J Mol Model (2012) 18:2673–2687
DOI 10.1007/s00894-011-1290-2

ORIGINAL PAPER

Predicting functional residues of the *Solanum lycopersicum* aspartic protease inhibitor (SLAPI) by combining sequence and structural analysis with molecular docking

Yasel Guerra · Pedro A. Valiente · Colin Berry · Tirso Pons

Received: 16 September 2011 / Accepted: 25 October 2011 / Published online: 20 November 2011

Abstract The *Solanum lycopersicum* aspartic protease inhibitor (SLAPI), which belongs to the STI-Kunitz family, is an effective inhibitor of the aspartic proteases human cathepsin D and *Saccharomyces* proteinase A. However, in contrast with the large number of studies on the inhibition mechanism of the serine proteases by the STI-Kunitz inhibitors, the structural aspects of the inhibition mechanism of aspartic proteases from this family of inhibitors are poorly understood. In the present study, we have combined sequence and structural analysis methods with protein-protein docking to gain a better understanding of the SLAPI inhibition mechanism of the proteinase A. The results suggest that: *i*) SLAPI loop L9 may be involved in the inhibitor interaction with the proteinase A´s active site, and *ii*) the residues I144, V148, L149, P151, F152 and R154 are implicated in the difference in the potency shown previously by SLAPI and another STI-Kunitz inhibitor isolated from *Solanum tuberosum* to inhibit proteinase A. These results will be useful in the design of site directed mutagenesis experiments to understand more thoroughly the aspartic protease inhibition mechanism of SLAPI and other related STI-Kunitz inhibitors.

Keywords Aspartic protease inhibitor · Comparative 3D modeling · Functional residue identification · Protein-protein docking · Protein-protein interface · STI-Kunitz inhibitor

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1290-2) contains supplementary material, which is available to authorized users.

Y. Guerra (✉) · P. A. Valiente · T. Pons
Centro de Estudio de Proteínas (CEP), Facultad de Biología,
Universidad de La Habana,
Calle 25 No. 455 % J e I,
Vedado CP 10400, La Habana, Cuba
e-mail: yaselg@gmail.com

C. Berry
Cardiff School of Biosciences, Cardiff University,
Cardiff CF10 3AT, Wales, UK

Present Address:
T. Pons (✉)
Structural Biology and Biocomputing Programme,
Spanish National Cancer Research Centre (CNIO),
C/Melchor Fernández Almagro 3,
Madrid 28029, Spain
e-mail: tpons@cnio.es

Introduction

Aspartic proteases are involved in a wide range of biological processes, including the pathogenesis of many diseases as well as many physiological roles [1, 2]. Since aspartic proteases play major roles in HIV infection, malaria, fungal infections such as candidiasis, and cancer, inhibitors to these enzymes are considered as potential therapeutic agents [1, 3]. However, in contrast with the wealth of data and widespread distribution of naturally-occurring proteinaceous protease inhibitors of serine and cysteine proteases, known proteinaceous aspartic protease inhibitors (APIs) are rare and unevenly distributed among classes of organisms [4]. To date, APIs are distributed in at least six families of proteinaceous inhibitors in the MEROPS database: Kunitz_legume inhibitors (family I3) [5–7], the Ascaris inhibitors (family I33) [8], the yeast inhibitor IA_3 (family I34) [9], a domain of the sea anemone *Actinia equina* inhibitor Equistatin (family I31) [10], the pig serpin inhibitor (family I4) [11] and the squash inhibitor SQAPI (familyI25) [12]. Nevertheless, only the yeast inhibitor IA_3 and the *Ascaris sum* inhibitor PI-3 are uniquely inhibitors of the aspartic proteinase family with

the other four being recruited from families with different inhibitory specificities [13, 14]. It has been suggested that the inhibitory mechanism developed for the new target in the latter APIs could retain similarities with those used by the families from which they were recruited [14]. The only proteinaceous API three-dimensional (3D) structures available are for IA_3, PI-3 and SQAPI, showing very different folds and inhibition mechanisms [13, 15, 16]. In the case of SQAPI, mutagenesis and docking results [13] suggested that the mechanism used by this molecule to inhibit pepsin is similar to that used by cystatin, with which it shares structural similarity, to inhibit cysteine proteases.

APIs belonging to the Kunitz_legume family are examples of recruitment of a serine protease inhibitor to develop inhibitory activity against a new protease target. These inhibitors show activity against the serine proteases trypsin (EC 3.4.21.4), chymotrypsin (EC 3.4.21.1) and aspartic proteases such as cathepsin D (EC 3.4.23.5) and yeast proteinase A (EC 3.4.21.41) [6, 7, 17]. At least ten isoforms have been isolated from *Solanum tuberosum* (potato) and three from *Solanum lycopersicum* (tomato). Most of them have been sequenced and in some cases biochemically characterized, exhibiting inhibition constants in the nanomolar range against cathepsin D and proteinase A [7]. However, neither a representative 3D-structure of these inhibitors nor studies related to their inhibitory mechanism have been reported. In fact, all the residues suggested to be implicated in the inhibition of the aspartic proteases have been identified based exclusively on sequence differences from other Kunitz inhibitors (particularly the soybean Kunitz inhibitor (STI)) [5, 6].

In the present work we predicted the amino acid residues involved in the inhibition of the aspartic protease proteinase A (EC 3.4.21.41) by a Kunitz_legume inhibitor isolated from *Solanum lycopersicum* (SLAPI, UniProt ID: Q9LEC1_SOLLC) [7]. This prediction was based on a combination of sequence analysis, comparative protein modeling and protein-protein docking. To the best of our knowledge, this is the first time that SLAPI-Loop L9 is proposed as involved in the interaction with the proteinase A´s active site. We also extended these predictions to other APIs from the Kunitz_legume family.

Materials and methods

The methodology followed here to identify new functional residues from the inhibitor SLAPI is presented in a flowchart (Fig. 1). We combined the information derived from sequence and structural analysis with protein-protein docking to increase the reliability of these results.

Sequence and structure analysis

Sequences and 3D structures of Kunitz_legume family inhibitors were retrieved from the UniProt (http://www.uniprot.org/) and PDB (http://www.rcsb.org/pdb/home/) databases, respectively. Position-specific iterated BLAST (PSI-BLAST) against the NCBI non-redundant database (nrNCBI) (http://www.ncbi.nlm.nih.gov) was used to identify SLAPI-related sequences. Sequence alignments with $E_{value}<10^{-3}$ and with a bit score >100 were considered significant. Multiple sequence alignment (MSA) of 532 sequences and the phylogenetic tree for the Kunitz_legume were downloaded from Pfam database (http://www.sanger.ac.uk/Pfam/) with the accession code PF00197.

A total of 22 sequences related to SLAPI (UniProt ID: Q9LEC1_SOLLC) were selected to generate a MSA using the profile menu of the program ClustalX [18]. The 22 sequences selected were taken according to: *i*) the seed_alignment available for the Kunitz_legume family in the Pfam database, *ii*) the Kunitz_legume family members with crystal structures available, and *iii*) two biochemically characterized APIs isolated from *Solanum tuberosum* (API-9 and API-11). The seed_alignment is provided by the Pfam database as a representative alignment for a particular family. The resulting MSA was manually parsed by analyzing the gaps, conserved amino acid regions and secondary structure information using SeaView software [19]. The crystal structures of nine of the 23 proteins used in the MSA were analyzed to extract the secondary structure information (PDB IDs: 3iir, 2qn4, 1r8n, 1avw, 1eyl, 1ava, 1tie, 1wba, 2dre). A consensus method implemented by the Phyre web server [20] available at (http://www.sbg.bio.ic.ac.uk/~phyre/) was used to predict the secondary structure in those proteins without 3D structures solved. The consensus method includes PsiPred [21], Jnet [22] and SSPRO predictions [23].

A parallel search was conducted in the InterPro(http://www.ebi.ac.uk/interpro/ISpy) and Prosite (http://www.expasy.org/prosite) databases looking for specific sequence motifs or fingerprint annotations.

With the aim of predicting potential functional residues we used the TreeDet web server [24] (http://www.pdg.cnb.uam.es/servers/treedet) and the sequence hypervariability analysis [25]. The TreeDet web server integrates the results from three methods: level entropy (S-method) [26], mutational behavior (MB-method) [26] and S3Det [27].

Comparative three-dimensional modeling

The 3D models of the tomato inhibitor (SLAPI) were generated using the MODELLER software [28] and the template crystallographic structures of the Kunitz_legume inhibitors: miraculin-like protein from *Murraya koenigii*

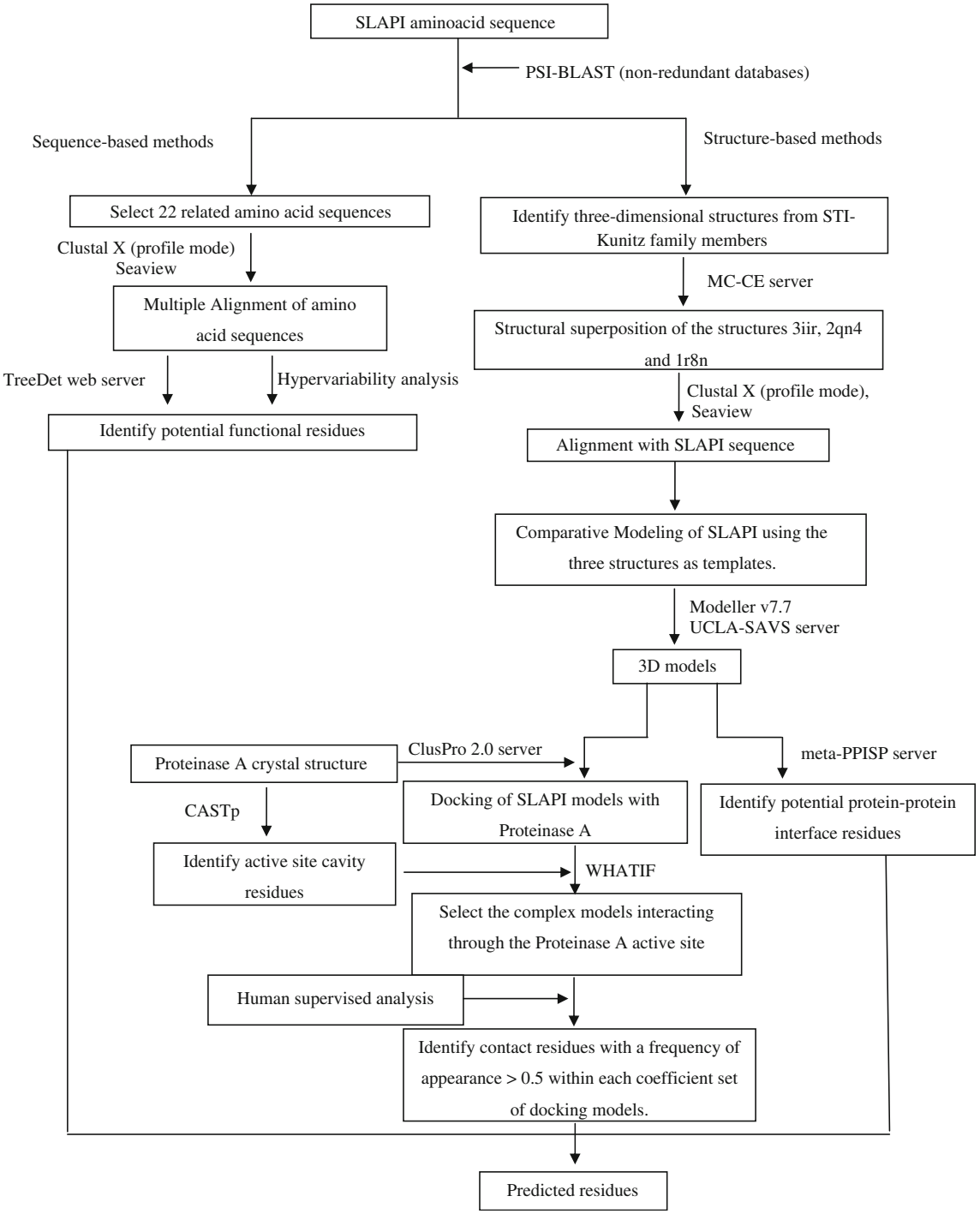


Fig. 1 Flowchart of the methodology followed to identify putative SLAPI functional residues. Residues were identified by combining the information derived from the sequence analysis, comparative modeling, protein-protein interface residue prediction and docking

(PDB ID: 3iir), α-amylase/subtilisin inhibitor from *Oryza sativa* (PDB ID: 2qn4) and serine protease inhibitor from *Delonix regia* (PDB ID: 1r8n). First, a profile was generated by the structural superposition of the crystallographic structures using MC-CE [29] (http://pathway.rit.albany.edu/~cemc/). Second, the SLAPI sequence was aligned by respect to the profile using CLUSTAL_X [18]. The multiple sequence alignment obtained was manually parsed by analyzing the gaps, conserved amino acid positions and the SLAPI secondary structure prediction. Twenty-five models per template were calculated with the spatial restraints extracted from the target-template alignment. In the case of template 2qn4, two alternative alignments with SLAPI were used (see supplementary

material for details). The 3D models obtained were evaluated using the UCLA Structural Analysis and Verification Server tools: PROCHECK [30], WHAT_CHECK [31] and VERIFY_3D [32, 33] available at (http://www.doe-mbi.ucla.edu/Services/SV/). All the models that fulfilled the following requirements were selected: no errors in the Rachamandran plot from PROCHECK, more than 65% of the residues with a sequence-to-structure compatibility (3D-1D) score higher than 0.2 in VERIFY_3D and a 2nd generation packing quality value better than −3.0 in WHAT_CHECK.

Principal component analysis (PCA) was performed to describe how heterogeneous are the 3D structures of the chosen models. Briefly, the input is an n by p coordinate matrix, X, where n is the number of structures and p is three times the number of the atoms. Each row in X represents the backbone atoms of each model. From X, the elements of the covariance matrix, C, were calculated as:

$$C_{ij} = < (x_i - < x_i >)(x_j - < x_j >) >, \quad (1)$$

where averages over the n structures are indicated by the brackets <>. The covariance matrix, C, can be decomposed as:

$$C = P \Delta P^T [2], \quad (2)$$

where the eigenvectors, P, represent the principal components (PCs) and the eigenvalues are the elements of the diagonal matrix, Δ. The eigenvalues are sorted in descending order. Each eigenvalue is directly proportional to the variance it captures in its corresponding PC.

Protein-protein interface residues prediction using meta-PPISP web server

To predict potential protein-protein interaction sites the selected 3D models of SLAPI were analyzed using the meta-PPISP web server [34] (http://pipe.scs.fsu.edu/meta-ppisp.html). The meta-PPISP web server is built on three individual methods: cons-PPISP, a neural network predictor that uses sequence profiles and solvent accessibilities of spatially neighboring residues as input; Promate, which uses a composite probability calculated from properties such as secondary structure, atom distribution, amino-acid pairing, and sequence conservation, and PINUP, based on an empirical energy function consisting of a side-chain energy term, a term proportional to solvent accessible area, and a term accounting for sequence conservation [35–37]. In meta-PPISIP, the three methods are combined in a linear regression analysis with the raw scores as input. All the residues with a final score >0.34 (default score used by the server) were considered as a positive prediction. According to the developers of the meta-PPISP server, the threshold value used (0.34) gives an equal number of predicted and actual interface residues [34].

Protein-protein docking

The SLAPI 3D models selected previously were docked with a proteinase A crystal structure (PDB ID: 1dpj, after removal of the chain corresponding to the IA_3 inhibitor), using the web server ClusPro version 2.0 (http://nrc.bu.edu/cluster) [38]. The fully automated docking server ClusPro was executed with the default parameters and, as a result, it provided four sets of models based on four different coefficients: Balanced, Electrostatic, Hydrophobic and WdV+Electrostatic [39]. The docking models calculated for all coefficients were downloaded and analyzed with the WHATIF program [40]. We also used the CASTp web server (http://sts-fw.bioengr.uic.edu/castp) to examine the crystallographic structure of proteinase A in order to identify the amino acid residues forming the active site cavity. All amino acid residues of SLAPI with at least one atom at less than 4 Å from any atom from proteinase A were considered as contact residues. First, the solutions where the inhibitor interacts with the proteinase A active site cavity residues were selected and only docking models with at least one contact residue in the proteinase A active site cavity were considered as valid.

Second, within each coefficient set of docking models the frequency of appearance (f_i) of every SLAPI contact residue among all the valid models was calculated. Frequency of appearance is the ratio of No. of models where residue i is a contact residue and the No. valid models. Also, all the inhibitor contact residues within 4 Å distance from the proteinase A catalytic aspartic acids (D32 and D215 –pepsin numbering) were identified.

Results and discussion

Comparative protein modeling of the STI-Kunitz inhibitor SLAPI

It is generally accepted that the 3D structures of proteins within a family are more conserved than their sequences [41]. Therefore, if similarity between two proteins is detected at sequence level, structural similarity can usually be assumed. Based on the Kunitz_legume inhibitor family 3D-structures annotated in the PDB, we selected the crystal structures of the bifunctional alpha-amylase/subtilisin inhibitor from *Oryza sativa* (PDB ID: 2qn4) and trypsin inhibitor from *Delonix regia* (PDB ID: 1r8n) as templates, considering their sequence identity (>25%) and high resolution (1.8 and 1.75 Å, respectively). We also selected the crystal structure of the miraculin-like protein from

Murraya koenigii (PDB ID: 3iir), despite its medium resolution (2.9 Å), since it is the only 3D-structure available for a member of this family containing 3 disulfide bridges. As result of the selection process (see Materials and methods for details), only five of 100 3D models calculated, were selected. The structural comparison of the five 3D models showed that they keep the typical β-trefoil architecture of the STI-Kunitz inhibitor family. The highest differences (RMSD>4 Å) are localized in the N-terminal region, and in the loops connecting the secondary structure elements, mainly loops L2, L6, L7, L8 and L9 (Fig. 2).

The 3D models selected were named according to the template used and their sequential number in the 25 models calculated per template. The five models selected were: 3iir-4, 3iir-19, 3iir-22, 2qn4-9 and 2qn4-19 and the parameters obtained in the validation process are shown in Table 1. None of the models calculated based on the template 1r8n passed the validation process (Supplementary material S1). In the case of model 3iir-19, despite its error in the stereochemical analysis (Table 1), it was selected because it was the only one with more than 80% of the residues having a sequence-to-structure compatibility (3D-1D) score >0.2 in the analysis with the Verify_3D tool. The quality values obtained during the validation process support the (good/higher) quality of the 3D models proposed here.

To investigate how heterogeneous are the five models chosen, we performed a PCA of the ensemble of these 3D structures (Supplementary material S2). PCA is a powerful linear technique used to aid in the comprehension of complex multidimensional systems by reducing the phase space while retaining essential degrees of freedom [42]. Notably, the three first PCs account for 97.87% of the total variance in the models structure (Supplementary material S2). The projection of the SLAPI models onto the subspace spanned by PC1 and PC2 provide a clear separation of the structures into three clusters (Supplementary material S2). The most populated cluster is defined by the models built using the 3D structure 3iir as template (Supplementary material S2). However, the 3D models 2qn9 and 2q19 are separated in two clusters due to the higher fluctuation of loop L9 in the PC1 and PC2 modes (Supplementary material S2).

Two other automatically generated 3D models for the SLAPI (QLEC1_SOLLC) have been provided by MODBASE and annotated in the protein model portal at http://www.proteinmodelportal.org/. These two models were created on October 12, 2004 and May 28, 2003, using the crystal structures 1r8n (29% ide, model TR Q9LEC1) and 1ava (26% ide, model GI8648959) as templates, respectively. The evaluation of MODBASE models using the UCLA Structural Analysis and Verification Server tools showed that 3D models proposed here have better quality parameters compared to the MODBASE ones (Table 1). Another important difference between them is that in all the MODBASE models available for the SLAPI inhibitor the C149 and C152 amino acid residues are not linked by a disulfide bridge. We consider that the presence of two disulfide bridges in loop L9 could make the modeling of this loop more reliable taking into account its length.

The use of several models to predict functional residues using a combination of sequence-based and structural prediction methods could give us more reliable results considering the low resolution of the 3D models calculated for SLAPI. Furthermore, the different conformations adopted by the loops would enable us to explore the effect of these structural differences on the prediction of the potential functional residues (see discussion below).

Prediction of functional residues implicated in the aspartic protease inhibitory specificity in some Kunitz_legume family members

To predict functional residues implicated in the aspartic protease inhibitory specificity we first generated a MSA for the Kunitz_legume family, which enables us to identify regions with different degrees of variability. Conserved regions or positions indicate residues supposedly under

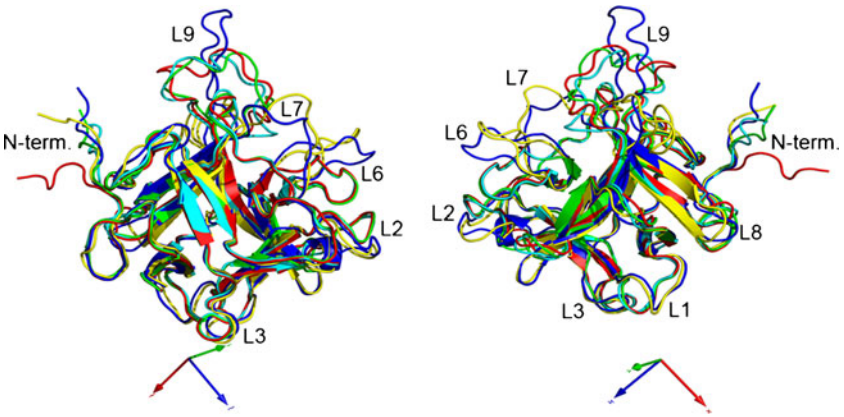


Fig. 2 Superposition of 3D models for SLAPI: 3iir-4 (green), 3iir-19 (cyan), 3iir-22 (red), 2qn4-9 (yellow) and 2qn4-19 (blue). The figure was created using the PyMOL software [68]

Table 1 Quality parameters for the SLAPI 3D models and templates

Criteria	Characteristic	3iir	3iir-4	3iir-19	3iir-22	2qn4	2qn4-9	2qn4-19	1r8n	TR Q9LEC1	1ava:C	GI 8648959
PROCHECK	% most favorable regions	80.3	80.6	80.0	81.3	87.8	85.2	85.8	88.2	81.0	83.6	77.8
	% additional allowed	14.7	14.8	17.4	16.8	11.8	12.3	10.3	8.6	13.1	15.8	19.0
	% generally allowed	2.5	4.5	0.6	1.9	0.4	2.6	3.9	3.3	4.6	0.7	3.3
	% disallowed regions	2.5	0.0	1.9	0.0	0.0	0.0	0.0	0.0	1.3	0.0	0.0
WHATCHECK	2nd generation packing quality	1.6	−2.8	−2.9	−2.7	0.2	−2.7	−2.9	−1.0	−2.6	−2.1	−2.9
VERIFY_3D	3D-1D score >0.2	94.8	72.0	82.0	69.3	95.6	72.0	65.1	92.5	73.5	99.45	63.2

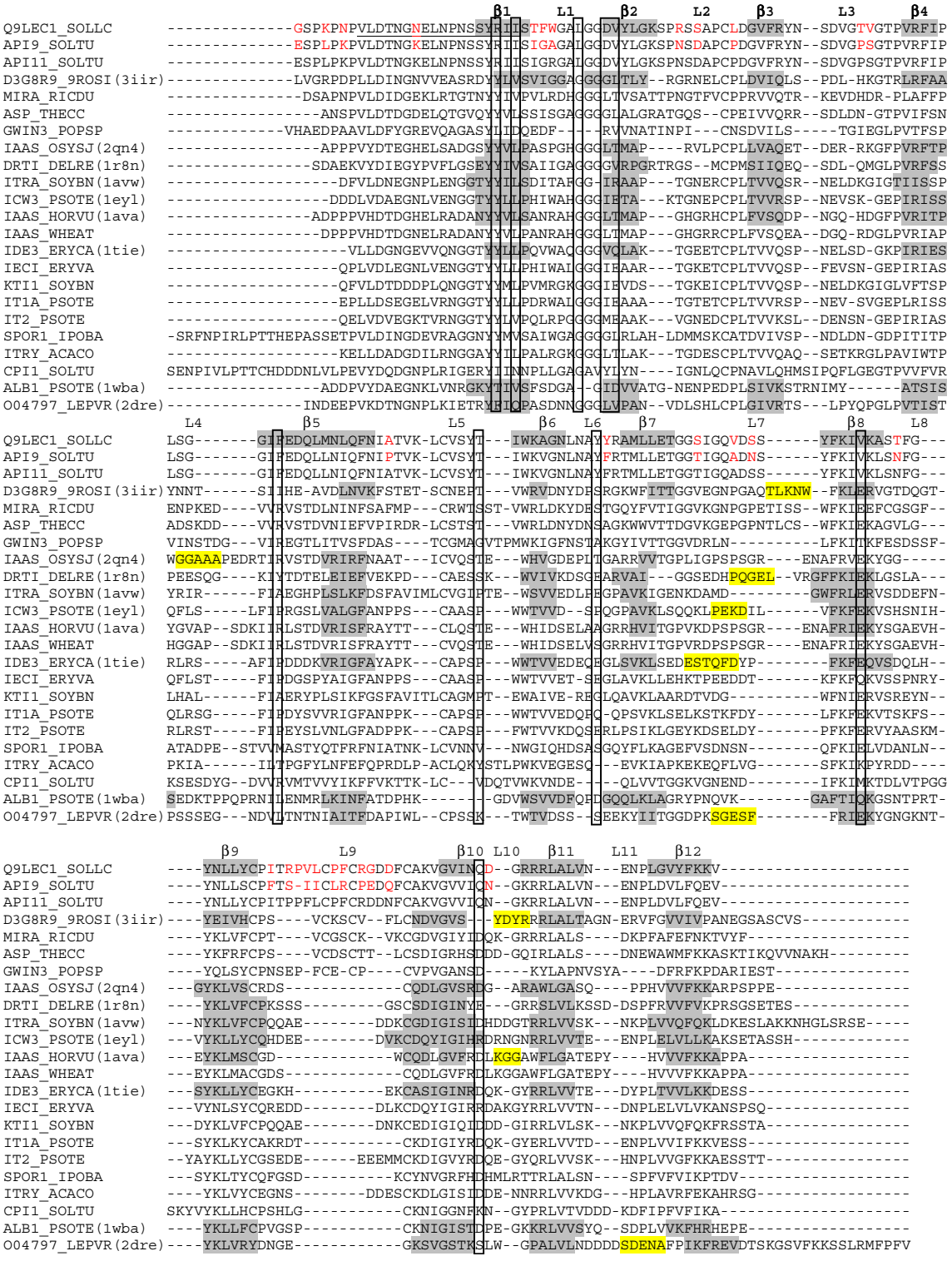
stronger evolutionary constraints and which, thus, might be more important for the protein to fulfil its function. Moreover, residues that are specifically conserved in subfamilies point to sequence changes that occurred during the divergence of a common ancestor, and they imply functional changes or the acquisition of modified specificity [43].

Combined PSI-BLAST and Pfam database searches, using the SLAPI amino acid sequence as query, yielded more than 500 sequences belonging to the STI-Kunitz (MEROPS inhibitor family I3) and Kunitz_legume (Pfam code: PF00197) families. This result led us to make a selection to generate a useful non-redundant MSA. Therefore, we selected 22 amino acid sequences related to SLAPI from nrNCBI and Pfam database searches (for details see Materials and methods section). The most conserved region in this family is located at the N-terminal region, which contains a representative sequence motif annotated in PROSITE (Fig. 3).

Although, conserved regions in a MSA are good candidates for functionally important sites, other residues showing alternative family-dependent conservation patterns may reveal key aspects of the evolution of the functional specificity [26]. One particular type of family-dependent conservation is residues showing conservation trends within subfamilies but differing between subfamilies. These positions are called *tree determinant* and their detection is the aim of the methods implemented in the TreeDet web server. The analysis of the MSA generated for the Kunitz_legume family by the TreeDet web server revealed some *tree determinant* positions (Fig. 3). These positions were predicted only by the S3det method with no positions predicted by the other two methods (S- and MB-methods). Most of the residues predicted as *tree determinant* are located in loop regions of SLAPI: mainly loops L1, L5, L6 and L10, which might be related with the functional specificity of the API subfamily.

On the other hand, hypervariability has been described in many protease inhibitor (PI) families as positive Darwinian selections at regions of the molecules where interaction with proteases occurs [14, 25]. It has been found that regions of hypervariability within a PI indicate the presence of external loops that are involved in protease binding [44, 45]. The analysis of all the predicted loops in the APIs SLAPI, API-9 and API-11 showed that there are some loops with Functional Divergence ratio (FDR) values significantly larger than 1 (Table 2). As was recommended by Creighton and Darby, (1989) [25], we also made pair-wise comparisons in order to know whether the apparent hypervariability of some regions is only the result of hypervariability of one of the proteins of the set. The data in Table 2 show that loops L1, L7 and L9 have FDR values larger than 1 in all the comparisons made and are similar to FDR values obtained for other protease inhibitors [25, 46]. We also noted the extremely large FDR value obtained for loop L9 in the API-9/API-11 pair-wise comparison. To our knowledge the only other API family where hypervariability has been analyzed is the SQAPI family [46]. In the SQAPI family the predicted protease binding regions

Fig. 3 Multiple sequence alignment of the Kunitz_legume family. Sequences are labeled by their UniProt identifier and PDB IDs where appropriate. The predicted secondary structure elements of SLAPI (Q9LEC_SOLLC) and the inhibitors with crystallographic structure are shown. *Yellow* shading denotes the α-helices and the *grey* shading the β-strands. The STI-Kunitz family signature identified by the PROSITE database is underlined. Residues that differ between SLAPI and API-9 are highlighted in red. Boxes represent the predicted functional residues by the TreeDet web server. Only the mature sequences are shown. The Kunitz_legume family inhibitor sequences are: API9_SOLTU (*Solanum tuberosum* aspartic protease inhibitor 9), API11_SOLTU (*Solanum tuberosum* aspartic protease inhibitor 11), D3G8R9_9ROSI (*Murraya koeniggi* miraculin-like protein from), MIRA_RICDU (miracle fruit miraculin protein), ASP_THECC (cacao 21 kDa seed protein), GWIN3_POPSP (poplar tree wound-responsive protein), IAAS_ORYSJ (*Oryza sativa* α-amylase/subtilisin inhibitor), DRTI_DELRE (*Delonix regia* trypsin inhibitor), ITRA-SOYBN (soybean trypsin inhibitor), ICW3_PSOTE (winged bean chymotrypsin inhibitor 3), IAAS_HORVU (barley α-amylase/subtilisin inhibitor), IAAS_WHEAT (wheat α-amylase/subtilisin inhibitor), IDE3_ERYCA (*Erythrina caffra* trypsin inhibitor), IECI_ERYVA (*Erythrina variegata* chymotrypsin inhibitor), KTI1_SOYBN (Kunitz-type inhibitor KTI1), IT1A_PSOTE (winged bean trypsin inhibitor 1A), IT2_PSOTE (winged bean trypsin inhibitor 2), SPOR1_IPOBA (sweet potato sporamin A), ITRY_ACACO (*Acacia confusa* trypsin inhibitor), CPI1_SOLTU (*Solanum tuberosum* cysteine protease inhibitor 1), ALB1_PSOTE (winged bean albumin-1), O04797_LEPVR (water-soluble chlorophyll-binding protein)



showing hypervariability, were confirmed later by mutagenesis studies [13, 46].

It is important to point out that the predictions made by the two methods used, revealed different aspects of the subfamily specificity. The residues predicted by the TreeDet web server represent those positions conserved within the subfamily while the hypervariability shows the opposite. However it does not mean that you can find both types of residues in a potential interacting region. For instance, loop 1 showed FDR values larger than 1 and also a residue (L32) predicted as functional by the TreeDet web server. In the case of the other two loops (L7 and L9) the

Table 2 Hypervariability in the aspartic protease inhibitors SLAPI, API-9 and API-11

Region	Residue no.(SLAPI numbering scheme)	FDR values[a]			
		SLAPI/API-9/API-11	API-9/API-11	SLAPII/API9	SLAPI/API-11
N-ter.	1-20	0.8	0.0	0.9	1.2
L1	27-34	2.1	1.9	1.6	2.2
L2	41-50	1.2	0.0	1.4	1.8
L3	55-65	0.7	0.0	0.8	1.0
L4	70-74	0.0	0.0	0.0	0.0
L5	87-97	1.0	1.3	0.4	0.0
L6	104-109	1.3	0.0	0.7	0.9
L7	116-125	1.6	1.4	1.4	1.2
L8	133-136	1.0	0.0	1.1	2.9
L9	143-162	3.5	25.0	2.6	1.2
L10	167-169	1.3	0.0	1.5	1.9
L11	177-179	0.0	0.0	0.0	0.0
C-ter.	188	0.0	0.0	0.0	0.0

[a] The functional divergence ratio (FDR, Creighton and Darby, 1989) [25] for each region was calculated as the ratio of the average variability of the region with respect to the remainder of the protein. The variability at each position is the number of replacements (the number of different amino acids minus one) divided by the number of sequences compared. The average of the variability for all the positions of each region gives the variability of the region. Internal deletions or insertions were counted as an amino acid

fact that they have several insertions and deletions in the MSA might be a reason for the absence of predicted residues by the TreeDet, taking into account the sensitivity of this kind of method to the alignment quality. Anyway, the predictions made by both methods can be considered as potential contact regions with proteases.

It has been recognized that protein-protein interface prediction methods which use structural information generally improve the results compared to those based solely on sequence features [47, 48]. In view of several results which indicated that combination of different predictors will improve the final protein-protein interface prediction, we used the metaserver meta-PPISP that showed increased prediction accuracy compared to the three individual web servers used to build it [34].

Although the number of protein-protein interface residues predicted by the meta-PPISP web server for the five SLAPI 3D models selected varied, we identified some residues which are predicted in every case (Table 3 and Fig. 4). Residues I144, P147, V148, P151, F152 and D156 were predicted as functional residues for the five SLAPI 3D models analyzed while residues C142, P143, L149, C150, C153, R154, G155, D157 and F158 were predicted in four of the five 3D models. All the aforementioned residues are located in loop L9, highlighting this region as a potential interacting loop. It is noteworthy that residues C142, P143, C150, C153, D156 and F158 are strictly conserved among the APIs from the STI-Kunitz family (Fig. 3). On the other hand, residues I144, P147, V148, L149, P151, F152, R154, G155 and D157, which are some of the amino acid positions that differ between SLAPI and API-9 inhibitors (Fig. 3), could be proposed as responsible for the difference in the inhibitors´ K_i values against proteinase A.

Protein-protein docking of SLAPI with aspartic protease proteinase A

Protein-protein interface residue prediction methods have been used in docking studies to limit the initial search as

Table 3 Predicted protein-protein interface residues of SLAPI 3D models by meta-PPISP. The table shows the predicted residues with a score >0.34 in at least four of the five 3D models analyzed

3iir-4	3iir-19	3iir-22	2qn4-9	2qn4-19
C142	C142	C142	C142	-
P143	P143	-	P143	P143
I144	I144	I144	I144	I144
P147	P147	P147	P147	P147
V148	V148	V148	V148	V148
L149	-	L149	L149	L149
C150	C150	C150	-	C150
P151	P151	P151	P151	P151
F152	F152	F152	F152	F152
C153	C153	C153	-	C153
-	R154	R154	R154	R154
G155	G155	G155	-	G155
D156	D156	D156	D156	D156
D157	-	D157	D157	D157

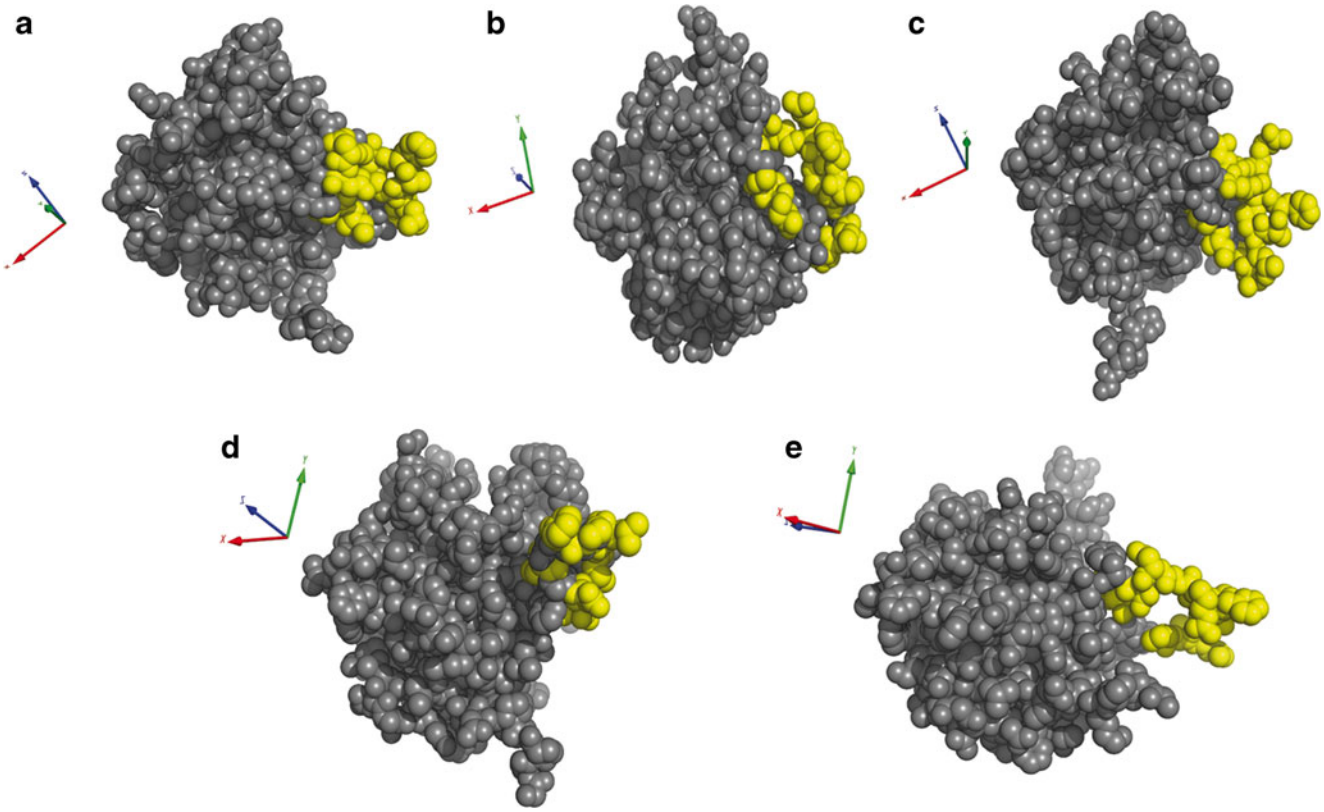


Fig. 4 Predicted protein-protein interface residues of SLAPI Residues are mapped onto the 3D models: 3iir-4 (**a**), 3iir-19 (**b**), 3iir-22 (**c**), 2qn4-9 (**d**), 2qn4-19 (**e**). Residues colored in yellow are predicted in at least four of the five 3D models (Table 3). The figure was created using the PyMOL software

well as to assist scoring docking solutions [49, 50]. However at this stage we did not apply the results from the meta-PPISP web server to filter protein-protein docking results, but instead compared both procedures. In the case of the protein-protein docking, we were more interested in identifying inhibitor residues with higher *preference* in the complex interface instead of a specific conformation adopted by the interacting molecules in the complex. To predict the regions of the inhibitor that may be able to interact specifically with the aspartic protease proteinase A, we performed protein-protein docking using the web server ClusPro 2.0. ClusPro´s authors suggest the use of the balanced coefficient in those cases where the nature of the interface is unknown. However, in the present work, we decided to analyze the docking models for all coefficients and compare their results.

To select the docking 3D models where the interaction occurs through the proteinase A active site cavity, we used the CASTp web server that allowed us to identify atoms forming protein pockets, to calculate the volumes and areas of the pockets, to identify atoms forming the "rims" of the pocket mouth(s), to calculate the number of mouth openings for each pocket as well as the area and circumference of the mouth openings [51]. The following residues were identified as forming the active site pocket: Y9, L10, A12, Q13, Y15, I30, D32, G34, S35, S36, N37, W39, I73, Q74, Y75, G76, T77, L110, T111, F112, A113, F114, G115, K116, F117, I120, Y189, D215, G217, T218, S219, L220, T222, K239, G243, Q244, Y245, D273, T275, L276, I283, S284, A285, I286, T287, P288, M289, D290, P292, I300, A304 and R307 (for more details see Supplementary material S4).

In a second step, we selected only the docking 3D models with at least one contact residue belonging to the proteinase A active site cavity. From a total of 563 models analysed, 410 were considered valid as a result of the selection process (see Materials and methods). The results for each coefficient set of docking models were analyzed separately. For the WdV+Electrostatic coefficient set, only 2 docking 3D models, from a total of 134, fulfilled the selection requirements. The analysis of the balanced coefficient showed that in 97% of docking models the inhibitor-protease interaction involved the protease active site cavity while the percentage of docking models considered as valid for the electrostatic coefficient was 89%: still, in the majority of the models its was predicted the SLAPI-proteinase A interaction occurs through the protease active site cavity. Although the chemical character

Table 4 Predicted functional residues of SLAPI 3D models by protein-protein docking

Coefficient	No. of SLAPI 3D models used for protein -protein docking	Predicted residue ($f_i \geq 0.5$)	% valid models/ total models
Balanced	Five	P151; F152	97%
	Four of five	I144; V148; C153; R154	
Electrostatic	Five	I144; V148; P151; F152; C153; R154	89%
	Four of five	L149; D157	
Hydrophobic	Five	L149; C150; P151; F152; C153	100%
	Four of five	I144; R146; V148; R154	

of the SLAPI-proteinase A complex interface is unknown, previous studies have shown that most protease-inhibitor interfaces have a predominantly hydrophobic chemical character [52]. This is in accordance with the percentages of valid docking models obtained in this work for the different coefficients: hydrophobic (100%), balanced (97%) and electrostatic (89%), which clearly suggest a preference for a hydrophobic driven association between the protein-ase A and SLAPI. In addition, the number of contact residues with a frequency (f_i) higher than 0.5 varied depending on which SLAPI 3D model was used as the ligand for docking (Supplementary material S5). Nevertheless, we found that a reduced number of residues had $f_i \geq 0.5$ according to docking models obtained using the five SLAPI 3D models or even with four of the five models. These results are summarized in Table 4 and Fig. 5.

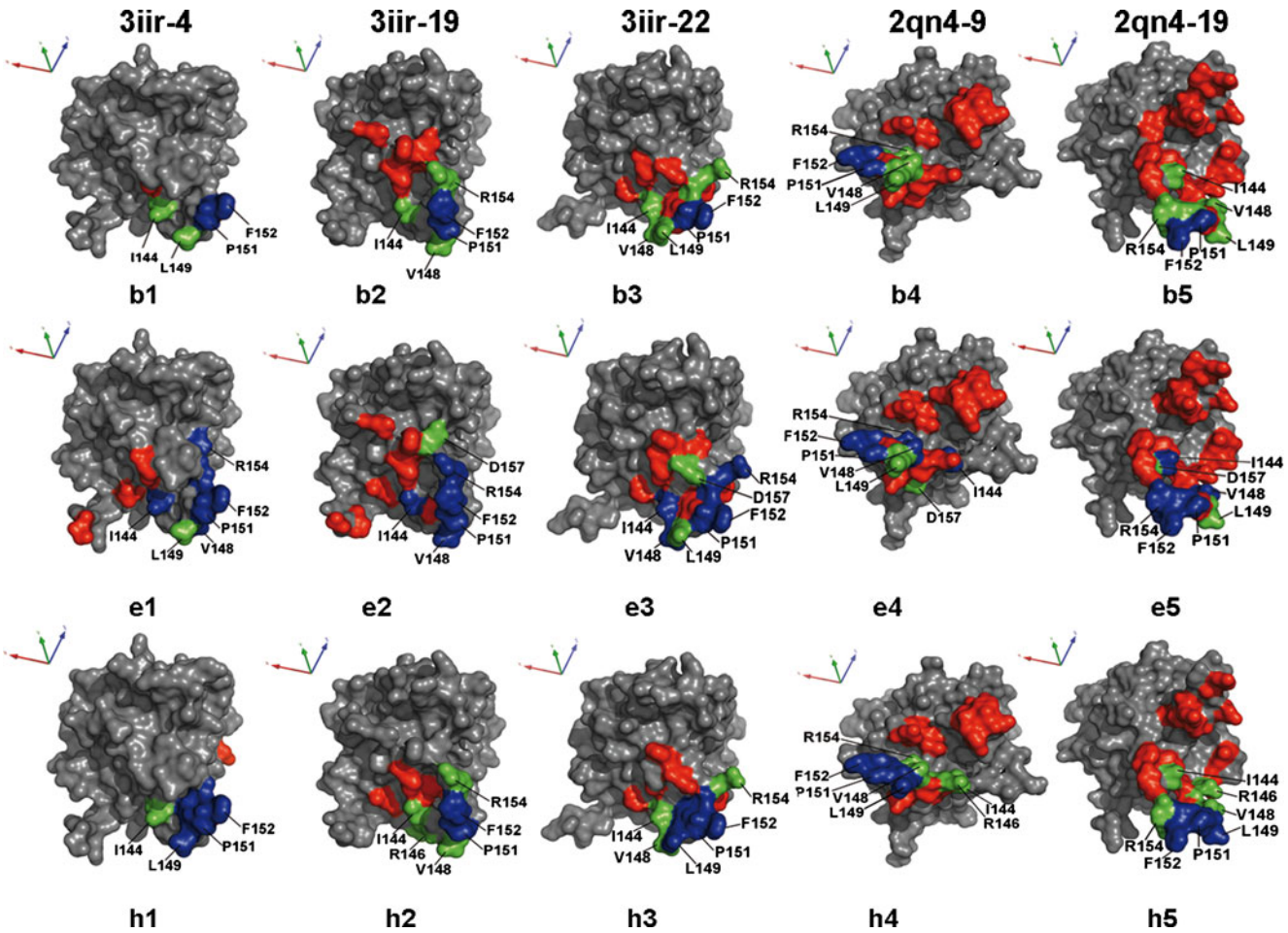


Fig. 5 Functional residues predicted by protein-protein docking. Residues are mapped on the SLAPI 3D models. Columns display the results from docking models generated with each SLAPI 3D model (labeled on the top) and rows display the results for each coefficient set of docking models: balanced (b1-b5), electrostatic (e1-e5) and hydrophobic (h1-h5). Residues are colored according to their f_i value and the number of 3D models where they have $f_i \geq 0.5$: gray ($f_i < 0.5$), blue ($f_i \geq 0.5$ in all the 3D models), green ($f_i \geq 0.5$ in four 3D models), red ($f_i \geq 0.5$ in three 3D models). The figure was created using the PyMOL software

Among the contact residues with frequency values higher than 0.5 we identified some in common between the different coefficient sets of docking models. That is the case for residues I144, V148, L149, P151, F152, C153 and R154; all of them located in loop L9. These results support those obtained previously by the meta-PPISP server where all the aforementioned residues were predicted as protein-protein interface residues for at least four of the five TI 3D models analyzed (Table 4 and Fig. 5). Consistent with this, loop L9 showed the highest values in the hypervariability analysis (Table 2). Moreover, all but one (C153) are among the residues that differ between the SLAPI and API-9 amino acid sequences and that could be related with the difference in their K_i values against proteinase A [7]. In this respect, the hydrolysis of the peptide bond between SLAPI residues R154-G155 resulted in the loss of its proteinase A inhibitory activity [7], suggesting that loop L9 could be involved in the inhibitory mechanism. The analysis of the proposed SLAPI 3D models showed that the cleaved bond (R154-G155) is located close to the end of loop L9, between two cysteine residues (C153 and C159) involved in two disulfide bridges at the beginning (C142-C159) and the end (C150-C153) of the loop. Clearly, the presence of these disulfide bonds limits the conformational freedom of loop L9. Considering this evidence, it could be possible that the effect of the hydrolysis of the peptide bond between residues R154-G155 has only a *local* effect over the conformation of loop L9 instead of on the entire molecule. Structural studies with miraculin-like protein from *Murraya koenigii* showed that the equivalent loop in this protein (L10) had lower B-factors compared to the crystal structure of other Kunitz_legume inhibitors, suggesting a lower degree of flexibility [53]. The presence of three disulfide bridges in the miraculin-like protein from *Murraya koenigii* has also been proposed as the reason for its remarkable structural stability against proteolysis [53, 54].

In a recent study, another Kunitz_legume family inhibitor isolated from *Solanum lycopersicum* (Uniprot ID: Q9LEG1_SOLLC) showed no inhibition of the aspartic proteases cathepsin D and proteinase A, despite its high sequence identity with the SLAPI inhibitor [55]. Actually, the alignment of the mature sequence of both inhibitors revealed that they just differ in the first amino acid (G→A) (results not shown). Based on this, it might be suggested that the N-terminal region is involved in the inhibition mechanism and this single difference seems to be enough to eliminate the aspartic protease activity present in the SLAPI inhibitor. However, the analysis of the sequence of the inhibitors API-9 and API-11, which showed activity against cathepsin D, revealed that their first amino acid residue is conserved (E) but very different from those found in the inhibitors from *Solanum lycopersicum* (Fig. 3). The residue G1 appears as a potential functional residue showing a $f_i \geq 0.5$ only in the docking models corresponding to the electrostatic coefficient for the SLAPI 3D models 3iir-4 and 3iir-19. In view of this, we consider that there is more

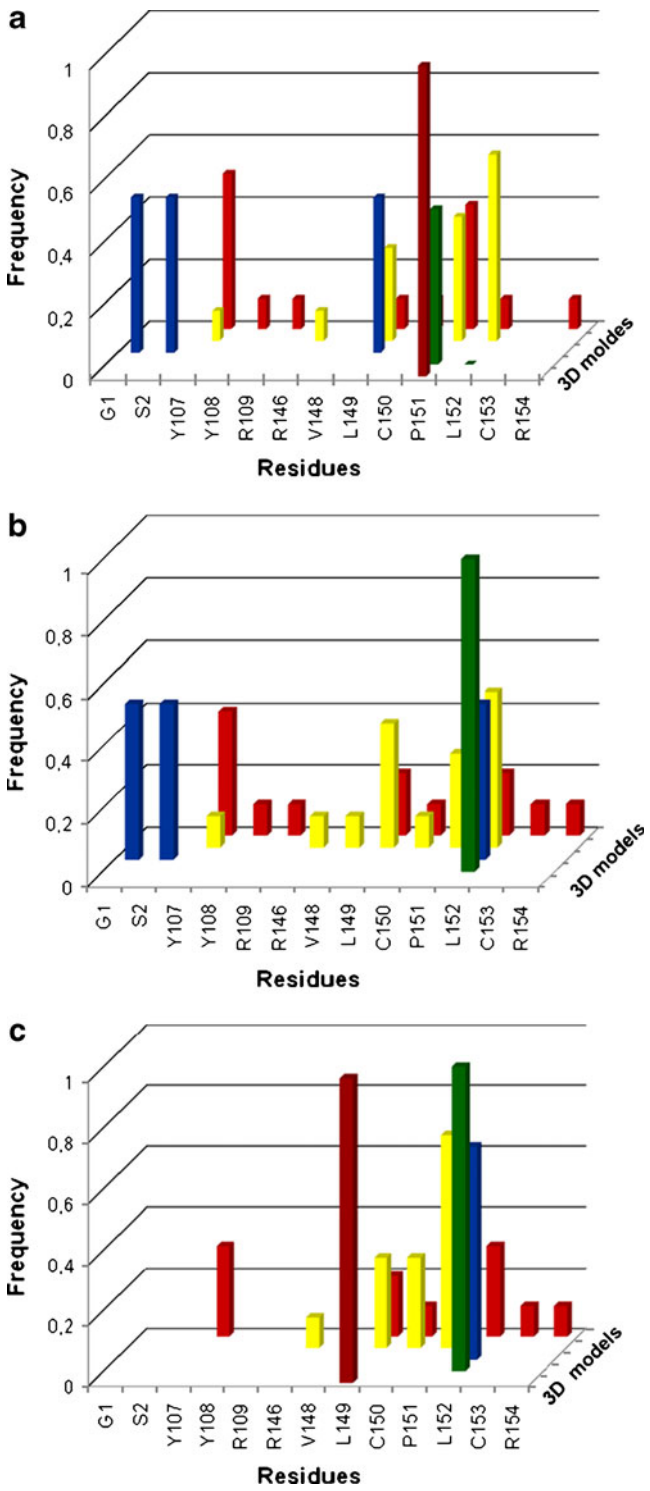


Fig. 6 SLAPI residues contacting with the proteinase A catalytic residues (D32 and D215). Balanced (**a**), Electrostatic (**b**) and Hydrophobic (**c**) sets of docking models. Color code: brown (3iir-4), green (3iir-19), blue (3iir-22), yellow (2qn4-9) and red (2qn4-19)

evidence supporting loop L9 as a potential region of interaction with the proteinase A for the APIs from the Kunitz_legume family respect to the N-terminal region.

We are aware that the specific features of the inhibition mechanism will require an experimentally-determined protease-inhibitor complex structure. In spite of that, we were interested to explore the possibility that some inhibitor residues could be able to interact directly with the proteinase A catalytic aspartic acids (D32 and D215), as a way to block the enzymatic activity. The Kunitz_legume family inhibitors form a noncovalent protease-inhibitor complex with serine proteases, highly similar to the enzyme-substrate interaction [56]. There is evidence that this type of protease inactivation is also used to inhibit cysteine and metallo proteases [56].

The percentage of docking models with SLAPI residues interacting with the proteinase A catalytic residues D32 and D215 ranged from 3.4 to 14.3%. The analysis of the docking models showed similar results for those generated with the SLAPI 3D models 3iir-4, 3iir-19 and 3iir-22, independent of the coefficient analyzed. However, for the docking models obtained using the SLAPI 3D models 2qn4-9 and 2qn4-19 as ligand, higher percentage values were obtained (37.9 to 60%). Despite these differences in percentage, we found again that residues L149, P151 and F152 were the most frequent inhibitor contact residues with the catalytic aspartic acids D32 and D215, independently of the SLAPI 3D model used as ligand or the coefficient selected (Fig. 6). Indeed, as mentioned previously, residues P151 and F152 are located between two cysteines bonded by a disulfide bridge (C150-C153) which could restrain the possible conformations of these residues as resulting of a hydrolysis of the peptide bond by the proteinase A. Such a situation could facilitate a mechanism based on hydrolysis/resynthesis of a single peptide bond as occurred in the canonical serine protease inhibitors [57, 58].

We also noted that the residue Y107 is predicted quite frequently as an inhibitor residue interacting with D32 and D215 (Fig. 6), however, it is restricted to the docking models generated using SLAPI 3D models 2qn4-9 and 2qn4-19 as ligand (Supplementary material S5). Besides, this residue was only predicted as a potential protein-protein interface residue by the meta-PPISp server for one of the SLAPI 3D models (2qn4-9) (Supplementary material S3).

The visual analysis of the docking models revealed that the inhibitor may be able to insert the L9 loop to interact with the proteinase A catalytic residues in different conformations (Fig. 7). For instance, the inhibitor inserts loop L9 from the top of the active site blocking access to the active site cavity (Fig. 7a). Another conformation places the L9 loop longitudinally at the bottom of the active site with residue P151 interacting with the catalytic residues D32 and D215 (Fig. 7b). In both resulting complexes, the proteases will be potentially unable to cleave a substrate.

The potential role of a very stable loop in the inhibition of aspartic proteases by a member of the Kunitz_legume family is in agreement with previous studies suggesting that proteins that have recruited one family of protease inhibitors to inhibit a second class of proteases, keep similarities in the inhibitory elements of both protease targets [14]. Within the Kunitz_legume family an inhibitor from *Prosopis juliflora* has been proposed to have an overlapped binding site for the serine protease Trypsin (EC 3.4.21.4) and the cysteine protease Papain (EC 3.4.22.2), located in the canonical loop responsible for the serine protease inhibition [59]. Among the APIs that have been suggested to recruit other families of inhibitors, the best studied is the pepsin inhibitor from squash *Cucurbita maxima* (SQAPI) [12, 13, 46, 60]. Comparative modeling and hypervariability studies showed that the binding loop in cystatin and SQAPI coincided, suggesting a recruitment of the inhibitory mechanism of cystatin by SQAPI [46]. Recently, the determination by Heady et al., [13] of the solution structure of SQAPI, in combination with mutagenesis studies, enabled them to obtain models of the complex Pepsin-SQAPI using docking procedures. The mutagenesis and docking results suggest that SQAPI appears to retain a

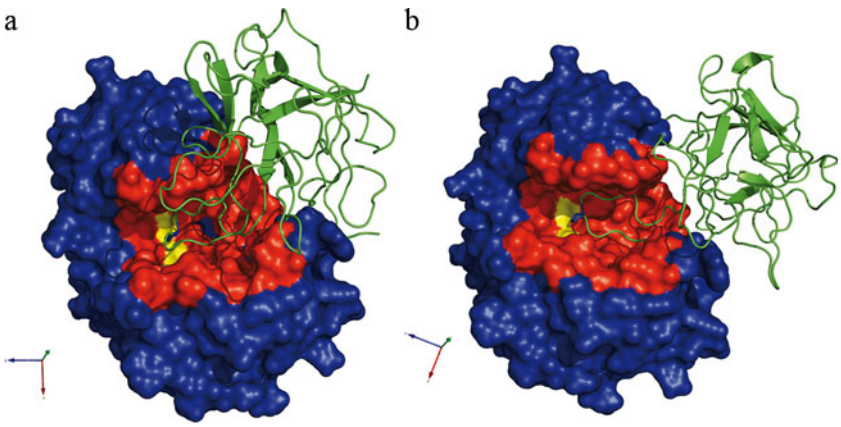


Fig. 7 Possible conformations predicted for the proteinase A: SLAPI complex. Two possible conformations are presented. Proteinase A is colored in blue and its active site cavity in red. The proteinase A catalytic aspartic acids D32 and D215 are colored in yellow. The inhibitor (SLAPI) is colored in green and the residue interacting with the proteinase A catalytic aspartic acids is colored in blue and represented in sticks. The figure was created using the PyMOL software

similar protease inhibitory mechanism as cystatin despite their different targets [13]. Unfortunately, the other two aspartic protease inhibitors with structures solved (IA_3 and PI-3) do not have other family members active against other protease classes. However, based on the results obtained in the present work combining sequence analysis, comparative modeling, protein-protein interface residue prediction and protein-protein docking, we hypothesize that loop L9 is involved in the inhibition of the proteinase A by the APIs for the Kunitz_legume family. A similar approach that combines sequence and structural analyzes, has been previously used by our group and by other researchers [61–67], to predict functional residues and protein-protein interfaces.

Conclusions

In this work we generated 3D models of the SLAPI inhibitor, which show better quality parameters compared to those available to date. Also, using a combination of sequence and structural analysis, together with protein-protein docking, we propose that residues I144, V148, L149, P151, F152 and R154 (SLAPI numbering scheme) are involved in the inhibition of proteinase A by the APIs of the STI-Kunitz family. All these residues are located in loop L9, which has not been predicted previously as a functional region in APIs of the STI-Kunitz family. In addition, the results obtained support in general the hypothesis that inhibitor families, which were recruited from inhibitors of another protease family, keep similarities in their inhibition mechanisms. Mutagenesis experiments are likely to be the easiest way to verify the function of the residues predicted in this work. All this information will be useful in the design of novel aspartic protease inhibitors with potential in biotechnology and biomedicine.

Acknowledgments This research was supported by the International Foundation for Science (IFS), Stockholm, Sweden, through a grant to YG (grant No. F/4927-1). YG would also like to thank MSc. Alexandra Nárvaez from the Pontificia Universidad Católica del Ecuador (PUCE) for providing lab facilities.

References

1. Cooper JB (2002) Aspartic proteinases in disease: a structural perspective. Curr Drug Targets 3:155–173
2. Dunn BM (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. Chem Rev 102:4431–4458
3. Abbenante G, Fairlie DP (2005) Protease inhibitors in the clinic. Med Chem 1:71–104
4. Laing WA, McManus MT (2002) Proteinase inhibitors. In: McManus MT, Laing WA, Allan AC (eds) Annual plant reviews, vol 7. Protein interactions in plants. Sheffield Academic, Sheffield, pp 77–119
5. Mareš M, Meloun B, Pavlík M, Kostka V, Baudygš M (1989) Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. FEBS Lett 251:94–98
6. Ritonja A, Križajl I, Meškol P, Kopitar' M, Lučovnik' P, Šhrukeljl B, Pungerčarl J, Buttle DJ, Barrett AJ, Turk V (1990) The amino acid sequence of a novel inhibitor of cathepsin D from potato. FEBS Lett 267:13–15
7. Cater SA, Lees WE, Hill J, Brzin J, Kay J, Phylip LH (2002) Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. Biochim Biophys Acta 1956:76–82
8. Martzen MR, McMullen BA, Smith NE, Fujikawa K, Peanasky RJ (1990) Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode Ascaris suum. Biochemistry 29:7366–7372
9. Dreyer T, Valler MJ, Kay J, Charlton P, Dunn BM (1985) The selectivity of action of the aspartic-proteinase inhibitor IA3 from yeast (Saccharomyces cerevisiae). Biochem J 231:777–779
10. Lenarčič B, Turk V (1999) Thyroglobulin type-1 domains in Equistatin inhibit both Papain-like Cysteine Proteinases and Cathepsin D. J Biol Chem 274:563–566
11. Mathialagan N, Hansen TR (1996) Pepsin-inhibitory activity of the uterine serpins. Proc Natl Acad Sci USA 93:13653–13658
12. Christeller JT, Farley PC, Ramsay RJ, Suvillan PA, Laing WA (1998) Purification, characterization and cloning of an aspartic proteinase inhibitor from squash phloem exudate. Eur J Biochem 254:160–167
13. Headey SJ, MacAskill UK, Wright MA, Claridge JK, Edwards PJB, Farley PC, Christeller JT, Laing WA, Pascal SM (2010) The solution structure of the squash aspartic acid proteinase inhibitor (SQAPI) and mutational analysis of pepsin inhibition. J Biol Chem 286:27019–27025
14. Christeller JT (2005) Evolutionary mechanisms acting on proteinase inhibitor variability. FEBS J 272:5710–5722
15. Ng KKS, Petersen JFW, Cherney MM, Garen C, Zalatoris JJ, Rao-Naik C, Dunn BM, Martzen MR, Peanasky RJ, James MNG (2000) Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3. Nat Struct Biol 7:653–657
16. Li M, Phylip LH, Lees WE, Winther JR, Dunn BM, Wlodawer A, Kay J, Gustchina A (2000) The aspartic proteinase from Saccharomyces cerevisiae folds its own inhibitor into a helix. Nat Struct Biol 7:113–117
17. Keilova H, Tomasek V (1976) Isolation and properties of Cathepsin D inhibitor from potatoes. Collect Czech Chem Commun 41:489–497
18. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882
19. Galtier N, Guoy M, Gautier C (1996) SEAVIEW and PHYLO_-WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 6:543–548
20. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. Nat Protocols 4:363–371
21. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16:404–405
22. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197–W201
23. Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 47:228–235

24. Carro A, Tress M, de Juan D, Pazos F, Lopez-Romero P, del Sol A, Valencia A, Rojas AM (2006) TreeDet: a web server to explore sequence space. Nucleic Acids Res 34:W110–W115
25. Creighton TE, Darby NJ (1989) Functional evolutionary divergence of proteolytic enzymes and their inhibitors. Trends Biochem Sci 14:319–324
26. del Sol MA, Pazos F, Valencia A (2003) Automatic methods for predicting functionally important residues. J Mol Biol 326:1289–1302
27. Rausell A, Juan D, Pazos F, Valencia A (2010) Protein interactions and ligand binding: from protein subfamilies to functional specificity. Proc Natl Acad Sci USA 107:1995–2000
28. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
29. Guda C, Lu S, Scheeff ED, Bourne PE, Shindyalov IN (2004) CE-MC: a multiple protein structure alignment server. Nucleic Acids Res 32:W100–W103
30. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
31. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structure. Nature 381:272
32. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170
33. Luthy R, Bowie JU, Eisenberg D (1991) Assesment of protein models with three-dimensional profiles. Nature 356:83–85
34. Qin S, Zhou H-X (2007) meta-PPISP: a meta web server for protein-proein interaction site prediction. Bioinformatics 23:3386–3387
35. Chen H, Zhou H-X (2005) Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. Proteins 61:21–35
36. Neuvirth H, Raz R, Schreiber G (2004) Promate: a structure based prediction program to identify the location of protein-protein binding sites. J Mol Biol 338:181–199
37. Liang S, Zhang C, Liu S, Zhou Y (2006) Protein binding site prediction using an empirical scoring function. Nucleic Acids Res 34:3698–3707
38. Comeau SR, Kozakov D, Brenke R, Shen Y, Beglov D, Vajda S (2007) ClusPro: performance in CAPRI rounds 6–11 and the new server. Proteins 69:781–785
39. Kozakov D, Brenke R, Comeau SR, Vajda S (2006) PIPER: an FFT-based protein docking program with pairwise potentials. Proteins 65:392–406
40. Vriend G (1993) What if: a molecular modeling and drug design program. J Mol Graph 8:52–56
41. Lesk AM, Chotia C (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J Mol Biol 136:225–270
42. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425
43. López-Romero P, Gómez MJ, Gómez-Puertas P, Valencia A (2004) Prediction of functional sites in proteins by evolutionary methods. In: Kamp RM, Calvete JJ, Choli-Papadopoulou T (eds) Principles and practice. Methods in proteome and protein analysis. Springer, Berlin, pp 319–340, Chapter 22
44. Laskowski MJ, Kato I, Ardelt W, Cook J, Denton A, Empie MW, Kohr WJ, Park SJ, Parks K, Schatzley BL et al. (1987) Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. Biochemistry 26:202–221
45. Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature 326:96–99
46. Christeller JT, Farley PC, Marshall RK, Anandan A, Wright MM, Newcomb RD, Laing WA (2006) The Squash Aspartic Proteinase Inhibitor SQAPI Is widely present in the Cucurbitales, comprises a small Multigene family, and is a member of the Phytocystatin family. J Mol Evol 63:747–757
47. Ezkurdia I, Bartoli L, Fariselli P, Casadio R, Valencia A, Tress ML (2009) Progress and challenges in predicting protein-protein interaction sites. Brief Bioinform 10:233–246
48. de Vries SJ, Bonvin AMJJ (2008) How proteins get in touch: interface prediction in the study of biomolecular complexes. Curr Protein Peptide Sci 9:394–406
49. van Dijk ADJ, de Vries SJ, Domínguez C, Chen H, Zhou H-X (2005) Data-driven docking: HADDOCK´s adventures in CAPRI. Proteins 60:232–238
50. Tress ML, de Juan D, Grana O, Gomez MJ, Gomez-Puertas P, Gonzalez JM, Lopez G, Valencia A (2005) Scoring docking models with evolutionary information. Proteins 60:275–280
51. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7:1884–1897
52. Lo Conte L, Cyrus C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285:2177–2198
53. Gahloth D, Selvakumar P, Shee C, Kumar P, Sharma AK (2010) Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from Murraya koenigii. Arch Biochem Biophys 494:15–22
54. Shee C, Islam A, Ahma F, Sharma AK (2007) Structure–function studies of Murraya koenigii trypsin inhibitor revealed a stable core beta sheet structure surrounded by α-helices with a possible role for α-helix in inhibitory function. Int J Biol Macromol 41:410–414
55. Lisón P, Rodrigo I, Conejero V (2006) A novel function for the Cathepsin D inhibitor in tomato. Plant Physiol 142:1329–1339
56. Otlewski J, Jelen F, Zakrzewska M, Oleksy A (2005) The many faces of protease-inhibitor interaction. EMBO J 24:1303–1310
57. Helland R, Otlewski J, Sundheim O, Dadlez M, Smalas AO (1999) The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. J Mol Biol 287:923–942
58. Ardelt W, Laskowki JM (1985) Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same … Leu18-Glu19 … reactive site. Biochemistry 24:5313–5320
59. Franco OL, Grossi de Sá MF, Sales MP, Mello LV, Oliveira AS, Rigden DJ (2002) Overlapping binding sites for Trypsin and Papain on a Kunitz-type proteinase inhibitor from Prosopis juliflora. Proteins 49:335–341
60. Farley PC, Christeller JT, Sullivan ME, Sullivan PA, Laing WA (2002) Analysis of the interaction between the aspartic peptidase inhibitor SQAPI and aspartic peptidases using surface plasmon resonance. J Mol Recognit 15:135–144
61. Arenas NE, Salazar LM, Soto CY, Vizcaíno C, Patarroyo ME, Patarroyo MA, Gómez A (2011) Molecular modeling and in silico characterization of mycobacterium tuberculosis TlyA: possible misannotation of this tubercle bacilli-hemolysin. BMC Struct Biol 11:16
62. Pons T, Naumoff DG, Martínez-Fleites C, Hernández L (2004) Three acidic residues are at the active site of a β-propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. Proteins 54:424–432
63. Perera E, Pons T, Hernandez D, Moyano FJ, Martínez-Rodríguez G, Mancera JM (2010) New members of the brachyurins family in lobster include a trypsin-like enzyme with amino acid substitutions in the substrate-binding pocket. FEBS J 277:3489–3501
64. Valiente PA, Batista PR, Pupo A, Pons T, Valencia A, Pascutti PG (2008) Predicting functional residues in Plasmodium falciparum plasmepsins by combining sequence and structural analysis with molecular dynamics simulations. Proteins 73:440–457

65. Pons T, González B, Ceciliani F, Galizzi A (2006) FlgM anti-sigma factors: identification of novel members of the family, evolutionary analysis, homology modeling, and analysis of sequence-structure-function relationships. J Mol Model 12:973–983
66. Rahi A, Rehan M, Garg R, Tripathi D, Lynn AM, Bhatnagar R (2011) Enzymatic characterization of catalase from bacillus anthracis and prediction of critical residues using information theoretic measure of relative entropy. Biochem Biophys Res Commun 411:88–95
67. Sikora S, Godzik A (2004) Combination of multiple alignment analysis and surface mapping paves a way for a detailed pathway reconstruction—the case of VHL (von Hippel-Lindau) protein and angiogenesis regulatory pathway. Protein Sci 13:786–796
68. DeLano WL (2004) The PyMOL molecular graphics system, 097th edn. DeLano Scientific LLC, San Carlos