Solution Structure of the Squash Aspartic Acid Proteinase Inhibitor (SQAPI) and Mutational Analysis of Pepsin Inhibition

Stephen J. Headley†1, Ursula K. MacAskill‡1, Michele A. Wright‡, Jolyon K. Claridge‡3, Patrick J. B. Edwards†4, Peter C. Farley†, John T. Christeller†, William A. Laing†5, and Steven M. Pascal‡5

From the †Institute of Fundamental Sciences and ‡Institute of Molecular Biosciences, Massey University, Palmerston North 4442, New Zealand, ‡The New Zealand Institute for Plant & Food Research Limited, Auckland 1142, New Zealand, and †The New Zealand Institute for Plant & Food Research Limited, Palmerston North 4442, New Zealand

Received for publication, April 22, 2010, and in revised form, May 20, 2010 Published, JBC Papers in Press, June 9, 2010, DOI 10.1074/jbc.M110.137018

The squash aspartic acid proteinase inhibitor (SQAPI), a proteinaceous proteinase inhibitor from squash, is an effective inhibitor of a range of aspartic proteinases. Proteinaceous aspartic proteinase inhibitors are rare in nature. The only other example in plants probably evolved from a precursor serine proteinase inhibitor. Earlier work based on sequence homology modeling suggested SQAPI evolved from an ancestral cystatin. In this work, we determined the solution structure of SQAPI using NMR and show that SQAPI shares the same fold as a plant cystatin. The structure is characterized by a four-strand antiparallel β-sheet gripping an α-helix in an analogous manner to fingers of a hand gripping a tennis racquet. Truncation and site-specific mutagenesis revealed that the unstructured N terminus and the loop connecting β-strands 1 and 2 are important for pepsin inhibition, but the loop connecting strands 3 and 4 is not. Using ambiguous restraints based on the mutagenesis results, SQAPI was then docked computationally to pepsin. The resulting model places the N-terminal strand of SQAPI in the S’ side of the substrate binding cleft, whereas the first SQAPI loop binds on the S side of the cleft. The backbone of SQAPI does not interact with the pepsin catalytic Asp32–Asp215 diad, thus avoiding cleavage. The data show that SQAPI does share homologous structural elements with cystatin and appears to retain a similar protease inhibitory mechanism despite its different target. This strongly supports our hypothesis that SQAPI evolved from an ancestral cystatin.

Known proteinaceous aspartic acid proteinase inhibitors (PIs) are rare and unevenly distributed among classes of organisms in contrast to proteinaceous inhibitors of serine and cysteine proteinases (1). A novel proteinaceous aspartic acid proteinase inhibitor (SQAPI) has been characterized from squash phloem exudate (Cucurbita maxima Duchesne) (2, 3). The protein is distinct from the other six families of proteinaceous aspartic PIs that have been identified and had their genes cloned: the potato plant Kunitz inhibitors (4), the Ascaris inhibitors (5), the yeast proteinase A inhibitor-3 (6), a domain of the sea anemone thyroglobulin type-1 inhibitor (7, 8), the pig serpin inhibitor (9), and the Bacillus sp. peptide, ATBI (10). The structures for some of these proteins have been solved (8, 11–13) and are, to date, very different from each other and exhibit distinct, and in some cases, novel inhibitory mechanisms. However, no structural information on SQAPI has been published to date, except that based on modeling (14).

It is likely that plant PIs represent examples of recent evolutionary change. Several exclusively plant PIs are restricted highly in their distribution, often to a single family (1, 15). The squash serine PI family appears limited to the Cucurbitaceae, the Ryan family (proteinase inhibitor II family) to the Solanaceae, the trypsin/α-amylase inhibitor family to the Graminaceae, the mustard seed inhibitor family to the Brassicaceae. Most recently, it has been proposed that the SQAPI family also is restricted highly in distribution to Cucurbitales (14). Such limited distribution among plant families may indicate a recent evolutionary history. However, the ancestral noninhibitor proteins from which neofunctionalization has occurred are unknown. Two inhibitor families unique to plants have a wider distribution (15), with the plant Kunitz and the Bowman-Birk inhibitor families being found in legumes and cereals and possibly other species. Only the Bud family (proteinase inhibitor I family) serpins and cystatin inhibitors have homologs outside of the plant kingdom and are distributed widely in plants (15).

An additional feature of proteinase inhibitors is the extent of subfunctionalization (i.e. evolution from a different class of PI) that has occurred (15). This is true particularly of the aspartic PIs, where four of the six known families appear to have evolved from PIs of other classes: Kunitz and serpin aspartic proteinase inhibitors belong to already established families of serine proteinase inhibitors (16); thyroglobulin type-1 inhibitors appear to be more common as a cysteine inhibitor family (MEROPS, the peptidase database), and we recently proposed that the
Solution Structure of SQAPI

SQAPI family has evolved from the phytocystatin inhibitor family (14).

The evidence for the hypothesis that SQAPI evolved from an ancestral cystatin was based on amino acid sequence homology, phylogenetic analysis, and three-dimensional modeling (14). Moreover, evidence from hypervariability in the putative contact regions and retention of predicted secondary structural elements suggested that the cysteine proteinase inhibitor mechanism also was recruited. Nevertheless, final acceptance of such a hypothesis must rest on direct structural determination, for understanding both the evolutionary history and the inhibitory mechanism of SQAPI. Here, we report the NMR solution structure of recombinant SQAPI and show that it does indeed have a close similarity to the structure of phytocystatin. Despite their different targets, our mutational data support a similar inhibitory mechanism. SQAPI, therefore, represents a model system that may provide an understanding of the evolutionary process of subfunctionalization.

EXPERIMENTAL PROCEDURES

Expression and Purification of rSQAPI for NMR Studies—Escherichia coli BL21 (DE23) pLysS cells (17) containing the previously described His-tagged rSQAPI (HDVA isoform GenBank™ AAT67162) expression vector (3) were grown at 37 °C in LB medium containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml). When the culture reached an A₆₀₀ of 0.5, solid glucose was added (5 g/liter), and the culture was incubated for a further hour. Cells were harvested, transferred to a modified M9 minimal medium (18) and incubated for 30 min to deplete nutrients carried over from the rich medium. The cells were transferred to fresh minimal medium (300 ml) containing glucose (8 g/liter) and NH₄Cl (1.67 g/liter) in a baffled flask to ensure good aeration. For production of 15N-labeled protein, 15N-enriched NH₄Cl (99%, product code: NLM-467-1, Spectra Stable Isotopes) was used. For production of 13C-labeled protein, 15N-enriched NH₄Cl, and 13C-labeled d-glucose (100% [U-13C]C-6, 99% product code: CLM-1396-1 Cambridge Isotope Laboratories) were used, and the glucose concentration was 4 g/liter. Once the A₆₀₀ began to increase, protein expression was induced with isopropyl 1-thio-D-galactopyranoside (0.32 mM), and when the A₆₀₀ reached a plateau, the cells were harvested (typically ~2 h after induction).

Cells were disrupted using a French press in 20 mM sodium phosphate buffer, pH 7.8, containing 500 mM NaCl. The cell-free supernatant was applied to a nickel affinity column (Pharmacia Biotech; HiTrap chelating), and the column was washed successively with 10 bed volumes of loading buffer, 20 bed volumes of 20 mM sodium phosphate buffer, pH 6.0, containing 500 mM NaCl, and 20 bed volumes of 20 mM sodium phosphate buffer, pH 6.0, containing 500 mM NaCl and 300 mM imidazole. His-tagged rSQAPI was eluted from the column with 20 mM sodium phosphate buffer, pH 6.0, containing 500 mM NaCl and 50 mM EDTA. Fractions containing His-tagged rSQAPI were pooled and concentrated, and the buffer was exchanged into 10 mM KH₂PO₄, pH 3.0, containing 0.2% NaN₃.

For removal of the N-terminal fusion peptide, His-tagged rSQAPI (<3 mg/ml) in 50 mM Tris-HCl buffer, pH 8.0, containing 1 mM CaCl₂ and 0.1% Tween was incubated for 18 h at 37 °C with 1 unit of enterokinase (Ekmax; Invitrogen) per 20 µg of fusion protein. Purification of rSQAPI from the free His tag, the enterokinase, and any unprocessed His-tagged rSQAPI exploited the intrinsic affinity of rSQAPI for the nickel column. The reaction mixture was applied to the nickel affinity column (1 ml bed volume) pre-equilibrated with 20 mM sodium phosphate buffer, pH 7.8, containing 500 mM NaCl. The column was washed with 10 bed volumes of the equilibration buffer to remove the enterokinase, 10 bed volumes of 20 mM sodium phosphate buffer, pH 6.0, containing 500 mM NaCl, and 10 bed volumes of 20 mM sodium phosphate buffer, pH 6.0, containing 30 mM imidazole and 500 mM NaCl, and then rSQAPI was eluted with 20 mM sodium phosphate buffer, pH 6.0, containing 500 mM NaCl and 300 mM imidazole. Fractions containing rSQAPI were identified by SDS-PAGE and pooled, and the buffer was exchanged into 10 mM KH₂PO₄, pH 3.0, containing 0.2% NaN₃, and concentrated to between 0.06 and 0.2 mM rSQAPI. The concentration of rSQAPI was determined using the calculated extinction coefficient (3).

Preparation and Analysis of SQAPI Mutants—The gene for the HDVA variant of SQAPI was recloned into pET30 (19) and expressed with a His tag to aid purification (14). SQAPI was assayed as described previously (2, 3). Equilibrium inhibition constants were measured by titration of the inhibitor against pepsin in 100 µl 0.75% lactic acid-HCl, pH 2.0, using 500 ng of fluorocasein (Enzchek, Molecular Probes) at 480 nm excitation and 520 nm emission at room temperature in a fluorometric microplate reader (Wallac Victor 1420 multichannel counter). Alternatively, the synthetic peptide 4-methylcoumarin-Arg-Lys-Pro-Lle-Glu-Phe-Ile-Leu-Lys(dinitrophenol)-Arg-OH (custom synthesized by Auspep, Parkville, Victoria, Australia) was used to assay pepsin at 0.2 mM. Substrate cleavage C-terminal of the second lysine in this synthetic substrate was followed by measuring the increase in fluorescence at 340 nm excitation and 430 nm emission. Pepsin and SQAPI were preincubated at 20 °C for 15 min before pepsin assay. Each mutant and the wild-type SQAPI was quantified by electrophoresis using an Experion (Bio-Rad) electrophoresis unit. Pepsin was calibrated by titration against pepstatin (Sigma) at high pepsin concentrations. The pepsin concentration used was constant within a titration but varied from 1 to 7 nM depending on the replicate experiment. Inhibition constants were not corrected for the presence of substrate, as initial rates were measured, and previous assays had shown that the dissociation rate constant for SQAPI with pepsin is extremely slow (3). Data were fitted to equation (2), \( v/v_0 = 1 - ((E_r + I_r + K_r) - \sqrt{(E_r + I_r + K_r)^2 - 4 E_r I_r)})/(2 E_r) \) using nonlinear regression in the Origin (Northampton, MA) graphics package. \( E_r \) and \( I_r \) represent the total concentration of proteinase and inhibitor in the assay, whereas \( v/v_0 \) are the rates of proteinase activity in the presence and absence of inhibitor, respectively. \( K_r \) is the apparent inhibition constant.

NMR—NMR experiments were performed in Shigemi tubes containing 0.6 mM SQAPI in 95% H₂O, 5% ²H₂O with 10 mM K₂HPO₄, pH 3, 0.2% NaN₃, and 1 mM EDTA. Spectra were recorded on a 700-MHz Bruker Avance spectrometer using a triple resonance Bruker CryoProbe™. A series of 15N-HSQC spectra were recorded at 10 °C, 25 °C, 40 °C, and 50 °C. Subse-
quent spectra for structure determination were recorded at 50 °C. The following spectra were acquired for assignment and structure determination purposes: 2D 1H,1H-NOESY (mixing time 120 ms), 15N-HSQC, 13C-HSQC, HNCACB, HNCA, HNCO, HN(CA)CO, HN(CO)CAB, HCCCONH TOCSY (time 12 ms), HBHA(CO)NH, HCCCH-TOCSY, HBHA(CO)NH, (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, 13C-NOESY-HSQC (mixing time 120 ms) and a 15N-NOESY-HSQC (mixing time 120 ms). Standard pulse sequences were used for data acquisition, and water suppression was achieved using either WATERGATE (20) or field gradient coherence selection schemes (21). The chemical shifts were referenced according to published methods (22), where the 1H chemical shift is referenced to the water peak, whereas the 13C and 15N chemical shifts were referenced by the 13C/1H and 15N/1H gyromagnetic ratios. NMR data were processed in TOPSPIN (version 2.0; Bruker Biospin). The processed spectra were analyzed using XEASY (version 1.4) software (23). 1H, 13C, and 15N chemical shifts assignments have been deposited in the BioMagResBank Database (24) with accession no. 16913.

Hydrogen bond restraints were assigned based on exchange rates of 15N-bound protons measured by running a series of two-dimensional 15N-HSQC spectra at 8 min, 20 min, 1 h, and 2 h after dissolving the protein in 2H2O at 10 °C. Hydrogen bonds only were included in later rounds of structure calculations in regions of a well defined structure that initially converged on the basis of NOE restraints alone.

Structure Determination—Automated NOE cross-peak picking and structure determination were performed with the Atnos/Candid program (25, 26) using the CNS (27) simulated annealing algorithm. Initial structures were generated from an extended strand conformation using simulated annealing with torsion angle dynamics for the high temperature and fast cooling stages, followed by Cartesian dynamics for a second slow cooling stage. The Atnos/Candid program incorporated 64 dihedral restraints derived from the Cα and Cβ chemical shifts. These initial structures were then used as starting structures for a second round of structure generation in Atnos/Candid using Cartesian dynamics for both the high temperature and cooling phases. During this phase 137 backbone dihedral restraints generated from TALOS (28) were included. The structures were then refined in CNS with hydrogen bond restraints added for those residues with slow exchanging amides that had a unique hydrogen bond acceptor as determined by structural convergence. The final structures were refined in a layer of TIP3 water molecules in CNS. The 10 lowest energy structures with no NOE violations <0.25 Å, bond violations >0.05 Å, and no angle, improper or dihedral violations >5° were selected to represent the solution structure of SQAPI.

Docking Procedures—SQAPI was docked with porcine pepsin (Protein Data Bank code 5PEP (29)) using the program HADDOCK (30). This program allows the specification of intermolecular ambiguous distance restraints between residues and the solution structure of SQAPI.
on each docking partner and then uses randomization and energy minimization in CNS. Residues for which ambiguous restraints are generated are divided into two classes, termed active and passive. Active residues are usually those directly implicated in binding, whereas passive residues are near neighbors to active residues. Intermolecular ambiguous distance restraints are generated for each active residue such that the restraint is satisfied if the active residue on one binding partner is close to either an active or passive residue on the other. For SQAPI, the active residues were selected based on the mutagenesis data and comprised Ala\textsuperscript{5}, Ile\textsuperscript{6}, and Gly\textsuperscript{7} of the N-terminal region and lle\textsuperscript{48}, Pro\textsuperscript{90}, Trp\textsuperscript{52}, and Asp\textsuperscript{53} of the first loop. Residues Gly\textsuperscript{32}, Pro\textsuperscript{34}, Tyr\textsuperscript{189}, Ile\textsuperscript{213}, Asp\textsuperscript{215}, Gly\textsuperscript{217}, Thr\textsuperscript{218}, Ser\textsuperscript{219}, and Ile\textsuperscript{231} were not visible at 37 °C. Moreover, no peaks were present in the misleading \textsuperscript{15}N-HSQC peak tally. This minor sample heterogeneity had no significant effect on structure determination, as NOEs from both N-terminal variants were indicative of an unstructured N-terminal region with only intra or sequential NOEs present. HSQC spectra were then recorded at 10, 25, 37, and 50 °C. At 50 °C, peaks that were ultimately assigned to residues His\textsuperscript{51}—Ser\textsuperscript{61} became visible in the spectrum. Spectra for assignments and structure determination were then acquired at this new temperature. An assay at 50 °C showed that SQAPI retains its ability to inhibit pepsin at this temperature (data not shown). Subsequently near complete (>99%) backbone and side chain assignments were achieved for nonlabile resonances of SQAPI at 50 °C.

RESULTS

Protein Expression and Purification—Single- and double-labeled SQAPI was expressed and purified with a final yield of ~30 mg/liter of minimal media. Purity was assessed as >95% by SDS-PAGE. Enterokinase was used to cleave the His\textsubscript{8} tag containing leader sequence from SQAPI and was expected to cleave the fusion protein 10 residues upstream from the beginning of the native SQAPI sequence.

Resonance Assignment—A preliminary peak count in the \textsuperscript{15}N-HSQC spectra recorded at 37 °C found 99 cross-peaks. This tally was sufficiently close to the expected 101 peaks to commence acquisition of three-dimensional spectra for assignments. However, during the backbone assignments, it became apparent that the peaks corresponding to the region His\textsuperscript{51}—Ser\textsuperscript{61} were not visible at 37 °C. Moreover, no peaks were present that could be assigned to the 10 residues of the vector encoded leader sequence. In addition, multiple resonances were assigned for the first five residues of SQAPI. To clarify the cause of this N-terminal peak doubling, an aliquot of the \textsuperscript{15}N SQAPI sample was subjected to N-terminal sequencing. This revealed that 74% of the rSQAPI had been cleaved, by enterokinase, at the arginine immediately preceding the native SQAPI sequence, and 26% had been cleaved at Met\textsuperscript{1}, the two forms therefore differing by one amino acid. This sample heterogeneity accounts for both the duplication of resonances observed for the first five residues of SQAPI in the NMR spectra and the misleading \textsuperscript{15}N-HSQC peak tally. This minor sample heterogeneity had no significant effect on structure determination, as NOEs present. HSQC spectra were then recorded at 10, 25, 37, and 50 °C. At 50 °C, a structure of SQAPI at 50 °C. An assay at 50 °C showed that SQAPI retains its ability to inhibit pepsin at this temperature (data not shown). Subsequently near complete (>99%) backbone and side chain assignments were achieved for nonlabile resonances of SQAPI at 50 °C.

Structure—Stereoviews of the 10 lowest energy SQAPI structures refined in water and a schematic view of the closest to average structure are shown in Fig. 1. Structural statistics for the ensemble are given in Table 1. SQAPI assumes a cystatin-like fold consisting of a twisted four-strand antiparallel \(\beta\)-sheet gripping an \(\alpha\)-helix like the fingers of a hand gripping a tennis racquet. With the exception of the first 10 residues, which are relatively disordered, the solution structure is well defined with a backbone heavy atom root mean square deviation of 0.44 Å for residues 11–95. There is a well defined turn from Gly\textsuperscript{11} through to the \(\alpha\)-helix that extends from Pro\textsuperscript{17}—Gln\textsuperscript{29}. A second structured turn bridges the \(\alpha\)-helix to the first \(\beta\)-strand, lle\textsuperscript{36}—lle\textsuperscript{46}. The first loop region, Lys\textsuperscript{47}—Asn\textsuperscript{54}, links the first strand with the second \(\beta\)-strand, Tyr\textsuperscript{55}—Lys\textsuperscript{63}. At the other extreme of the molecule, another structured turn His\textsuperscript{80}—Ser\textsuperscript{87} links the second strand with the third \(\beta\)-strand, Lys\textsuperscript{71}—Lys\textsuperscript{80}. A four-residue hairpin loop, Ala\textsuperscript{81}—Asn\textsuperscript{84}, links the third stand with the fourth \(\beta\)-strand, Ser\textsuperscript{85}—Phe\textsuperscript{95}, that extends to the C terminus. Coordinates have been deposited in the Protein Data Bank (Protein Data Bank code 2KXG).
Pepsin Inhibition Assays—To obtain further information on the interactions of SQAPI with pepsin, a series of SQAPI point mutants with stepwise glycine substitution along the Lys47–Asn54 and Ala81–Asn84 loops were prepared, and the inhibition constants were determined by titration of SQAPI against a constant concentration of pepsin. A typical assay is shown in Fig. 2A for wild-type SQAPI and one mutation and the effect of each residue change on the $K_I$ values is shown in Fig. 2B. The $K_I$ values were determined using two different substrates that had little effect on the $K_I$ values obtained, with those with the peptide substrate (listed first) generally being marginally lower: for I49G (14.9 and 15.2), P50G (11.3 and 14.4), H51G (1.3 and 1.8), D53G (14.9 and 13.5), S83G (both 0.8), D84G (both 1.0), and N85G (0.7 and 0.8). W52G was only measured with the synthetic substrate. Thus, whereas substitutions in loop 1 increased the $K_I$ by ~10–20 fold at every position except His51, substitutions in loop 2 had little effect.

Two other mutants were made by truncation of the N-terminal residues up to and including Ile6 or Gly7 (Fig. 2C). For the ∆-Ile6 construct, partial inhibition of pepsin was only evident at higher SQAPI concentrations. The ∆-Gly7 construct was almost completely inactive.

Docking—Earlier studies (2, 3) showed that SQAPI was a strong inhibitor of pepsin. Therefore, a simulated interaction was performed as described under “Experimental Procedures.” The 10 lowest energy solutions depicted SQAPI bound to pepsin in a similar orientation. The lowest energy conformation was selected to best represent the structure of the pepsin-SQAPI complex (Fig. 3, A and B). The structure shows the N-terminal region of SQAPI binding the S′ side of the substrate binding cleft, whereas the Lys47–Asn54 loop binds on the S side of the cleft. Whereas the N terminus of isolated SQAPI is disordered in solution, in the SQAPI-pepsin model, the N terminus forms a β-strand. In the model, the pepsin catalytic site Asp32–Asp215 diad does not contact the peptide backbone of SQAPI, a point that will be further discussed below.

DISCUSSION

The solution structure of aspartic proteinase inhibitor SQAPI (Fig. 1) that we determined in the present study shares its fold with cystatin (31), suggesting the PI is derived from the ancestral cystatin protein, a family of cysteine proteinase inhibitors that is widely distributed through plants and animals (32). SQAPI, on the other hand, has been found only within the Cucurbitales (14), indicating a recent origin. This is the first example of recruitment of the cystatin structure to an inhibitor of a different proteinase class. Six families of proteinaceous aspartic PIs have been identified and are distinct from SQAPI and from one another, both in structure and in mechanism. SQAPI appears to have retained a similar inhibitory mechanism to its
We have generated a model of the inhibition of pepsin by SQAPI using the program HADDOCK, which utilizes ambiguous interaction restraints (Fig. 3, A and B). The docking was restrained based on our mutagenesis results for SQAPI, whereas for pepsin, in the absence of mutagenesis data, the residues within 10 Å of the active site were used to generate restraints. This approach assumes that SQAPI is an active site inhibitor of pepsin as opposed to an allosteric inhibitor. The resultant model supports the proposal that the mechanism for SQAPI binding is similar to that of cystatins, where the N-terminal strand binds on the S’ side of the active site cleft and loop 1 to the S side (Fig. 3, A and B). The backbone of SQAPI does not interact with the catalytic diad, thus avoiding SQAPI being inactivated by catalytic cleavage. The majority of the intermolecular contacts are hydrophobic in nature with the exception of SQAPI Asp^{53}, for which hydrogen bonding between its carboxylate side chain and the backbone amides of pepsin residues Ser^{110} and Phe^{111} are predicted.

The NMR resonances for residues His^{51}–Ser^{61}, which include the important binding determinants Trp^{52} and Asp^{53}, are only visible in the ^15N-HSQC spectra at elevated temperatures. This indicates a degree of dynamic broadening at lower temperatures attributable to intermediate conformational exchange. Intermediate exchange occurs when the NMR frequency differences arising between states are commensurate with the lifetime of the states and is usually in the order of milliseconds. Such intermediate timescale motions are commonly described for regions of proteins involved in catalysis (33) or protein–protein interactions (34). This raises the prospect of multiple conformations occurring in this region of the protein. Elevating the temperature to 50 °C may have either averaged the states by inducing fast exchange or may simply favor one particular state. If multiple states of this loop exist, they cannot be separately characterized via the present analysis. Nevertheless, we have shown that SQAPI retains activity at 50 °C, the temperature at which the structure was determined.

Although loop 2 of cystatin, analogous to the Ala^{81}–Asn^{84} loop of SQAPI, contributes binding energy to interactions with some cysteine proteinases, it does not interact strongly with others (35). The sequence Ala^{81}–Lys^{87} is invariant across all 11 SQAPI variants for which a C-terminal sequence is available (3), leading us to hypothesize a role for SQAPI loop 2 in aspartic peptidase inhibition. To investigate this possibility, we mutated the SQAPI residues Ser^{82}, Asp^{83}, and Asn^{84}. The absence of variation in $K_i$ upon glycine substitution suggests that these invariant residues are not important for pepsin binding. Our docking results utilized ambiguous interaction restraints based upon loss of inhibition seen for N-terminal and Lys^{17}–Asn^{84} loop mutants. As a consequence, our model includes these binding determinants at the interface. In agreement with the mutagenesis data, our model does not support the direct involvement of the Ala^{81}–Asn^{84} loop in inhibition. However, peptide contacts are predicted by our model between the nearby residues Leu^{78} and Lys^{100} from the third $\beta$-strand of SQAPI and Gly^{109} and Ser^{110} of pepsin.

With SQAPI, four of the seven known proteinaceous aspartic proteinase inhibitors have been recruited from ancestral and widely distributed forms of inhibitors of other protease
classes (36). The other three inhibitors, each a unique protein, have highly restricted distribution, suggestive of recent evolution. The paucity of aspartic proteinase inhibitors and their recent evolution may indicate that aspartic proteinases themselves are a more recent evolutionary event than serine and cysteine proteinases.

REFERENCES

1. Laing, W. A., and McManus, M. T. (2002) in Protein Interactions in Plants (McManus, M. T., Laing, W. A., and Allan, A. C., eds), pp. 77–119, Sheffield Academic Press, Sheffield, UK
2. Christeller, J. T., Farley, P. C., Ramsay, R. J., Sullivan, P. A., and Laing, W. A. (1998) Eur. J. Biochem. 254, 160–167
3. Farley, P. C., Christeller, J. T., Sullivan, M. E., Sullivan, P. A., and Laing, W. A. (2002) J. Mol. Recognit. 15, 135–144
4. Mares, M., Meloun, B., Pavlik, M., Kostka, V., and Baudys, M. (1989) FEBS Lett. 251, 94–98
5. Martzen, M. R., McMullen, B. A., Smith, N. E., Fujikawa, K., and Peanasky, R. J. (1990) Biochemistry 29, 7366–7372
6. Schu, P., Suarez Rendueles, P., and Wolf, D. H. (1991) Eur. J. Biochem. 197, 1–7
7. Galesa, K., Pain, R., Jongsma, M. A., Turk, V., and Lenarcic, B. (2003) FEBS Lett. 539, 120–124
8. Lenarcic, B., and Turk, V. (1999) J. Biol. Chem. 274, 563–566
9. Mathialagan, N., and Hansen, T. R. (1996) Proc. Natl. Acad. Sci. U.S.A. 93, 13653–13658
10. Dash, C., Phadtare, S., Deshpande, V., and Rao, M. (2001) Biochemistry 40, 11525–11532
11. Li, M., Phlyip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, A., Kay, J., and Gustchina, A. (2000) Nat. Struct. Biol. 7, 113–117
12. Ng, K. K., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., Dunn, B. M., Martzen, M. R., Peanasky, R. J., and James, M. N. (2000) Nat. Struct. Biol. 7, 653–657
13. Petersen, J. F., Chernea, M. M., Rao-Naik, C., Zalatoris, J. L., Dunn, B. M., and James, M. N. (1998) Adv. Exp. Med. Biol. 436, 391–395
14. Christeller, J. T., Farley, P. C., Marshall, R. K., Anandan, A., Wright, M. B., Newcomb, R. D., and Laing, W. A. (2006) J. Mol. Evol. 63, 747–757
15. Christeller, J., and Laing, W. (2005) Protein Pept. Lett. 12, 439–447
16. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) Biochem. J. 378, 705–716
17. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) Methods Enzymol. 185, 60–89
18. Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., and Craigie, R. (1998) J. Biomol. NMR 11, 97–102
19. Laing, W. A., Bulley, S., Wright, M., Cooney, J., Jensen, D., Barralclough, D., and MacRae, E. (2004) Proc. Natl. Acad. Sci. U.S.A. 101, 16976–16981
20. Piotto, M., Saudek, V., and Sklenář, V. (1992) J. Biomol. NMR 2, 661–665
21. Tolman, J. R., Chung, J., and Prestegard, J. H. (1992) J. Magn. Res. 98, 462–467
22. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) J. Biomol. NMR 6, 135–140
23. Bartels, T., Xia, H., Billetter, M., Güntert, P., and Wüthrich, K. (1995) J. Biomol. NMR 6, 1–10
24. Marin, A., Malliavin, T. E., Nicolas, P., and Delsuc, M. A. (2004) J. Biomol. NMR 30, 47–60
25. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) J. Biomol. NMR 24, 171–189
26. Herrmann, T., Güntert, P., and Wüthrich, K. (2002)) J. Mol. Biol. 319, 209–227
27. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and G. L., W. (1998) Acta Cryst. D54, 905–921
28. Cornilescu, G., Delaglio, F., and Bax, A. (1999) J. Biomol. NMR 13, 289–302
29. Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J., and Blundell, T. L. (1990) J. Mol. Biol. 214, 199–222
30. Domínguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) J. Am. Chem. Soc. 125, 1731–1737
31. Nagata, K., Kudo, N., Abe, K., Arai, S., and Tanokura, M. (2000) Biochemistry 39, 14753–14760
32. Margis, R., Reis, E. M., and Villeret, V. (1998) Arch. Biochem. Biophys. 359, 24–30
33. McElheny, D., Schnell, J. R., Lansing, J. C., Dyson, H. J., and Wright, P. E. (2005) Proc. Natl. Acad. Sci. U.S.A. 102, 5032–5037
34. Yao, S., Headey, S. J., Keizer, D. W., Bach, L. A., and Norton, R. S. (2004) Biochemistry 43, 11187–11195
35. Pol, E., and Björk, I. (1999) Biochemistry 38, 10519–10526
36. Christeller, J. T. (2005) FEBS J. 272, 5710–5722