Journal of Protein Chemistry, Vol. 15, No. 6, 1996

Amino Acid Sequence of the *Phaseolus vulgaris* var. "Fogo na Serra" Inhibitor and Interactive Surface Modeling for the Enzyme–Inhibitor Complex

Patricia G. B. de Carvalho,[1] Carlos Bloch, Jr.,[1,2] Lauro Morhy,[1] Maria C. M. da Silva,[3] Luciane V. de Mello,[3] and Goran Neshich[3]

Received May 22, 1996

A trypsin and chymotrypsin inhibitor from seeds of *Phaseolus vulgaris* var. "Fogo na Serra" (PFSI) was purified and its complete amino acid sequence was determined using Edman degradation methods. The inhibitor was found to belong to the Bowman–Birk family of enzymatic inhibitors; it has 82 amino acid residues and a 8.985-kDa molecular mass. The PFSI/α-chymotrypsin binary complex has been modeled using the Turkey ovomucoid inhibitor third domain (OMTKY3) bound to α-chymotrypsin [Fujinaga *et al.* (1987), *J. Mol. Biol.*, **195**, 397–418. template. The model allowed identification of the binding surface.

KEY WORDS: Chymotrypsin inhibitor; enzyme–inhibitor complex; amino acid sequence; molecular modeling.

1. INTRODUCTION

Bowman–Birk inhibitors are small globular proteins constituted of a single polypeptide chain about 70–90 amino acid residues long. Most of these inhibitors have 14 cysteine residues involved in seven disulfide bonds (Richardson, 1991). Sequence and three-dimensional structural homology studies revealed that these proteins have two independent inhibitor sites, hence the name "double-headed" inhibitors. Each reactive site is distinguishable by its specificity to either trypsin or chymotrypsin. In the majority of the Bowman–Birk family, the N-terminal portion of the polypeptide chain is involved in trypsin inhibition, while the C-terminus is responsible for chymotrypsin inactivation (Werner and Wemmer, 1992*a*,*b*). Bowman–Birk inhibitors are among the most-studied families of proteins due to their convenient structural and physiological features. The highly stable tertiary structures and the ability to inhibit two larger molecules such as trypsin and chymotrypsin at the same time make these proteins an interesting target for protein chemists, crystallographers, NMR spectroscopists, and molecular modelers. A number of amino acid sequences of these inhibitors have been reported (Richardson, 1991), and five entries of molecular coordinates are now available on the Brookhaven Protein Data Bank (August 1995). Inhibitors are also of interest because of their possible involvement in the protection mechanisms of seeds against insects and microorganisms (Ryan, 1973) and also because of the importance of these inhibitors from pharmacological and nutritional points of view.

[1] Centro Brasileiro de Seqüenciamento de Proteínas, Departamento de Biologia Celular, Universidade de Brasília, 70970-910, Brasília-DF, Brazil.
[2] Departamento de Química, Universidade de Brasília, 70970-910, Brasíla-DF, Brazil.
[3] EMBRAPA/CENARGEN, National Center for Genetic Resources and Biotechnology, 70770-900, Brasilia-DF, Brazil.
[4] To whom correspondence should be addressed; e-mail: cblochjr@guarany.cpd.unb.br.

0277-8033/96/0800-0591$09.50/0

2. MATERIALS AND METHODS

2.1. Materials

Seeds of *Phaseolus vulgaris* var. "Fogo na Serra" were harvested in Brasilia, D. F., and stored at 4°C until utilization. Other reagents and materials used were as follows: low-molecular-weight markers from Pharmacia; trypsin, TPCK-trypsin, α-chymotrypsin, *Staphylococcus aureus* protease strain V8, pepsin A, acrylamide, N-N-methylenebisacrylamide, TEMED (N-N-N′-N′-tetramethylethylenediamine), TFA (trifluoracetic acid), iodoacetic, acid and 4-vinylpyridine from Sigma Chemicals Company; PITC (phenylisothiocyanate) and DABITC (4-N,N′-dimethylamineazobenzene-4′-isothiocyanate) from Pierce; and acetonitrile HPLC grade, from Merck. All other reagents were of analytical grade.

2.2. Purification and Inhibitory Activity Assays

Trypsin and chymotrypsin inhibitor of *Phaseolus vulgaris* var. "Fogo na Serra" (PFSI) was isolated from seeds and purified according to Morhy and Ventura (1987), with an additional step of reverse-phase HPLC on a C_{18} Vydac column using a gradient (0–30%) of acetonitrile in aqueous TFA 0.1%. Trypsin and chymotrypsin inhibitory activities were assayed on the highly purified sample of PFSI as described by Erlanger *et al.* (1961, 1966).

2.3. Reduction and S-Alkylation

Carboxymethylation was carried out according to Crestfield *et al.* (1963), yielding CM-PFSI. S-Vinylpiridylethylation was done as described by Allen (1983), yielding VP-PFSI. From 0.2 to 1 mg of PFSI was used each time. Excess reagents and buffer were removed using Sep-Pak C_{18} cartridges (Waters-Millipore).

2.4. Enzymatic Digestions, Peptide Purification, and Protein Sequencing

Enzymatic PFSI peptides were produced by controled hydrolysis using TPCK-trypsin (2% enzyme/substrate ratio for 3 hr at 37°C), pepsin (1% enzyme/substrate ratio for 4 hr at 37°C), and *S. aureus* V8 protease (2% enzyme/substrate ratio for 24 hr at 37°C) (Aitken *et al.*, 1989). Peptides obtained from enzymatic hydrolysis were purified by reverse-phase HPLC on an ABI PTC-222 C_{18} analytical column (5 μm particle size, 220 × 2.1 mm), using a gradient (0–30%) of acetonitrile in aqueous TFA 0.1%.

Protein sequencing was performed using both manual (Yarwood, 1989) and automatic (Applied Biosystems 477A Protein Sequencer) methods on both intact and CM-PFSI and on the peptides produced by enzymatic digestion of alkylated PFSI.

2.5. Sequence Comparisons in Data Banks

The PFSI sequence was compared to the nucleotide and amino acid sequence data banks. Positions having identical residues and conservative substitutions were identified. The sequences selected were submited to automatic multiple alignment, which was performed by using the Genetics Computer Group (GCG) Sequence Analysis Software Package and its PILEUP algorithm, giving highest weight to the conserved cysteines and to the residues within the reactive sites of all members of the Bowman–Birk family.

2.6. PFSI Model Building

The three-dimensional structure of PFSI was modeled based on the coordinates of the NMR structure for the trypsin/chymotrypsin Bowman–Birk Inhibitor from soybean (BBI-I) (Werner and Wemmer, 1992) and on the 1PI2 X-ray crystal structure (Chen *et al.*, 1992). Molecular modeling and molecular mechanics were carried out with the PFSI (82 residues) sequence using Insight II with Discover (Biosym Technologies, San Diego, CA. 1991) on Silicon Graphics 4D70GT, INDIGO ELAN, and SGI ONYX VTX workstations.

Model building included such computational techniques as position-specific rotamer search for the residues different from the parent protein model, quality analysis of the resulting model based on the dihedral angle values, analysis of hydrogen bond energy values, and packing quality analysis of the resulting model and of the initial parent protein. For the complex building, hard-body docking with parallel energy minimization calculation was used.

The final model had all the amino acid residues which differed between PFSI/BBI-I and PFSI/PI2

exchanged by a semiautomatic procedure available in the Insight II "Biopolymers" module. The bad steric contacts were removed and the side-chain angles of the exchanged residues were analyzed. The χ_1 and χ_2 were adjusted in order to obtain the best position, based on stereochemical parameters derived from high-resolution protein structure (McGregor *et al.*, 1987; Ponder and Richard, 1987; Schrauber *et al.*, 1993), according to the Verify algorithm described by the What If program (Vriend, 1990). The quality of the models was checked by the Quality module of the What If package (Vriend and Sander, 1993). Energy minimizations were carried out using Discover (Insight II-Biosym).

2.7. Modeling of PFSI/α-Chymotrypsin Complex

The binding surface of the PFSI in contact with α-chymotrypsin was modeled by homology to the complex of Turkey ovomucoid inhibitor third domain (OMTKY3) bound to chymotrypsin (Fuginaga *et al.*, 1987). The atomic structure of this complex, solved by X-ray crystallography, was taken from the 1cho.pdb of the Brookhaven PDB (Bernstein *et al.*, 1977). Docking the inhibitor into the respective enzyme required the proper orientation of the PFSI model with respect to bound α-chymotrypsin, achieved by superimposing the PFSI loop, residues Thr-42 to Ile-45, and the OMTKY3 loop, residues Thr-17 to Arg-21. The side chains of the binding β-hairpin of PFSI were modeled in order to correspond to the same side-chain orientation of residues belonging to the binding loop of OMTKY3. The χ_1 and χ_2 angles of the PFSI loop side chains were corrected according to OMTKY3 values, in order to avoid undesirable contact bumps. Once the initial model was prepared, further energy minimization of the complex was done using Discover (Insight II). For the minimization procedure, a subset of amino acid residues, within 5 Å of the molecular complex interacting surface, was formed on both molecules. This subset was allowed to accommodate coordinates for equilibrium energy. All other atoms were fixed in space, allowing significant reduction in the CPU time required for completion of the computation of the minimum energy. In addition, constraints were used on the position of atoms of the amino acid residues at the molecular complex interface judged as essential for the proper positioning of the inhibitor in the protease active pocket. These distances were mainly fixed at positions which allowed hydrogen bond formation between the PFSI Phe-43 and the catalytic triad, active pocket, and nonspecific binding amino acids of α-chymotrypsin. Energy minimization was carried out until the maximum derivative of the energy with respect to the atomic positions was less than 0.001 kcal/mol/Å. A short molecular dynamics simulation was run on 500 K, and the conformation with minimum energy was further allowed to equilibrate using the conjugate gradient algorithm, converging down to 0.0001 kcal/mol/Å. Then the docking procedure of proteins was adjusted by moving the PFSI three-dimensional model in a random direction around the interface. The best position was chosen by monitoring the decrease of the total energy values. Additional energy minimization was done to remove any resulting steric contacts.

2.8. Identification of the Enzyme–Inhibitor Interaction Surface

The Connolly surface (Connolly, 1983) was generated by Insight II using a water probe radius of 1.4 Å.

3. RESULTS AND DISCUSSION

3.1. Protein Sequencing

Manual and automatic sequencing (Fig. 1) of native, alkylated, and hydrolyzed samples of PFSI revealed an 82-amino-acid single-chain polypeptide with a molecular weight of 8.985 kDa.

3.2. Sequence Comparisons

The alignment of the amino acid sequences of various inhibitors from the Bowman–Birk family is shown in Fig. 2. Each sequence shows approximately 55% identity to PFSI. The intercysteine loop pattern is conserved throughout this family. The reactive-site residues for trypsin and chymotrypsin are marked by an asterisk and they are in identical positions in the PFSI, PI2, and BB-I chains. The sequence homology of PFSI, PI2, and BBI-I also

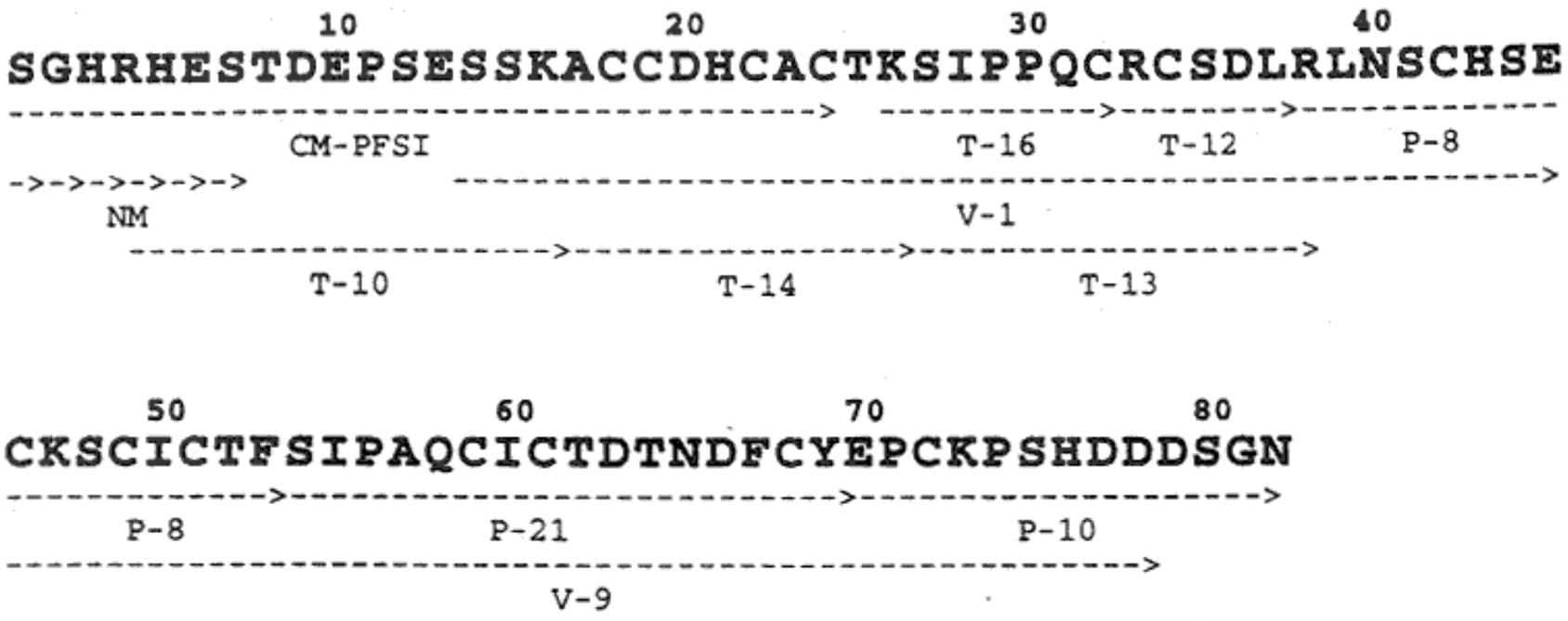

Fig. 1. The amino acid sequence of the trypsin and chymotrypsin inhibitor of *Phaseolus vulgaris* var. "Fogo na Serra" (PFSI). Arrows indicate peptides obtained from digestions of the alkylated PFSI with trypsin (T), pepsin (P), and the protease from *S. aureus* V8 (V). CM-PFSI indicates S-carboxymethylated inhibitor, which was automatically sequenced without enzymatic digestion. NM indicates the residues of the inhibitor in the native form which were sequenced manually.

shows preservation of the position of polar as well as nonpolar residues.

3.3. Proposed Model for PFSI

Figure 3 shows the energy-minimized 3D structure of the PFSI molecule with color coding for the two respective active sites. They are depicted as CPK surfaces: Phe-43 chymotryptic inhibitory site and Lys-16 tryptic inhibitory site. The residue Phe-43 has substituted a Leu present in the BBI-I molecule. Both inhibitory sites are located on the tips of the corresponding β-hairpin loop structures.

The final PFSI model presents some unusual structural features: aromatic residues appear as

	1		*		50
1.	SVHHQDSSDE	PSESSHPCCD	LCLCTKSIPP	QCQCADIRLD	SCHSACKSCM
2.	....	.DEYSKPCCD	LCMCTRSMPP	QCSCED.RIN	SCHSDCKSCM
3.	....	DDESSKPCCD	QCACTKSNPP	QCRCSDMRLN	SCHSACKSCI
4.	SGHHEHSTDZ	PSZSSKPCCB	HCACTKSIPP	QCRCTDLRLD	SCHSACKSCI
5.	..DHHHSTDE	PSESSKPCCD	ECACTKSIPP	QCRCTDVRLN	SCHSACSSCV
6.	SGHHDETTDE	PSESSKPCCD	QCSCTKSMPP	KCRCSDIRLN	SCHSACKSCA
7.	SGHHZBSTBZ	ASZSSKPCCR	ZCACTKSIPP	ZCRCSZVRLN	SCHSACKSCA
8.	...HEHSSDE	SSESSKPCCD	LCTCTKSIPP	QCHCNDMRLN	SCHSACKSCI
9.	SGHRHESTDE	PSESSKACCD	HCACTKSIPP	QCRCSDLRLN	SCHSECKSCI

	51 *			83
1.	CTRSMPGQCR	CLDTHDFCHK	PCKSRDKD..	...
2.	CTRSQPGQCR	CLDTNDFCYK	PCKSR.....	...
3.	CALSYPAQCF	CVDITDFCYE	PCKPSEDDKE	N..
4.	CTLSIPAQCV	CBBIBDFCYE	PCKSSHSDDD	NNN
5.	CTFSIPAQCV	CVDMKDFCYA	PCKSSHDD..	...
6.	CTYSIPAKCF	CTDINDFCYE	PCKSSRDDDW	DN.
7.	CTFSIPAZCF	CGBIBBFCYK	PCKSSHSBBB	BWN
8.	CALSEPAQCF	CVDTTDFCYK	SCHNNAEKD.	...
9.	CTFSIPAQCI	CTDTNDFCYE	PCKPSHDDDS	GN.

Fig. 2. Alignment of Bowman–Birk type protease inhibitors. These sequences were aligned based on the position of the half-cysteine residues. Asterisks indicate the interaction sites for trypsin (in BBI-I it is a Lys residue) and for chymotrypsin (in BBI-I it is a Leu residue). Aligned sequences from Bowman–Birk family: (1) IBB2_PHAAN (*Phaseolus angularis*), (2) 1PI2 (*Glycine max*), (3) BBI-I (*Glycine max*), (4) IBB_PHALU (*Phaseolus lunatus*), (5) IBB3_DOLAX (*Dolichos axillaris*), (6) IBB1_PHAAN (*Phaseolus angularis*), (7) IBB_VIGUN (*Vigna unguiculata*), (8) IBB4_DOLAX (*Dolichos axillaris*), (9) PFSI.

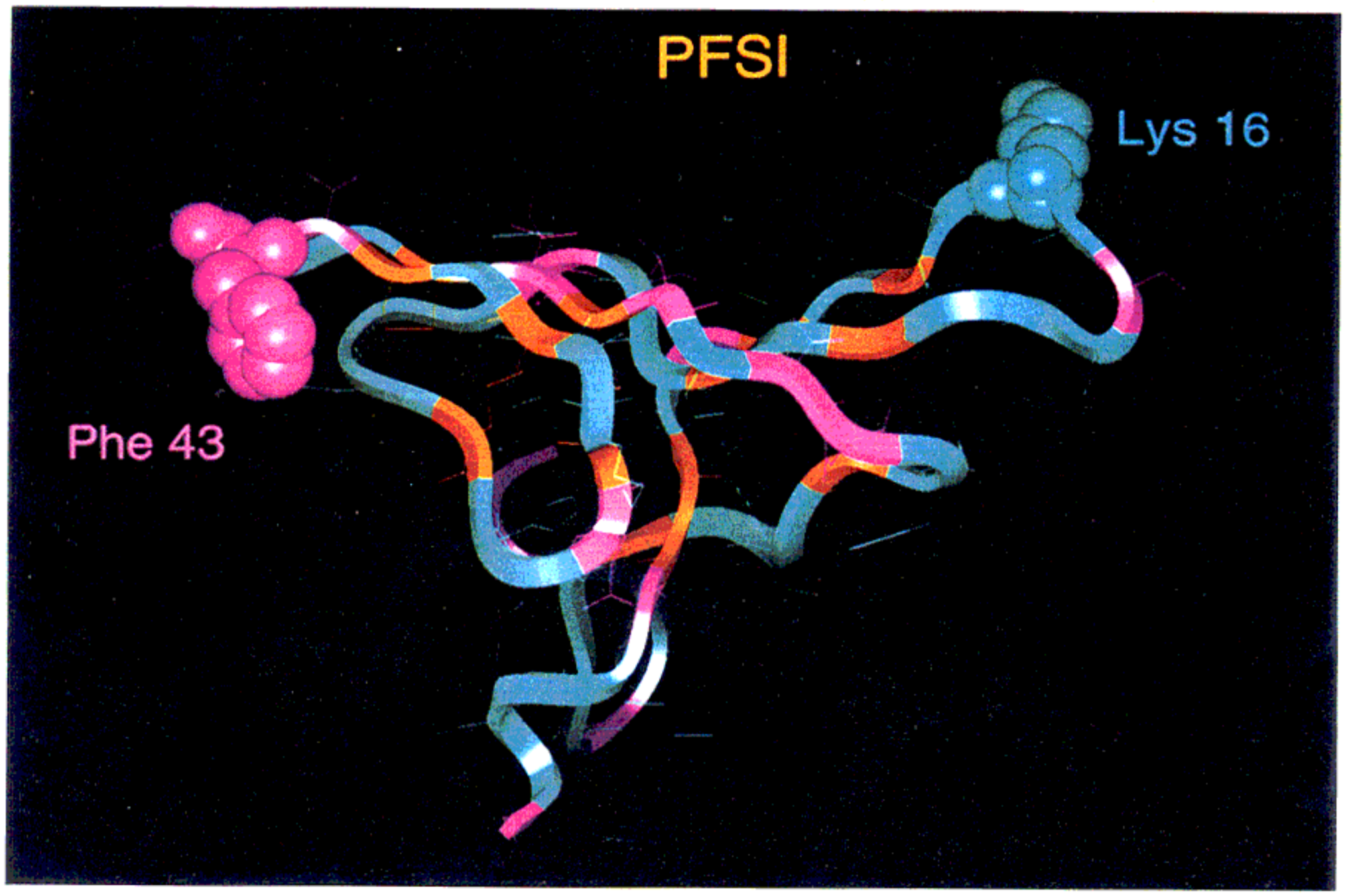

Fig. 3. Energy-minimized structure of the 3D PFSI model with color coding for the substituted residues (pink), residues originally present in the BBI-I molecule (turquoise), Cys residues (orange), and two active sites (cpk, chymotryptic inhibitory site, Phe-43; and tryptic inhibitory site, Lys-16). Both inhibitory sites are located on the tips of the corresponding β-hairpin structures.

solvent-exposed, and some hydrophilic residues are hidden inside the core. The quality of the models was checked by the Quality module of the What If package (Vriend and Sander, 1993), which can evaluate the dihedral peptide angle value distribution (Ramachandran plot), chirality, peptide bond planarity, side-chain conformation, and nonbonded contacts. Our final model does not show departure from the expected statistical values for any of these parameters; the model propagated the same "quality" characteristics as the parent structure. However, the techniques commonly used to verify "normal" proteins could not be used to check the quality of the model because of the peculiar exposed aromatic residues and hidden hydrophilic residues. A probable answer to the "abnormal" aromatic residue packing is dimerization of the peptide chain in such a way that exposed aromatics interact favorably between the multiple peptide units.

The What If Quality test gave us the value of −2.7 for the contact quality-index for both parent protein structures and for the model. This value is considered as an indication of poor quality (Vriend and Sander, 1993). Further consideration of the data presented by Werner and Wemmer (1992*a*,*b*) indicated the existence of regions in the protein whose structure might be considered of different reliability, based on the overall number of NOE constraints obtained for BBI-I. The regions containing the inhibitory sites of BBI-I were found to be well-defined β-hairpins and considered more accurate. These parts of the molecule proved to be more reliable than the rest, due to a higher number of NOEs per residue and to the slow exchanged amide constraints observed.

3.4. PFSI/α-Chymotrypsin Complex

An important goal of the present study was to provide a model for the interaction between α-chymotrypsin and the PFSI. Figure 4 shows the model of the catalytic triad; the active-site pocket amino acids of α-chymotrypsin are depicted in

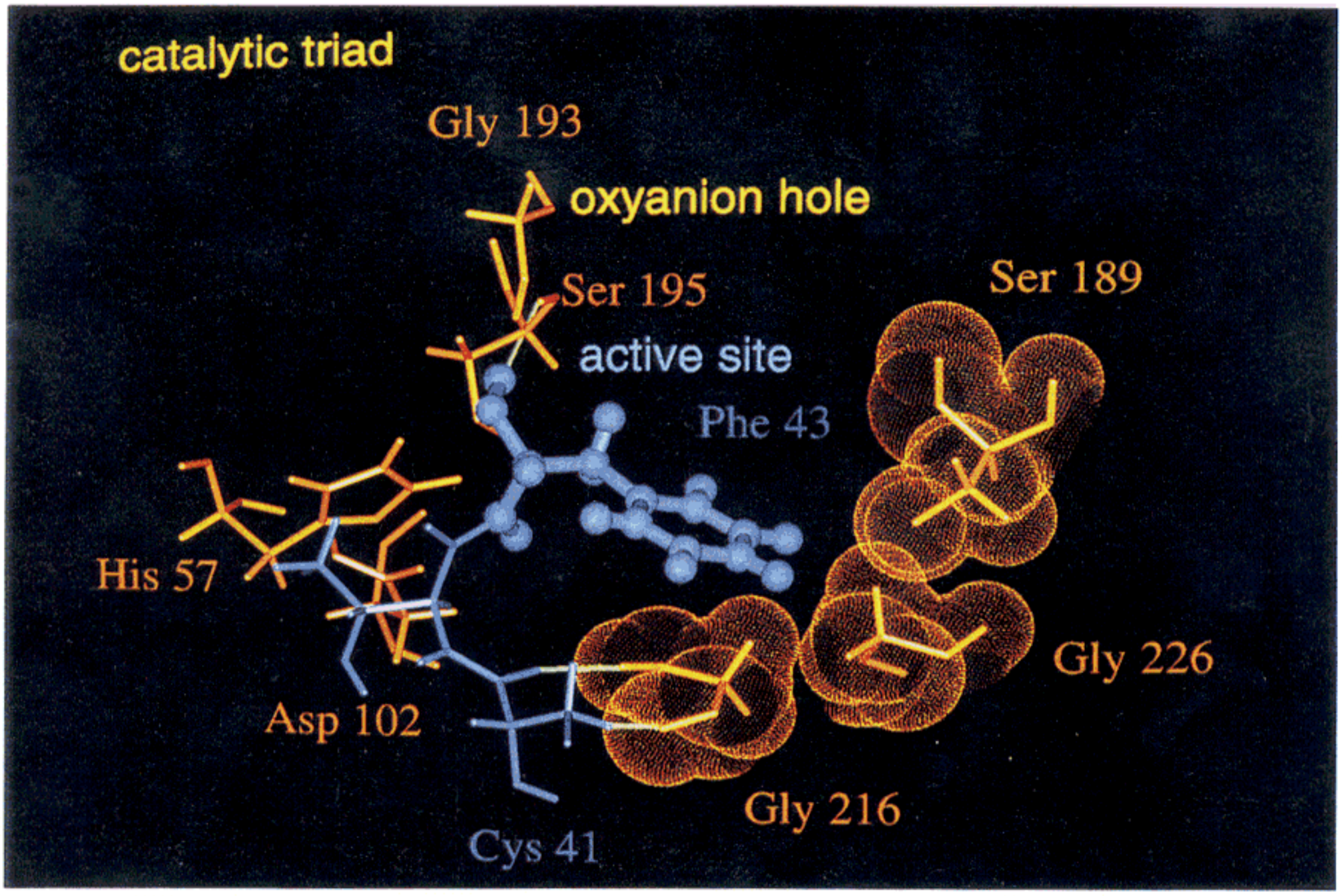

Fig. 4. PFSI/α-chymotrypsin complex: view looking into the binding site. Active site of chymotrypsin (yellow) with a bound inhibitor (blue). The configuration depicted illustrates positioning of the Phe-43 of PFSI and how it binds in relation to the catalytic triad, the substrate specificity pocket, the oxyanion hole, and the nonspecific substrate binding region. Hydrogen bonds between PFSI and α-chymotrypsin are stripped. The active-site pocket amino acids of α-chymotrypsin are shown in ball-and-stick configuration.

ball-and-stick representation. The chymotrypsin-active site on the PFSI is also depicted (Phe-43 in blue). This model was obtained by superimposing the PFSI chymotrypsin-binding loop (residues Thr-42 to Ile-45) and the OMTKY3 binding loop. An rms of 0.84 Å was found before energy minimization. The χ_1 and χ_2 angles of the PFSI loop side-chain residues were corrected to the same values of the OMTKY3 structure in order to avoid undesirable contact bumps. Once the initial model was obtained, further energy minimization of the complex was carried out using Discover module algorithms (from Insight II). The conformation of the resulting enzyme–inhibitor complex has a satisfactory steric arrangement for hydrogen bonding of interacting amino acids, as shown in Fig. 4. Furthermore, superposition of OMTKY3 and PFSI loops (residues Thr-42 to Ile-45) indicates a remarkable similarity to the backbone conformation of the OMTKY3, since rmsd of only 0.48 Å has been achieved for all homologous C_α atoms. At the same time, the rms value for chymotrypsin C_α atoms after the described procedure was only 0.2 Å.

3.5. The Solvent-Accessible Surface of the Active Site

The solvent-accessible surface (Connoily, 1983) at the interface of the two interacting molecules is presented in Fig. 5. The surface of the protease is represented in yellow (produced by taking into account all residues within 5 Å of the PFSI surface). Specific pocket space complementarity is well seen in the case of Phe-43 (cpk). Green surface area identifies the PFSI Connolly surface. The P_1 residue (Phe-43) (Schechter and Berger, 1967) is solvent-exposed and protrudes away from the β-hairpin loop, playing an important role in the specificity. The conformation of the P3–Pe′ reactive loop is consistent with that of a proteinase docking conformation as proposed in several crystal

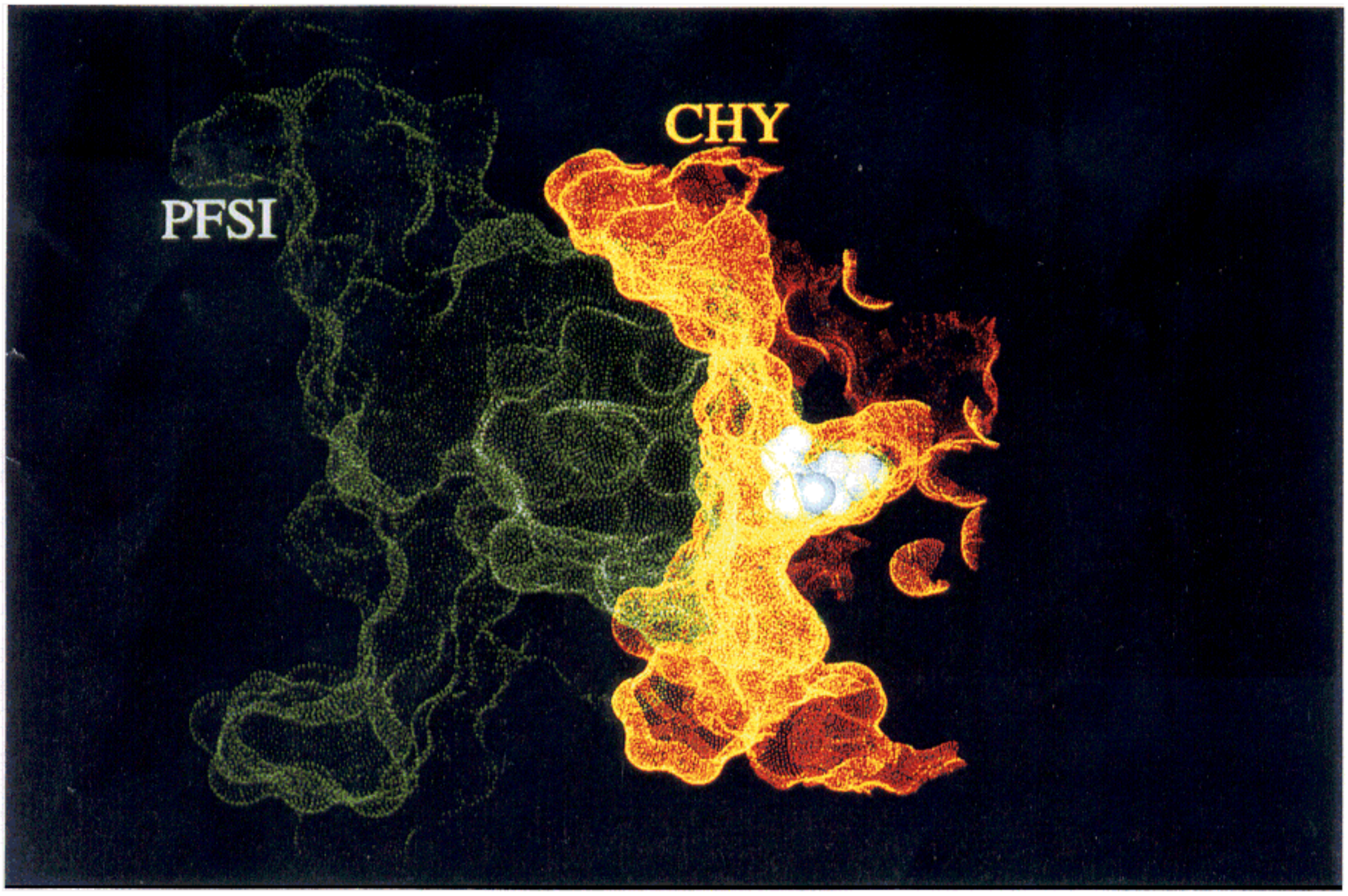

Fig. 5. PFSI/α-chymotrypsin binary complex interactive Connolly surface. Profile view of the Phe-43 pocket. Inhibitor surface is in green and protease surface is in yellow. The Connolly surface was generated by all amino acid residues within 5 Å of the binding interface.

structures of enzyme–inhibitor complexes (Shoichet and Kuntz, 1991; Werner and Wemmer, 1992; Katz and Christianson, 1993). The primary recognition event occurs between the P_1 residue of the inhibitor and the S_1 (Schechter and Berger, 1967) pocket of a target proteinase. It has been demonstrated that the structure of the resulting complex may be consistent with a "locked conformation." A scheme illustrating this model is shown in Fig. 5. Several residues participate in one or more interactions with side chains projecting out from the β-hairpin loop. The convex surface of the β-hairpin loop is largely hydrophobic and mediates most of the contacts. In the specific pocket there are hydrogen bonds, which can be formed directly with residue side chains or peptide backbones that surround the oxyanion hole, involving directly the Ser-195 residue. Details of the hydrogen bonding contacts are represented by the green lines in Fig. 4. Our modeling studies indicated that the model complex could be formed between the PFSI and the target enzyme.

4. CONCLUSIONS

A newly described sequence of PFSI has been reported here and classified under the family of the Bowman–Birk inhibitors. We found that the BBI-I model, which was taken as template for modeling the structure of the PFSI in this work, shows poor packing quality by directional atomic contact analysis. At first sight, we could conclude that as a result homology modeling would have the propagation of such conformational characteristics. However, we believe that exposed aromatic residues are involved in hydrophobic interactions within the multimer peptide formed under physiological conditions.

In addition, the hairpin loops containing the active sites were considered the most best-defined region in BBI-I. We have proposed a model for PFSI, in the absence of an experimentally determined 3D structure, as a contribution to the understanding of enzyme–inhibitor interactions.

Figure 4 shows that the model obtained for the α-chymotrypsin/PFSI complex accounts for some functional aspects of both proteins, including a

remarkable structural fitting at the binding interface seen in Fig. 5. By identification of the interaction surface of the complex, we open possibilities for the experimental design of a specific serine-protease inhibitor suitable for increasing plant defense potential. The knowledge of interacting molecules could significantly improve our understanding of thermodynamic parameters involved in inhibitor binding to the protease. The presence of the peculiar negative potential surface at the very tip of the PFSI active-site hairpin, while PFSI as a whole is a generator of a positive potential field, might explain how two molecules with net positive charges bind to each other.

ACKNOWLEDGMENTS

The authors wish to dedicate this work to the memory of Professor L. F. G. Labouriau. We express our gratitude to CNPq, CAPES, FAPDF, and PADCT for financial support.

REFERENCES

Aitken, A., Geisow, M. J., Findlay, J. B. C., Hoimes, C., and Yarwood, A. (1989). In *Protein Sequencing—A Practical Approach* (Findlay, J. B. C., and Geisow, M. J., eds.), Oxford University Press, Oxford, pp. 43–67.

Allen, G. (1983). In *Sequencing of Proteins and Peptides—Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S., and Burdon, R. H., eds.), Elsevier Science Publishers B. V., Oxford, pp. 28–31.

Bernstein, F. C., *et al.* (1977). *J. Mol. Biol.* **112,** 535–542.

Chen, P., Rose, J., Love, R., Wei, C. H., and Wang, B. C. (1992). *J. Biol. Chem.* **267,** 1990–1994.

Connolly, M. L. (1983). *Science* **221,** 709–713.

Crestfield, A. M., Moore, S., and Stein, W. H. (1963). *J. Biol. Chem.* **238,** 622–627.

Erlanger, B. F., Kokowsky, N., and Cohen, E. (1961). *Arch. Biochem. Biophys.* **95,** 271–278.

Erlanger, B. F., Edel, F., and Cooper, A. G. (1966). *Arch. Biochem. Biophys.* **115,** 206–210.

Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, Jr., and James, N. G. (1987). *J. Mol. Biol.* **195,** 397–418.

Katz, D. S., and Christianson, D. W. (1993). *Protein Eng.* **6,** 701–709.

McGregor, M. J., Islam, S. A., and Sternberg, M. J. E. (1987). *J. Mol. Biol.* **198,** 295–310.

Morhy, L., and Ventura, M. M. (1987). *An. Acad. Brasil. Ci.* **59,** 71–81.

Ponder, J. W., and Richard, F. M. (1987). *J. Mol. Biol.* **193,** 775.

Richardson, M. (1991). In *Methods in Plant Biochemistry*, Vol. 5 (Rodgers, L., ed.), Academic Press, New York, pp. 259–305.

Ryan, . (1973).

Schechter, I., and Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27,** 157.

Schrauber, H., Eisenhaber, F., and Argos, P. (1993). *J. Mol. Biol.* **230,** 592–612.

Shoichet, B. K. and Kuntz, I. D. (1991). *J. Mol. Biol.* **221,** 327–346.

Vriend, G. (1990). *J. Mol. Graph.* **8,** 52–

Vriend, G., and Sander, C. (1993). *J. Appl. Cryst.* **26,** 47–60.

Yarwood, A. (1989). In *Protein Sequencing—A Practical Approach* (Findlay, J. B. C., and Geisow, M. J., eds.), Oxford University Press, Oxford, pp. 119–145.

Werner, M. H., and Wemmer, D. E. (1991). *Biochemistry* **30,** 3356–3364.

Werner, M. H., and Wemmer, D. E. (1992*a*). *Biochemistry* **31,** 999–1010.

Werner, M. H., and Wemmer, D. E. (1992*b*). *J. Mol. Biol.* **225,** 873–889.