# Inhibition of Trypsin by Cowpea Thionin: Characterization, Molecular Modeling, and Docking

**Francislete R. Melo,**[1,2*] **Daniel J. Rigden,**[2] **Octávio L. Franco,**[2,3] **Luciane V. Mello,**[2] **Maria B. Ary,**[4]
**Maria F. Grossi de Sá**[2] **and Carlos Bloch Jr.**[2]

[1]*Departamento de Biologia Celular, Universidade de Brasília, Brasília-DF, Brasil*
[2]*Centro Nacional de Recursos Genéticos e Biotecnologia, Embrapa/Cenargen, Brasília-DF, Brasil*
[3]*Pós-Graduação em Ciências Genômicas, Universidade Católica de Brasília, Brasília-DF, Brasil*
[4]*Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará, Fortaleza-CE, Brasil*

***ABSTRACT*** **Higher plants produce several families of proteins with toxic properties, which act as defense compounds against pests and pathogens. The thionin family represents one family and comprises low molecular mass cysteine-rich proteins, usually basic and distributed in different plant tissues. Here, we report the purification and characterization of a new thionin from cowpea (*Vigna unguiculata*) with proteinase inhibitory activity. Cowpea thionin inhibits trypsin, but not chymotrypsin, binding with a stoichiometry of 1:1 as shown with the use of mass spectrometry. Previous annotations of thionins as proteinase inhibitors were based on their erroneous identification as homologues of Bowman-Birk family inhibitors. Molecular modeling experiments were used to propose a mode of docking of cowpea thionin with trypsin. Consideration of the dynamic properties of the cowpea thionin was essential to arrive at a model with favorable interface characteristics comparable with structures of trypsin-inhibitor complexes determined by X-ray crystallography. In the final model, Lys11 occupies the S1 specificity pocket of trypsin as part of a canonical style interaction. Proteins 2002;48:311–319.** © 2002 Wiley-Liss, Inc.

**Key words: thionin; cowpea; trypsin inhibitor; molecular modeling**

## INTRODUCTION

Plants produce substances that act as natural inhibitors for a variety of enzymes and metabolic processes.[1] Among known plant protection compounds are the low molecular mass substances, polyanions, and many different proteins with antifungal and/or antimicrobial activity. These include the chitinases, glucanases, lipid transfer proteins, chitin-binding lectins, and thionins. Thionins are proteins of about 5 kDa found in different tissues of many plant species.[2] They are part of a group of usually basic, cysteine-rich peptides with toxic and antimicrobial properties.[3,4] Thionins occur in the seed endosperm, stems, roots, and in etiolated or pathogen-stressed leaves of a number of plant species.[2]

The three-dimensional structures of thionins have been studied in detail, both by X-ray crystallography and NMR.[5] The molecules are amphipathic and consist of a two-layer

α-β sandwich.[6] Based on disulfide-bond structure, all known thionins can be classified into three groups: a group with four disulfide bounds, which would include types I and II, a group with only three of the above disulfide bounds (types III and IV), and a group that has only two of the above bonds.[5]

Members of the thionin family possess broad activity against gram-positive and/or gram-negative bacteria[7] and fungi.[8,9] Their toxicity is due to an electrostatic interaction of thionin with the negatively charged membrane phospholipids, followed by either pore formation or a specific interaction with a membrane.[3] It has been reported that thionins are able to inhibit certain other enzymes possibly by covalent attachment through the formation of disulfide bonds.[10] The enzymes β-glucuronidase and neomycin phosphotransferase II, for example, were inactivated by thionins in a concentration and time-dependent manner, in a way that could be prevented and reversed by dithiothreitol (DTT).[10] A further activity—inhibition of α-amylases—is associated with barley thionin.[11]

Several thionin sequences are annotated as trypsin inhibitors or probable trypsin inhibitors on the basis of reported sequence similarity with Bowman-Birk type proteinase inhibitors.[12] Here we show this relationship to be entirely non-existent but, nevertheless, find bovine pancreatic trypsin (BPT) inhibitory activity associated with the newly characterized cowpea thionin (Cp-thionin). Furthermore, docking experiments, allied to consideration of the dynamic properties of Cp-thionin, lead to a highly plau-

sible predicted binding mode of cowpea thionin in complex with trypsin.

## MATERIALS AND METHODS
### Protein Purification

A crude extract was obtained from cowpea seeds (*Vigna unguiculata* cv. Epace-10) with 0.1 M HCl, 0.15 M NaCl followed by a fractionation with ammonium sulfate (0–60%). The fraction 0–60 was applied on a column Red-Sepharose according to Melo et al.,[13] and the retained peak was applied on a HPLC reverse-phase analytical column (Vydac 218 TP 1022 C-18) at a flow rate of 1.0 mL/min. The material eluted in individual peaks was collected, lyophilized, and stored at −20°C.

### MALDI-TOF Mass Spectrometry Analysis

To study the complex formed by Cp-thionin and trypsin, both proteins were incubated at 37°C for 60 min in water in the ratio of 1 mole of inhibitor to 1 mole of trypsin. The complex and the freeze-dried samples corresponding to peak from HPLC were prepared for matrix-assisted laser desorption time of flight analysis (MALDI-TOF) on a Voyager-DE STR Bioworkstation (PerSeptive Biosystems, Framingham, MA) according to Melo et al.[13]

### Amino Acid Sequence Determination and Analysis

A sample of purified Cp-thionin was reduced according to Crestfield et al.[14] Reversed phase HPLC (Vydac C-18) using a linear gradient of 0–100% acetonitrile in 0.1% (v/v) aqueous trifluoroacetic acid was used to desalt the sample. The complete sequence was determined from 2–5 pmol of alkylated sample by using an automatic standard Edman degradation in a PPSQ-23 sequencer from Shimadzu Co.

### Proteinase and Inhibitory Proteinase Assays

Proteolytic and proteolytic inhibitory activities were tested with 10 μM of CBZ-Phe-Arg-7-amidomethylcoumarin (AMC) for trypsin activity and CBZ-Leu-Leu-Val-Tyr-AMC for chymotrypsin-like activity in buffer (2.0 mM Tris-EDTA, pH 6.5, 2.5 mM DTT) in a 100-μL reaction volume according Solomon et al.[15] with minor modifications. The reaction was stopped after 30 min with 0.2 M $Na_2CO_3$. Measurements were made in a DyNA Quant 500 fluorescence reader (Pharmacia-Biotech), with excitation at 365 nm and emission at 460 nm. TPCK (N-tosyl-L-phenylalanine chloromethyl ketone), PMSF (phenylmethylsulfonyl fluoride), SKTI (soybean Kunitz-type trypsin inhibitor), BTCI (black-eye pea trypsin/chymotrypsin inhibitor), purified according de Freitas et al.[16] and Cp-thionin were used at a standard concentration of 50 μg/mL. Pancreatic bovine trypsin, substrates and inhibitors, with the exceptions of BTCI and Cp-thionin itself, were all purchased from Sigma Co. (St. Louis, MO). Assays were carried out in triplicate. Triplicate inhibition values differed by no more than 10%.

### Construction of a Model of Cp-Thionin

Comparison of the Cp-thionin sequence with the SWISSPROT database was carried out by using the FASTA3 program.[17] The thionin fold is shared my many different classes of protein, including insect toxins, amylase inhibitors, antimicrobial proteins, and antifungal proteins.[3] The best templates for Cp-thionin model construction was determined by threading methods[18–21] to be the structures of gamma-1P and gamma-1S thionins (PDB files 1gps and 1gpt, respectively) as determined by NMR.[22] These shared around 50% sequence identity in an ungapped alignment with Cp-thionin.

The most representative of the eight NMR structures deposited in each PDB file was used for model construction. The OLDERADO website (http://neon.chem.le.ac.uk/olderado/)[23] showed these to be model 1 in the case of 1gps and model 7 for 1gpt. With use of these templates, 10 models were constructed by using MODELLER,[24,25] and PROSA II[26] was used to choose the model showing the most favorable packing and solvent exposure characteristics. PROCHECK[27] was used for additional analysis of stereochemical quality. High-quality models are characterised by low PROSA II scores and high PROCHECK G-factors.

### Dynamics Simulations

Modes of conformational freedom of the Cp-thionin structure were explored by using 1000 structures generated by CONCOORD[28,29] and subsequently subjected to Essential Dynamics (ED) analysis.[30,31] Maximum and minimum projected structures were obtained for each of the largest six eigenvectors, together responsible for 88% of conformational freedom.

### Docking Studies

Two different approaches were applied to analyze possible modes of binding of Cp-thionin to bovine trypsin. In the first, the possibility was tested that Cp-thionin would inhibit trypsin in the canonical fashion[32] seen for various other families of trypsin inhibitors. For this inhibition mode, the key factor is the presence of a lysine or arginine residue (P1) in an exposed loop of extended conformation,[33] which binds in the S1 specificity pocket of trypsin. Complexes of trypsin bound to representatives of four different classes of canonical inhibitor (bovine pancreatic trypsin inhibitor-2ptc[34]; *Streptomyces* subtilisin inhibitor-2tld[35]; squash trypsin inhibitor-1ppe[36]; Bowman-Birk type inhibitor from adzuki bean-1tab[37]) were superposed by using enzyme coordinates alone. The superposition showed strong similarity in backbone structure for five inhibitor residues with the P1 lysine or arginine occupying the central position (RMS differences for these five residues, calculated by using 2ptc as a base and Cα atoms alone, were 0.66, 0.43, and 0.7 Å for 2tld, 1ppe, and 1tab, respectively). Each lysine or arginine present in the model of Cp-thionin, along with two flanking residues on either side, was then fit to the canonically binding five residues of the inhibitor molecule present in 2ptc.

The second docking approach applied was the use of the program GRAMM.[38] This approach made no assumptions about binding mode so that possible noncanonical modes, such as have been observed recently in thrombin complexes,[39,40] could be explored.

Putative docking models were refined with MULTI-DOCK.[41] a program specially designed for refinement of complex coordinates and X-PLOR 3.851.[42] A constant elevated dielectric constant of 5 was used because protein interfaces are known to be more polar than protein interior.[43] Analysis of refined structures made use of the Protein-Protein Interaction Server (http://www.biochem.ucl.ac.uk/bsm/PP/server/).[43] In addition, enzyme-inhibitor interaction energy was calculated by using X-PLOR, and HBPLUS[44] was used to count hydrophobic contacts between enzyme and inhibitor.

Initial predictions of binding mode from either approach were further explored by using the predicted conformational flexibility of Cp-thionin. For the canonical binding-based approach, the minimum and maximum projections for the first 6 eigenvectors were superposed as above. For GRAMM-derived results, these 12 projections were superposed on refined putative inhibitor-binding mode coordinates. Because these projections consisted of Cα coordinates alone, it was necessary to regenerate corresponding complete protein structures. This was carried out by using MODELLER with two structures used as input. The first was one of the 12 ED-derived projections from which alone Cα restraints were derived. The other input structure was either the original model, in studies assuming canonical binding or the refined GRAMM-derived putative docking mode. This second structure was used for the calculation of other main- and side-chain restraints. By manipulating the ranges over which these restraints were calculated, it proved possible to regenerate structures with similar Cα coordinates to the projections (RMS values of 0.195–0.344 Å; mean 0.268 Å) but with side-chain conformations, visually examined, closely resembling those of the complete structure. After docking of these variants, refinement and analysis were carried out as described above.

## RESULTS
### Purification and Characterization of Cp-Thionin

The purification of Cp-thionin involved the use first of a Red-Sepharose column, as was also used for purification of α-amylase inhibitors.[13] The retained peak was dialyzed with distilled water, lyophilised, and subjected to analytical reversed-phase HPLC, yielding three peaks (Fig. 1). These peaks were monitored at 216 nm and identified generically as peaks A, B, and C. The major peak, denominated B, was resolved at 47% acetonitrile and eluted at 24 min (Fig. 1). By mass spectrometry analysis, this peak contained a monomeric peptide with molecular mass of 5229.03 Da (Fig. 2). MALDI-TOF mass spectrometry results clearly showed the presence of homodimers with 10459.84 Da along with smaller quantities of other multimers (Fig. 2).

Sequencing of Cp-thionin revealed a protein of 47 residues including eight cysteines (Fig. 3). The three residues preceding the final C-terminal Cys were not as clearly identified as the bulk of the sequence but, nevertheless, represent strongly favored amino acids.

The inhibitory activity of Cp-thionin toward trypsin and chymotrypsin from bovine pancreas was tested in enzymatic assays with the results shown in Figure 4. Cp-
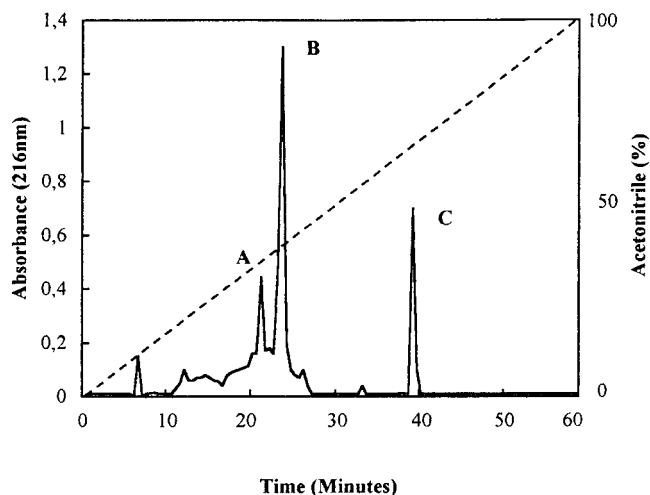


Fig. 1. Analytical reversed-phase HPLC of retained peak from Red-Sepharose chromatography.[13] The separation was carried out on a Vydac 218TP analytical column using a flow rate of 1 mL/min. TFA (0.1%) was used as ion-pairing agent, and the dashed line indicates the acetonitrile gradient. The sample contained 1.0 mg of protein.
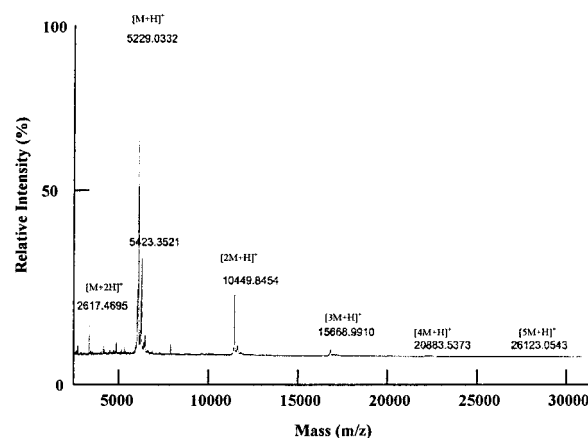


Fig. 2. MALDI-TOF spectrum of peak B purified from reversed phase HPLC.

thionin at a concentration of 50 μg/mL produced inhibition of around 73%, somewhat less than observed for the positive controls—soybean Kunitz trypsin inhibitor (SKTI) and PMSF [Fig. 4(A)]. Cp-thionin did not inhibit chymotrypsin [Fig. 4(B)]. This last result showed that Cp-thionin differs in this respect to Bowman-Birk inhibitors, which inhibit both trypsin and chymotrypsin.

To confirm the binding of Cp-thionin to bovine pancreatic trypsin, mass spectrometry experiments were performed (Fig. 5). In Figure 5(A), trypsin alone is seen to have a molecular mass of 23565 Da. When thionin was added to enzyme solution, at the same concentration, a new peak was observed with a molecular mass of 28795 Da, corresponding to the sum of the molecular masses of trypsin (23565 Da) and Cp-thionin (5229 Da).

### Modeling and Dynamics Studies

The templates used to construct the model of Cp-thionin shared around 50% sequence identity with Cp-thionin in

```
                              10              20              30              40
1gps        51%  K I C R R R S A G F K G P C M S N K N C A Q V C Q Q E G W G G G N C D G P F R R C K C I R Q C
1gpt        53%  R I C R R R S A G F K G P C V S N K N C A Q V C M Q E G W G G G N C D G P L R R C K C M R R C
Cp-thionin       R V C E S Q S H G F K G A C T G D H N C A L V C R N E G F S G G N C R G F R R R C F C L T K C
Q39894 SRP  81%  R V C E S Q S H G F H G L C N R D H N C A L V C R N E G F S G G R C K G F R R R C F C T R I C
Q39182 PI   72%  R T C E S Q S H R F K G T C V S A S N C A N V C H N E G F V G G N C R G F R R R C F C T R H C
Q9XG53 GT   70%  R T C E S Q S H R F K G P C V S E K N C A S V C E T E G F S G G D C R G F R R R C F C T R P C
AAF72042 D  72%  R T C E S Q S H K F K G T C L S D T N C A N V C H S E R F S G G K C R G F R R R C F C T T H C
P81008 GZT  57%  R V C R R R S A G F K G V C M S D H N C A Q V C L Q E G Y G G G N C D G I M R Q C K C I R Q C
1gps sec str  →→→→→→→→→→→→       →→→→→→→→→→→→→→→→→→→
```
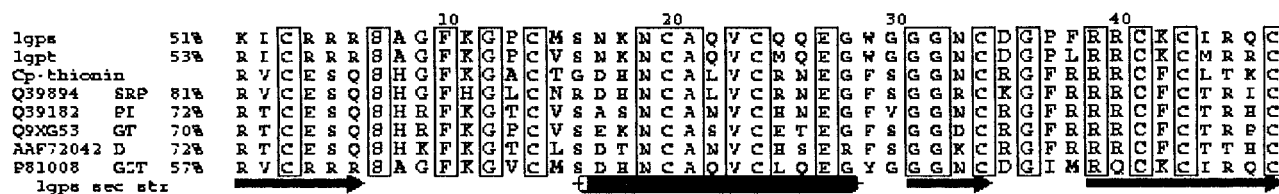
Fig. 3. Alignment of the Cp-thionin sequence with the sequences of the known structures used to model Cp-thionin (above Cp-thionin itself) and selected homologues found in the Swissprot database (below). The database hits are labeled with Swissprot code, Swissprot annotation, and degree of sequence similarity to Cp-thionin. The annotations are as follows: SRP: sulfur-rich protein[65]; PI: proteinase inhibitor (direct submission to Swissprot); GT: gamma-thionin (direct submission to Swissprot); D: defensin[66]; GZT: gamma-zeathionin.[67]
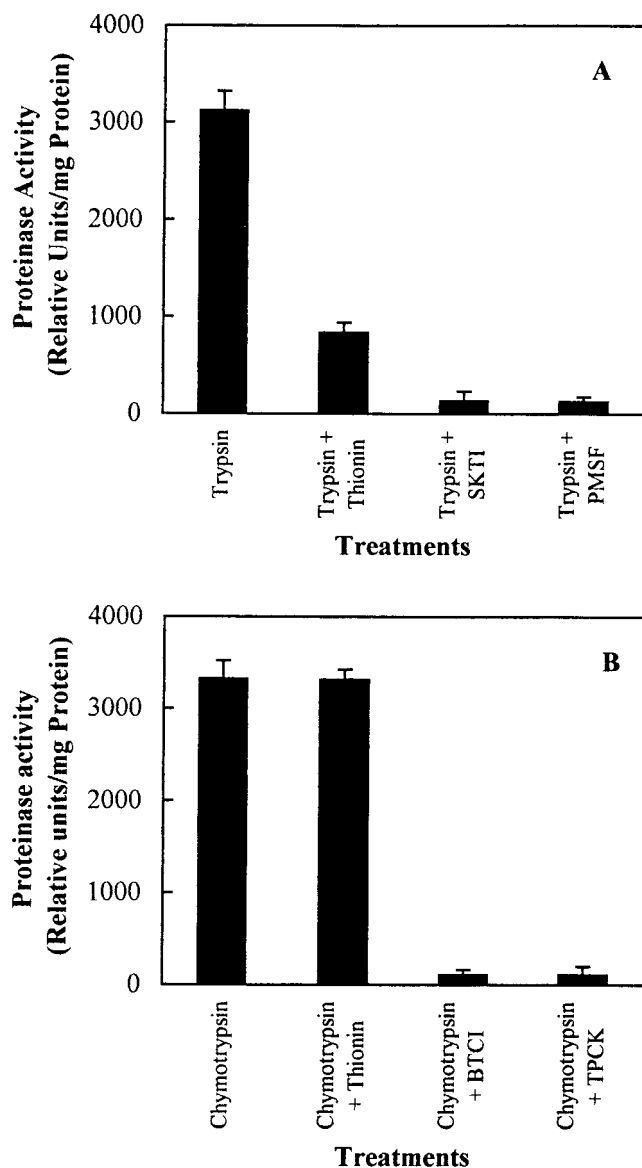


Fig. 4. **A:** Inhibitory activities of purified Cp-thionin, SKTI, and PMSF toward BPT. **B:** Inhibitory activities of purified Cp-thionin, BTCI, and TPCK toward bovine pancreatic chymotrypsin. Each measurement was performed in triplicate. Error bars show standard deviations.
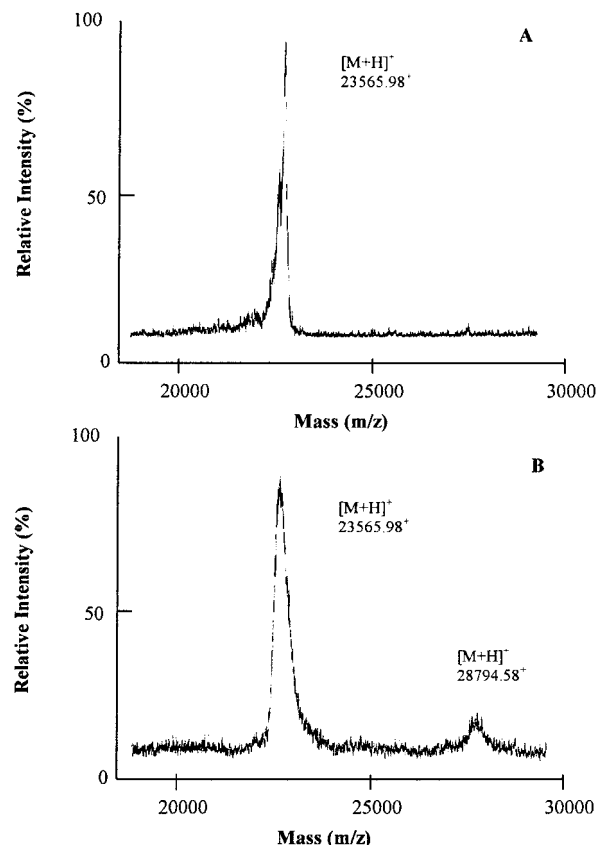


Fig. 5. MALDI-TOF spectra of semipurified BPT (**A**) alone and (**B**) after incubation with purified Cp-thionin in distilled water for 3 h at 37°C.

score of $-9.36$ compared with the ranges $-7.39$ to $-9.04$ and $-7.01$ to $-8.65$ seen for the several NMR structures of 1gps and 1gpt, respectively. Similarly, the overall G-factor for the final model was $-0.30$ compared with $-0.35$ to $-0.25$ and $-0.33$ to $-0.15$ for 1gps and 1gpt, respectively. Hence, by analysis both of packing and solvent exposure and of stereochemistry, the model of Cp-thionin performed as well as or better than the template structures.

The predicted structural flexibility of the Cp-thionin, visualized after ED analysis of the CONCOORD-derived structures, was compared with the RMS deviation of backbone torsion angles reported for the different NMR template structures.[22] In the template structures the regions between residues 7–9, 12–16, 30–31, and 37–39 are the most variable. For the Cp-thionin model, these flexible regions are visible in analysis of the contribution of

an ungapped alignment. Therefore, the good quality expected for of the model was confirmed by PROSA II and PROCHECK analysis. The final model had a PROSA II
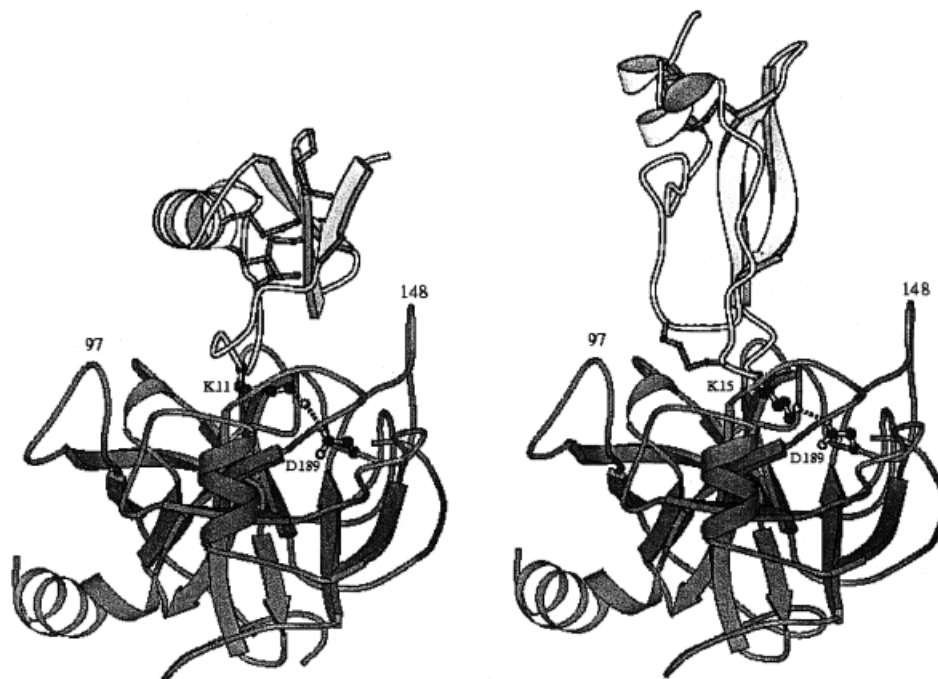
Fig. 6. Ribbon diagram comparison of the predicted mode of Cp-thionin binding to BPT (**left**) with the crystallographically observed mode of inhibition of BPTI (**right**). BPT is colored gray. The P1 lysine residues and residue Asp189 of the enzyme at the base of the S1 specificity pocket of BPT are shown in ball-and-stick representation, as are disulfide bonds present in the inhibitors. Water molecules bridging the interaction between the P1 lysine and Asp189 are shown as isolated spheres, and their hydrogen bonds as dotted lines. Enzyme loops containing residues 97 and 148, which interact differentially in the two binding modes, are labeled.

different residues to the most important eigenvectors resulting from ED analysis. For example, peaks at positions 12, 26, and 37 are visible in the by-residue analysis of eigenvector 1.

## Docking Studies

The same analyses were applied to all putative inhibitor binding modes, irrespective of their origin in manual, canonical docking or in the results of GRAMM. Various analyses have been published of protein-protein interfaces[43,45] and docking methods with different emphases developed. Although some studies have suggested that special consideration of hydrophobicity,[46] electrostatics,[47,48] or desolvation[49] is important, other work has emphasized the predominant involvement of simple complementarity.[50] In our analyses, complementarity, measured as gap volume index in the output of the Protein-Protein Interaction Server, was given a key role. The interaction energy, number of H-bonds, and number of hydrophobic contacts were also monitored closely. Even the most advanced docking programs can only be expected to produce the correct solution among a variable number of similarly scoring solutions so that application of biochemical knowledge remains crucial.[47] This consideration prompted not only the testing of canonical binding modes for Cp-thionin but also the discarding of GRAMM solutions for inhibitor positions lying distant from the active site cleft of the enzyme.

The sequence of Cp-thionin contains eight lysine or arginine residues. When each was tested through superposition on the canonically binding loop of the inhibitor in 2ptc, only one complex structure lacked serious steric clashes between Cp-thionin and trypsin: that with Lys11 as the P1 residue. In addition, fitting Lys11 gave a better RMS match to the canonical loop of 2ptc (1.24 Å for residues 9–13) than any other lysine or arginine (with the exceptions of Arg1 and Lys46, situated near the N- and C-termini, respectively, for which only three and four residues, respectively, could be fit). Also supporting the choice of Lys11 was its location on a relatively extended, exposed loop, as invariably observed for proteinase inhibitors.[33]

After refinement of this initial structure with MULTI-DOCK and X-PLOR, analysis revealed a complex with large buried surface area and favorable complementarity characteristics, as shown by it low gap volume index (Table I). To take into account possible conformational changes of inhibitor on enzyme binding and possible errors in our final model, models corresponding to maximum and minimum projections of the ED analysis were also docked with Lys11 occupying the S1 specificity pocket. In general, these modeled complexes exhibited less favorable characteristics, particularly poorer complementarity, than the original complex, but there was one interesting exception. The model complex corresponding to the maximum projection of eigenvector 3 had a slightly smaller interface area

**TABLE I. Analysis of Possible Cp-Thionin Binding Modes and Comparison With Some Known Trypsin-Inhibitor Complex Structures[†]**

| | Cp-thionin models | | | | | Experimentally determined structures | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Canonical | Canonical MAX3 | GRAMM model 2 | GRAMM model 3 | GRAMM model 5 | 2ptc | 1ppe | 1tab | All enzyme-inhibitor structures |
| Interface surface area (Å²) | 935 | 921 | 797 | 564 | 640 | 769 | 923 | 790 | 785 ± 75 |
| Gap volume index | 1.83 | 1.70 | 2.89 | 3.84 | 3.00 | 2.09 | 1.19 | 3.05 | 2.2 ± 0.5 |
| X-PLOR interaction energy | −27.5 | −27.1 | −17.4 | −18.0 | −23.3 | −32.9 | −33.1 | +34.9 | n/a |
| Interface H-bonds | 5 | 11 | 0 | 6 | 3 | 12 | 12 | 7 | 1.4 ± 0.5 per 100Å² interface |
| Interface hydrophobic contacts | 23 | 42 | 21 | 20 | 33 | 37 | 43 | 35 | n/a |
| Planarity | 3.05 | 2.38 | 1.95 | 1.99 | 2.56 | 1.92 | 2.90 | 2.35 | 2.7 ± 0.4 |
| Length/breadth ratio | 0.69 | 0.70 | 0.62 | 0.74 | 0.57 | 0.56 | 0.58 | 0.47 | 0.6 ± 0.1 |
| Number of segments | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 3 ± 0.9 |

[†]The structure of 2tld could not be analyzed because only Cα coordinates were deposited in the PDB. Results for model structures are after refinement with MULTIDOCK and X-PLOR. Averaged data for all enzyme-inhibitor structures are taken from the Protein-Protein Interaction Server.[43]

and comparable X-PLOR interaction energy. However, its complementarity was improved and, strikingly, both the number of hydrogen bonds and the number of hydrophobic contacts approximately doubled, compared to the original complex, reaching levels typical of actual trypsin-inhibitor complexes (Table I). Compared to the original complex, the number of inhibitor segments contributing to the interaction dropped from three to two, but further examination showed that the third interacting segment of the original complex consisted of just one residue; in fact, actual trypsin-inhibitor complexes show just two interacting segments (Table I).

When the top 10 best scoring GRAMM solutions were visualized, three were particularly interesting because lysine or arginine residues were placed near the S1 specificity pocket of trypsin. In the second best scoring solution, this residue was Arg1; this finding is interesting because of the precedents of thrombin inhibitors, which also position their N-terminal residues in the S1 specificity pocket.[39,40] Solutions 3 and 5 placed Lys11 near the S1 pocket, but the loop in which this residue resided ran approximately perpendicular to its direction in the canonical-style modeled complex. Solutions 3 and 5 were structurally similar (RMS difference of 2.2Å over 47 Cα atoms) but considered different enough to justify separate analysis.

After refinement, all three of these complexes showed characteristics worse than the canonical style model, in particular smaller interface surface area, worse complementarity (higher gap volume indices) and worse interaction energies. In each case, the complementarity was significantly worse than generally observed for enzyme inhibitor complexes (Table I), which, despite the precedent of the Bowman-Birk-type inhibitor complex—1tab—with its high gap volume index, suggested that the canonical style model was more likely to be correct. Superposition of models corresponding to the 12 eigenvector projections onto each of the three solutions yielded many variant complex structures that were refined and analyzed. In only three cases were modest improvements in complementarity obtained, and each time the improvement was associated with reduced interface surface area and much worse interaction energy compared with the respective parent structures (data not shown). In none of the potential complex structures analyzed in Table I do the less confidently determined three residues preceding the C-terminal Cys (see above) contribute significantly to the interface.

## DISCUSSION

Comparisons of the Cp-thionin sequence with others (Fig. 3) show it to be at most 81% identical to known sequences but clearly a member of the thionin superfamily. In the range 57–81% sequence identity with Cp-thionin molecules with five different annotations are found: sulfur-rich protein, proteinase inhibitor, γ-thionin, γ-zeathionin, and defensin (Fig. 3). The length of Cp-thionin (47 residues) and its four disulfide bonds most closely the characteristics of γ-thionins.[5,51] Mass spectrometry data for Cp-thionin showed the formation of homodimers (Fig. 2), apparently the first observation of oligomerization in the thionin family.

Cp-thionin is capable of inhibiting BPT, but not bovine pancreatic chymotrypsin (Fig. 4). As well as confirming the formation of a complex between Cp-thionin and BPT, mass spectrometry data (Fig. 2) showed that the stoichiometry of the interaction is 1:1, as observed for crystallographically for other trypsin inhibitors.[32] Hence, mass spectrometry has the potential to be a useful tool for the direct study of protein-protein interactions.

In principle, the binding of Cp-thionin to BPT could represent specific proteinase-inhibitor interaction or could be a nonspecific association mediated by the formation of disulfide bonds, as previously observed for other thionin/enzyme combinations.[10] There are three main reasons for supposing the latter unlikely. First, the presence in the enzyme assay conditions of 2.5 mM DTT would be suffi-

cient to prevent putative nonspecific inhibition via disulfide bond formation.[10] Second, free cysteine residues, more commonly seen in nonspecifically inhibited proteins,[10] are present in neither BPT nor Cp-thionin. Third, BPT and bovine pancreatic chymotrypsin, which share homology, around 45% sequence identity, and similar patterns of cysteine residues, behave differently toward Cp-thionin: the former strongly inhibited, and the latter entirely unaffected (Fig. 4).

Proteinase inhibitory activity has only been reported once before, for a distant homologue of Cp-thionin (sharing 34% sequence identity[52]). Nevertheless, several other thionin family sequences are currently annotated as proteinase inhibitors or probable proteinase inhibitors, but this seems to be based on false reasoning. The source of this annotation seems to lie with the similarity, 30% identity in a stretch of 47 residues, of a potato thionin sequence to a soya member of the Bowman-Birk proteinase inhibitor family, reported in 1988.[12] It is now known that these figures are not sufficient to infer homology[53,54] and, indeed, Bowman-Birk type inhibitor sequences were totally absent from the results of database searches or threading programs, whether Cp-thionin or the potato thionin[12] was used as input. Furthermore, BLAST 2 sequences[55] failed to give any significant alignment for the two sequences. More conclusively, by using knowledge of the three-dimensional structures of thionins and Bowman-Birk type inhibitors, the published alignment of potato thionin and the soya Bowman-Birk type inhibitor[12] leads to different disulfide bridge patterns for aligned cysteine residues. The significance mistakenly ascribed to that alignment is in accord with known difficulties encountered by some alignment programs when dealing with cysteine-rich sequences.[21] The perpetuation of this annotation error in the sequence databases is in line with general concerns recently expressed about annotation mechanisms,[56] despite precautions taken by some databases,[57] and echoes a similar error recently revealed for a group of amylase inhibitors.[58]

A high-quality model is a prerequisite for docking studies. Therefore, protein structure verification tools were applied to the final Cp-thionin model, showing it to compare very favorably with the templates used for its construction. The qualitative patterns of conformational flexibility resulting from CONCOORD/ED analysis were also checked against the structural variation reported for the template structures determined by NMR.[22] The good agreement found is in accord with the known ability of the CONCOORD method to qualitatively reproduce modes of structural freedom observed either in rigorous molecular dynamics experiments or in ensembles of NMR structures.[28] This observation may also be compared with results showing dynamics similarities in proteins sharing the same fold, even at low levels of sequence identity.[59]

A variety of indicators were used during the assessment of putative Cp-thionin/BPT complex structures in response to the continued uncertainty over the key contributors to the stability of protein-protein interfaces.[46−49] The complementarity of the putative interface was given the highest weighting, but interaction energy, number of H-bonds, and number of hydrophobic contacts were also closely monitored. The chosen complex (Fig. 5) exhibits comparable statistics for all of these quantities to known BPR-inhibitor complexes (Table I) and to enzyme-inhibitor interfaces in general.[43] The additional geometric characteristics, planarity, and length/breadth ratio,[43] of the final complex are also in line with experimentally determined structures (Table I). It is important that the final modeled complex was obtained not with the original Cp-thionin model but with a structure derived from consideration, by using the ED technique, of large concerted motions predicted for the model. Given the known conformational changes of some proteins on formation of complexes.[41] this result has general relevance. The ED-derived structure can be fit to the original model with RMS deviations of 0.88 Å for all Cα atoms and 1.46 Å for all atoms.

According to the final model, Cp-thionin binds to BPT in a canonical fashion with Lys11 of the inhibitor occupying the S1 specificity pocket of the enzyme. Therefore, it was interesting to compare the final model to experimentally determined canonical BPT-inhibitor complexes. Analysis of residues of trypsin contacting inhibitor residues in the final model showed high similarity to contact patterns observed in the different inhibitor classes represented by 2ptc, 1ppe, and 1tab. Overall, seven segments of BPT contribute to inhibitor binding in the final model compared to nine, seven, and seven in the three crystal structures. The number of BPT residue contacts shared by the crystal structures with the final model was 19 of 28, 21 of 30, and 19 of 31 for 2ptc, 1ppe, and 1tab, respectively, but with different patterns of similarities. In Figure 5, the final model is compared with the BPT-BPTI complex of 2ptc. The largest differences in the patterns of inhibitor contacts between these structure lie around residue 96, where the contacts in 2ptc are more extensive than those of the final Cp-thionin model (five residues involved, two H-bonds vs one residue involved, no H-bonds), and around residue 146 where the opposite is found (one residue, no H-bonds in 2ptc; eight residues, two H-bonds in the final model). Despite the similarities in enzyme residues contacting inhibitor, Cp-thionin bears no significant structural resemblance to other trypsin inhibitors, as shown by searches in the FSSP database.[60] Searches of the SCOP database[61] show thionins, squash trypsin inhibitor (1ppe), and Bowman-Birk inhibitors (e.g., 1tab) to be grouped together at the "knottin-fold" level, as characterized by a β-hairpin with two adjacent disulfides. Nevertheless, each reside in a different superfamily and overall structural similarity is absent.

In the final model, Lys11 of the inhibitor is not close enough to Asp189 at the base of the S1 pocket (5.2 Å from atom NZ to atom OD2) to form a direct interaction. This suggested that the interaction might be water-mediated as seen in the case of BPTI.[62] Indeed, a water molecule may be positioned between Lys11 and Asp189 to interact favorably with both (2.7 Å H-bonds) and also with the backbone carbonyl groups of Ser 190 and Gly 219 (3.0 Å H-bonds) (Fig. 5). Some other characteristic H-bonds seen

in canonically binding inhibitors are also seen in the final model (Fig. 6). For example, an H-bond is formed between the backbone N atom of the P3 residue (Gly9 in this case) and Gly216 of BPT. An H-bond is also formed between the backbone O atom of the P2 residue (here Phe10) and the side-chain of Gln192 of BPT. Other characteristic interactions are not present, perhaps because of local errors in the model.

In conclusion, we have shown specific trypsin inhibition associated with Cp-thionin. We predict that the interaction of Cp-thionin with BPT occurs in a canonical style fashion with Lys11, situated on an exposed, extended loop. Crucial to the finding of the highly favorable binding mode of the final model was the consideration of the dynamic properties of the inhibitor molecule. The novel use of CONCOORD followed by ED analysis, both relatively computationally inexpensive, should find wide application. Proteinase inhibitors are of great practical interest in biotechnology (e.g., in the production of transgenic plants resistant to pest attack).[63,64] The discovery and characterization of new inhibitor classes will facilitate these and other studies.

## REFERENCES

1. Broekaert WF, Cammue BPA, De Bolle MCF, Thevissen K, De Samblanx GW, Osborn RW. Antimicrobial peptides from plants. Crit Rev Plant Sci 1997;16:297–323.
2. Bohlmann H, Apel K. Thionins. Ann Rev Plant Physiol Plant Mol Biol 1991;42:227–240.
3. Florack DE, Stiekema WJ. Thionins: properties, possible biological roles and mechanisms of action. Plant Mol Biol 1994;26:25–37.
4. Bohlmann H. The role of thionins in plant protection. Crit Rev Plant Sci 1994;13:1–16.
5. Garcia-Olmedo F, Molina A, Alamillo JM, Rodriguez-Palenzuela P. Plant defense peptides. Biopolymers 1998;47:479–491.
6. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. CATH—a hierarchic classification of protein domain structures. Structure 1997;5:1093–1108.
7. Caaveiro JMM, Molina A, Gonzalez-Manas JM, Rodriguez-Palenzuela P, Garcia-Olmedo F, Goni FM. Differential effects of five types of antipathogenic plant peptides on model membranes. FEBS Lett 1997;410:338–342.
8. Thevissen K, Ghazi A, De Samblanx GW, Brownlee C, Osborn RW, Broekaert WF. Fungal membrane responses induced by plant defensins and thionins. J Biol Chem 1996;271:15018–15025.
9. Bohlmann H, Vignutelli A, Hilpert B, Miersch O, Wasternack C, Apel K. Wounding and chemicals induce expression of the *Arabidopsis thaliana* gene Thi2.1, encoding a fungal defense thionin, via the octadecanoid pathway. FEBS Lett 1998;437:281–286.
10. Pineiro M, Diaz I, Rodriguez-Palenzuela P, Titarenko E, Garcia-Olmedo F. Selective disulphide linkage of plant thionins with other proteins. FEBS Lett 1995;369:239–242.
11. Bloch C Jr, Richardson M. A new family of small (5 kDa) protein inhibitors of insect alpha-amylases from seeds or sorghum (*Sorghum bicolor* (L) Moench) have sequence homologies with wheat gamma-purothionins. FEBS Lett 1991;279:101–104.
12. Stiekema WJ, Heidekamp F, Dirkse WG, van Beckum J, de Haan P, tem Bosch C, Louwerse JD. Molecular cloning and analysis of four potato tuber mRNAs. Plant Mol Biol 1988;11:255–269.
13. Melo FR, Sales MP, Pereira LS, Bloch C Jr, Franco OL, Ary MB. α-Amylase inhibitors from cowpea seeds. Protein Pept Lett 199;6:385–390.
14. Crestfield AM, Moore S, Stein WH. Alkylation and identification of the histidine residues at the active site of ribonuclease. J Biol Chem 1963;238:622–627.
15. Solomon M, Belenghi B, Delledone M, Menachem E, Levine A. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. Plant Cell 1999;11:431–443.
16. de Freitas SM, Ikemoto H, Ventura MM Thermodynamics of the binding of chymotrypsin with the black-eyed pea trypsin and chymotrypsin inhibitor (BTCI). J Protein Chem 1999;18:307–313.
17. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85:2444–2448.
18. Fischer D, Barret C, Bryson K, Elofsson A, Godzik A, Jones D, Karplus KJ, Kelley KA, Maccallum RM, Pawowski K, Rost B, Rychlewski L, Sternberg M. CAFASP-1: critical assessment of fully automated structure prediction methods. Proteins 1999; Suppl 3:209–217.
19. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815.
20. Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. In: Altman RB, Dunker AK, Lauderdale K, Klein TE, editors. Pacific Symp Biocomputing. World Scientific; 2000. pp 119–130.
21. Karplus K, Barrett, C, Hughey, R. Hidden Markov models for detecting remote protein homologies. Bioinformatics 1998;14:846–856.
22. Bruix M, Jimenez MA, Santoro J, Gonzalez C, Colilla FJ, Mendez E, Rico M. Solution structure of gamma 1-H and gamma 1-P thionins from barley and wheat endosperm determined by 1H-NMR: a structural motif common to toxic arthropod proteins. Biochemistry 1993;32:715–724.
23. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 1996;9:1063–1065.
24. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
25. Sanchez R, Sali A. Evaluation of comparative protein structure modeling by MODELLER-3. Proteins 1997;1:50–58.
26. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
27. Laskowski R, MacArthur M, Moss D, Thornton, J. PROCHECK: a program to check stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–290.
28. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen, HJC. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:240–251.
29. de Groot BL, Vriend G, Berendsen HJC. Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism. J Mol Biol 1999;286:1241–1249.
30. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
31. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11:205–217.
32. Bode W, Huber R. Structural basis of the endoproteinase-protein inhibitor interaction. Biochim Biophys Acta 2000;477:241–252.
33. Tyndall JDA, Fairlie DP. Conformational homogeneity in molecular recognition by proteolytic enzymes. J Mol Recog 1999;12:363–370.
34. Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystallographic refinement at 1.9 Angstroms. J Mol Biol 1974;89:73–79.
35. Takeuchi Y, Nonaka T, Nakamura KT, Kojima S, Miura K, Mitsui Y. Crystal structure of an engineered subtilisin inhibitor complexed with bovine trypsin. Proc Natl Acad Sci USA 1992;89:4407–4411.
36. Bode W, Greyling HJ, Huber R, Otlewski J, Wilusz T. The refined 2.0 Å X-ray crystal structure of the complex formed between bovine beta-trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Cucurbita maxima*). Topological similarity of the squash seed inhibitors with the carboxypeptidase A inhibitor from potatoes. FEBS Lett. 1989;242:285–292.

37. Tsunogae Y, Tanaka I, Yamane T, Kikkawa J, Ashida T, Ishikawa C, Watanabe K, Nakamura S, Takahashi K. Structure of the trypsin-binding domain of Bowman-Birk type protease inhibitor and its interaction with trypsin. J Biochem 1986;100:1637–1646.
38. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
39. Rydel TJ, Ravichandran KG, Tulinsky A, Bode W, Huber R, Roitsch C, Fenton JW. The structure of a complex of recombinant hirudin and human alpha-thrombin. Science 1990;249:277–280.
40. van de Locht A, Stubbs MT, Bode W, Friedrich T, Bollschweiler C, Hoffken W, Huber R The ornithodorin-thrombin crystal structure, a key to the TAP enigma? EMBO J 1996;15:6011–6017.
41. Jackson RM, Gabb HA, Sternberg MJE. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276:265–285.
42. Brünger AT. Available on the WWW at http://atp.slac.stanford.edu; 1999.
43. Jones S, Thornton JM. Principles of protein-protein interactions derived from structural studies. Proc Natl Acad Sci USA 1996;93:13–20.
44. McDonald IK, Thornton JM Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
45. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
46. Vakser IA, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
47. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
48. Sheinermann FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol 2000;10:153–159.
49. Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein-protein docking. Proteins 2000;40:525–537.
50. Norel R, Petrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. Proteins 1999;15:307–317.
51. Bloch C Jr, Patel SU, Baud F, Zvelebil MJJM, Carr MD, Sadler PJ, Thornton JM. H-NMR structure of an antifungal γ-thionin protein Siα1: similarity to scorpion toxins. Proteins 1998;32:334–349.
52. Wijaya R, Neumann GM, Condron R, Hughes AB, Polya GM. Defense proteins from seed of *Cassia fistula* include a lipid transfer protein homologue and a protease inhibitory plant defensin. Plant Sci 2000;159:243–255.
53. Sander C, Schneider R. The HSSP data base of protein structure-sequence alignments. Nucleic Acids Res 1993;21:3105–3109.
54. Rost B. Twilight zone of protein sequence alignments. Protein Eng 1999;12:85–94.
55. Tatusova TA, Madden, TL BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 1999;174:247–250.
56. Lewis S, Ashburner M, Reese MG. Annotating eukaryoted genomes. Curr Opin Struct Biol 2000;10:349–354.
57. Junker VL, Apweiler R, Bairoch A. Representation of functional information in the SWISS-PROT data bank. Bioinformatics 1999;15:1066–1067.
58. Iulek J, Franco OL, Silva M, Slivinski CT, Bloch C Jr, Rigden DJ, Grossi de Sá MF. Purification, biochemical characterization and partial primary structure of a new alpha-amylase inhibitor from *Secale cereale* (Rye). Int J Biochem Cell Biol 2000;32:1195–1204.
59. Mello LV, van Aalten DM, Findlay JB. Dynamic properties of the guanine nucleotide binding protein alpha subunit and comparison of its triphosphate hydrolase domain with that of ras p21. Biochemistry 1998;37:3137–3142.
60. Holm L, Sander C Mapping the protein universe. Science 1996;273:595–603.
61. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
62. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr B 1983;39:480–486.
63. Gatehouse AMR, Gatehouse JA. Identifying proteins with insecticidal activity: use of encoding genes to produce insect-resistant transgenic crops. Pest Sci 1998;52:165–175.
64. Schuler TJ, Poppy GM, Kerry BR, Denholm I. Insect-resistant transgenic plants. Trends Biotech 1998;16:168–175.
65. Choi Y, Ahn JH, Choi YD, Lee JS. Tissue-specific and developmental regulation of a gene encoding a low molecular weight sulfur-rich protein in soybean seeds. Mol Gen Genet 1995;246:266–268.
66. Urdangarin MC, de la Canal L. A defensin gene expressed in sunflower inflorescence. Plant Physiol Biochem 2000;38:253–258.
67. Castro MS, Fontes W, Morhy L, Bloch C Jr. Complete amino acid sequences of two gamma-thionins from maize (*Zea mays* L.) seeds. Protein Pept Lett 1996;3:267–274.