Molecular Basis for the Partition of the Essential Functions of Thrombin among Snake Venom Serine Proteinases

The Case of Thrombin-Like Enzymes

R.C. Maroun
Unité des Venins, Institut Pasteur, Paris, France

Key Words
Thrombin · Snake venoms · Thrombin-like serine proteinases · Molecular modeling · Molecular recognition

Abstract
Thrombin is a mammalian serine proteinase that plays a prominent role in the maintenance and regulation of hemostasis through its interaction with various substrates and/or ligands. The venoms of several snakes contain glycosylated serine proteinases that have been recognized to possess one or more of the essential activities of thrombin on fibrinogen (Fg) and/or platelets. These proteinases share about 60% sequence identity. One class of snake venom serine proteinases are those known as thrombin-like (TLE), named after their ability to directly clot Fg in order to preferentially produce fibrinopeptide A, fibrinopeptide B or both. To understand the molecular basis of this phenomenon, the corresponding amino acid sequences and molecular structures need to be analyzed. Given the absence of experimentally determined tertiary structures of snake venom, TLEs, three-dimensional molecular models should prove useful in this context. Towards this goal, we obtained models of snake venom TLEs that used TSV-PA as template, TSV-PA being the only snake venom serine proteinase whose crystal structure is known to date. Along with a comparative sequence analysis the models contribute to the identification and description of thrombin-homologous or alternative binding sites, helping thus to understand differences in specificity.

Introduction

Serine proteinases of a given fold share an identical catalytic mechanism, but markedly differ in substrate specificity [1–3], a property that is the outcome of evolutionary diver-
Thrombin is a multifunctional mammalian serine proteinase that plays a major role in the maintenance and regulation of hemostasis. It stimulates plasma coagulation and platelet aggregation and thus possesses a procoagulant action. It acts on fibrinogen (Fg) by cleaving two peptide bonds, producing fibrin monomers that polymerize to form a clot. Thrombin is also the most potent activator of platelets [5]. It mediates platelet activation and aggregation by hydrolysis of a platelet thrombin receptor [5, 6]. It also has anticoagulant properties, since after its interaction with thrombomodulin it activates protein C, which is a major anticoagulant factor [7]. The crystal structure of human α-thrombin bound to specific inhibitors revealed prominent structural features of the thrombin molecule [8–10]. These features are the location of the catalytic triad within a deep canyon-like active-site cleft, and the presence of two extended surfaces. One surface, composed of positively charged and hydrophobic residues, is referred to as exosite I (ESI) or Fg recognition exosite. The other surface (exosite ESII) is solely composed of positively charged residues. ESI is required for thrombin binding to several substrates (Fg, proteinase activated thrombin receptor, and heparin cofactor II), ligands (thrombomodulin and glycoprotein 1b) [11] or the inhibitor hirudin [12]. For ESI, apolar as well as electrostatic interactions appear to be important determinants of complex formation [13]. ESII, which is located close to the carboxy-terminal B-chain helix, is involved in heparin [14], haemadin [15] and prothrombin fragment 2 binding [18].

Snake venom serine proteinases belong to the trypsin family of serine proteinases (family S1 of clan SA, [17]) and exhibit a high degree of sequence identity (60–68%). They have been characterized for their activity on a variety of substrates and it appears that each one is quite specific for a given macromolecular substrate, particularly the components of the blood coagulation cascade [18, 19]. The so-called snake venom thrombin-like serine proteinases (SVTLE) possess primarily one of the essential functions of thrombin: the clotting of Fg. The enzymes may thus clot Fg by splitting off preferentially either fibrinopeptide A, fibrinopeptide B or both. They are named venombin A, B or AB, respectively. Hirudin has no effect or an only negligible effect on their activity. Although to a much lower extent than thrombin, some may activate other blood coagulation factors.

The biological significance of the display of the essential Fg-clotting function of thrombin in SVTLEs is not clear. In an attempt to better understand the determinants of enzyme specificity, we obtained and analyzed their three-dimensional (3D) molecular models. We considered most appropriate to use the crystal structure of the one-chain nonzymogen plasminogen activator of Trimeresurus stejnegeri snake venom (TSV-PA) as the template molecule for the knowledge-based approach used. Indeed, TSV-PA, a snake venom serine proteinase, possesses functional, amino acid sequence (61–73% sequence identity) and structural features in common with the SVTLEs. Among the structural features are the pattern of Cys residues and, most of all, the C-terminal extension characteristic of snake venom serine proteinases. A comparative sequence analysis along with the resulting 3D models allows us to highlight and locate spatially alternative and/or equivalent residues that could be involved in the interaction of these enzymes with their natural substrates.
Materials and Methods

Sequence Analysis

The proteins used were those in table 1 of Pirkle [19] for which complete amino acid sequences exist, plus contortrixobin, whose sequence has been made available recently. The SVTLEs and their accession numbers are as follows: ancrod B (SwissProt P47797), batroxobin (SwissProt P04971), bilineobin (PIR S65621), bothrombin (PIR A54361), calobin (SwissProt Q91053), contortrixobin (TrEMBL P82981), crotalase [20], flavoxobin (SwissProt P05620), and the gyroxin analog from Lachesis m. muta (SwissProt P33589). The other two sequences are TSV-PA (SwissProt Q91516) and human α-thrombin heavy chain (PDB code 1ppb; since we deal only with the heavy chain, the letter H used to denote it will not be used in this text). The statistical analysis of the protein sequences was performed with the SAPS software [21].

Sequence Homology Modeling and Structure Prediction

3D structure visualization, modeling and analysis of the 9 SVTLEs were performed with the Homology module of the insightII 98.0 package (Accelrys, San Diego, Calif., USA) using as template TSV-PA complexed to the EGRcmk inhibitor [22]. The sequence alignment was that in figure 1 of Pirkle [19] with slight modifications in order to optimize the alignment to TSV-PA. The sequence alignment between TSV-PA and α-thrombin is structure based. The program ClustalW 1.6 was then applied to the alignment in figure 1 to obtain alignment scores and define more precisely, with the help of the crystal structures, the variable sequence regions that were subsequently mapped onto the structure of TSV-PA.

The putative secondary sites of the SVTLEs were searched through visual comparison to the structure of thrombin and its defined exosite residues [13]. The amino acid residue numbering is based on that of α-chymotrypsinogen.

Results

Sequence Analysis

The SAPS composition and distribution sequence analysis of the SVTLEs indicated that the composition in the major hydrophobic residues Leu, Val, Ile, Phe and Met is of 23.9–28.5%. The acidic residues significantly outnumber the basic residues for contortrixobin, flavoxobin and crotalase, just like TSV-PA. In analogy to α-thrombin, the basic residues outnumber the acidic ones in ancrod and bilineobin. For batroxobin, bothrombin, calobin and the L. m. muta gyroxin analog, both kinds of residues are in approximately the same proportions. SAPS detects significant mixed-charge linear clusters of amino acid residues for six of the enzymes. These clusters vary in length from 25 to 44 residues, as compared to TSV-PA with a cluster of 31 residues. Thrombin, batroxobin, calobin and crotalase contain no significant mixed-charge clusters according to SAPS. Weak mixed-charge clusters could nevertheless be detected by inspection. For example, for thrombin, the cluster spans Asp60e and Lys81, which encompasses the Fg-binding exosite (FBE) segment 67–80 [13].

The multiple-sequence alignment in figure 1 shows a number of variable regions (underlined) common to the SVTLEs. These are the 34–39, 60–67, 72–76, 90–101, 146–148, 165–167, 171–178 fragments (A–G), as well as the 245–245e C-terminal portion (H). With respect to the 7 venomins A (crotalase to calobin), and outside the variable regions, six residues (Val17, Phe34, Ser42, Arg111, His137 and Tyr159) are unique to contortrixobin (a venombin B), whereas only 3 are unique to bilineobin, a venombin AB (Ile16, Val158 and Ser229). It is also possible to detect the residues shared by all venomins and α-thrombin, excluding Cys, catalytic residues and conserved segments around molecular markers (see below; ref. [23]; green residues in the sequence of thrombin in figure 1). Those residues that are unique to all SVTLEs are indicated by an X. For example, at position 50 and 225 all SVTLEs possess a Glu and a Pro, respectively, as compared to an Arg and a Tyr for thrombin. Ancrod B contains a free sulphydryl at Cys193.
Fig. 1. Multiple alignment for the SVTLEs, TSV-PA and α-thrombin. First, α-thrombin and TSV-PA were aligned topologically. Thereafter, the alignment in figure 1 of Pirkle [19] for SVTLEs was added and optimized with respect to TSV-PA. Exosite I basic, hydrophobic and extension residues are in blue, magenta and gold, respectively. ESII residues are in red. Residues common to all SVTLEs and thrombin are in green, excluding Cys, catalytic residues and sequence motifs around molecular markers. Variable regions A–H (see text) for all 9 SVTLEs together are underlined. PDB file 1PPB [14] was used as the crystallographic reference of thrombin for structure and sequence. The sequences are arranged as venombin A (crotalase to calobin), venombin B (contortrixobin) and venombin AB (bilineobin). Chymotrypsinogen numbering for thrombin is shown.
Structure Analysis

Certain residues occupied positions in space equivalent to the corresponding thrombin residues, other residues were particular to the SVTLEs (not shown). As compared to thrombin, with 9 nonneutralized basic residues composing ESI, the SVTLEs contain up to 5 residues (fig. 1). Hydrophobic residues prone to forming a hydrophobic cluster are on the average less numerous than thrombin, except for bothrombin, 7. Calobin presents only 2 hydrophobic residues. In brief, it is possible to detect the presence of an exosite in SVTLEs analogous to ESI of thrombin and composed of basic and hydrophobic residues. However, ESII of thrombin is poorly represented in the SVTLEs, none of which presents the ‘heparin-binding’ motif of the ESII of thrombin [13], and whose basic residues are outnumbered by those of thrombin.

Figure 2 shows the variable regions underlined in figure 1 as they are mapped onto the 3D structure of TSV-PA. Some correspond to well-known regions or loops (36, 60, 99, 174) and to the segment on the N-terminal side of the autolysis loop of thrombin. The five regions are located east, northeast, northwest, west and south of the catalytic site in the canonical view. Segment 72–76 is southeast of the catalytic cavity and relatively far from it. Segment 165–167 is part of helix 2 of the trypsin fold, east of the active site. The remaining segment comprises the C-terminal extension unique to snake venom serine proteinases.

A number of sequence motifs surrounding serine proteinase evolutionary markers and active site residues for the S1 family of clan SA have been identified [23]. Thus, for the motif surrounding catalytic His57 (54TAAHC58), ancrod and L. m. muta gyroxin

Fig. 2. MOLSCRIPT [36] representation of the TSV-PA crystallographic structure in which the detected variable regions in SVTLEs (underlined in fig. 1) are labeled A–H. Active site residues His57, Asp102 and Ser195 and disulfide bonds are in ball and stick representation. Arrows: ß-strands; ribbons: α-helices.
analog present a residue other than the expected Ala56. Interestingly enough, all SVTLEs present a DIML motif at Asp102, instead of the $^{102}$DIAL$^{105}$ motif. The gyroxin analog differs even further in the sense that it is missing the 2 last residues of this tetrad. As for the motif around Ser195 (193$^{GDSGGP198}$), position 193 is occupied by a variety of residues, instead of the conserved Gly in mammalian serine proteinases. In addition, flavoxobin presents a Thr at 197. As concerns the Ser214 sequence pattern (211$^{GIVSW215}$), apart from bilineobin, either position 213 or position 215 shows variations. Ser214 itself is not conserved in crotalase, ancredo and the $L$. $m$. muta gyroxin analog. Finally, for the triad following position 225, 225(P/Y)GV(Y/F)228, Gly226 and Val227 turn into Ala and (Leu/Phe/Ile/Ala), respectively.

It is interesting to note that, of the 6 thrombin single-point mutants that retain less than 5% of the wild-type Fg-clotting activity (Trp60d, Lys70, His71, Tyr76, Glu217, Arg221) [24] and are not located in the active site cleft (Trp60d), only Lys70 and Tyr76 belong to the canonical Fg recognition exosite. As for the remaining residues, amino acid changes in His71 and Arg221 may indirectly disturb Fg binding, given that they destabilize loops that are involved in the binding. On the other hand, Glu217 makes direct contacts with Fg. Of the 6 positions, only His71 is strictly conserved throughout the SVTLEs. Tyr76 becomes a Val and the positive charge at position 70 is conserved only in batroxobin and ancredo B, indicating a certain modification of the corresponding exosite. The charge at position 217 is preserved in five cases (positions 217 or 218, fig. 1); however, Arg221 becomes a Glu or a Gln.

Inspection of the structure of TSV-PA reveals that a cluster of hydrophobic residues is indeed present in the region corresponding to ESI, but that the basic residues number three and do not constitute a cluster. Five basic residues represent ESII. Four of them form a cluster, with one residue poorly exposed. None of the 5 residues are located on the C-terminal $\alpha$-helix (fig. 1). TSV-PA has an acidic segment $^{96}$DDE$^{98}$ in the active site cavity. This tripeptide, and especially Asp97, has been shown to play a key role in plasminogen activation and thus substrate recognition [25]. Contortrixobin and flavoxobin contain an Asp-Glu peptide in the same region – the remaining SVTLEs show no acidic residues here.

**Conclusion and Discussion**

Overall, no quantitative correlation can be found between the specificity of venombin (A, B or AB) and either the hydrophobic or charged-residue content or the described sequence motifs around the molecular markers. In spite of these facts, it is possible to gather some useful information from figure 1 concerning those residues that are unique to each class of enzyme and that could contribute to the subfamily or specific enzyme specificity. Thus, certain residues are unique to each venombin B, venombin AB and all SVTLEs. Others are shared by thrombin and the SVTLEs. Furthermore, several of the variable-sequence segments are located around the active site, whereas one is localized relatively far away from it. Yet another comprises a C-terminal extension characteristic of snake venom serine proteinases. This region is part of the C-terminal portion of the sequence in the proteinase domain that encodes function in serine proteinases [26].

It is interesting to note that TSV-PA, which does not clot Fg, indeed lacks a patch of basic residues in the ESI region. The 3–4 basic residue cluster in the ESII region could present a binding surface to heparin.
The diversity in the functionality of thrombin is thus no doubt a reflection of the presence of specific residues or regions, as well as allosteric changes that may affect its functions [27]. Furthermore several residues, by overlapping other subsites, may fulfill more than one functional role. The modified or reduced functionality of some SVTLEs partially resides in the absence or variation of those specific residues or regions. Whereas in some cases the latter may be detected by sequence comparison, the allosteric changes during binding, if any, may not and would have to be modeled in order to study their influence. On the other hand, the SVTLEs are generally not purely monofunctional, i.e., do not act uniquely on a substrate, which makes it more difficult to precisely define the key residues. An example is crotalase, which in addition to being a SVTLE, displays kallikrein-like activity [28].

Position 193 plays a significant role in determining substrate specificity and resistance to inhibition [29]. As in thrombin, residue 193 is a Gly in calobin, bothrombin and batroxobin. These SVTLEs should be then sensitive to Kunitz-type inhibitors, such as bovine pancreatic trypsin inhibitor. The other SVTLEs contain either an Arg or the bulky residues Phe and Tyr. Because of the restricted accessibility of the S\textsubscript{2} pocket that these side chains cause, the remaining SVTLEs should not be sensitive to the inhibitor, just like TSV-PA, which has a Phe, or mesotrypsinogen with an Arg.

All the SVTLEs possess a Pro at position 225. This position has been shown to influence Na\textsuperscript{+} binding near the primary specificity site and to contribute to its architecture. Binding of Na\textsuperscript{+} allosterically enhances the catalytic activity of the enzyme. Pro at position 225 abrogates Na\textsuperscript{+} binding through an effect on the architecture of the water channel around the primary specificity site. In addition, this Pro indicates that the SVTLEs evolved from the TCN/P225 lineage with a trypsin-like ancestor [30–32]. It follows then that the molecular mechanism of recognition and binding of ligands by SVTLEs is different from those of thrombin and not mediated by allosteric transitions. Some of the factors contributing then in a combined fashion to these differences include: (1) the absence of the cleft formed by the 60-loop of thrombin (making the active site more accessible); (2) the number of acidic residues lining the primary-specificity pocket; (3) the number and type of residues in the apolar binding site near the catalytic center; (4) the number and location of the basic residues constituting the ESI [33].

On the basis of a model, an alternative Fg recognition site was proposed for crotalase [20, 34]. Given that a number of basic and hydrophobic residues are present in SVTLEs at a region equivalent to the ESI in thrombin, the alternative site seems to be rather an extension of the ESI specific to SVTLEs, that includes Lys84\textsubscript{a} and Lys86. These residues are indeed conserved among SVTLEs and could interact with the acidic Fg sequence on the C-terminal side of the cleavage site. Surprisingly, the ESI extension residues are also present in TSV-PA.

Carbohydrates may play an important role in the structure and stability of a glycoenzyme, and sometimes lead to considerable loss of enzymatic activity. The presence of carbohydrates is usually not taken into account when modeling. Nevertheless, theoretical tools are now available that make it possible to model these moieties if their locations, nature and content are known. In spite of this, molecular models of SVTLEs obtained by sequence homology modeling, by keeping the overall serine proteinase fold and bringing together residues that are far in the sequence, ease the task of identifying regions involved in biomolecular recognition. Several structural
features were thus reported and validated experimentally on the basis of a model of the gyroxin analog based on α-thrombin and trypsin [35]. Finally, the TSV-PA structure and the models derived from it should provide a basis for the study of structure-function relationships of snake venom serine proteinases. In a forthcoming paper we intend to develop in detail the structural features of the SVTLEs obtained here.

Acknowledgments

I am grateful to Prof. Wolfram Bode (Max-Planck Institut für Biochemie, Martinsried, Germany) for having encouraged me to start this work. I am a research scientist at the Institut National de la Santé et de la Recherche Médicale (INSERM), France.

References

13 Bode W, Huber R, Rydel TJ, Tulinsky A: Structure-function relationships of snake venom serine proteinases. In a forthcoming paper we intend to develop in detail the structural features of the SVTLEs obtained here.

Partition of the Functions of Thrombin between Snake Venom Proteinases

Haemostasis 2001;31:247–256

255


