Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 2571–2587
Bioorganic & Medicinal Chemistry

S1 subsite in snake venom thrombin-like enzymes: can S1 subsite lipophilicity be used to sort binding affinities of trypsin-like enzymes to small-molecule inhibitors?☆

Floriano P. Silva, Jr.[a,b] and Salvatore G. De-Simone[a,c,*]

[a]*Laboratório de Bioquímica de Proteínas e Peptídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, 21045-900 Rio de Janeiro, RJ, Brazil*
[b]*Departamento de Química Orgânica, Instituto de Química, Universidade Federal do Rio de Janeiro, Cidade Universitária, 21949-900 Rio de Janeiro, RJ, Brazil*
[c]*Departamento de Biologia Celular e Molecular, Instituto de Biologia, Universidade Federal Fluminense, 24001-970 Niterói, RJ, Brazil*

Received 21 May 2003; revised 1 March 2004; accepted 14 March 2004

Available online 9 April 2004

Abstract—Thrombin-like enzymes isolated from snake venoms comprise a group of serine proteinases responsible for many important coagulation disorders in the envenomed victims. Besides, these proteinases have great biotechnological interest as antithrombotic agents and as diagnostic tools. However, in spite of the recent overflow of snake venom thrombin-like enzymes (SVTLEs) on protein sequence databases, there is a lack of three-dimensional (3D) structural information on this family. Without such 3D structures available many aspects of the biological function and biochemical properties of these enzymes still remain obscure. Therefore, we have gone through a series of computational techniques, which enabled us to identify the set of residues involved in molecular recognition of inhibitors bound to the S1 subsite of snake venom thrombin-like enzymes (SVTLEs) and ultimately conclude that nonpolar (van der Waals) intermolecular interactions and ligand's hydrophobicity are the most important factors affecting binding affinities to the S1 subsite of a SVTLE isolated from the venom of *Lachesis muta muta* (Lmm-TLE). Consequently, we have proposed that S1 subsite lipophilicity may be used to sort binding affinities of trypsin-like enzymes to small molecules by showing that the inhibitory potency of several S1-directed compounds follows subsite lipophilicity among Lmm-TLE and other three homologous proteases. Noteworthy, in the course of our analyses we determined that thrombin's S1 subsite should, in fact, be considered less lipophilic than that of trypsin if we account for the presence of the sodium-controlled water channel communicating with the S1 subsite in the coagulant enzyme.

1. Introduction

A group of proteolytic enzymes have evolved to present a common catalytic strategy, which consists of employing the hydroxyl moiety of an activated serine residue to promote nucleophilic attack on the carbonyl group of peptide bonds. Substrate specificity is achieved through recognition of the side chains near to the bond effectively cleaved by complementary cavities on the enzyme's surface. This large group of enzymes, known as the serine peptidases, can be classified in families according to sequence similarity and further assembled in clans when these are considered to have a common ancestry.[1] One of the largest and most characterized of these families has as an archetype the digestive enzyme chymotrypsin, whose fold is characterized by the

Keywords: Snake venom thrombin-like enzymes; Trypsin; Molecular modeling; S1 subsite lipophilicity; Binding affinity.
Abbreviations: RMSD, root mean square deviation; PLS, partial least-square; LV, latent variable; MLR, multiple linear regression; t-PA, tissue-plasminogen activator; HCA, hierarchical cluster analysis; QSAR, quantitative structure–activity relationship; LP, lipophilic potential.
☆ This work was supported by the Brazilian Council for Scientific and Technological Development (CNPq) and Oswaldo Cruz Foundation (PAPES-FIOCRUZ). FPSJr was a CNPq Dsc. fellowship recipient.
* Corresponding author. Tel.: +55-21-3865-8157; fax: +55-21-2590-3-495; e-mail: dsimone@ioc.fiocruz.br

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.03.024

presence of two similar β-barrel domains asymmetrically packed against each other, with the catalytic site lying between them.[2] The catalytic apparatus is composed from one side by Ser195 from the C-terminal domain and by His57 and Asp102 from the N-terminal domain, on the other side (following chymotrypsinogen numbering system).[3] As recently reviewed,[4] this triad is aligned in such a way that the Ser195 hydroxyl is polarized to attack the carbonyl group of the scissile bond with concomitant capture of the hydroxylic proton by His57, which is stabilized by Asp102, resulting in a tetrahedral intermediate/transition state that is readily stabilized by the backbone N–H groups of Ser195 and Gly193 (forming the so-called 'oxyanion-hole'). As the latter collapses to the acylated enzyme, a proton is transferred from His57 to the leaving group (the polypeptide fragment C-terminal to the labile bond). The enzyme is regenerated by a similar mechanism where a water molecule acts as the nucleophile.

The S1 subsite of the chymotrypsin-like proteases is responsible for the so-called 'primary specificity' of these enzymes [here we follow Schecter and Berger nomenclature,[5] where P1–P1′ represent the residues N- and C-terminal to the scissile bond, respectively; Distal residues are denoted outward; S1, S1′, etc. denote the corresponding binding sites]. This cavity in delineated by three surface loops (loop 1: 184–195, loop 2: 213–228, and loop 3: 169–175) in the C-terminal domain. There is a remarkable structural match between this pocket on the enzyme surface and the amino acid occupying the P1 position of its substrates.[6] Indeed, in the digestive enzyme trypsin,[7] the negatively charged Asp189, positioned at the base of the S1 pocket, determines the absolute requirement for a positively charged residue such as Arg and Lys in the P1 position of its substrates. The same electrostatic interaction accounts for the competitive inhibitory activity of positively charged ammonium, amidinium, and guanidinium derivatives. Nevertheless, site-directed mutagenesis experiments have shown that the S1 subsite is not the single determinant of substrate specificity in serine proteinases.[8] Thus, in more specialized enzymes such as thrombin,[9] binding affinities are defined by the cooperative contribution of more distal sites in the protein surface.

Serine peptidases from some Viperidae venoms are able to convert fibrinogen into fibrin by releasing fibrinopeptides A and/or B. However, in most cases, these fibrin monomers are unable to be cross-linked by Factor XIII, resulting in aberrant clots that are independent of plasmin activity for dissolution.[10] The in vivo consequence of this activity is dependent upon the plasmatic concentration of the enzyme but in ordinary envenomations it usually causes a marked anticoagulant effect. Although thrombin has many activities,[9] these proteinases have been termed thrombin-like enzymes (TLE) because of their skill to mimic the fibrinogen clotting ability of this enzyme, which is central to the coagulation process in mammalians. Moreover, a number of snake venom thrombin-like enzymes (SVTLEs) can mimic other thrombin functions such as activation of coagulation factors,[11] even though the mechanisms by which they work may differ from those of the mammalian enzymes. Having a M_r ranging from 41 to 47 kDa, which is much greater than the predicted molecular weight of 25 kDa (~230 residues), it has been shown that SVTLEs are extensively glycosilated. These proteins possess hydrolytic activity over synthetic substrates (arginyl and lysyl esters/amides), as well as an inhibitory profile against amidines and guanidines similar to trypsin.

The venoms of all *Lachesis muta* (*muta*, *rhombeata*, and *stenophyrs*) pit vipers cause considerable hemorrhage and the *L. m. rhombeata* (Lmr-TLE) and *L. m. muta* (Lmm-TLE) subspecies have had, to the present, at least one SVTLE characterized.[12,13] Protein sequencing of Lmm-TLE has revealed a high degree of sequence similarity (~60%) to other SVTLEs such as flavoxobin, bathroxobin, and ancrod.[14–17] All these SVTLEs, including Lmm-TLE and Lmr-TLE, are able to release only fibrinopeptide A after cleavage of the Arg16-Gly17 peptide bond in the Aα-chain of fibrinogen. Because of this specific feature, ancrod, and batroxobin have been used as defibrinogenating agents for a number of clinical conditions.[10] Although not functionally related, the tissue plasminogen activator-like enzyme from the *Trimeresurus stejnejeri* venom (TSV-PA) is structurally related to Lmm-TLE (54% identity). TSV-PA, in turn, is the only snake venom serine proteinase to have its structure determined to date.[18] Lmm-TLE is also homologous to the mammalian serine proteases trypsin (40% identity) and thrombin (30% identity). Thus, sequence similarity allied to the biochemical data available has allowed Lmm-TLE and other SVTLEs to be classified in the chymotrypsin family. Interestingly, it has been recently proposed, based on a homology model, that Lmm-TLE has hybrid properties unifying the activities and structural features of the mammalian thrombin and trypsin enzymes.[19]

In view of the potential biotechnological and clinical application of SVTLEs as well as the importance of the S1 subsite in determining the binding affinities of small synthetic substrates and inhibitors of trypsin-like serine proteases, we have performed a series of computational analyses aiming to point out the dominating contributions to the inhibitory profile of small amidines/guanidines targeting the S1 subsite of a representative SVTLE—Lmm-TLE. In order to accomplish this, we have generated a homology model of this enzyme and then used it in docking simulations of a set of 21 known S1 specific reversible inhibitors of Lmm-TLE[20] (Fig. 1). We have mapped the S1 subsite interactions of the docked complexes so as to find, which residues and type of interactions can be important for dictating the distinct properties of this subsite among SVTLEs. The proposed formalism for calculation of binding affinities includes the van der Waals (vdW) component of the intermolecular interaction energy of the docked complexes, the hydration energy (Δ_{Hyd}) of the inhibitors and accessible area buried (AAB) upon formation of the ligand–protein complex. The relative importance of each of these terms has been determined by a statistical regression method. According to the regression model obtained, the vdW as well as the Δ_{Hyd} terms are the most important in

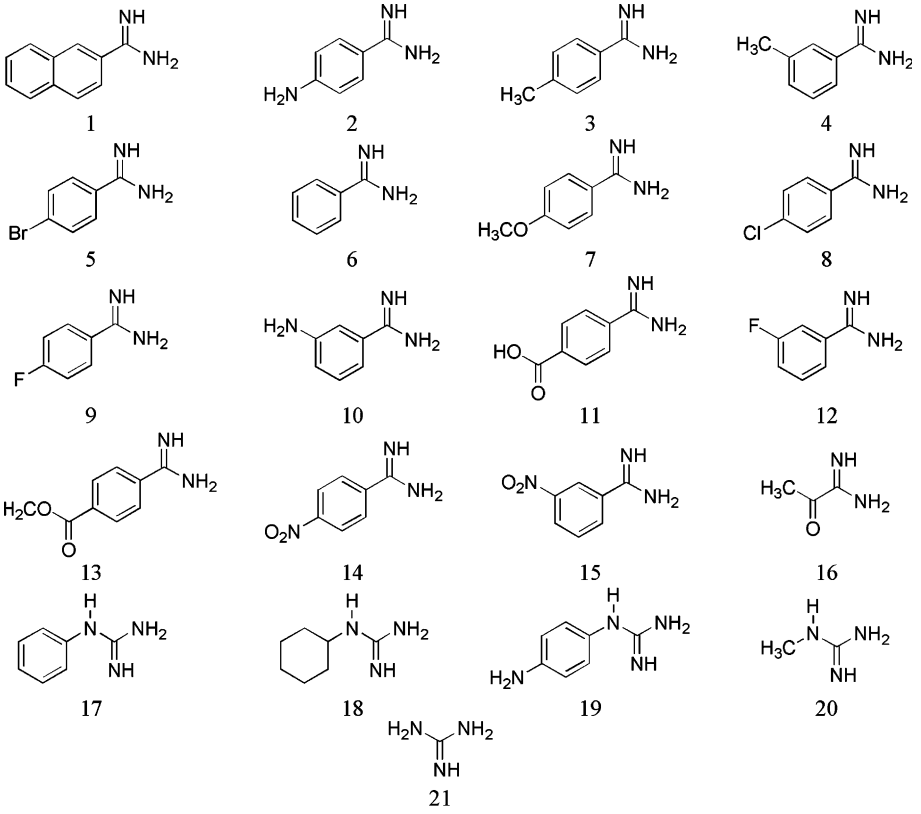

Figure 1. Structure of S1-directed known inhibitors of *Lachesis muta muta* thrombin-like enzyme. This set of small amidines and guanidines has been reported in the work of Magalhães et al.[20]

describing observed binding affinities. The relevance of hydrophobic interactions for achieving high binding affinity to the S1 subsite of Lmm-TLE and other trypsin-like proteases has then been further validated by showing that the inhibitory potency of several S1-directed compounds follows subsite lipophilicity among Lmm-TLE and other three homologous proteases. Noteworthy, in the course of our analyses we determined that thrombin's S1 subsite should, in fact, be considered less lipophilic than that of trypsin if we account for the presence of the sodium-controlled water channel communicating with the S1 subsite in the coagulant enzyme.

2. Results

2.1. Template selection, alignments, and Lmm-TLE modeling

A BLAST search over the Ex-PDB database has returned several serine peptidases with considerable sequence similarity to Lmm-TLE. Among the highest scoring proteases, TSV-PA B chain (Ex-PDB code 1BQYB; e-value 5.4×10^{-67}) and rat anionic trypsin (Ex-PDB code 1ANE_; e-value 1.4×10^{-38}) were selected for the homology modeling of Lmm-TLE. In order to have a contribution from a serine protease presenting a closer function to Lmm-TLE, the structure of human α-thrombin H chain (Ex-PDB code 1H8DH; e-value 10^{-14}) was selected for the homology model construction. After several cycles of preliminary model construction, analysis, and refinement, the final alignment between Lmm-TLE and its templates was produced (Fig. 2). One major difficulty in deriving this alignment was posed by a three-residue deletion in the loop 90. This region is responsible for correct positioning of Asp102 in the catalytic site. The majority of this loop presents residue variations among SVTLEs but this deletion is unique to Lmm-TLE as shown by an alignment of almost 50 sequences available in public databases (Fig. 3).

2.2. Overall structure and validation of Lmm-TLE model

The ribbon diagram depicted in Figure 4 shows the secondary structure of the Lmm-TLE 3D model along with the predicted disulphide bonds. The transition between the two domains in Lmm-TLE can be perceived by the color changing from blue (N-terminal) to red (C-terminal) in Figure 4. Each domain is comprised by a six stranded β-barrel packed against an α-helix. Most of each domain is assembled continuously, but the polypeptide chain crossing the domains twice afford the N-terminal strand to compose the C-terminal domain and the C-terminal α-helix to pack against the N-terminal β-barrel. The catalytic site residues His57, Asp102, and Ser195 show the usual hydrogen bonding pattern confirming the proper folding of the predicted model. This can be further accessed by the RMSD at the spatial position of backbone atoms in the templates used in model construction: TSV-PA, RMSD = 0.78 Å (816 atoms); rat anionic trypsin, RMSD = 0.95 Å (772 atoms) and α-thrombin, RMSD = 1.10 Å (764 atoms).

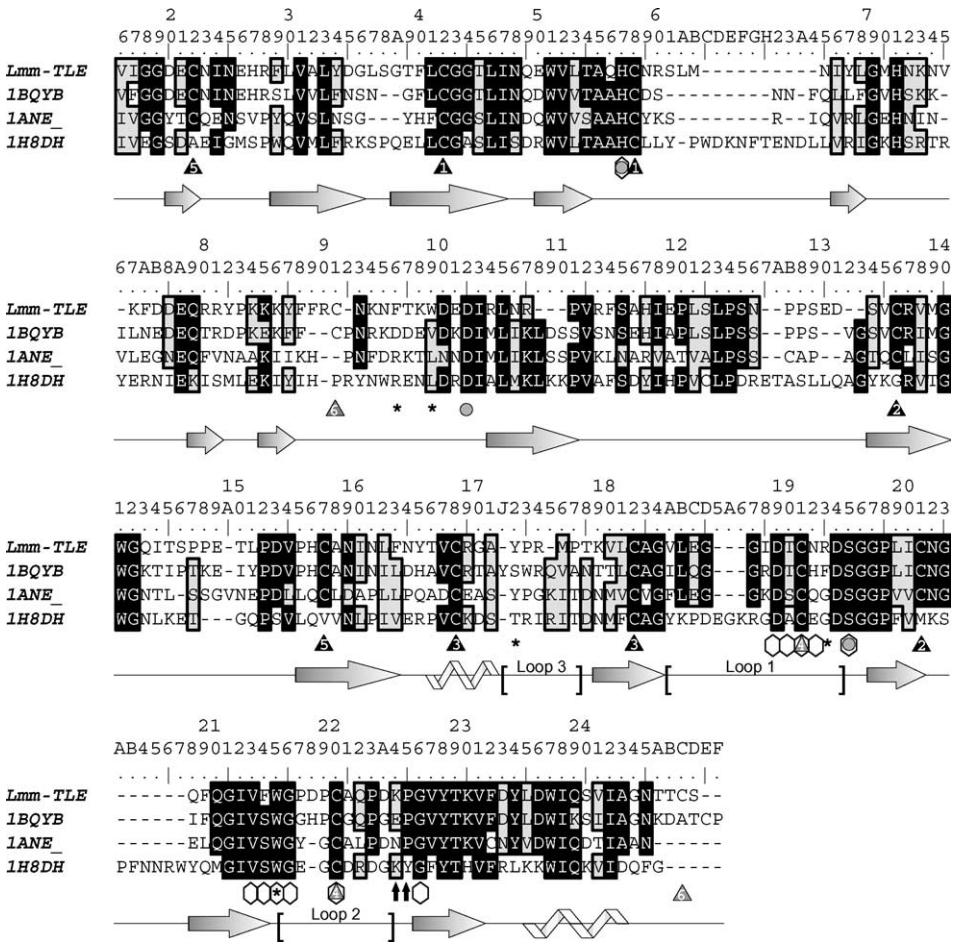

Figure 2. Structural alignment used on homology model construction. Lmm-TLE—*Lachesis muta muta* thrombin-like enzyme;[14] 1BQYB—*Trimeresurus stejnegeri* plasminogen activator (TSV-PA);[18] 1ANE_—Anionic rat trypsin;[57] 1H8DH—human α-thrombin.[25] The alignment was boxed according to the conservation threshold of 70% and numbered after the chymotrypsinogen sequence.[3] After each alignment block is represented the predicted secondary structure of Lmm-TLE: arrow—β-strand; helix—α-helix; straight line—nondefined (loop). Loops 1–3 are delimited by brackets. Residues with a structural, catalytic, substrate recognition, or any particular function in the chymotrypsin family are detached: catalytic triad residues—gray circles; residues lining the S1 subsite walls—beehive-like polygons; Na^+ binding site—black arrows; aryl binding site—asteristics; conserved cystines—triangles 1–4; and cysteines residues in Lmm-TLE possibly involved in disulphide bridges—triangles 5 and 6. The black filled triangles mark disulphide bridges predicted by the Lmm-TLE 3D model while gray filled triangles spot these that were not.

In most members of the family, intradomain disulphide bonds help to keep the active conformation of the peptidase. Indeed, the proposed Lmm-TLE model predicted four cysteine pairs: 22–157, 42–58, 136–201, and 168–182. Other four Cys residues present in the Lmm-TLE sequence (191, 220, 91, 245) were not modeled as disulfide-bridged cysteines. Nevertheless, careful inspection of the model reveals that Cys191 and Cys220 are sufficient close to each other to afford a disulphide bond linking them. Loops 1, 2, and 3 as detached in Figure 4, set the limits of the S1 subsite in the C-terminal domain of the enzyme. At the bottom of loop 1 is displayed the Asp189 residue responsible for the primary specificity of this family of serine proteinases (Fig. 4).

2.3. Structural features of the docked complexes between Lmm-TLE and its inhibitors

Inhibitors docked into Lmm-TLE S1 binding pocket are displayed in Figure 1 and listed in Table 1 followed by the observed inhibitory activities and the decomposition of the energetic scores (intermolecular and intramolecular) given by the dock program. The flexible docking calculations have confirmed the presumed binding mode of these amidine and guanidine derivatives. The conserved interaction observed in all simulated complexes is a face-to-face electrostatic attraction between the positively charged amidine or guanidine moiety and the negatively charged carboxyl group of Asp189 localized at the bottom of the S1 pocket. This salt-bridge governs the orientation of the inhibitors at the enzyme pocket as can be perceived from the superposition of inhibitors **13**, **17**, and **19** in Figure 5A. A diagram of the extensive nonpolar contacts and hydrogen bonds at the S1 subsite made by the most active compound in the series, 2- nafthamidine, is depicted in Figure 5B.

A set of Lmm-TLE residues—Asp189, Thr190, Asn192, Ser195, Phe214, Trp215, Gly216, Cys220, and Gly226 were involved in several contacts with the majority of the inhibitors docked. Among these, only Asp189,

Figure 3. Multiple sequence alignment of snake venom thrombin-like enzymes (SVTLEs). Sequences were retrieved from protein sequence databanks using the sequence retrieval system (SRS at ExPASy server, http://www.expasy.ch/). SVTLEs discussed in the text or that could only be extracted from publications are identified by the ordinary names. The other sequences are denoted by the codenames found in databases. The alignment was boxed according to the conservation threshold of 85% and numbered after the chymotrypsinogen sequence.[3]

Thr190, Gly216, and Cys220 were unanimous to all complexes analyzed. Moreover, four extra residues—His57, Cys191, Val213, and Asp218 were also encountered making contacts with twelve of the studied inhibitors. Two types of hydrogen bonds were observed in all simulated complexes—the first involving one of the amidine or guanidine nitrogen atoms and the Oγ of Thr190; the second one involving the other amidine or guanidine nitrogen atom and the Oγ2 of Asp189. These interactions, combined, form the basic attachment points for these inhibitors at the S1 subsite of Lmm-TLE. Nevertheless, some benzamidine inhibitors (**7**, **11**, **13**, and **14**) bearing acceptor groups in position 4 of the aromatic ring are able to hydrogen bond to the Ser195 hydroxyl (Fig. 5C). On the other hand, compounds **2** and **19**, which carry acceptor groups at the position 4 of the phenyl ring, are able to hydrogen bond to the Asn192 Oδ1 and to the Asp218 Oδ1 atoms, respectively. A summarized account of the interactions between inhibitor moieties and enzyme residues is presented in Table 2. No obvious trend has been observed between the extended hydrogen bonding capacity of some inhibitors and the experimental binding affinities of these compounds. The accessible area of Lmm-TLE S1

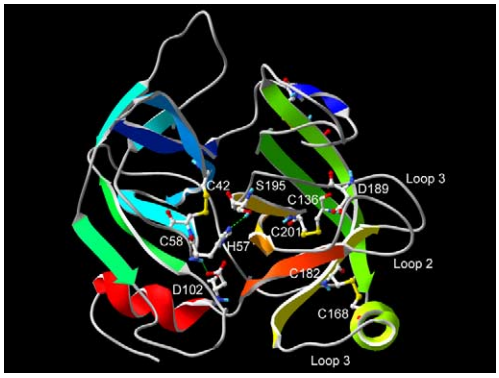

Figure 4. *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE) homology model. Ribbon diagram colored by secondary structure succession from the N-terminal (blue) to the C-terminal (red). The catalytic and Cys residues involved in disulphide bridges are depicted in sticks and colored according to the CPK model. Loops 1–3 composing the S1 subsite are also identified.

subsite buried upon complexation is also listed in Table 2 and is barely proportional to the molecular size of the inhibitors.

2.4. Calculation of binding affinities for the amidine and guanidine inhibitors of Lmm-TLE

In the search for the relative contribution of each component of the formalism proposed to describe binding affinity to Lmm-TLE, we have performed PLS regression (PLSR) over three parameters: E_{vdw}, ΔE_{hyd} and AAB. The PLSR model could be derived from the first LV (as assigned by crossvalidation), which is able to explain 76% of the variance in activity data (pK_i). This model presented thoroughly statistical meaning as indicated by the accompanying statistics in the MLR-like equation shown below:

$$pK_i = 5.364 \pm 0.701 - 0.100 \pm 0.012E_{vdw} + 0.090 \pm 0.016\Delta E_{hyd} + 0.008 \pm 0.002\text{AAB}$$

$$n = 21, \quad s = 0.41, \quad R^2 = 0.76, \quad F_{(1,19)} = 37.2, \quad p_{(1,19)} < 0,001, \quad Q^2 = 0.60, \quad S_{PRESS} = 0.62$$

Judging from the regression coefficients of the equation above, it is clear that the major contributions for the observed binding affinities come from the complementary fit of the inhibitors in the S1 subsite pocket (represented by E_{vdw}) followed by the hydrophobicity of the inhibitors (measured by the positive ΔE_{hyd} values). A negative coefficient for the vdW component occurs because the more negative is this term more favorable are the nonpolar interactions of the inhibitor with the enzyme residues. On the other hand, as a measure of the free energy of solvation in water, the more positive the value of ΔE_{hyd} the less favorable is this process, that is, more hydrophobic is the inhibitor. In turn AAB is a measure of the gain (mainly entropic in nature) in freeing bound water molecules from the interface of the protein–ligand complex upon association.

From crossvalidated standard deviation (S_{PRESS}), which is well below the standard deviation from mean pK_i values (0.96), we see that the model is able to improve

Table 1. Observed pK_i values and calculated energetic descriptors for the S1-directed inhibitors of *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE)

S1-directed inhibitors	pK_i[a]	Intermolecular interaction energy[b]			ΔE_{hyd}[c] (kcal/mol)
		Total (kcal/mol)	VdW (kcal/mol)	Electrostatic (kcal/mol)	
2-Naphtamidine (1)	5.21	−38.0	−22.0	−12.3	−55.6
4-Aminobenzamidine (2)	4.89	−40.7	−21.2	−12.8	−56.7
4-Methylbenzamidine (3)	4.70	−37.3	−21.3	−12.6	−54.9
3-Methylbenzamidine (4)	4.80	−34.9	−20.0	−11.7	−55.0
4-Bromobenzamidine (5)	4.30	−36.5	−22.1	−12.3	−59.2
Benzamidine (6)	4.27	−35.2	−20.9	−11.8	−56.1
4-Methoxybenzamidine (7)	4.23	−34.2	−18.0	−12.5	−56.1
4-Chlorobenzamidine (8)	4.11	−36.2	−21.6	−12.3	−58.5
4-Fluorobenzamidine (9)	3.95	−37.3	−21.3	−12.6	−57.6
3-Aminobenzamidine (10)	3.80	−37.3	−21.7	−12.4	−58.5
4-Carboxybenzamidine (11)	3.40	−36.5	−22.4	−9.3	−60.7
3-Fluorobenzamidine (12)	3.39	−36.6	−21.1	−13.0	−57.7
4-Carboethoxybenzamidine (13)	3.24	−35.5	−19.1	−12.1	−60.8
4-Nitrobenzamidine (14)	3.12	−37.2	−22.5	−13.0	−66.8
3-Nitrobenzamidine (15)	3.04	−37.2	−21.6	−12.8	−65.5
Acetamidine (16)	1.37	−22.9	−17.2	−9.3	−64.4
Phenylguanidine (17)	3.72	−36.4	−14.5	−11.8	−56.0
Cyclohexylguanidine (18)	3.66	−42.6	−20.8	−10.7	−55.1
4-Aminophenylguanidine (19)	2.79	−38.8	−14.4	−13.2	−57.6
Methylguanidine (20)	2.66	−33.8	−14.6	−12.0	−61.8
Guanidine (21)	2.03	−39.8	−12.2	−13.2	−68.2

[a] Experimental values taken from Magalhães et al.[20]
[b] Scores produced by the dock program.[65]
[c] Calculated using the SM5.4/A solvation model,[64] implemented in PC Spartan PRO release 1.05, Wavefunction, Inc.

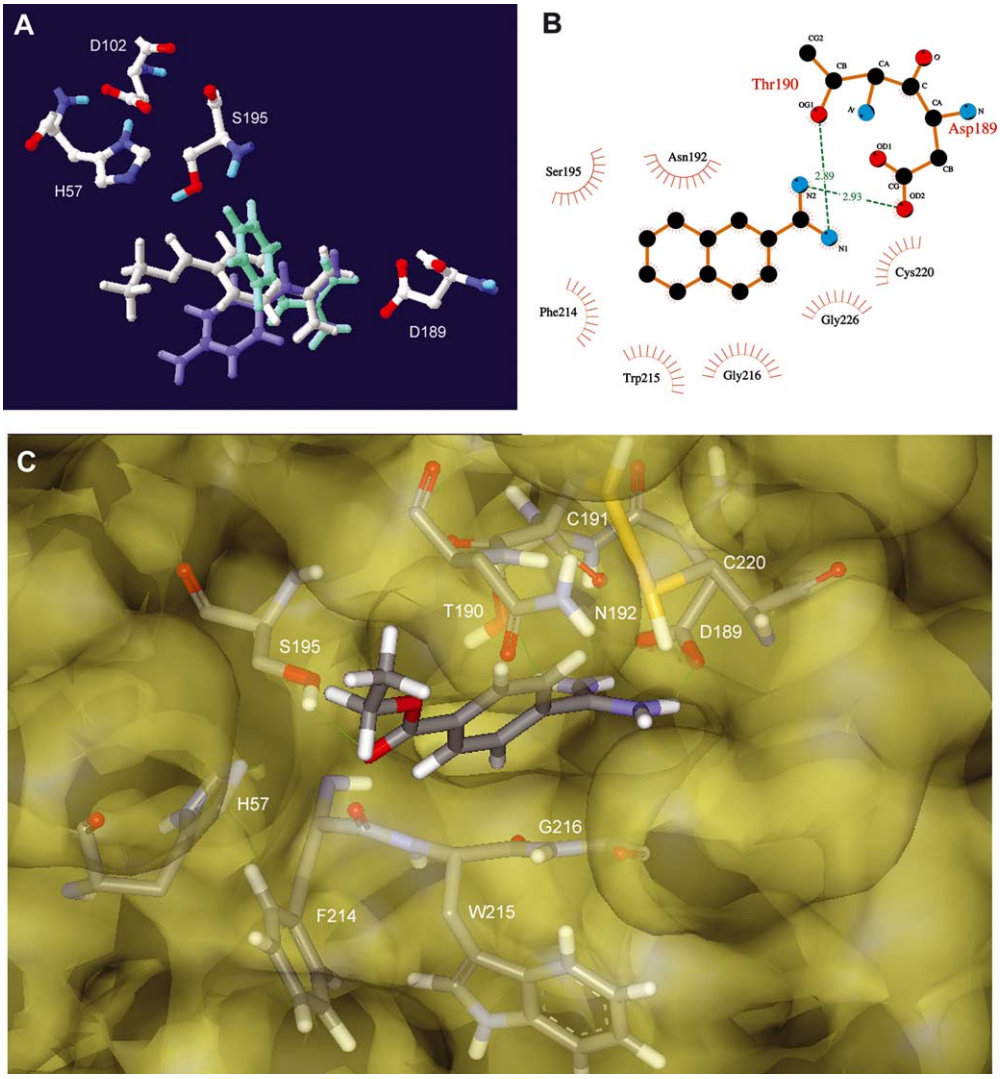

Figure 5. Binding mode of amidine and guanidine inhibitors within the S1 subsite of *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE). (A) Superposed inhibitors **13** (4-carboethoxybanzamidine, gray), **17** (phenylguanidine, green) and **19** (4-aminophenylguanidine, blue) are represented in sticks. (B) Schematic representation of the interactions made by inhibitor 1 (2-naphitamidine); contacting atoms are detached with red marks and hydrogen bonds are indicated in dashed green lines along with donor/acceptor pairs distances. (C) Inhibitor **13** docked at the Lmm-TLE S1 pocket; the solvent accessible surface (1.4 Å probe) was generated transparent in order to make interacting residues visible whereas the hydrogen bonds are indicated by green lines.

our basic predictive ability on the inhibitory activity of the set of compounds analyzed. Other feature of this model is its progress in relation to the published model of Magalhães et al.[20] Its now possible to account for the order of binding affinities for the role set of 21 assayed inhibitors in contrast to the previous model, which only included the 14 benzamidines. The later also presented an inferior fit to data (F-ratio = 21.3) and a slightly worse predictive ability ($Q^2 = 0.56$). Analysis of fitted versus experimental pK_i plots shown in Figure 6 detected no severe outliers in the data set.

2.5. Analysis of the relative lipophilicity of S1 subsites of trypsin-like enzymes

As noted above, the regression coefficients of the PLSR equation have pointed out the importance of the inhibitor hydrophobicity in order to attain strong association to Lmm-TLE. Therefore, it should be reasonable to ask if there is a correlation between the differential binding affinities of small S1-directed inhibitors of trypsin-like serine proteinases and the lipophilicity of this subsite on the enzymes surfaces. As a first step to test this hypothesis, we have determined the lipophilic potential (LP) over the molecular surfaces of the S1 subsite cavities of 20 structures from seven trypsin-like enzymes: trypsin, thrombin, t-PA, plasmin, Factor Xa, the catalytic domain of CIs component of complement and Lmm-TLE (Fig. 7). Subsequently, we devised a scheme to quantify the relative lipophilicity of the S1 subsite that was based on an arbitrary division of the LP scale given by the program MOLCAD (Table 3). HCA was applied as an objective methodology to obtain an independent grouping of the analyzed structures according to the relative lipophilicity values. It can be

Table 2. Structural data for the docked complexes

Inhibitor	Contacts[a]								AAB[c] (Å^2)
	Total number	Extra residues	Absent residues	Hydrogen Bonds[b] (Å)					Bumps	
				A	B	C	D	E		
1	41	—	—	2.9	2.9	—	—	—	1	245.9
2	28	—	—	3.1	2.9	—	—	3.2	2	237.0
3	31	—	—	3.0	2.9	—	—	—	3	240.6
4	40	V213	—	2.8	3.1	—	—	—	3	259.2
5	28	—	—	3.0	2.9	—	—	—	2	251.4
6	33	C191	F214 W215	3.2	3.0	2.9	—	—	1	227.2
7	45	C191	—	2.9	3.1	—	2.87	—	3	245.5
8	26	C191	W215	3.0	2.9	—	—	—	1	237.4
9	8	—	W215	3.1	2.9	—	—	—	1	232.5
10	38	V213	F214 W215	3.3	3.0	2.9	—	—	1	254.3
11	49	—	S195	3.1	3.0	—	—	—	3	232.9
12	34	V213 C191	F214 W215	3.2	2.9	2.9	—	—	3	232.3
13	44	C191	G226	2.8	3.2	—	2.8	—	2	237.7
14	33	—	S195	3.1	3.0	—	3.1	—	1	228.5
15	32	V213	—	2.7	3.2	—	—	—	2	229.5
16	22	—	F214 S195 N192	3.1	2.9	—	—	—	0	129.4
17	33	V213 C191	G226	3.2	2.8	—	—	—	2	255.4
18	28	C191	F214	3.0	3.2	—	—	—	2	246.2
19	29	D218	F214 G226 S195	2.9	3.2	—	—	2.7	3	212.5
20	18	—	F214 S195 N192	3.0	3.0	—	—	—	0	167.5
21	13	—	S195 N192	2.7	3.1	—	—	—	0	108.2

[a] A set of enzyme residues are able to make contacts with at least 60% of the studied inhibitors, these are: D189, T190, N192, S195, F214, W215, G216, C220, and G226. Any additional residue making contact within inhibitors are classified as 'extra' and any missing major contacting residues are listed under 'absent'.

[b] Hydrogen bonds detected between inhibitors moieties and enzyme groups are classified in five types according to the following donor/acceptor heavy atom pairs observed: A—amidine or guanidine N (arbitrarily set 1)/T190 Oγ1; B—amidine/guanidine N (arbitrarily set 2)/D189 Oγ2; C—amidine/guanidine N1/D189 Oγ2; D—S195 Oγ/benzamidine 4-nitro or 4-carbonyl O; E—benzamidine or phenylguanidine 4-amino N/N192 Oδ1 (for inhibitor **2**) or D218 Oδ1 (for inhibitor **19**).

[c] Enzyme accessible area buried upon inhibitor complexation.

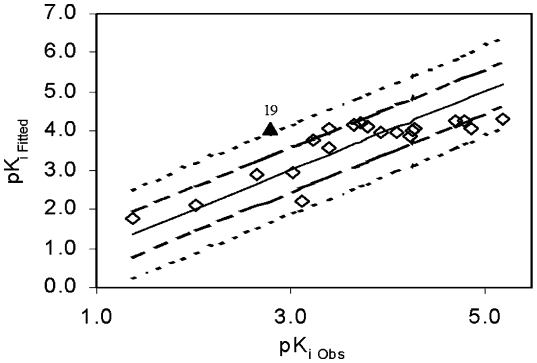

Figure 6. Calculated versus experimental pK_i values of *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE) inhibitors. The central thicker line represents the experimental pK_i values. The first pair of dashed lines delimits the interval for one standard deviation from mean (*s*) while the second pair delimits the interval for 2*s*. A single data point near these limits was identified by a black filled triangle.

perceived from the complete-linkage dendrogram depicted in Figure 8 that four natural groups arose from the analysis. Group 1 is a heterogeneous, medium S1 lipophilicity cluster and is highlighted by the presence of two plasmin structures and by Lmm-TLE. Group 2 is solely composed of thrombin structures of low S1 subsite lipophilicity. Group 3 concentrates three trypsin structures as well as two thrombins of medium-to-high S1 subsite lipophilicity. Finally, group 4 gathers the structures showing highest S1 subsite lipophilicity being mainly constituted by the t-PAs, Factor Xa and CIs.

2.6. Sorting binding affinities of trypsin-like enzymes to small-molecule inhibitors according to the relative lipophilicity of their S1 subsites

Discrepancies on the lipophilicity measures for the different structures of thrombin, trypsin, t-PA, and plasmin were observed. Nevertheless, even with these deviations, which may be attributed to different experimental conditions (e.g., presence of inhibitors and/or cofactors), inter-species variation (i.e., bovine vs rat trypsins and bovine vs human thrombins) and post-translational processing (ε vs α bovine thrombins and γ vs α vs α slow form human thrombins), the HCA

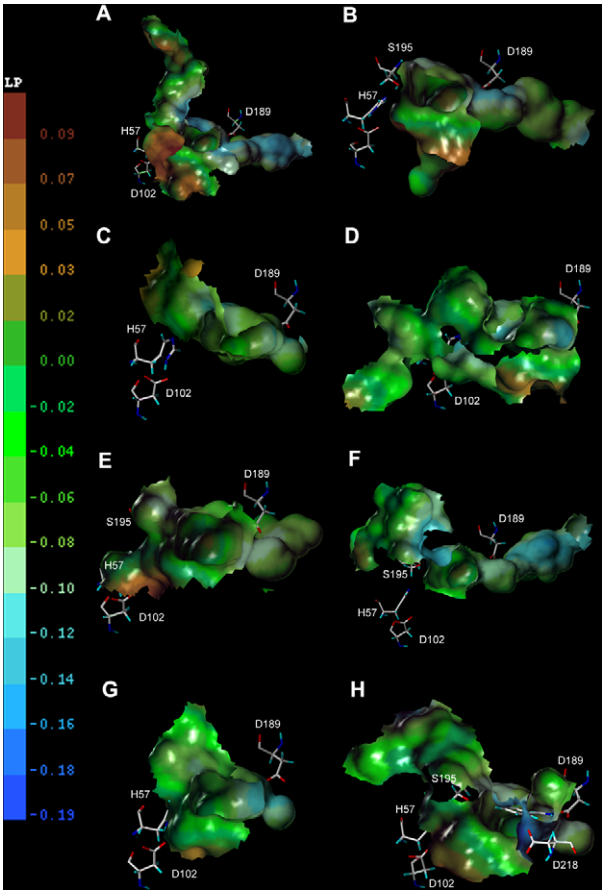

Figure 7. Analysis of lipophilicity potentials over S1 subsite surfaces from 8 trypsin-like enzymes. (A) Human α-thrombin (extracted from the complex with hirugen and *p*-amidinophenylpyruvate, PDB code 1AHT); (B) Human tissue-type plasminogen activator (extracted from the complex with Dansyl-Glu-Gly-Arg Chloromethyl Ketone, PDB code 1BDA); (C) Human plasmin (extracted from the ternary complex—Microplasmin-Staphylokinase-Microplasmin, PDB code 1BUI); (D) Catalytic domain of human CIs complement component (PDB code 1ELV); (E) Human Factor Xa (PDB code 1HCG); (F) Human α-thrombin 'slow' form (PDB code 1MHO); (G) Bovine β-trypsin (extracted from the benzamidine complex, PDB code 3PTB); (H) *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE, this work) with benzamidine docked. Visible active site residues are labeled in each panel along with Asp189. The lipophilicity scale, shown in the left, ranges from blue (less lipophilic) to brown (more lipophilic).

method has enabled a classification consistent with the following order of relative S1 subsite lipophilicity: thrombin ≪ [Plasmin < Lmm-TLE < trypsin] ≪ [t-PA < CIs < Factor Xa]. After carefully determining the order of S1 subsite lipophilicity among the trypsin-like enzymes analyzed we could investigate if there was a trend in binding affinities of these proteases to S1-directed inhibitors that would follow this order. Indeed, as the plot in Figure 9 shows, the order of S1 subsite lipophilicity can be used to correctly sort the binding affinities of four trypsin-like enzymes—thrombin, plasmin, Lmm-TLE, and trypsin—to several small-molecule inhibitors. This result reinforces our conclusion that the hydrophobic character of S1-directed inhibitors, in addition to the overall shape complementarity to the cavity and the deep salt bridge to Asp189, has a major contribution to the observed binding affinities of small S1-directed inhibitors.

3. Discussion

The lack of experimental structural data for the SVTLEs poses a further difficulty in understanding the structural features governing the activity of their S1-directed inhibitors. We have tried to circumvent this problem by generating a homology model of Lmm-TLE. Although most residues with a putative function have also been assigned in the alignment of Figure 2, most of the general features of the Lmm-TLE structure have been discussed before, through the model constructed by Castro et al.[19] Thus, considering that our primary goal in this work was to study the significant aspects of S1 subsite recognition and binding affinities, other features of the Lmm-TLE model will not be discussed in depth. Nevertheless, one point concerning the possible existence of a 'fibrinogen recognition exosite' (FRE) in Lmm-TLE deserves attention. Noteworthy, Castro et al. have hypothesized a FRE homologous to the one found in mammalian thrombin (Arg60, Lys73, Lys76, Arg81, Arg82, Lys85, Lys86, Lys87, Arg110, Arg113), but their experimental data have not supported this proposal. Considering the distribution of basic residues over the Lmm-TLE surface, a FRE analogous to that proposed for crotalase,[21] formed by a dislocated groove surrounded by Arg80 and Arg113 from one side and Lys85 and Arg107 from the other, should represent an alternative for the FRE location in Lmm-TLE.

The S1 subsite in serine proteases has been the target of a plethora of studies since the identification of the complementary aspect of the interaction of these enzymes with their polypeptide substrates. The availability of the first X-ray crystal structures of serine proteases of pharmaceutical and biological importance has greatly contributed to elucidate the structural features of the complementary fit between side chains of amino acids residues occupying the P1 position of substrates and the structure of this pocket on the enzyme surface.[22] These early structural studies revealed that the S1 subsite architecture is composed by three surface loops from the C-terminal domain.[2] These segments in Lmm-TLE are represented by loops 1 (184–195), 2 (215–225), and 3 (172–178) as detached in the alignments of Figures 2 and 3 as well as in the 3D model of Figure 4. As can be perceived from the secondary structure representation of Lmm-TLE in Figure 4, loops 1 and 2 offer the main organization of the pocket while loop 3 contributes with more distal interactions to extended ligands of this subsite. Analysis of the predicted configuration of the complexes formed by the 21 S1-directed inhibitors of Lmm-TLE shown in Figure 1 and listed in Table 1 has disclosed the set of protein residues essential for binding (Table 2). Position 189 has been by far the most characterized in view of its primary role in the preference displayed by the enzyme for the P1 residue.[4,6]

Table 3. Analysis of lipophilicity potentials over S1 subsite surfaces from selected trypsin-like enzymes

Enzymes (PDB code)[a]	Area of S1 subsite containing cavity									
	Total[b]	Core[c]		Lipophilicity potential[d]						
				Low		Medium		High		Total lipophilic[e]
	(Å^2)	(Å^2)	%	(Å^2)	%	(Å^2)	%	(Å^2)	%	%
α-Thrombin—human (1MHO)	670.2	293.0	43.7	170.2	25.4	461.8	68.9	38.2	5.7	74.6
ε-Thrombin—bovine (1ETR)	665.0	303.1	45.6	153.1	23.0	429.9	64.6	82.0	12.3	77.0
α-Thrombin—human (1AHT)	916.7	352.3	38.4	197.0	21.5	507.5	55.4	212.2	23.1	78.5
α-Thrombin—bovine (1HRT)	662.7	333.3	50.3	132.2	19.9	372.4	56.2	158.1	23.9	80.1
Plasmin—human (1BUI)	280.2	232.5	83.0	55.1	19.7	200.8	71.7	24.3	8.7	80.3
β-Trypsin—bovine (1BTY)	316.9	265.6	83.8	56.1	17.7	239.8	75.7	20.9	6.6	82.3
Plasmin—human (1BLM)	326.6	276.1	84.5	55.4	17.0	271.2	83.0	0.0	0.0	83.0
Lmm-TLE—*L. m. muta*	523.7	316.7	60.5	86.6	16.5	392.9	75.0	44.1	8.4	83.4
γ-Thrombin—human (2HNT)	716.8	335.5	46.8	113.8	15.9	511.5	71.4	91.5	12.8	84.1
β-Trypsin—bovine (3PTB)	298.7	226.7	75.9	45.4	15.2	227.9	76.3	25.3	8.5	84.8
α-Thrombin—human (1AWH)	858.4	378.0	44.0	121.3	14.1	571.2	66.5	165.8	19.3	85.9
Anionic-2-Trypsin—rat (1EZS)	432.6	251.4	58.1	53.8	12.4	328.4	75.9	50.4	11.7	87.6
Anionic-2-Trypsin—rat (1ANE)	423.5	221.8	52.4	46.8	11.1	341.3	80.6	35.5	8.4	89.0
t-PA—human (1BDA)	558.9	325.2	58.2	52.8	9.4	413.2	73.9	92.9	16.6	90.6
t-PA—human (1A5H)	523.3	289.2	55.3	42.6	8.1	428.4	81.9	52.3	10.0	91.9
β-Trypsin—bovine (1CLO)	310.2	228.7	73.7	23.9	7.7	251.4	81.0	34.9	11.3	92.3
CIs—human (1ELV)	580.0	292.8	50.5	41.3	7.1	442.1	76.2	96.3	16.6	92.8
t-PA—human (1RTF)	583.9	314.4	53.8	40.8	7.0	475.9	81.5	67.2	11.5	93.0
Anionic-2-Trypsin—rat (1DPO)	487.3	258.1	53.0	23.7	4.9	378.2	77.6	85.4	17.5	95.1
Factor Xa—human (1HCG)	463.9	338.2	72.9	11.3	2.4	410.5	88.5	42.1	9.1	97.6

[a] Species from where proteins were derived follows the enzyme name.
[b] Areas calculated from the total cavity detected by the MOLCAD module in Sybyl v.6.8, Tripos Inc.
[c] Surface area within 4.0 Å from any of the benzamidine atoms. For the purposes of this analysis, benzamidine was docked into Lmm-TLE and thrombin subsites by superposition of the crystallographic complex with bovine trypsin.
[d] Areas under the following lipophilicity ranges—low: −0.19 to −0.10; medium: −0.10 to −0.01 and high: −0.01 to 0.09. Percentages are expressed in relation to total cavity area.
[e] Calculated from the sum of areas within medium and high LP.

It can be perceived from the panels in Figure 5 that once the basic complementary salt-bridge has been established, other residues in loops 1 and 2 may afford additional interactions that could determine the observed binding affinities to a number of small ligands able to fit the S1 subsite. For instance, the structural data listed in Table 2 indicate that Thr190, Asn192, and Ser195 are able to provide such additional interactions in the form of hydrogen bonds or electron pair acceptor–donor interactions. Position 190 has been recognized as an important point of differentiation between enzymes presenting additional hydrogen bonding capability, due to serine or threonine residues in this position, as in urokinase type plasminogen activator (uPA) and factor VIIa, and those presenting an alanine, like in thrombin and tPA.[23] On the other hand, position 192 has recently been questioned, as its role in macromolecular substrate recognition in TSV-PA has not confirmed the results obtained with mammalian coagulant enzymes.[18] Among other SVTLEs, it can be seen from the alignment in Figure 3 that almost any kind of residue can be found in this position. Nevertheless, within certain groups some conservation may still be encountered: the two database entries for ancrod (VSP1_AGKCO, VSP1_AGKRH)[17] present either an asparagine or lysine in this position; Lmm-TLE; batroxobin (VSPA_BOTAT)[16], bothrombin (VSPA_BOTJA),[24] and flavoxobin (VSP1_TRIFL)[15] display a smaller glycine residue while calobin (VSP1_AGKCA)[25] and bilineobin (VSP2_AGKBI)[26] present a basic amino acid in this position.

The analysis of the docked complexes also revealed that Cys191 and Asp218 could also provide further stabilization to inhibitors bearing appropriately positioned polar groups. Surprisingly, position 218 has received little attention besides being located in the entrance of the S1 pocket. This residue is absolutely nonconserved among TSV-PA, rat anionic trypsin and human thrombin (see the alignment in Fig. 2). Considering the

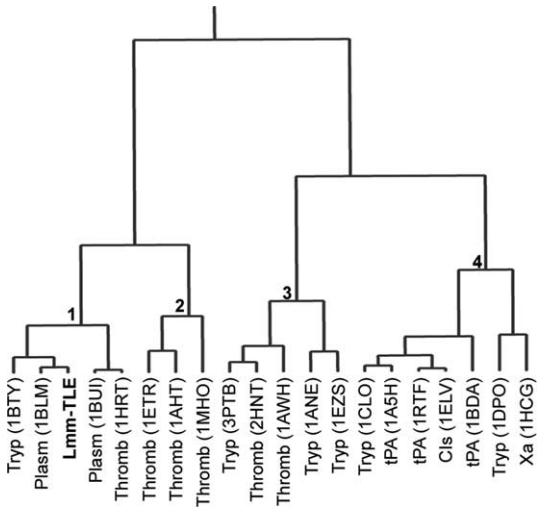

Figure 8. Dendrogram of trypsin-like enzymes characterized by the relative lipophilicity of the S1 subsite. Enzymes names are abbreviated next to the respective branch and are followed by the PDB code of the structure used in the analysis. Significant clusters are identified by the numbers 1–4 next to them. The hierarchical clustering analysis was made using complete linkage.

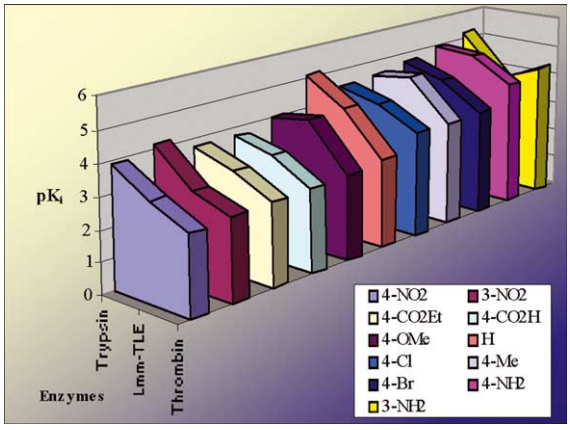

Figure 9. Trend in binding affinities of trypsin, *Lachesis m. muta* thrombin-like enzyme (Lmm-TLE), Plasmin, and thrombin to S1-directed benzamidine derivatives. For each inhibitor, the negative logarithm of the experimental inhibition constants (pK_i) are plotted according to the color code exhibited in the inset. Lmm-TLE data was retrieved from Magalhães et al.[20] For thrombin, plasmin, and trypsin, data relative to benzamidine and its 3-NO_2, 4-NO_2, 4-OMe, 3-NH_2 derivatives was obtained from Andrews et al.[79] while data corresponding to the remaining 4-substituted derivatives (Me, Cl, CO_2H, CO_2Et, Br, NH_2) was extracted from Coats et al.[80]

different biological substrates and more specifically the distinct kinetic properties of these enzymes, a relevant role for position 218 could be envisaged for these proteins. However, besides being occupied by polar residues, position 218 displays little conservation between SVTLEs (Fig. 3). For example, calobin, bothrombin, batroxobin, and Lmm-TLE display an aspartate; flavoxobin, like TSV-PA, exhibits a histidine while ancrod and bilineobin present an asparagine and lysine, respectively.

The structural data listed in Table 2 along with the inspection of the binding mode of Lmm-TLE inhibitors (Fig. 5) also indicate that the Val213 and Cys220 side chains along with the main chain atoms from Phe214, Trp215, Gly216, and Gly226 can supply mainly non-specific (vdW) interactions. Based on a theoretical 3D model of crotalase, Massova et al.[21] have proposed that the Phe214 side chain, which pointed in the direction of the catalytic His57, would turn this region of the substrate binding-groove narrower than in trypsin and thrombin that have a smaller serine in this position. Supporting this prediction, our Lmm-TLE model displays the same kind of geometry for the Phe214 side chain. Interestingly, this feature seems to be rare among SVTLEs because a serine is conserved in this position for the large majority of the enzymes aligned, as can be observed in the alignment of Figure 3. Position 213, in turn, presumably contributes significantly to the hydrophobicity of the S1 subsite since a nonpolar amino acid (valine, leucine, or alanine) is conserved among all SVTLEs (Fig. 3), trypsin, thrombin, and TSV-PA (Fig. 2). A nonpolar residue in this position should be important for increasing the binding affinities of S1 subsite ligands by increasing the lipophilicity of the S1 subsite.

Although being unable to present sufficient size to achieve high selectivity to close-related serine proteases, ligands large enough to fill just the S1 subsite are still object of much research work because the perfect anchoring in this pocket is the first step to design a potent and selective inhibitor.[27,28] In this way, the obvious applicability of a tight-binding noncovalent inhibitor has promoted numerous studies aiming to predict the binding affinities of small ligands directed to the S1 subsite of these proteases. In the case of trypsin-like serine proteases, there is a 10^5-fold preference for substrates containing an arginine or lysine in the P1 position, in relation to the next favored residue.[6] In a seminal work, Mares-Guia and Shaw have investigated this amazing specificity, using amidines and guanidines as structural analogues of the arginine side chain.[29] Following works have suggested the lipophilic feature of the interaction site based on the hydrophobic nature of the most potent amidine and guanidine derivatives as well as synthetic substrates.[30,31]

The subsequent availability of X-ray crystal structures has greatly influenced these studies.[32] For instance, in other pioneering work, Mares-Guia et al.[33] have conducted a QSAR study that has rationalized the activity of 4-substituted benzamidine derivatives by pointing out the importance of electron-releasing substituents (ERS) in position 4 of the aromatic ring in order to maximize the inhibitory potency through formation of an additional dipolar interaction with the Oγ atom of the catalytic Ser195 residue. This model has until recently represented a major paradigm in the binding of benzamidines to trypsin-like proteases. Later, the group of Hansch et al.[34] has proposed a QSAR model where hydrophobic and steric terms were added in order to improve the correlation to inhibition data. Interestingly, these authors have proposed another role for the required electrophilic character of ERS, which was to re-establish the hydrophobicity of the aromatic ring

diminished by the electron-withdrawing character of the protonated benzamidine moiety.

Recent publications have, in general, supported the importance of the hydrophobicity and nonpolar interactions of S1-directed inhibitors of trypsin-like enzymes.[35–39] Indeed, the regression coefficients of the PLSR model of Lmm-TLE inhibitors clearly show the importance of the vdW interaction term and hydration energy in order to achieve maximum affinity to the S1 subsite of this SV-TLE. This is a curious observation since the S1 subsite provides many acceptor and donor groups able to participate in more specific, directional interactions with corresponding moieties on the inhibitors structures. The Jorgensen's group has conducted a Monte Carlo simulation, in order to compute the binding affinities of 4-substituted benzamidines to trypsin, which has helped to clarify this issue.[36] The simulation revealed that mutation of the 4-amino substituent to methyl, chlorine or hydrogen introduces no net gain in the number of hydrogen bonds formed between inhibitor moieties and S1 surface groups because the number of potential hydrogen bonds lost in passing the inhibitors from the aqueous phase to the S1 pocket is constant among the studied inhibitors. Hence, the total number of hydrogen bonds involving protein atoms in the active site is nearly independent of the inhibitor and hydrogen bonding cannot therefore be invoked to rationalize the trends in binding affinities.

Based on purely continuum electrostatic grounds some authors have suggested that substitution by ERS decreases the dipolar character of the closely related 4-substituted benzamidines, decreasing in turn the free-energy penalty in desolvating these inhibitors upon association with the enzyme.[36,37] Nevertheless, when some measure of the ERS character of the substituents, like Hammett's sigma constants,[40] are plotted against pK_i only a poor correlation is obtained.[20,33,37] Thus, independent of the interpretation employed, the electronic properties of the substituent in these benzamidines cannot be invoked to rationalize binding affinities in this case. Hence, it is now possible to ask if models combining sigma constants and hydrophobicity descriptors, such as LogP,[34] have only improved the correlation to activity data by a misleading collinearity (correlation) of the descriptors employed. In fact, it has been argued that as there is only a poor relationship between LogP and inhibitory activity, one should not rationalize the binding affinities of S1-directed inhibitors in terms of their hydrophobicity.[20,37] Nevertheless, evidence for the role of a hydrophobic effect[41] in the binding of S1-directed inhibitors has been observed in a number of thermodynamic studies.[37,38,42] The results of our analysis have also clearly indicated a role for the hydrophobicity of the inhibitors if high binding affinity to the S1 subsite in Lmm-TLE is to be achieved.

Our results also show that when more structural diverse S1-directed inhibitors are considered, the overall shape complementarity of the inhibitors to the S1 cavity becomes important to maximize the binding affinity. This contribution is represented by the vdW component of the intermolecular interaction energy present in the PLSR model obtained in this work. Results from more accurate calculations with thrombin inhibitors[39] also support this finding. It is interesting to note that inclusion of the electrostatic contribution to the intermolecular interaction energy is not strictly required to describe the great majority of the variance in inhibition data. It is curious since it should be expected that the requirement for a positively charged moiety like the amidinium or guanidinium groups would be readily translated into an important electrostatic contribution to binding affinities. The analysis of Checa et al.[43] concerning the inhibition of trypsin by flavonoids has helped to shed some light in this apparent contradiction. It was concluded that vdW interactions would be dominant in describing binding affinities because there is little variation in the electrostatic term, which is largely solvent independent. Any attempt to include the electrostatic interaction term has not been successful in improving the statistics of our PLSR model. The above discussion indicates that the hydrophobic character of S1-directed inhibitors, in addition to the overall shape complementarity to the cavity and the deep salt bridge to Asp189, possesses a considerable contribution to the observed differential binding affinities among small S1-directed inhibitors of Lmm-TLE and related serine proteinases.

If such a relationship between strong binding to the S1 subsite and the hydrophobic nature of the inhibitor is as true for other trypsin-like enzymes as it is true for Lmm-TLE, it should be possible to observe a direct correlation between the order in binding affinities of these enzymes to their inhibitors and the lipophilicity of this cavity on their surfaces. A comparison between the overall order in the LP over the surfaces of S1 subsite boundaries in Lmm-TLE, trypsin, plasmin, and thrombin (Table 3) and the trend in binding affinities to S1-directed inhibitors shared by these enzymes (Fig. 9) provides general support for this hypothesis as the order of binding affinities, as expressed by pK_i, among simple 4- and 3-substituted benzamidine inhibitors follows, in most cases, the order of lipophilicity of the S1 subsite from the four proteases: thrombin $\ll$ plasmin $<$ Lmm-TLE $<$ trypsin.

It is important to note that the proposed order of lipophilicity in S1 subsites of thrombin and trypsin is the inverse of the usual order described in the literature.[44] In view of the importance of our conclusions regarding S1 subsite lipophilicity to a number of scientists working in the fields of thrombin inhibition and haemostasis control, we decided to discuss this issue in deep. At this point, it is wise to note that an inherent handicap of our lipophilicity quantification scheme is that its relative, since the range of the LP scale depends on which structures entered the analysis. Hence, our lipophilicity measure, although simple, is not absolute. For instance, the lower limit of the LP scale applied on our analysis was imposed by the high hydrophilic region on the entrance of the Lmm-TLE S1 subsite that could be attributed to an aspartate residue at position 218 (Fig. 7, panel H). A negatively charged residue at this position is common to many SVTLEs (Fig. 3) and thrombin.

The rationalization for the proposed order of lipophilicity between trypsin and thrombin that is commonly found in the literature has been primarily based on the substitution in position 190 of a serine residue in the former enzyme by an alanine residue in the later. Nevertheless, it became clear from the results of our analysis, which was based on the 3D LP (calculated from atomic hydrophobicity constants[45]), that not only the composition of the loops giving the S1 subsite architecture but also the orientation of the residues present may influence the lipophilicity of this cavity. In fact, other factors may affect the effective lipophilicity of the S1 subsite. For instance, recent crystallographic studies have evidenced that the result of the S190A (trypsin → thrombin) mutation may be counterbalanced by the presence of a conserved water molecule bridging Ala190 to the benzamidinium moiety of inhibitors.[23] Similarly, even more pronounced should be the effect of the presence of a water channel communicating with the S1 subsite, as in fact is found in thrombin. This channel contains a Na^+ binding site that has been shown to serve as an allosteric switch between the so-called 'slow' and 'fast' forms of thrombin.[46,47] Notably, for the purposes of our analysis, this channel was considered as an integral part of the S1 subsite of the thrombin structures analyzed, as evidenced in Figure 7 (panels A and F). Indeed, such channel was clearly responsible for the great proportion of low LP over the S1 subsites surfaces of thrombins (Table 3).

Dang and Di Cera[48] observed that for several serine proteases, the presence of a proline residue at position 225 could be successfully correlated with the inability of the enzyme to be allosterically regulated by Na^+. In accordance with this proposal, our analytic procedure could detect a channel communicating with the S1 subsite of factor Xa, which, like thrombin, has a Y225. Interestingly, analysis of the single CIs structure available (PDB code 1ELV) could not reveal such channel, although this enzyme also presents a tyrosine residue at position 225 (Fig. 7, panel D). Accordingly, careful inspection of the results from the Na^+ regulation experiments reported by Dang and Di Cera[48] reveals that not all serine proteases not carrying a P225 can be as much activated by Na^+ as thrombin and factor Xa can, suggesting possible drastic structural differences in the Na^+ binding sites among such enzymes. Conversely, we could detect the presence of a channel starting from the surface opposite to the active site and ending in the S1 subsite of tPA that would not be predicted by the rule of Dang and Di Cera, since this serine protease presents a P225. Whether this channel would be able to shelter a Na^+ binding site or not needs experimental confirmation.

Another interesting observation from our analysis concerning the structural differences between the fast and slow forms of thrombin was that, in agreement with the literature, we detected a much smaller accessible surface area for the S1 subsite in a putative slow form of thrombin (PDB code 1MHO) compared to the sodium-bound fast form (see Fig. 7, panels A and F). A handful of biochemical experiments collected along the years[46–51] have crystallized the idea that the slow form of thrombin would have a much tighter substrate binding cleft in order to account for the drastic decrease in the rate of substrate binding upon the fast → slow transition. In this manner, our result is consistent with the claim of Di Cera and co-workers[52] that the sodium-free structure reported by them (PDB code 1MHO) would correspond to the slow form of thrombin, although a recent publication disagrees from that.[53]

It is now clear that electrostatic interactions or differential hydrogen bonding cannot be used to explain the observed binding affinities of simple amidines/guanidines inhibiting trypsin-like enzymes. On the other hand, the combination of all nonpolar contacts that can be made by these compounds in the S1 subsite of trypsin-like enzymes, as an example of the lock-and-key concept, is the primary force contributing to maximization of binding affinities. Destabilization of the inhibitors in the aqueous phase by the hydrophobic effect also makes an important contribution to affinity for the lipophilic S1 subsite. In consideration of the stated above, it is tempting to think that in the course of evolution, SVTLEs, may have been shaped to present the catalytic efficiency of trypsin while achieving the required specificity of thrombin by different strategies, like the maximization of lipophilicity through specific mutations in residues composing the S1 subsite architecture or by employing an alternative FRE.

SVTLEs have been successfully applied in a number of clinical conditions.[10] Their therapeutic action is probably related to the fibrinogen depletion in vivo, which results in decreased blood viscosity and better blood flow.[10] However, the prolonged usage of these enzymes is inefficient due to the development of immunological resistance by the patient. Fortunately, replacement by a different SVTLE is able to by-pass this handicap in successive courses of long-term anticoagulant therapy because of the lack of cross-reactivity between two SVTLEs, for example, ancrod, and batroxobin.[10] Related studies with other STVLEs should be able to disclose such beneficial lacks of cross-reactivity within these specific venom components. Other use of SVTLEs should be envisaged in the clinical research of anomalies in the coagulation cascade of patients, since these enzymes act in more specific points of this pathway. Thus, considering all possible applications of SVTLEs, the design of potent small-molecule inhibitors aimed to manipulate the activity of these enzymes should be of great relevance. The results from this study should have application in the development of small compounds for controlling the enzymatic activity of SVTLEs in whatever biotechnological or clinical research application that these enzymes can have, in special the prevention of possible hemorrhagic counter-effects of the clinical use of these enzymes and the exploitation as chemical probes for functional-proteomics studies of snake venoms. With these straightforward computational analysis we hope to have contributed to clarify the debate over the structural features responsible for S1 binding affinities in the trypsin family, in special among the structure scarce SVTLEs.

4. Methods

4.1. Sequence analysis and multiple sequence alignment

The sequences of Lmm-TLE and other similar SVTLEs were retrieved from the SWISS-PROT or Tr-EMBL protein data banks using the Expert Protein Analysis System WWW server (ExPASy; http://www.expasy.ch/). An initial alignment of these proteinases was obtained in the ClustalW v.1.81 software,[54] using program's default parameters. This alignment was then inspected and manually corrected. An analysis of the secondary structure potential of Lmm-TLE sequence was also carried out using the PSI-PRED[55] software.

4.2. Template selection and model building

Templates were retrieved after similarity (BLAST) search with Lmm-TLE sequence against the ExNRL-3D database available through the SWISS-MODEL server (http://www.expasy.ch/swissmod).[56] Among the highest scoring structures, chain B of TSV-PA,[18] rat anionic trypsin bound to benzamidine[57] and chain H of human α-thrombin in complex with a tripeptide phosphonate inhibitor[58] (Ex-PDB codes 1BQYB, 1ANE_ and 1H8DH, respectively) were selected as templates for the homology modeling procedure. An initial alignment between Lmm-TLE and its templates was accomplished by combining the information derived from the multiple sequence alignment of SVTLEs described above to the S1A chymotrypsin family alignment retrieved from the MEROPS database of peptidases.[1] Improvement of this alignment was accomplished by maximizing the superposition of secondary structural elements and conserved functional residues within the chymotrypsin family. Following this optimal alignment, the sequence of Lmm-TLE was folded over its templates structures using the SwissPDB Viewer v.7.2b2 software.[59] Preliminary model building was accomplished by an automated procedure through the use of the optimize mode in SWISS-MODEL server. Several cycles of alignment refinement and model construction were carried out. The modeling progress was accompanied through the use of Ramachandran plots generated with the PROCHECK software.[60] All operations conducted in SwissPDB Viewer were performed on a Pentium III 800 MHz personal computer.

4.3. Model refinement and validation

All energy minimization operations described in this section were conducted with the GROMOS96 v.43B1 force field[61] implemented in SwissPDB Viewer v.3.7b2 running on a Pentium III 800 MHz personal computer. At first, any missing peptide bonds in the Lmm-TLE model were introduced and locally minimized by 20 steps of steepest descent followed by 100 steps of conjugated gradients minimization methods, until the energy difference between two steps was below 0.01 kJ/mol. Two successive refinements of the Lmm-TLE model were performed using a similar minimization protocol in which the type of residues constrained during the minimization were varied: initially, only residues out of the most favored regions of the Ramachandran plot (53 total) were allowed to move; then residues with high model B-factor and/or force field energy were included (37 total). The final refined theoretical structure of Lmm-TLE was achieved by an energy minimization with a harmonic constraint over all residues showing a low model B-factor. Refinement progress was accompanied by the distribution of Lmm-TLE residues over the most favored and allowed regions of the Ramachandran plot. The geometry of Asp189 and the residues composing the catalytic triad (His57, Asp102, and Ser195) received special attention due to their importance is this work. When necessary, manual torsions were imposed on the dihedral angles of these residues in order to reproduce the conserved architecture of the S1 subsite and active site of trypsin-like serine proteinases. A thorough stereochemical validation of the final Lmm-TLE model was achieved by submitting the structure to the analytical softwares contained in the PROCHECK and WHATCHECK packages.[60,62] An extensive functional crossvalidation of the model was also performed by assuring that any biochemical information available to Lmm-TLE was in accordance to the structural features predicted by the proposed 3D model.

4.4. Small compound modeling and calculation of hydration energy

The structures for the set of 21 amidine and guanidine inhibitors (Fig. 1) reported in the work of Magalhães et al.[20] were constructed in PC Spartan Pro release 1.05 (Wavefunction, Inc., Irvine, CA, USA; kindly made available by Dr. Ricardo B. de Alencastro and Dr. Magaly G. Albuquerque from Lab. de Modelagem Molecular, DQO/IQ, UFRJ, Brazil) using the semi-empirical AM1 Hamiltonian.[63] All amidinium and guanidinium groups were protonated and the carboxylic group of inhibitor **11** was constructed in the ionized form. Atomic point charges needed for the docking procedure were calculated through fitting of the electrostatic potential. Hydration energy (ΔE_{hyd}) was calculated using the SM5.4/A model[64] implemented in PC Spartan Pro. All calculations were performed in a Pentium II-MMX 350 MHz personal computer.

4.5. Docking calculations

Automated flexible molecular docking simulations were carried out using the DOCK[65] v4.0 program suite, which takes into account only the flexibility of the ligand. This seems a reasonable approximation here, since no major conformational change has been observed in the bound structures of small S1-directed inhibitors with a number of serine proteases of the chymotrypsin family.[4] The docking procedure can be divided in four general stages: ligand/receptor preparation, site characterization, scoring grid calculation, and docking itself. The set of ligands to be docked consisted of the 21 amidine and guanidine derivatives,[20] shown in Figure 1, constructed

as described above. The final refined Lmm-TLE modeled structure was used as the receptor. In order to prepare the receptor for energetic grid calculation, all hydrogen atoms were added and then kollman-type[66] atomic point charges were assigned. Preliminary site characterization was accomplished by running the autoMS script distributed within the DOCK package, which automatically generates a complementary sphere representation of cavities found in the receptor surface calculated by the *ms* program (program number 429, Quantum Chemistry Program Exchange, Indiana University). A better characterization of the S1 subsite in Lmm-TLE model was achieved by molecular graphics assisted editing of the spheres cluster generated by autoMS. This procedure was made in SwissPDB Viewer v.3.7b2 software and was guided by the analysis of the volume occupied by several S1-directed inhibitors bound to trypsin-like enzymes. In order to obtain a set of input parameters for the grid and dock programs able to reproduce the observed binding mode for the class of inhibitors considered in this work we have used as a guide the crystallographic complex of benzamidine and rat anionic trypsin (PDB code 1ANE)[57]. Input parameters governing ligand flexibility, orientation, receptor matching, scoring, and minimization were varied until the crystallographic binding mode of benzamidine in rat trypsin could be reproduced with an RMS deviation of less than 1.0 Å. Docking calculations were performed in the SGI Origin 2000 server installed at FIOCRUZ/Scientific Computation Program (PROCC).

4.6. Analysis of docked complexes of amidine and guanidine competitive inhibitors of Lmm-TLE

Besides intermolecular energy terms (vdW and electrostatic) furnished by the dock program, the simulated complexes had their nonbonding interactions (hydrogen bonds and nonpolar contacts) mapped using the HBPLUS program.[67] Schematic plots of these interactions were generated using the Ligplot software.[68] In addition, the docked complexes were inspected using molecular graphics in SwissPDB viewer v7.2b2 and the receptor solvent accessible area buried (AAB) upon binding was calculated for each inhibitor according to the following formula: $\mathrm{AAB} = (A + B) - AB$, where A is receptor solvent accessible area, B is the ligand solvent accessible area and AB is the solvent accessible area of the complex. A probe radius of 1.4 Å, representing the vdW radius of water, was used in these calculations.

4.7. Binding affinity calculation background

The methodology employed by us stems from the additivity principle in calculating binding affinities,[69,70] that is, different contributions to free energy of binding can be calculated separately and summed. The resulting 'master equation' usually accounts for the contributions due to solvent, conformational changes in the protein and ligand, the protein–ligand interactions and variation in the degrees of freedom of the diverse molecular motions involved.[70] Partial least square regression (PLSR) algorithm[71] can be used to weight out each of these terms in order to maximize the explanation of the observed property, in this case, $\mathrm{p}K_\mathrm{i}$. The latter approach has very straight relationship to quantitative structure–activity relationship studies (QSAR).[35]

It has been suggested that in the case of small S1-directed inhibitors of trypsin-like enzymes, the electrostatic portion of the intermolecular interaction energy is almost constant for a series of congeneric ligands, contributing very little to the variance in binding affinities.[43] Thus, we have opted to explore the ability of the scoring function from the docking program to approximate the major enthalpic component of the binding free energy in this case, that is the intermolecular vdW energy, while the other components related to the solvent contribution to the observed order in binding affinities are entered in the regression equation in the form of calculated ligand/complexes energetic and structural parameters like ΔE_hyd and AAB, respectively: $\mathrm{p}K_\mathrm{i} = k + aE_\mathrm{vdw} + b\Delta E_\mathrm{hyd} + c\mathrm{AAB}$, where E_vdw is the vdW portion of the intermolecular scoring function of dock 4.0 program; ΔE_hyd is the hydration energy of the ligand calculated according to the SM5.4/A model,[64] AAB is the solvent accessible area of the protein that is buried upon inhibitor binding, k is a constant and a, b, and c are the regression coefficients. It is interesting to note that we have used ΔE_hyd in place of cLogP[72] as a hydrophobicity parameter because ΔE_hyd has included ionic compounds in its parameterization and as a semi-empirical quantum chemical calculation, presents much less restriction about compound classes or ionization states to be submitted to calculations. On the other hand, cLogP has an intrinsic handicap with ionized compounds, since calculated values are derived from parameters fitted from experimental data with neutral or ionizable compounds in the neutral form from buffered solutions.[72]

The PLS method models the experimental data (Y or dependent variable) in terms of calculated descriptors of the complex/inhibitors (X block or independent variables). This robust chemometric method is superior to the classical multiple linear regression (MLR) in being able to deal with a great number of (even highly correlated) X variables by reducing this large data set to a few latent variables (LV) maximum correlated to the Y variable.[71] The importance of each original X variable to the LV is determined by its loading. The number of significant LVs to be used is usually determined by crossvalidation.[73] This regression approach, although not based in the rigorous statistical thermodynamics grounds of more realistic simulations[74] is high computationally efficient and has been successfully applied in a number of cases.[35,75]

4.8. Calculation of binding affinities

Experimental binding affinity data, expressed as the negative logarithm of the inhibitory constants ($\mathrm{p}K_\mathrm{i}$), for the series of amidine and guanidine competitive S1-directed inhibitors of Lmm-TLE were computed from the work of Magalhães et al.[20] (Table 1). The PLS method

implemented in the MSS/QSAR module of Sybyl v.6.8 from Tripos Inc. (St. Louis, MO, USA) was used to obtain the MLR-like regression coefficients of the model. The statistical significance of these was calculated by the bootstrap method.[76] Overall statistical relevance of the model was judged by the correlation coefficient (r^2), p value, standard deviation (s) and F-ratio. Predictive ability of the model was evaluated by the crossvalidation coefficient (Q^2) and standard error of predictions (SEP).

4.9. Calculation of the lipophilic potential over the surface around S1 subsite

For this analysis, besides the final Lmm-TLE model, 19 structures of trypsin-like enzymes were used: 6 trypsins (PDB codes 1BTY, 3PTB, 1CLO, 1ANE, 1EZS, and 1DPO), 6 thrombins (PDB codes 1ETR, 1HRT, 2HNT, 1AWH, 1MHO, and 1AHT), 3 t-PAs (PDB codes 1RTF, 1A5H, and 1BDA), 2 plasmins (PDB codes 1BUI and 1BLM), CIs (PDB code 1ELV), and Factor Xa (PDB code 1HCG). Structures were selected in order to account for possible effects due to experimental conditions, inter-species variation, and post-translational processing. All manipulations and calculations described in this section were conducted in the Sybyl v.6.8 modeling package from Tripos, Inc. (St. Louis, MO, USA). First, the PDB structures were freed from any ligands and water molecules followed by addition of all hydrogen atoms. Next, the molecular surface around the S1 subsite (constituting the major cavity in the surface of these structures) was calculated by the MOLCAD module. Then, the LP generated from the atomic hydrophobicity constants derived by Ghose and Crippen[77] was mapped on the surface surrounding the S1 subsite, according to the method of Furet et al.[45] implemented in MOLCAD.

4.10. Analysis of the binding affinities of trypsin-like enzymes to benzamidine derivatives according to the relative lipophilicity of the S1 subsite cavity

At first, in order to have a simple measure of the S1 subsite lipophilicity, the LP was put in the same scale for all enzymes analyzed and the areas belonging to three arbitrary property ranges were measured: low (−0.19 to −0.10), medium (−0.10 to −0.01), and high (−0.01 to 0.09); each of these was expressed as percent of the total cavity surface area and the total subsite lipophilicity was considered as the sum over the medium and high LP percentages. Following, a classification of the 20 structures analyzed was proposed based on a complete linkage, hierarchical-clustering analysis[78] (HCA) of S1 subsite lipophilicities (MSS module, Sybyl molecular modeling package). Natural groups were identified by careful visual inspection of the dendrogram produced by the analysis. Finally, the pK_i values[79,80] of several benzamidine-based inhibitors of trypsin, plasmin, thrombin, and Lmm-TLE were plotted to investigate the variation of binding affinities according to S1 subsite lipophilicities.

Acknowledgements

We thank Prof. R. B. De Alencastro and C. A. F. Oliveira for carefully reading this manuscript. We are in debt with Prof. O. A. C. Antunes and Prof. M. G. Albuquerque for valuable corrections and suggestions. We also thank the referee for generous suggestions.

References and notes

1. Rawlings, N. D.; O'Brien, E. A.; Barrett, A. J. *Nucleic Acids Res.* **2002**, *30*, 343.
2. Lesk, A. M.; Fordham, W. D. *J. Mol. Biol.* **1996**, *258*, 501.
3. Meloun, B.; Kluh, I.; Kostka, V.; Moravek, L.; Prusik, Z.; Vanacek, J.; Keil, B.; Sorm, F. *Biochim. Biophys. Acta* **1966**, *130*, 543.
4. Hedstrom, L. *Chem. Rev.* **2002**, *102*, 4501.
5. Schecter, A.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157.
6. Czapinska, H.; Otlewski, J. *Eur. J. Biochem.* **1999**, *260*, 571.
7. Colomb, E.; Guy, O.; Deprez, P.; Michel, R.; Figarella, C. *Biochim. Biophys. Acta* **1978**, *525*, 186.
8. Perona, J. J.; Craik, C. S. *Protein Sci.* **1995**, *4*, 337.
9. Stubbs, M. T.; Bode, W. *Trends Biochem. Sci.* **1995**, *20*, 23.
10. Markland, F. S. *Toxicon* **1998**, *36*, 1749.
11. Matsui, T.; Fujimura, Y.; Titani, K. *Biochim. Biophys. Acta* **2000**, *1477*, 146.
12. Aguiar, A. S.; Alves, C. R.; Melgarejo, A.; Giovanni-De-Simone, S. *Toxicon* **1996**, *34*, 555.
13. Silva, N. J.; Aird, S. D.; Seebart, C.; Kaiser, I. I. *Toxicon* **1989**, *27*, 763.
14. Magalhães, A.; Fonseca, B. C. B.; Diniz, C. R.; Gilroy, G. P.; Richardson, M. *FEBS Lett.* **1993**, *329*, 116.
15. Itoh, N.; Tanaka, N.; Funakoshi, I.; Kawasaki, T.; Mihashi, S.; Yamashima, I. *Nucleic Acids Res.* **1988**, *16*, 10377.
16. Burkhart, W.; Simth, G. F. H.; Su, J.-L.; Parikh, I.; Levine, H. *FEBS Lett.* **1992**, *297*, 297.
17. Zhang, Y.; Wisner, A.; Xiong, Y. L.; Bon, C. *J. Biol. Chem.* **1995**, *270*, 10246.
18. Parry, M. A.; Jacob, U.; Huber, R.; Wisner, A.; Bon, C.; Bode, W. *Structure* **1998**, *6*, 1195.
19. Castro, H. C.; Silva, D. M.; Craik, C.; Zingali, R. B. *Biochim. Biophys. Acta* **2001**, *1547*, 183.
20. Magalhães, A.; Monteiro, M. R.; Magalhães, H. P. B.; Mares-Guia, M.; Rogana, E. *Toxicon* **1997**, *35*, 1549.
21. Massova, I.; Pirkle, H.; Edwards, B. F. P.; Mobashery, S. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 3139.
22. Bartunik, H. D.; Summers, L. J.; Bartsch, H. H. *J. Mol. Biol.* **1989**, *210*, 813.
23. Katz, B. A.; Sprengeler, P. A.; Luong, C.; Verner, E.; Elrod, K.; Kirtley, M.; Janc, J.; Spencer, J. R.; Breitenbucher, J. G.; Hui, H.; McGee, D.; Allen, D.; Martelli, A.; Mackman, R. L. *Chem. Biol.* **2001**, *8*, 1107.
24. Nishida, S.; Fujimura, Y.; Miura, S.; Ozaki, Y.; Usami, Y.; Suzuki, M.; Titani, K.; Yoshida, E.; Sugimoto, M.; Yoshioka, A.; Fukui, H. *Biochemistry* **1994**, *33*, 1843.
25. Hahn, B. S.; Yang, K. Y.; Park, E. M.; Chang, I. M.; Kim, Y. S. *J. Biochem.* **1996**, *119*, 835.
26. Nikai, T.; Ohara, A.; Komori, Y.; Fox, J. W.; Sugihara, H. *Arch. Biochem. Biophys.* **1995**, *318*, 89.
27. Rawlings, N. D.; Barret, A. J. *Methods Enzymol.* **1994**, *244*, 19–61.
28. Sanderson, P. E. J. *Med. Res. Rev.* **1999**, *19*, 179.
29. Mares-Guia, M.; Shaw, E. *J. Biol. Chem.* **1965**, *240*, 1579.

30. Mares-Guia, M.; Figueiredo, A. F. S. *Biochemistry* **1970**, *9*, 3223.
31. Mares-Guia, M.; Shaw, E.; Cohen, W. *J. Biol. Chem.* **1967**, *242*, 5777.
32. Bode, W.; Schwager, P. *J. Mol. Biol.* **1975**, *98*, 693.
33. Mares-Guia, M.; Nelson, D. L.; Rogana, E. *J. Am. Chem. Soc.* **1977**, *99*, 2331.
34. Recantini, M.; Klein, T.; Yang, C.-. Z.; McClarin, J.; Langridge, R.; Hansch, C. *Mol. Pharmacol.* **1986**, *29*, 436.
35. Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681.
36. Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663.
37. Talhout, R.; Engberts, J. B. F. N. *Eur. J. Biochem.* **2001**, *268*, 1554.
38. Dullweber, F.; Stubbs, M. T.; Musil, D.; Stürzebecher, J.; Klebe, G. *J. Mol. Biol.* **2001**, *313*, 593.
39. Guimaräes, C. R. W.; De Alencastro, R. B. *J. Phys. Chem. B* **2002**, *106*, 466.
40. Hammett, L. P. *Chem. Rev.* **1935**, *17*, 125.
41. Sharp, K. A.; Madam, B. *J. Phys. Chem. B* **1997**, *101*, 4343.
42. Rogana, E.; Penha-Silva, N.; Mares-Guia, M. *Brazilian J. Med. Biol. Res.* **1989**, *22*, 1177.
43. Checa, A.; Ortiz, A. R.; Pascual-Tereza, B.; Gago, F. *J. Med. Chem.* **1997**, *40*, 4136.
44. Tapparelli, C.; Metternich, R.; Ehrhardt, C.; Cook, N. S. *Trends Pharmacol. Sci.* **1993**, *14*, 366.
45. Furet, P.; Sele, A.; Cohen, N. C. *J. Mol. Graphics* **1988**, *6*, 182.
46. Di Cera, E.; Guinto, E. R.; Vindigni, A.; Dang, Q. D.; Ayala, Y. M.; Wuyi, M.; Tulinsky, A. *J. Biol. Chem.* **1995**, *270*, 22089.
47. Zhang, E.; Tulinsky, A. *Biophys. Chem.* **1997**, *63*, 185.
48. Dang, Q. D.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10653.
49. Wells, C. M.; Di Cera, E. *Biochemistry* **1992**, *31*, 11721.
50. Guinto, E. R.; Vindigni, A.; Ayala, Y. M.; Dang, Q. D.; Di Cera, E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11185.
51. Lai, M.-T.; Di Cera, E.; Shafer, J. A. *J. Biol. Chem.* **1997**, *272*, 30275.
52. Pineda, A. O.; Savvides, S. N.; Waksman, G.; Di Cera, E. *J. Biol. Chem.* **2002**, *277*, 40177.
53. Huntington, J. A.; Esmon, C. T. *Structure* **2003**, *11*, 469.
54. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673.
55. Jones, D. T. *J. Mol. Biol.* **1999**, *292*, 195.
56. Peitsch, M. C. *Biochem. Soc. Trans.* **1996**, *24*, 274.
57. Perona, J. J.; Tsu, C. A.; Craik, C. S.; Fletterick, R. J. *J. Mol. Biol.* **1993**, *230*, 919.
58. Skordalakes, E.; Dodson, G. G.; Green, D. S.; Goodwin, C. A.; Scully, M. F.; Hudson, H. R.; Kakkar, V. V.; Deadman, J. J. *J. Mol. Biol.* **2001**, *311*, 549.
59. Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714.
60. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283.
61. van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.
62. Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272.
63. Dewar, M. J. S.; Zoebisch, E. G.; Horsley, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
64. Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.
65. Kuntz, I. D. *Science* **1992**, *257*, 1078.
66. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
67. McDonald, I. K.; Thornton, J. M. J. *Mol. Biol.* **1994**, *238*, 777.
68. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Prot. Eng.* **1995**, *8*, 127.
69. Dill, K. A. *J. Biol. Chem.* **1997**, *272*, 701.
70. Ajay Murko, M. A. *J. Med. Chem.* **1995**, *38*, 4953.
71. Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109.
72. Leo, A. *Chem. Rev.* **1993**, *93*, 1281.
73. Topliss, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, *21*, 1238.
74. Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.
75. Head, R. D.; Smythe, M. L.; Oprea, T. I.; Walker, C. I.; Green, S. M.; Marshall, G. R. *J. Am. Chem. Soc.* **1996**, *118*, 3959.
76. Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18.
77. Ghose, A. K.; Crippen, G. M. *J. Comp. Chem.* **1986**, *7*, 565.
78. Cormarck, R. M. *J. Royal Stat. Soc.* **1971**, *A134*, 321.
79. Andrews, J. M.; Roman, D. P., Jr.; Bing, D. H. *J. Med. Chem.* **1978**, *21*, 1202.
80. Coats, E. A. *J. Med. Chem.* **1973**, *16*, 1102.