Delft, The Netherlands) equipped with a DIP2020 image-plate detector (MacScience, Japan). To optimize the anomalous signal the derivative data frames were collected at a crystal rotation angle $\varphi$ and $\varphi + 180°$. A 2.2 Å native data set was collected at the EMBL Outstation (DESY, Hamburg) on beamline BW7B using a wavelength of 0.8825 Å and a large MAR Research image-plate detector (X-ray Research GmbH, Hamburg, Germany). For X-ray data reduction and scaling the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997) were used, respectively. Statistics of the X-ray data collection may be found in Table 1.

### 2.3. Structure determination

A self-rotation function was calculated with the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994) using data in the resolution range 4–7 Å. It revealed the presence of a non-crystallographic twofold axis.

Three heavy-atom sites were identified in the $K_2PtCl_4$ derivative using the program *RSPS* (Collaborative Computational Project, Number 4, 1994). Refinement of the heavy-atom parameters and all subsequent density-modification calculations were performed with the *PHASES* program

**Table 2**
Phase-calculation statistics.

| | Overall | In resolution shells | | | | |
|---|---|---|---|---|---|---|
| Average resolution of shell (Å) | — | 7.3 | 4.6 | 3.9 | 3.5 | 3.2 |
| No. of reflections | 10352 | 2070 | 2070 | 2070 | 2070 | 2072 |
| Figure of merit | 0.46 | 0.60 | 0.53 | 0.46 | 0.39 | 0.31 |
| Isomorphous phasing power† | 2.70 | 2.80 | 2.71 | 2.64 | 2.64 | 2.78 |
| Anomalous phasing power† | 1.27 | 1.76 | 1.56 | 1.27 | 1.02 | 0.85 |

† Phasing power is the ratio of the root-mean-square heavy-atom structure-factor amplitude to the root-mean-square lack of closure.

**Table 3**
Refinement statistics.

| | |
|---|---|
| Resolution limits (Å) | 20–2.2 |
| *R* factor (%) | 21.5 |
| Free *R* factor (%) | 24.0 |
| No. of reflections used | 27864 |
| No. of protein atoms | 2148 |
| No. of solvent atoms | 190 |
| R.m.s. deviations in *B* factors (Å$^2$) | |
| Main-chain bonds | 1.4 |
| Main-chain angles | 2.5 |
| Side-chain bonds | 1.8 |
| Side-chain angles | 2.7 |
| NCS-related atoms | 2.6 |
| Average *B* protein (range) (Å$^2$) | 41.0 (24.2–77.0) |
| Average *B* solvent (range) (Å$^2$) | 46.4 (24.5–73.8) |
| Root-mean-square deviations from target | |
| Bond lengths (Å) | 0.017 |
| Bond angles (°) | 1.9 |
| Dihedral angles (°) | 27.2 |
| Improper angles (°) | 1.2 |

package (Furey & Swaminathan, 1998). The derivative was of excellent quality as is illustrated by its isomorphous phasing power of about 2.7 over the whole resolution range (Table 2). The phasing power of the anomalous data was 1.26 on average and falls below 1.0 around 3.5 Å resolution. Solvent flattening was performed using a solvent fraction of 55% and an integration radius of 8 Å. The solvent-flattened map clearly resolved the protein and solvent regions. Combined with the known heavy-atom positions, the solvent-flattened map suggested that the asymmetric unit contains three LAPP molecules. Two tightly interacting molecules form a dimer with twofold non-crystallographic symmetry (NCS) and a third molecule forms a similar dimer with a molecule related by twofold crystallographic symmetry. The non-crystallographic dimer and the crystallographic dimer are in turn related by a second local twofold axis. In the self-rotation function only one twofold axis was observed, as space-group symmetry causes both peaks to coincide. For electron-density averaging, the orientation of both non-crystallographic twofold axes was taken from the self-rotation function; their positions were defined as the midpoint between appropriate twofold-related heavy-atom sites. A mask was generated using the programs *O* (Jones *et al.*, 1991) and *MAMA* (Kleywegt & Jones, 1994). A dimer mask was used because it was not possible to discern the boundary between the individual molecules within a dimer. A mask constructed around the non-crystallographic dimer was rotated to obtain a mask around the crystallographic dimer. The NCS operators were refined with the *LSQROT* program from the *PHASES* package. Using the refined NCS operators, a total of 48 rounds of density averaging and solvent flattening were performed. Averaging consisted of two steps. Firstly, the density within the non-crystallographic dimer was averaged. Next, the averaged density of the non-crystallographic dimer was averaged with the density of the crystallographic dimer. In rounds 1–16 phased reflections in the resolution range 20.0–3.1 Å were used. During cycles 17–32 calculated phases were gradually included for unphased reflections in the resolution range 20.0–3.1 Å. During cycles 33–48 phases were extended from 3.1 to 2.9 Å. After the 16th and the 32nd cycle a new solvent mask was calculated. After each cycle phases were combined with the original heavy-atom phases. The electron-density correlation within the non-crystallographic dimer was 0.66 before and 0.86 after density modification. These values are 0.63 and 0.84, respectively, when the electron density of the non-crystallographic dimer is compared with that of the crystallographic dimer. Density modification improved the electron-density map significantly (Fig. 1). Large segments of secondary structure could be identified readily. The presence of an $\alpha$-helix allowed the identification of the space group as $P4_322$.

Continuous main-chain density was observed for residues 37–124 and the large majority of side chains had clear density except for Glu44, Lys54, Lys73, Glu77, Val80 and Glu81. These side chains were modelled as alanines. No density was observed for residues 1–36.

### 2.4. Structure refinement

Crystallographic refinement was performed with the program *CNS* (Brunger *et al.*, 1998) using torsion-angle dynamics (Rice & Brünger, 1994), maximum-likelihood refinement (Pannu & Read, 1996) and bulk-solvent correction. Throughout structure refinement 7.5% of the data were used for cross-validation (Brünger, 1992). The starting model had an *R* factor of 46.4% and a free *R* factor of 46.1%. In the first refinement cycle the model was optimized with respect to the low-resolution 2.9 Å native data set and phases obtained from density modification. Strict non-crystallographic symmetry was imposed. This resulted in an *R* factor of 30.8% and a free *R* factor of 31.0%. During a second round of refinement phase restraints were omitted, yielding an *R* factor of 27.5% and a free *R* factor of 28.6%. For subsequent refinement the high-resolution 2.2 Å native data set was used. To compensate for differences in unit-cell parameters between the low- and high-resolution data sets rigid-body refinement was performed. This reduced the *R* factor from 53.8 to 40.5% and the free *R* factor from 52.9 to 40.1%. One round of simulated-annealing refinement reduced these values to 34.0 and 35.9%, respectively. Subsequently, five rounds of model building with *O* and refinement with *CNS* were carried out. During this process no density showed up for residues 1–36. A peak in the difference electron density was interpreted as a water if it had a height of at least 3.0$\sigma$ and a hydrogen-bonding partner at a distance between 2.6 and 3.2 Å. Waters were

removed from the model if the $B$ factor refined to a value higher than 75 Å$^2$. During refinement, harmonic NCS restraints were imposed with an energy constant of 1250 kJ mol$^{-1}$ Å$^{-2}$. In the final stages of refinement a continuous feature of electron density that was observed close to all three LAPP models was tentatively identified as an asparagine side chain. This density is located at an 8 Å distance from the N-terminus of the model, but is not connected to it. The atomic coordinates of LAPP and the structure factors used for refinement are available from the PDB under accession code 1i8n.

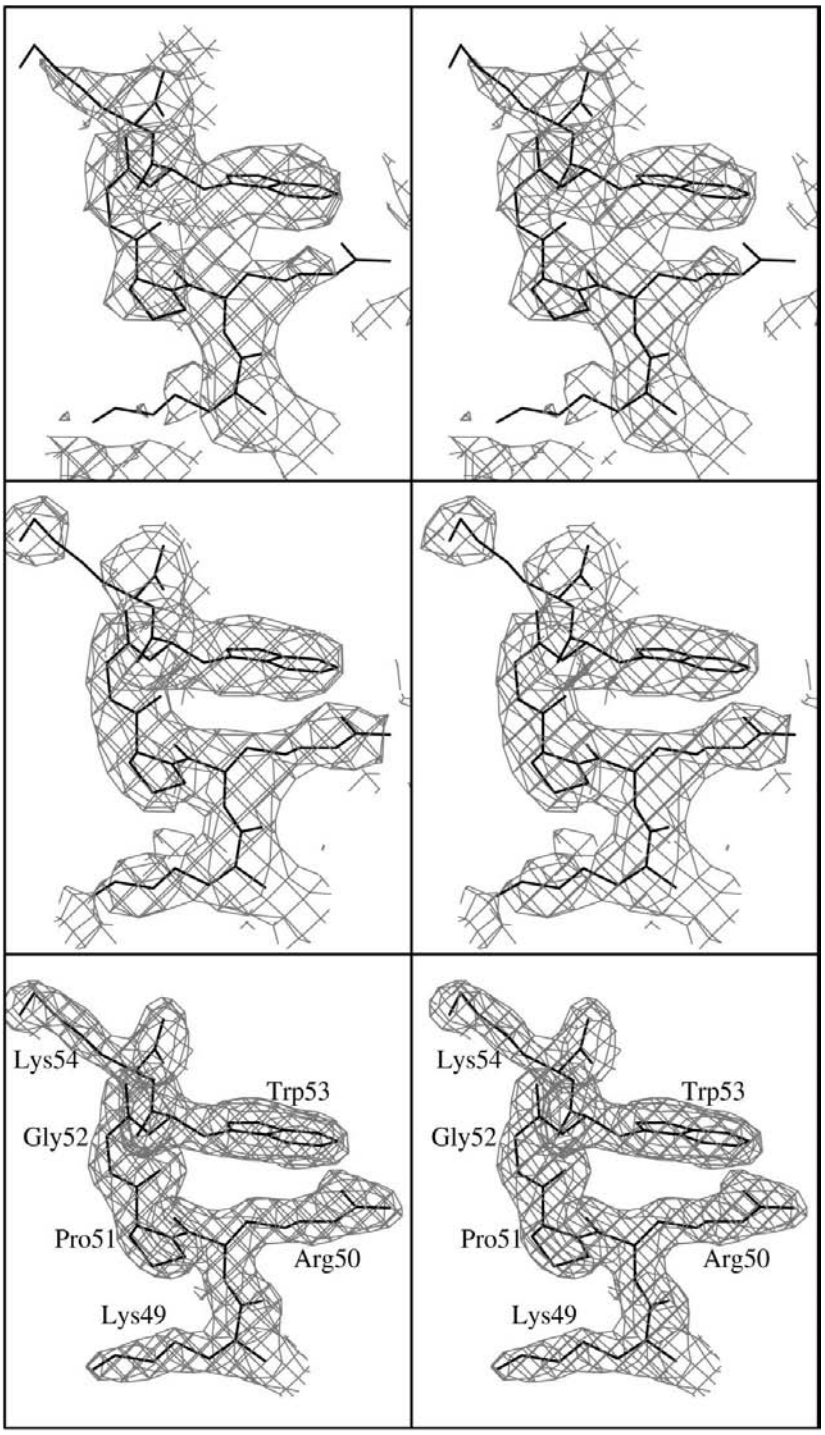


**Figure 1**
Electron density for residues 49–54 of the LAPP molecule $A$ at different stages of the structure determination. The electron densities shown were calculated from FOM-weighted $F_o$ and SIRAS phases (top), FOM-weighted $F_o$ and phases obtained after iterative density modification (middle) and $\sigma_A$-weighted $F_o$ and phases calculated from the refined model (bottom). The maximal resolution of the data used in map calculation was 3.1, 2.9 and 2.2 Å, respectively.

## 3. Results and discussion

### 3.1. Quality of the refined model

The refined model of LAPP has an $R$ factor of 21.5% and a free $R$ factor of 24.0% for data between 20.0 and 2.2 Å resolution. The cross-validated coordinate error as calculated by the program *CNS* using the *SIGMAA* procedure is 0.35 Å. The model has good stereo-geometry (Table 3). In the Ramachandran plot created by *PROCHECK*, (Laskowski *et al.*, 1993) 92.4% of all residues are located in the favoured regions and 7.6% in the additionally allowed regions. No residues are located in the generously allowed or disallowed regions.

The model consists of three LAPP molecules, each comprising residues 37–125 and 190 waters. Protein atoms were subjected to three-fold NCS restraints during refinement. Because of significant deviations from NCS, no restraints were imposed on residues 79–82, residues 111–112 and the side-chain atoms of residues 57, 62 and 95. Independent refinement of the NCS-related molecules did not seem to be justified as this increased the free $R$ factor to 24.9%, although it improved the $R$ factor to 20.9%.

A significant part of the protein chain including 36 residues at the N-terminus and one residue at the C-terminus is not visible in the electron-density map. The missing N-terminal region is discussed in more detail below. The average $B$ value of all protein atoms in the model is 41.6 Å$^2$. High average $B$ factors are found at the termini and in a surface loop comprising residues 78–82. This loop has rather weak electron density and was difficult to model; its conformation should therefore be considered tentative. Pro78, which is part of this loop, has been modelled in a *cis* conformation in molecules $A$ and $B$ and in a *trans* conformation in molecule $C$. Side-chain conformations of four more residues are ill defined. These include Glu37 and Glu44 for which no side-chain atoms have been modelled, Lys73 that has been modelled up to atom C$^{\gamma}$ and the side chain of Thr40 that may have

multiple conformations as judged from its electron density. In view of the resolution limit of the diffraction data, only one conformation of Thr40 was modelled. Apart for the exceptions described above the structure has well defined electron density (Fig. 1).

### 3.2. Overall protein structure

The core of LAPP consists of a six-stranded antiparallel $\beta$-sheet (Fig. 2). On one side this $\beta$-sheet is flanked by a two-and-a-half turn $\alpha$-helix and on the other side by two long loops. The $\alpha$-helix is connected to the core of the protein by disulfide bonds between Cys71 and Cys96 and between Cys75 and Cys85. A third disulfide bond is formed by Cys46 and Cys124.

The secondary-structure assignment for LAPP is shown in Fig. 3. The strands in the central $\beta$-sheet are arranged in the spatial order $\beta0$–$\beta1$–$\beta5$–$\beta3$–$\beta4$–$\beta2$. The connections between antiparallel strands ($\beta0$–$\beta1$, $\beta3$–$\beta4$) are rather short. The connections between parallel strands ($\beta1$–$\beta2$, $\beta2$–$\beta3$ and $\beta4$–$\beta5$) form the structural elements that cover either side of the central $\beta$-sheet. The $\beta2$–$\beta3$ connection forms the $\alpha$-helix. The $\beta1$–$\beta2$ and $\beta4$–$\beta5$ connections form two long loops that lie on the face of the $\beta$-sheet opposite to the $\alpha$-helix. Each of these loops contains a single turn of $3_{10}$-helix and a short stretch of residues with an extended conformation. The extended regions referred to as $\beta1'$ and $\beta4'$ in Figs. 2 and 3 run antiparallel and are connected by one main-chain hydrogen bond, thus forming a small $\beta$-sheet-like structure. The central $\beta$-sheet curves around the $\alpha$-helix, which interacts with residues from all strands except $\beta0$.

### 3.3. The N-terminal region

The 36 N-terminal residues of LAPP are not visible in the electron-density map. We have considered the possibility that this part of the molecule has been lost owing to proteolysis. However, sequencing of the material used for crystallization showed that the N-terminal residues are Gln-Asp-Glu-Asp-Ala as is expected for the intact protein.

To exclude the possibility that weak electron density of the N-terminus had disappeared as a consequence of solvent flattening and NCS averaging, the SIRAS-phased map was carefully inspected, but additional density was not observed.

In the map of the refined model, some additional electron density was observed close to the N-termini of all three LAPP molecules in the asymmetric unit. When this feature was modelled as a group of water molecules it gave rise to very short interatomic distances. Its position inside a pocket at the molecular surface close to the N-terminus of the model suggested that it might be a side chain originating from a residue in the N-terminal region of LAPP. Based on its shape and chemical environment, this density was tentatively modelled as an asparagine side chain.

The N-terminal region of LAPP that is invisible in the electron density has a peculiar amino-acid composition (Fig. 3). Of 36 residues, 11 are glycines and 12 are aspartates and glutamates; the absence of positively charged residues

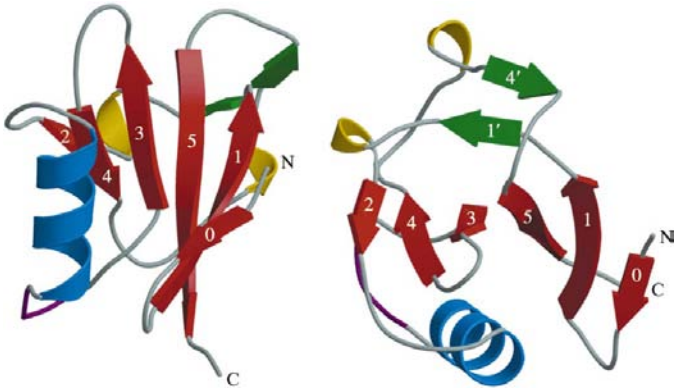


**Figure 2**
Ribbon representation of LAPP in two approximately perpendicular views. The N- and C-termini are labelled. $\beta$-strands are shown in red, $\alpha$-helices in blue and $3_{10}$-helices in yellow. Depicted in green are two stretches of extended residues that form a small $\beta$-sheet-like structure but are connected by only one main-chain hydrogen bond. A mobile loop comprising residues 78–82 is shown in magenta. Labelling of strands in the central $\beta$-sheet conforms to the labelling used for the NMR structure of hepatocyte growth factor (Zhou *et al.*, 1998). [Figure was drawn using *MOLSCRIPT* (Kraulis, 1991) and *RENDER* (Merritt & Bacon, 1997).]

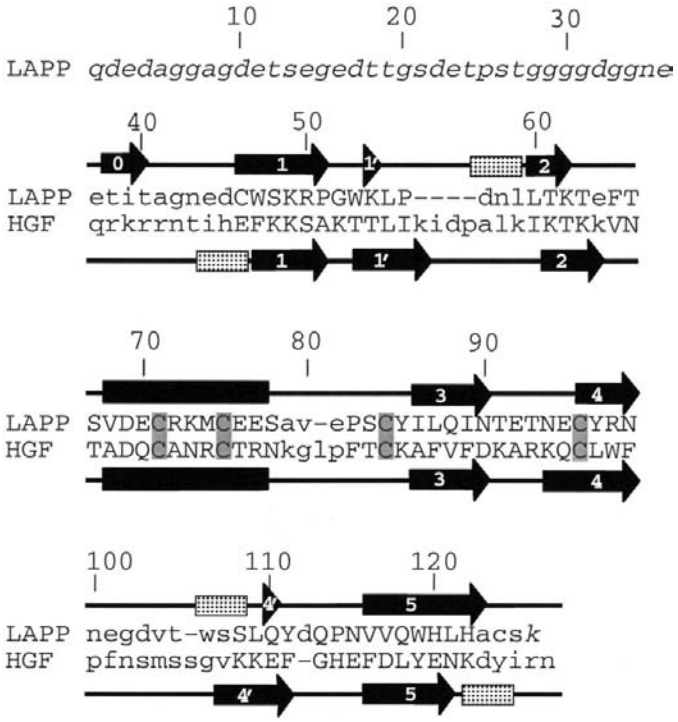


**Figure 3**
Secondary-structure assignment of LAPP as determined with the program *DSSP* (Kabsch & Sander, 1983) and structure-based sequence alignment of LAPP (SWISS-PROT Q01747) with the N-domain of HGF (SWISS-PROT P14210; PDB code 1bht). Residue numbers refer to LAPP. Residues of LAPP printed in italics are not visible in the electron-density map; this includes 36 N-terminal residues printed on the top line. No corresponding N-terminal residues are present in HGF. Lower-case letters denote residues whose $C^{\alpha}$ atoms cannot be superimposed within 3.5 Å. $\alpha$-helices are indicated by a black bar, $3_{10}$-helices by a dotted bar and $\beta$-strands by arrows. The triangles denote residues involved in an isolated $\beta$-bridge.

give it a large overall negative charge. The complete absence of large hydrophobic residues excludes the possibility that the N-terminus forms a compact globular domain. In view of its amino-acid composition, it is more likely to have an extended conformation. Because of the high solvent content (63%) there is ample room in the crystal for the N-terminal region.

### 3.4. Crystal packing

Crystals of LAPP can be described as consisting of LAPP dimers with internal twofold symmetry. The asymmetric unit contains three molecules contributed by one complete dimer and one dimer positioned on a crystallographic twofold axis.

The intradimer interface is the most extensive area of molecular interaction in the crystal and involves a prominent stacking of tryptophan side chains (Fig. 4). In the non-crystallographic dimer this interaction buries an accessible surface area of 1360 Å$^2$ and involves 49 interatomic distances shorter then 3.5 Å. In the crystallographic dimer these values are 1490 Å$^2$ and 50, respectively. The second tightest interaction occurs close to the local twofold axis that relates two LAPP dimers. In this interface strands $\beta$0 from adjacent molecules run antiparallel and are hydrogen bonded *via* the main-chain O and N atoms of Thr40. The surface area buried in this interface is about 1030 Å$^2$ and the number of contacts shorter than 3.5 Å is 28, which is about half of that observed for the intradimer surface. Other crystal packing interactions are significantly less extensive, involving buried surfaces of 500 Å$^2$ or less.

The extensive intradimer interface that is observed in crystals of LAPP suggests that the dimer may be relatively stable and could also exist in solution. Analytical ultracentrifuge experiments (T. M. Connolly, unpublished results) showed however that LAPP is monomeric at neutral pH.

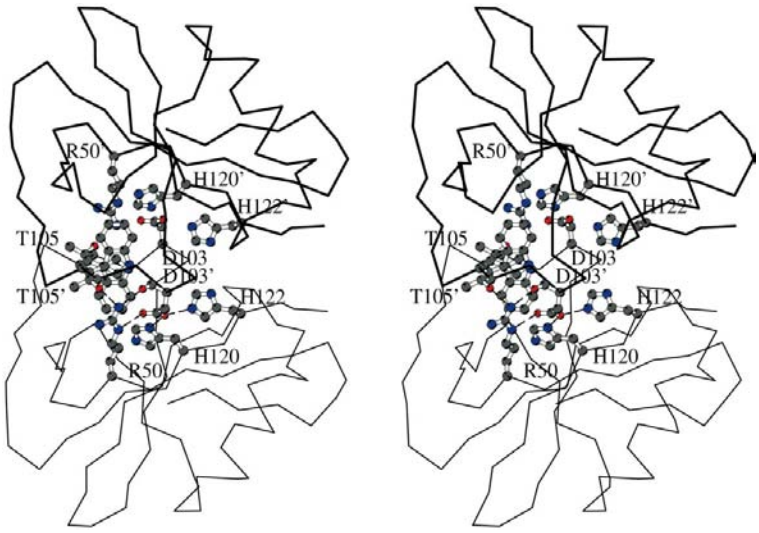


**Figure 4**
Stereographic representation of a tightly interacting LAPP dimer. Residues in the dimer interface are shown in ball-and-stick representation. Hydrogen bonds and salt bridges are indicated by dashed lines for half of the dimer interface. A third LAPP molecule in the asymmetric unit (not shown) forms a similar dimer with a molecule that is related by a crystallographic twofold axis.

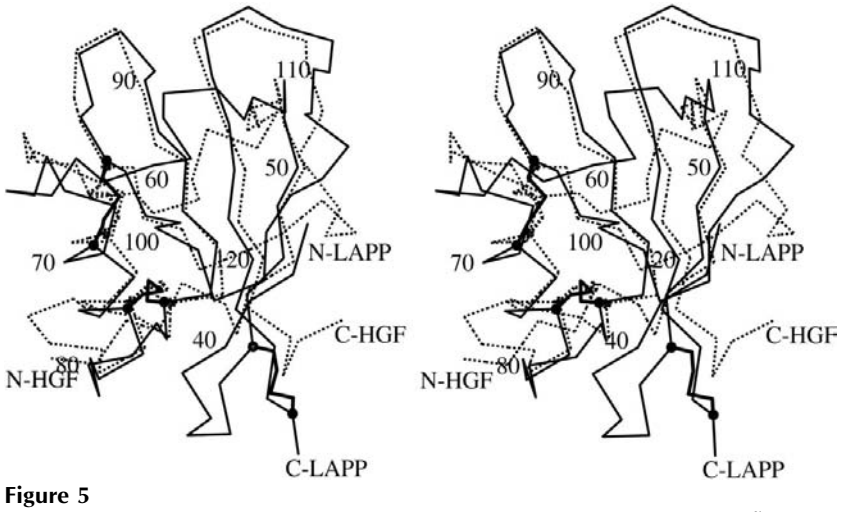


**Figure 5**
Superposition of LAPP and the N-domain from HGF (PDB code 1bht). The C$^\alpha$ trace of LAPP is shown as a solid line and that of HGF as a dashed line. Disulfide bonds are shown in slightly thicker lines. Spheres indicate C$^\alpha$ positions of cysteine residues. The N- and C-termini are labelled. Residue numbering refers to LAPP.

### 3.5. Structural comparison with the N domain of HGF

A search for LAPP-related folds with the program *DALI* (Holm & Sander, 1993) identified the structure of the N-domain of human hepatocyte growth factor (HGF) with a *Z* score of 8.1. HGF is the prototype of a family of growth factors that resemble the blood proteinase plasminogen in sequence, domain organization and mechanism of activation. The N-domain is located at the N-terminus of the 69 kDa $\alpha$-chain of HGF and is followed by four kringle domains. An NMR structure of the isolated human HGF N-domain (Zhou *et al.*, 1998) and two crystal structures of the N-domain together with the first kringle domain (Chirgadze *et al.*, 1999; Ultsch *et al.*, 1998) have been determined. A superposition of the N-domain of HGF and LAPP (Fig. 5) shows that the overall folds are very similar indeed. The r.m.s. fit for 69 aligned C$^\alpha$ positions is 2.0 Å. The best conserved parts are the central $\beta$-sheet and the $\alpha$-helix. When the superposition is limited to this region of the molecule, the r.m.s. difference is 1.4 Å for 36 aligned C$^\alpha$ positions. The two disulfide bonds that link the $\alpha$-helix to the core of the protein are conserved, indicating that these may form

an important stabilizing element of this type of fold. The disulfide bond that links the N- and C-terminal regions of LAPP has no equivalent in HGF. In LAPP the central $\beta$-sheet contains one additional strand, $\beta$0, formed by the N-terminus. In HGF the N-terminus points in a completely different direction. Both structures also diverge in the loop that connects the $\alpha$-helix to strand $\beta$3 and in the extended loops that cover one face of the central $\beta$-sheet. In HGF the latter loops contain more secondary structure; in particular, strands $\beta1'$ and $\beta4'$ contain five residues each instead of the isolated $\beta$-bridge observed in LAPP.

A comparison of the sequences of LAPP and 19 N-domains from HGF- and plasminogen-related proteins shows a strict conservation of four disulfide-bonded cysteines and a conservation of the hydrophobic nature of five residues located in the core of the domain. Four of these hydrophobic residues are located in the interface between the central $\beta$-sheet and the $\alpha$-helix; the fifth residue is located on the opposite side of the $\beta$-sheet (Fig. 6). The pattern of conserved cysteines and hydrophobic residues may define this type of fold.

Tordai *et al.* (1999) recently showed on the basis of sequence comparison and structure prediction that the N-domains of HGF and plasminogen are homologous with the apple domains found in prekallikrein and coagulation factor XI and with a novel domain found in numerous nematode proteins. The members of this superfamily are referred to as PAN domains. The authors did not identify LAPP as a PAN domain. Interestingly, the disulfide-bonding pattern of the apple and nematode domains is identical to that of LAPP. In particular, they all contain a disulfide bond that links the N- and C-termini of the domain. This disulfide bond is absent in HGF. The five hydrophobic residues that we found to be structurally conserved among LAPP and HGF (Fig. 6) are also present in all other PAN domains (Fig. 2 of Tordai *et al.*, 1999).

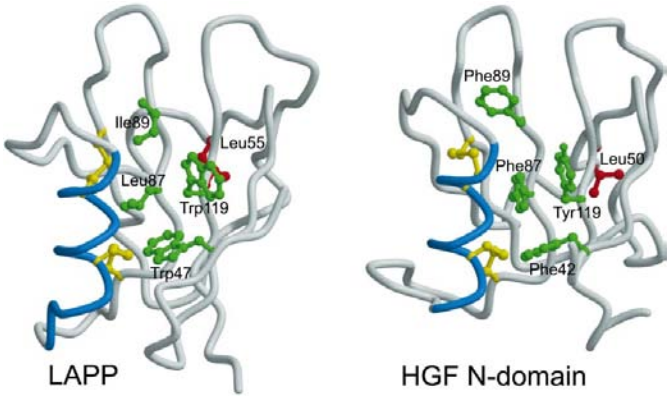


**Figure 6**
Residues with a conserved hydrophobic nature in LAPP and the human HGF N-domain. These residues are also conserved among all other PAN domains. Residues that have a conserved hydrophobic nature and are located at the interface of the central $\beta$-sheet and the $\alpha$-helix are shown in green. A conserved leucine residue located on the opposite site of the central $\beta$-sheet is coloured red. Shown in yellow are two conserved disulfide bonds. $C^{\alpha}$ traces are depicted in worm representation; the $\alpha$-helical region is coloured blue.

Additionally, two more hydrophobic residues are conserved in all PAN domains including HGF but not LAPP. In HGF these residues are Trp98 and Leu118, whereas arginine and glutamine are found at these positions in LAPP. This shows that LAPP is a unique and distant member of the PAN domain superfamily.

### 3.6. Collagen binding

It is not known whether the collagen-binding site of LAPP is located in its flexible N-terminal region or in its compactly folded C-terminal domain. Comparison of the structure of LAPP with those of the collagen-binding domains of vWF (Huizinga *et al.*, 1997) and integrin $\alpha_2\beta_1$ (Emsley *et al.*, 1997) did not reveal similar features that could identify a collagen-binding site. The collagen-binding site of LAPP does not necessarily have to be similar to the collagen-binding site of the vWF-A3 domain or the integrin $\alpha_2$-I domain. This is illustrated by the fact that the latter domains share the same fold but have entirely different collagen-binding sites. LAPP is likely to bind to a G*X*R motif in collagen ($X \neq$ E) (Verkleij, 1999). This binding motif is different to the amino-acid sequences recognized by vWF and $\alpha_2\beta_1$. Nevertheless, LAPP blocks the interaction of both proteins with their respective binding sites. It is most likely that LAPP does so by interacting with nearby G*X*R motifs that occur frequently in collagen types I and III. An extended conformation of the N-terminal region could allow LAPP to interfere with collagen binding of $\alpha_2\beta_1$ and vWF at sites located relatively distant from the primary G*X*R binding site of LAPP itself. Since the N-terminal region of LAPP has a large overall negative charge it is possible that it is involved in collagen binding by interacting with positively charged arginine residues of the G*X*R motif. However, the paucity of biochemical data and the fact that we do not see the N-terminal region of LAPP in the crystal structure preclude a conclusive analysis of the collagen-binding site of LAPP at this moment.

## 4. Conclusions

LAPP consists of a compactly folded C-terminal domain and a flexible N-terminal region that is disordered in the crystal. The C-terminal domain has a fold similar to the N-domain of HGF and is classified as a PAN domain. A pattern of four conserved cysteines forming two disulfide bonds and of five residues with a conserved hydrophobic character may represent the signature of this type of fold. The crystal structure suggests that a study of N- and C-terminal fragments consisting of residues 1–36 and 37–129, respectively, could define the localization of the collagen-binding activity of LAPP.

## References

Barnes, M. J., Knight, C. G. & Farndale, R. W. (1998). *Curr. Opin. Hematol.* **5**, 314–320.
Bienkowska, J., Cruz, M., Atiemo, A., Handin, R. & Liddington, R. (1997). *J. Biol. Chem.* **272**, 25162–25167.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chirgadze, D. Y., Hepple, J. P., Zhou, H., Byrd, R. A., Blundell, T. L. & Gherardi, E. (1999). *Nature Struct. Biol.* **6**, 72–79.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Connolly, T. M., Jacobs, J. W. & Condra, C. (1992). *J. Biol. Chem.* **267**, 6893–6898.
Deckmyn, H., Stassen, J. M., Vreys, I., Van Houtte, E., Sawyer, R. T. & Vermylen, J. (1995). *Blood*, **85**, 712–719.
Depraetere, H., Kerekes, A. & Deckmyn, H. (1999). *Thromb. Haemost.* **82**, 1160–1163.
Emsley, J., King, S. L., Bergelson, J. M. & Liddington, R. C. (1997). *J. Biol. Chem.* **272**, 28512–28517.
Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. (2000). *Cell*, **101**, 47–56.
Furey, W. & Swaminathan, S. (1998). *Methods Enzymol.* **277**, 590–620.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Huizinga, E. G., Van der Plas, R. M., Kroon, J., Sixma, J. J. & Gros, P. (1997). *Structure*, **5**, 1147–1156.
Jones, T. A., Zou, J. Y., Cowan, S. J. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Keller, P. M., Schultz, L. D., Condra, C., Karczewski, J. & Connolly, T. M. (1992). *J. Biol. Chem.* **267**, 6899–6904.
Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
Knight, C. G., Morton, L. F., Onley, D. J., Peachey, A. R., Messent, A. J., Smethurst, P. A., Tuckwell, D. S., Farndale, R. W. & Barnes, M. J. (1998). *J. Biol. Chem.* **273**, 33287–33294.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. J., MacArthur, M. W., Moss, D. J. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–290.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Munro, R., Jones, C. P. & Sawyer, R. T. (1991). *Blood Coagul. Fibrinolysis*, **2**, 179–184.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.
Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.
Romijn, R. A. P., Bouma, B., Wuyster, W., Gros, P., Kroon, J., Sixma, J. J. & Huizinga, E. G. (2001). *J. Biol. Chem.* **276**, 9985–9991.
Tordai, H., Bányai, L. & Patthy, L. (1999). *FEBS Lett.* **461**, 63–67.
Ultsch, M., Lokker, N. A., Godowski, P. J. & de Vos, A. M. (1998). *Structure*, **6**, 1383–1393.
Van Zanten, G. H., Connolly, T. M., Schiphorst, M. E., de Graaf, S., Slootweg, P. J. & Sixma, J. J. (1995). *Arterioscler. Thromb. Vasc. Biol.* **15**, 1424–1431.
Verkleij, M. W. (1999). PhD thesis. Utrecht University, The Netherlands.
Verkleij, M. W., IJsseldijk, M. J., Heijnen-Snyder, G. J., Huizinga, E. G., Morton, L. F., Knight, C. G., Sixma, J. J., De Groot, P. G. & Barnes, M. J. (1999). *Thromb. Haemost.* **82**, 1137–1144.
Zhou, H., Mazzulla, M. J., Kaufman, J. D., Stahl, S. J., Wingfield, P. T., Rubin, J. S., Bottaro, D. P. & Byrd, R. A. (1998). *Structure*, **6**, 109–116.