Anticarcinogenic Bowman Birk Inhibitor Isolated from Snail Medic Seeds (*Medicago scutellata*): Solution Structure and Analysis of Self-Association Behavior†

M. Catalano,‡ L. Ragona,*,‡ H. Molinari,‡,§ A. Tava,|| and L. Zetta‡

Istituto per lo Studio delle Macromolecole, Laboratorio NMR, CNR, Via Ampère 56, 20131, Milano, Italy, Dipartimento Scientifico e Tecnologico, Università degli Studi di Verona, Strada Le Grazie, 37134 Verona, Italy, and Istituto Sperimentale Colture Foraggere, Viale Piacenza 29, 26900 Lodi, Italy

Received September 6, 2002; Revised Manuscript Received December 16, 2002

ABSTRACT: The high-resolution three-dimensional structure of a Bowman Birk inhibitor, purified from snail medic seeds (*Medicago scutellata*) (MSTI), has been determined in solution by ^{1}H NMR spectroscopy at pH 5.6 and 27 °C. The structure of MSTI comprises two distinct symmetric domains each composed of a three-stranded β-sheet containing a VIb type loop, where the active sites are located. A characteristic geometry of three aromatic residues confers stability to this protein, and we observe that this feature is conserved in all the Bowman Birk inhibitors of known structure. The two active domains exhibit different conformational features: the second domain displays higher flexibility and hydrophobicity with respect to the first one, and these properties have been correlated to a lower trypsin inhibitory specificity, in agreement with titration studies that have shown a stoichiometric ratio MSTI:trypsin of 1:1.5. NMR analysis indicated that MSTI undergoes self-association at concentrations higher than 2 mM, and the residues involved in this mechanism are localized at opposite faces of the molecule, having the highest positive and negative potential, respectively, thus indicating that electrostatic intermolecular interactions are the driving forces for MSTI association. Most of the residues affected by self-association are highly conserved in BBIs from different seeds, suggesting a functional relevance for these charged superficial patches, possibly involved in the interaction with other enzymes or macromolecules, thus triggering anti-carcinogenic activity.

The Bowman-Birk inhibitors (BBI)[1] are small serine protease inhibitors found in seeds of legumes and in many other plants (*1*). Typically, their molecular masses range between 6 and 9 KDa, and they contain seven disulfide bonds with a prominent role in the stabilization of their active configurations. Most BBIs exist in various isoforms (*2*). All members of the BBI family have two tandem homology regions on the same polypeptide chain, each comprising a consensus motif of three β-strands with a kinetically independent reactive site on the outermost loop. It has been shown that splitting the homology regions by partial peptic digestion yields two active fragments (*3*) and hence BBIs have been called "double-headed" inhibitors for their capability of inhibiting simultaneously and independently two different serine proteases. Most BBIs inhibit trypsin at the first reactive site (N-terminal) and chymotrypsin at the second reactive site (C-terminal). The amino acid residues around the reactive site are usually designed as P3, P2, P1, P1′, P2′, and P3′. P1 residue is involved in the so-called "primary" contact region and confers inhibitory specificity: arginine and lysine for trypsin, leucine, phenylalanine, and tyrosine for chymotrypsin, and alanine for elastase (*1*). However, in contrast to primary specificity, trypsinolysis of soybean BBI (BBI-I) occurs not only in the predicted trypsin inhibitory domain but also in the predicted chymotryptic inhibitory domain (*4*).

Despite extensive studies of BBIs, only a few three-dimensional structures have been solved by X-ray or by NMR. They include structures of tracy soybean (*5*), peanut (*6*), pea seeds (*2*), and soybean inhibitor (*7*, *8*). The X-ray structures of trypsin complexes of inhibitors from azuki bean (*9*), mung bean (*10*), and soybean (*11*) have also been reported. Recently a novel 14 amino acid residue cyclic peptide, cyclo (RCTKSIPPICFPDG), isolated from sunflower seeds, was found to be a potent inhibitor of trypsin (K_d = 100 pM), and the X-ray structure of its complex with bovine trypsin showed both sequence and conformational similarity with the trypsin-reactive loop of the Bowman-Birk family of serine protease inhibitors (*12*). This inhibitor, however, is unique in being monofunctional, cyclic, and far shorter than inhibitors belonging to this family. It has been suggested that the high potency of this peptide is likely to arise from the considerable structural rigidity achieved through its cyclic nature which is further stabilized by a single internal disulfide bond.

† This work was supported by the Italian National Research Council "Progetto Finalizzato Biotecnologie" and "Progetto Agenzia 2000".

* To whom correspondence should be sent: Istituto per lo Studio delle Macromolecole, Laboratorio NMR, CNR, Via Ampère 56, 20131, Milano, Italy. Tel.: 0039 02 70643554. Fax: 0039 02 70643557. E-mail: laura.ragona@ismac.cnr.it.

‡ Istituto per lo Studio delle Macromolecole.

§ Università degli Studi di Verona.

|| Istituto Sperimentale Colture Foraggere.

[1] Abbreviations: BBIs, Bowman-Birk inhibitors; BBI-I, Bowman-Birk inhibitor from soybean; MSTI, inhibitor from *Medicago scutellata* seed; PI-II, Bowman-Birk inhibitor from tracy soybean; PsTI-IVb, Bowman-Birk inhibitor from pea seed; TC, temperature coefficients; TSP, 3-(trimethylsilyl)-propionic acid-d_4 sodium salt.

10.1021/bi020576w CCC: $25.00
Published on Web 02/20/2003

The 11-residue peptide NleCTASIPPQCY, designed to reproduce the short β-sheet segment that forms the reactive site loop of BBIs, cyclized via a disulfide bridge, retains the "canonical conformation" typical of BBIs, indicating that this sequence represents an independent structural β-hairpin motif, and maintains the biological activity of the native protein (*13*). The 11-residue synthetic fragments SCTKSIxyQCY, where x and y were systematically substituted with A or P, were found to be potent inhibitors only when a cis-proline was present at position x (*14*).

It has been suggested that BBIs have antitumoral and radio-protective activity, and it is still unknown whether these activities are correlated to the inhibitory activity of trypsin/chymotrypsin or other enzymes or to the capability of interacting with other macromolecules (*11*, *15*−*21*). In this respect research has been focalized into three main lines, namely: (i) cellular proteases involved in carcinogenesis and possibly inhibited by BBIs; (ii) the effect of the inhibitors on the nuclear excision repair system, through the stabilization of TP53 protein (*22*, *23*), or the expression regulation of some protooncogenes such as c-myc and c-fos, overexpressed in many cancer cell lines (*24*); and (iii) in vivo inhibition of extra-cellular trypsin, secreted by many cancer cells and involved, with other metalloproteases, in the proteolysis of the extracellular matrix, thus favoring metastasis (*16*). The effort of the research along these lines stresses the importance of structural studies which may provide a link between specific structural features and antitumoral function.

Here we report a NMR structural study of a Bowman Birk trypsin inhibitor purified from Medicago Scutellata seeds (MSTI, 62 a.a.) (*25*). MSTI can inhibit the catalytic activity of bovine β−trypsin but, different from other BBIs, does not exhibit antichymotriptic activity.

Preliminary investigations showed that MSTI has cytotoxic activity on human brest carcinoma (MCF7) and human cervical carcinoma (HeLa) cell lines and improves the citotoxic effects of cisplatin (*26*). These MSTI properties could be relevant for the treatment of cisplatin resistant tumoral cell lines. Interestingly, MSTI anticarcinogenic property indicates that antichymotryptic activity is not a strict requirement for antitumoral activity as often reported in the literature (*27*, *28*).

Structural features, charge distribution and hydrophobicity of MSTI will be discussed in the present paper and compared to data available for other BBIs in order to elucidate common structural features of this class of proteins and to identify regions of the molecule relevant for its anticarcinogenic activity.

MATERIAL AND METHODS

Sample Preparation. MSTI (MW 6926) was purified from *Medicago scutellata* seeds following the procedure previously reported (*25*). The stoichiometric ratio of MSTI:trypsin complex was evaluated on the pure MSTI sample following the procedure reported in reference (*25*).

NMR structural studies of MSTI were performed in aqueous and deuterated 20 mM sodium phosphate buffer at pH 5.6, in the temperature range 17−47 °C, and employing sample concentrations ranging from 0.6 to 2.0 mM, at 500 MHz. At concentrations lower than 1.2 mM, no amide chemical shift change was observed, while at 2.0 mM amide shifts higher than 0.2 ppm were observed, probably due to self-association. NMR experiments performed at 500, 700, and 800 MHz, on 1.2 mM samples, were employed for structural characterization.

NMR Spectroscopy. NMR spectra were acquired on Bruker AVANCE 500, 700, and 800 MHz spectrometers.

Standard homonuclear DQF-COSY (*29*), TOCSY (*30*), and NOESY (*31*) experiments were recorded. Data matrixes contained 4096 × 1024 points in f_2 and f_1, respectively, for data acquired at 700 and 800 MHz, and 2048 × 512 points for data acquired at 500 MHz. Water suppression was achieved with gradients by 3−9−19 pulse sequence (*32*) or using excitation sculpting sequence (*33*).

At 500 MHz, TOCSY experiments were acquired with two isotropic mixing times, 35 and 50 ms, while mixing times of 50 and 125 ms were employed for NOESY experiments.

At 700 and 800 MHz, TOCSY experiments were acquired with 50 ms isotropic mixing, and 60 and 125 ms mixing times were employed for NOESY experiments.

For amide exchange rate measurements, 12 sequential homonuclear 2D TOCSY spectra with 50 ms were acquired. The duration of each TOCSY experiment was 4 h.

Spectra were processed with XWINNMR. Analysis of spectra and cross-peaks volumes were performed using XEASY software (*34*).

All chemical shifts were referenced to the methyl resonance of 3-(trimethylsilyl)-propionic acid-d_4 sodium salt (TSP).

Secondary chemical shift maps were obtained using as reference the random coil chemical shifts reported by Wishart (*35*).

Hydrogen Exchange Rates. The hydrogen exchange rates (k_{ex}) were determined by fitting cross-peak volumes to a first-order exponential decay, $y = e^{-k_{ex}t}$, where y represents the measured cross-peak volume of the resonances $H^{\alpha}-H^{N}$ and t is the time in minutes (*36*). The time point for an experiment was taken to be the time elapsed from the addition of D_2O buffer to initiation of that NMR experiment plus half the duration of the acquisition. The cross-peak volumes were normalized to the cross-peak volume of the non exchangeable aromatic protons $H_{2,6}/H_4$ of F56.

All the data were fitted with the program Sigmaplot (Jandel Scientific).

Temperature Coefficient. Temperature coefficients were measured for amide proton chemical shifts by analyzing TOCSY spectra acquired at 10° intervals from 17 to 47 °C. Chemical shifts versus temperature data were fitted to a linear equation with the program Sigmaplot (Jandel Scientific).

Structure Calculations. Calculated structures were obtained from restrained molecular dynamics simulations using DYANA (*37*) followed by energy minimization using the AMBER force field as implemented by the program DISCOVER (Molecular Simulations, San Diego, CA). Volume integration was performed on the NOESY spectrum acquired at 800 MHz with mixing time of 125 ms. The calibration of the peak volumes was performed using caliba, a routine of DYANA package, and the obtained list of distance restraints was used as input for DYANA calculations.

ϕ angle restraints were derived from $^3J_{NH}$ constants estimated by the separation of extrema in the dispersive and absorptive plots of DQF−COSY spectra (*38*). ϕ angle

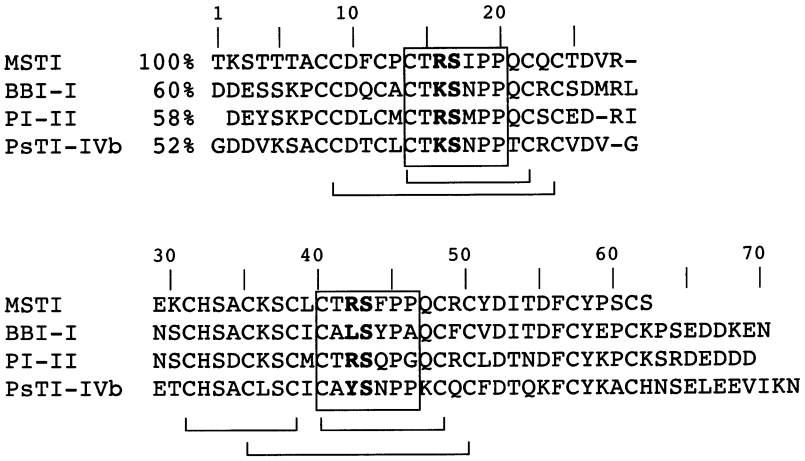

FIGURE 1: Amino acid sequence alignment of MSTI to three BBIs of known structure. Sequences of inhibitors from *Medicago Sculellata* (MSTI), soybean (BBI-I), tracy soybean (PI-II), and pea seed (PsTI-IVb) are shown. The percentage of sequence identity is reported. Residues in P1 and P1′ position are shown in bold, and a box defines the active loop of the inhibitors. The pattern of disulphide bonds is drawn on each tandem region. BBIs present two additional disulfide bonds involving C8−C61 and C12−C57 (MSTI numbering), not shown in the figure.

restraints of $-139° \pm 30°$ for $^3J_{HN-H\alpha}$ coupling constants greater than 8.0 Hz and $-60° \pm 30°$ for $^3J_{HN-H\alpha}$ coupling constants smaller than 5.0 Hz were used for calculations.

Amide proton exchange rates were estimated from residual amide proton signals observed in a TOCSY spectrum recorded at 27 °C and pH 5.6 after the sample was exchanged through AMICON in deuterated phosphate buffer. Hydrogen bond formation or solvent exclusion from the amide protons was assumed to account for the slow and medium exchange rate amide protons. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints. Each hydrogen bond was introduced as O−N distance of 2.00 Å and HN−O distance of 3.00 Å.

One hundred calculations were run employing DYANA starting from random polypeptide conformations, and the 20 conformers with the lowest residual target function values were analyzed. The restraints were re-examined in view of consistent violations and relaxed where necessary. This procedure was repeated until no consistent violations were found in half or more of the structures. After this step 100 of new DYANA calculations were started, and the 20 structures with the lowest target functions were further refined using the AMBER force field. Each structure was minimized performing 100 steps of steepest descent and 300 steps of conjugate gradient. The 15 structures with the lowest potential energy were selected for further analysis. RMSD values were calculated over the range of residues showing medium and long-range distance restraints (i.e., between amino acids at positions i and j, with $i - j \geq 2$), that is, between residue 6 and 61.

The INSIGHTII program (Molecular Simulations, San Diego, CA) was used to visually observe sets of structures, and the MOLMOL program (*39*) was employed for calculation and display of electrostatic potential surfaces.

RESULTS

Assignment of Spin Systems. The assignment of MSTI spectra was not straightforward due to the presence of (i) two symmetrical motives, namely, C_{14}TRSIPPQC_{22} and C_{40}TRSFPPQC_{48}, which represent the two tryptic inhibitory domains (Figure 1), and (ii) 14 cysteines, all involved in the following disulfide bridges C8−C61, C9−C24, C12−C57, C14−C22, C31−C38, C35−C50, and C40−C48 as previously determined (*25*). In addition the MSTI primary sequence presented 32 AMX spin systems that caused ambiguity in the first step of the residue assignment. Extensive overlap was observed at 500 and 700 MHz also for residues outside the two symmetric regions, and acquisition of NMR spectra at 800 MHz was necessary to complete the assignment of spin systems.

The starting points for the assignment were two isolated spin systems with amide chemical shift at low fields (11.52 and 11.54 ppm) corresponding to D10 and K36, respectively. The assignment of A7, showing β protons at high fields (0.50 ppm), and of some cysteine spin systems showing downfield shifted Hα (C12, C14, C22, C24, C38, C40, C48, and C50) could be easily performed. All these cysteines showed Hα−Hα NOEs, useful starting points for the assignment of β-sheet regions of MSTI (see later). The assignment of the 500 MHz spectra at 27 °C was in a first step interrupted in the loop regions (residues 16−18 and 41−44) and at the level of residues Q21, Q23, K30, C31, H32, C35, Q47, F56, C61, and S62 due to extensive overlap in the amide region. To resolve these ambiguities, experiments were performed at different temperatures, in the range 17−47 °C, but the shifts induced were not sufficient to disentangle the overlapping signals.

At 700 MHz the overlapping signals were better resolved but only some hypothesis could be put forward relative to residues in the loop regions.

TOCSY and NOESY spectra were therefore performed at 800 MHz; many new NOEs could be unambiguously identified and the assignment could be completed. S17 and S43, located in the two homologous loops, exhibited completely overlapped spin systems which could be distinguished on the basis of the NOE correlations S17Hβ-I18HN and S43Hβ-F44HN. In fact, S43Hβ-F44HN NOE, belonging to an intensely crowded region at 500 MHz, could be clearly identified at 800 MHz. Also at this field, S17HN−I18HN

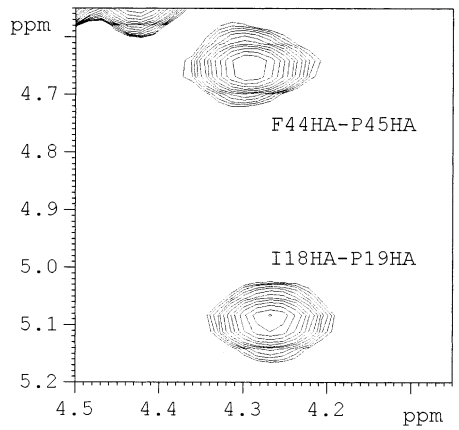

FIGURE 2: Selected region of NOESY spectrum of MSTI at 27 °C and pH 5.6 acquired at 800 MHz showing the NOE connectivities I18Hα−P19Hα and F44Hα−P45Hα.

and S43HN−F44HN cross-peaks could be assigned. R16 and R42 exhibited as well overlapping spin systems, except for the Hα protons which could be identified only at 800 MHz. The assignment of C35, showing the HN−Hα cross-peak superimposed to those of S43 and S17, was obtained at 800 MHz, on the basis of the identification of the following NOEs: C35Hα−K36HN, C35HN−K36HN, and A34Hβ's−C35HN. At 800 MHz, H32 could be assigned since it was possible to identify H_2 and H_4 long-range NOEs with the side chains of Y58 and A34. The higher dispersion and sensitivity in the region 7.10−7.60 ppm and 8.10−8.80 ppm, at 800 MHz, permitted the assignment of K30 and C31, which were identified through the sequential NOEs K30Hβ's−HNC31 and K30Hγ's−C31HN. At 800 MHz, the nearly degenerate $H_{2,6}$ and $H_{3,5}$ (7.37 and 7.35 ppm respectively) of F11 were unambiguously identified, together with the NH_2 resonances of all arginines, glutamines, and asparagines.

The observation of sequential $d_{\alpha\delta}$ (i,i+1) and $d_{\alpha\delta'}$ (i,i+1) (trans), or $d_{\alpha\alpha}$ (i,i+1) (cis) connectivities in NOESY spectra at short mixing times (60 ms) allowed the determination of the geometry of the X-Pro bonds. On these basis, I18−P19 and F44−P45 were shown to have cis geometry (Figure 2), while all the remaining X−Pro bonds were trans.

The complete assignment, as obtained at 800 MHz, is available as Supporting Information and has been deposited in BMRB (id code: BMRB−5617).

The assignments were tested using the automated assignment algorithm NOAH (*40*). The long-range ambiguous NOEs were peaked without assignment and used as input in a further NOAH run. The assignment suggested by NOAH was then analyzed, and a few reasonable NOESY peak assignments were added.

The assignment obtained at 800 MHz at 27 °C in turn allowed the assignment of the spectra acquired at 500 MHz at different temperatures (17, 37, 47 °C) and of the spectra acquired at a higher protein concentration (2 mM) at 27 °C (see later).

Secondary Structure. A summary of NOE connectivities and $^3J_{HN-H\alpha}$ coupling constants is reported in Figure 3, while secondary chemical shifts are reported in Figure 4. Conformational shift maps are in good agreement with the secondary structure determined on the basis of NOEs. Two triple stranded β-sheets, located at the level of residues 11−15 (A strand), 21−25 (B strand), 53−55 (F strand), and 27−29 (C strand), with 37−41 (D strand) 47−51 (E strand) defining the first and second domain, respectively, could be identified on the basis of secondary shifts, canonical nonsequential $d_{\alpha\alpha}$ (i,j), d_{NN} (i,i) and $d_{\alpha N}$ (i,j) NOEs and $^3J_{HN-H\alpha}$ couplings (Figure 5). Each β-sheet is characterized by a β-hairpin made of five-residue pairs of antiparallel β-sheets and a five-residue tight type VIb turn, together with an additional short strand forming a triple stranded β-sheet.

Each type VIb turn includes a P1−P1′ reactive site conferring the tryptic inhibitory activity, corresponding to peptide bonds R16−S17 and R42−S43, and contains a cis peptide X−Pro bond (I18−P19 for the first β-sheet and

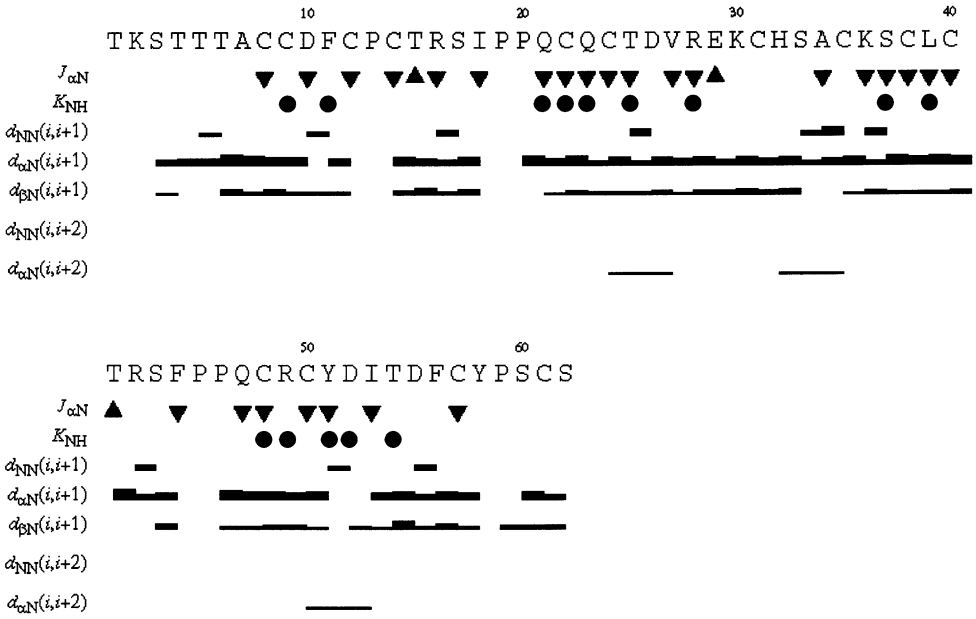

FIGURE 3: Summary of short and medium range upper distance limits and $^3J_{\alpha N}$ values. The thickness of the bars indicates the relative normalized intensity of the NOE. $^3J_{\alpha N}$'s are three-bond coupling constants between Hα and HN, where the symbols represent the following: ▼ for $^3J_{\alpha N} > 8$ Hz, ▲ for $^3J_{\alpha N} < 5$ Hz. In the row labeled k_{NH}, filled circles identify residues which present slow amide proton exchange rates (see Table 1).

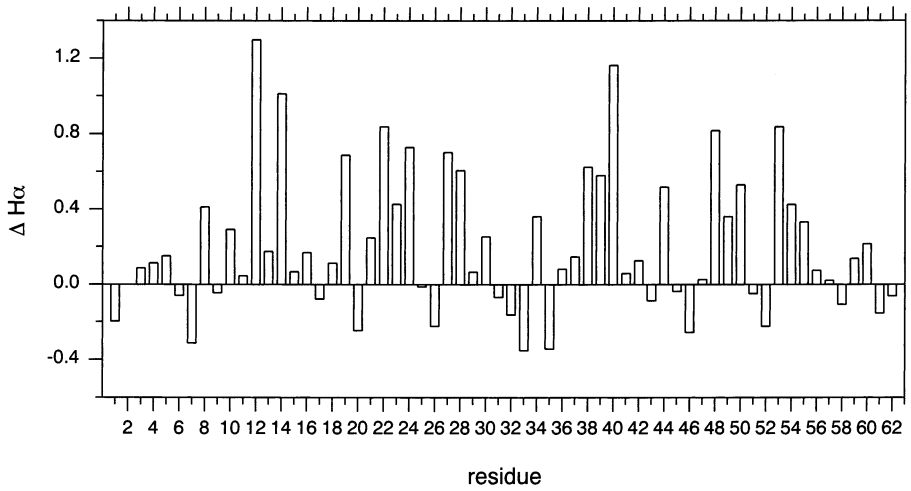

FIGURE 4: Secondary Hα chemical shifts of MSTI at 27 °C and pH 5.6.

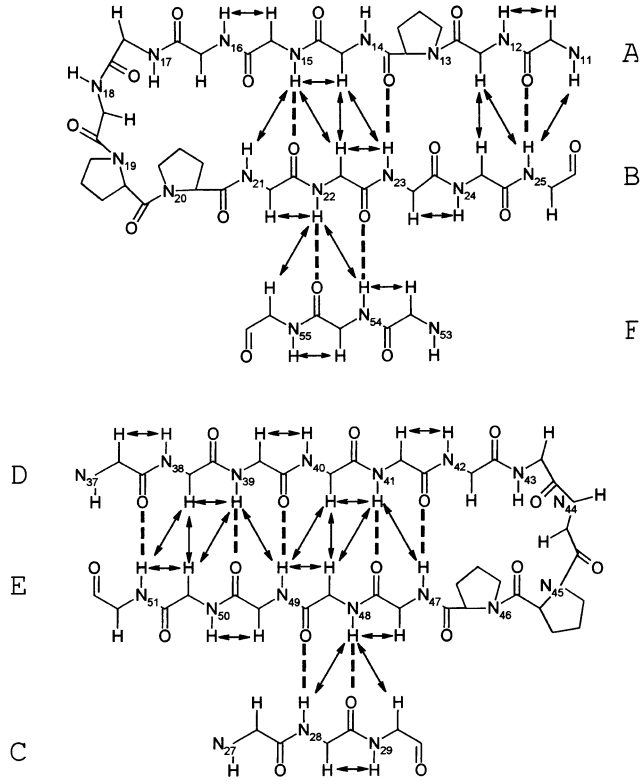

FIGURE 5: Secondary structure contacts and hydrogen bonds for MSTI at 27 °C and pH 5.6. β-structure elements A−F are indicated together with observed inter-strand connectivities (arrows). Dotted lines indicate H-bonds present in the final calculated structures.

F44−P45 for the second β-sheet) that is required for inhibitory activity (*14*). The following observations are consistent with a type VIb turn: (i) coupling constants $^3J_{HN-H\alpha} \leq 5$ Hz were measured for residues T15 and T41; (ii) strong NOEs $d_{\alpha N}$ (i,i+3) were observed between P20Hα and S17HN, in the first active loop, and between P46Hα and S43HN in the second active loop (*41*). The presence of two turn conformations was also confirmed by negative secondary shifts for residues S17 and P20 in the first active loop and S43 and P46 in the second one (Figure 4).

A few half-turns could be identified at the level of residues 24−26 (N-terminal β-sheet) and 50−52 (C-terminal β-sheet) on the basis of C24Hα−D26HN and C50Hα−D52HN NOEs complemented by negative secondary shifts observed for D26 and D52 (Figures 3 and 4). The presence of H32Hα−A34HN and S33HN−A34HN NOEs suggests that the polypeptide chain is bent, between the two domains, in a type I-like turn.

Amide Exchange Rates and Temperature Coefficients. Hydrogen/deuterium exchange experiments were performed at 27 °C and pH 5.6 in order to gain information on the hydrogen-bonded amides. The fitting of the measured Hα−HN cross-peak volumes at different times after addition of D_2O allowed the estimate of amide exchange rates (k_{ex}) for 14 residues (Table 1). The results are consistent with the β-sheet structure identified on the basis of NOE connectivities: all amide protons of residues located in the strand regions show low k_{ex} with the exception of T15, T41, and Q47.

Amide temperature coefficient and amide exchange rate are, to some extent, complementary tools to predict hydrogen bond donors, and a combined use of the two approaches produces a far more reliable indicator of hydrogen bonding than either alone (*42*). Temperature coefficients are easily measured for all residues, do not display a pH dependence as amide exchange, and are not strongly influenced by surface exposure as exchange rates. A general criterion is that all of the amide protons displaying slow exchange rates and temperature coefficients (TC) more positive than −4.5 ppb/K could be located in hydrogen-bonded regular secondary structures. Chemical shifts variations of MSTI as a function of temperature are reported in Table 1.

Exchange rates and TC are in good agreement. The criterion of TC more positive than −4.5 ppb/K (within experimental error) identifies more amide protons involved in H-bond, some of them possibly corresponding to exposed H-bonded residues. Only two residues, namely, C48 and T54, show low k_{ex} and TC around 6 ppb/K, and they could be involved in less stable H-bonds stabilizing the short third strand.

The indication of the presence of H-bonds, as identified by the two techniques, was exploited for structure calculations. The partners for all hydrogen bonds were assigned on the basis of preliminary structures obtained by imposing only NOE restraints, as described in Material and Methods.

Tertiary Structure. A summary of experimental restraints used for structure calculations is reported in Table 2. A set of 726 NOEs was supplemented by (i) 21 distance restraints for the 7 disulfide bridges, (ii) 22 distance restraints for 11

Table 1: Temperature Coefficients and Amide Exchange Rates for MSTI at pH 5.6[a]

residue	NH $\Delta\delta/\Delta T$	k_{ex} (10^{-5} min^{-1})
S3	−6.0 ± 2.0	
T4	−8.7 ± 0.5	
T5	−6.5 ± 0.9	
T6	**−2.9** ± 2.9	
A7	−8.6 ± 0.5	
C8	−5.5 ± 0.9	
C9	**−3.7** ± 0.5	154 ± 12
D10	**−4.2** ± 0.5	
F11	**−2.7** ± 0.6	46 ± 2
C12	−5.8 ± 0.4	
C14	−8.8 ± 0.6	
T15	**−4.2** ± 0.6	
R16/R42	−7.7 ± 0.5	
S17/S43	**−1.1** ± 0.8	
I18	−9.8 ± 0.6	
Q21	**−2.6** ± 0.6	196 ± 17
C22	**−5.0** ± 0.6	36 ± 12
Q23	**−0.9** ± 0.5	123 ± 5
C24	**−2.7** ± 0.6	
T25	**−3.7** ± 0.3	154 ± 8
D26	**−1.0** ± 0.7	
V27	−7.9 ± 0.7	
R28	**−5.0** ± 0.3	11.9 ± 0.3
E29	**−3.1** ± 0.6	
K30	**−3.7** ± 0.7	
C31	−10.0 ± 1.0	
H32	−6.1 ± 0.6	
S33	−6.9 ± 0.8	
A34	**−3.1** ± 0.8	
C35	**−2.6** ± 0.7	
K36	**−4.6** ± 0.5	
S37	**−2.3** ± 0.5	135 ± 5
C38	−9.2 ± 0.9	
L39	**−3.0** ± 0.6	76 ± 3
C40	−9.0 ± 0.7	
T41	**−4.8** ± 0.7	
R42/R16	−7.7 ± 0.4	
S43/S17	**−1.1** ± 0.8	
F44	−10.0 ± 0.7	
Q47	**−1.8** ± 0.1	
C48	−6.2 ± 0.3	118 ± 10
R49	**−1.5** ± 0.2	8 ± 1
C50	**−4.9** ± 0.6	
Y51	**−4.5** ± 0.5	126 ± 11
D52	**−1.1** ± 0.6	27 ± 10
I53	−7.5 ± 0.9	
T54	−5.8 ± 0.6	16 ± 1
D55	**−2.9** ± 0.3	
F56	**−2.7** ± 0.9	
C57	−7.7 ± 0.3	
Y58	−15.4 ± 0.3	
S60	−8.9 ± 0.6	
C61	**−1.1** ± 0.8	
S62	−6.4 ± 0.9	

[a] Temperature coefficients values lower than 5.0 ppb/K are shown in bold. Amide exchange rates were measured at 27 °C.

Table 2: Structural Statistics of the 15 Best Structures Obtained for MSTI

(a) Restraints	
total number of distance restraints	726
number of intraresidual restraints	257
number of sequential restraints	264
number of medium-range restraints	22
number of long-range restraints	183
number of torsion angle restraints	29
number of hydrogen bonds	11
(b) DYANA	20 Structures
target function (Å^2)	1.84 ± 0.19
$\langle rmsd\rangle_{bb}$ (6−61) (Å)	0.75 ± 0.18
$\langle rmsd\rangle_{heavy}$ (6−61) (Å)	1.51 ± 0.17
(c) DISCOVER	15 Structures
total potential energy (kcal/mol)	−115 ± 10
$\langle rmsd\rangle_{bb}$ (6−61) (Å)	0.84 ± 0.09
$\langle rmsd\rangle_{heavy}$ (6−61) (Å)	1.60 ± 0.12
$\langle rmsd\rangle_{bb}$ (11−15, 21−25, 53−55) (Å)	0.40 ± 0.05
$\langle rmsd\rangle_{bb}$ (37−41, 47−51, 27−29) (Å)	0.57 ± 0.09

backbone hydrogen bonds defined on the basis of deuterium hydrogen exchange studies and temperature coefficient values, and (iii) 29 ϕ angle restraints derived from $^3J_{N\alpha}$ coupling constants. The superposition of the 15 best structures is reported in Figure 6a, and the structures have been deposited in the PDB (PDB code 1MVZ). The structure of MSTI comprises two distinct domains composed by a three stranded β-sheet and the active site, located in a VIb type loop. The combination of the tight turn and the antiparallel β-sheet forms the characteristic β-hairpin inhibitory domain common to other Bowman Birk inhibitors (Figure 6b). The first domain comprises residues 11−25 and 53−55, while the second involves residues 27−29 and 37−51. The β-hairpins are planar and extended, and each strand is held in place by the disulfide cross-link across the sheet, namely, C14−C22 for the first domain and C40−C48 for the second domain. The mentioned disulfide bridges are functionally relevant since they hold the two peptide chains together after P1−P1′ peptide bond is cleaved by the protease (*1*). The other disulfide bridges connect residues belonging to one domain with residues at the border of the same domain (C9−C24 and C12−C57 for the first domain and C31−C38 and C35−C50 for the second one) or connect two residues at the border of the same domain (C8−C61). The regions of antiparallel β-sheets are quite well defined with $\langle rmsd\rangle_{bb}$ (11−15, 21−25, 53−55) = 0.40 ± 0.05 Å for the first domain and $\langle rmsd\rangle_{bb}$ (37−41, 47−51, 27−29) = 0.57 ± 0.09 Å for the second domain. The type VIb turns are less defined, due to the lack of NOEs especially for residues R16−S17 and R42−S43 which only show intra and sequential NOEs. The mentioned peptide bonds correspond to reactive sites and a hyper-exposed P1 residue (R in both domains) is a characteristic and important functional feature of the canonical conformation (*43*). The salt bridge between the basic P1 of the substrate with the acidic D189 in the interior of the S1 pocket of trypsin is responsible of the narrow preference of this enzyme for basic residues in position P1 (*11*).

Twelve backbone H-bonds typical of β structure were present in all the final structures (Figure 5). T15 and T41 amides are hydrogen bonded to backbone oxygen of residues Q21 and Q47, respectively. The two amides are located close to the first and second active sites, in a symmetrical position. They show NOEs typical of β-sheet structure, temperature coefficients lower than −5 ppb/K, but their hydrogen exchange rates were not measurable at pH 5.6. The fast exchange of the two amides underlines the exposure of the inhibitory region and the low stability of the H-bond pattern close to the active loops.

The amide proton of Q21 (P5′ position), belonging to the first domain, forms H-bond with the side-chain oxygen (Oγ) of T15 (P2 position) in all the structures. Q47 (P5′ position of the second domain) presents a low-temperature coefficient but a fast k_{ex}, and its amide forms H-bonds with Oγ of T41 (P2) only in 6 out of 15 structures. In the remaining struc-

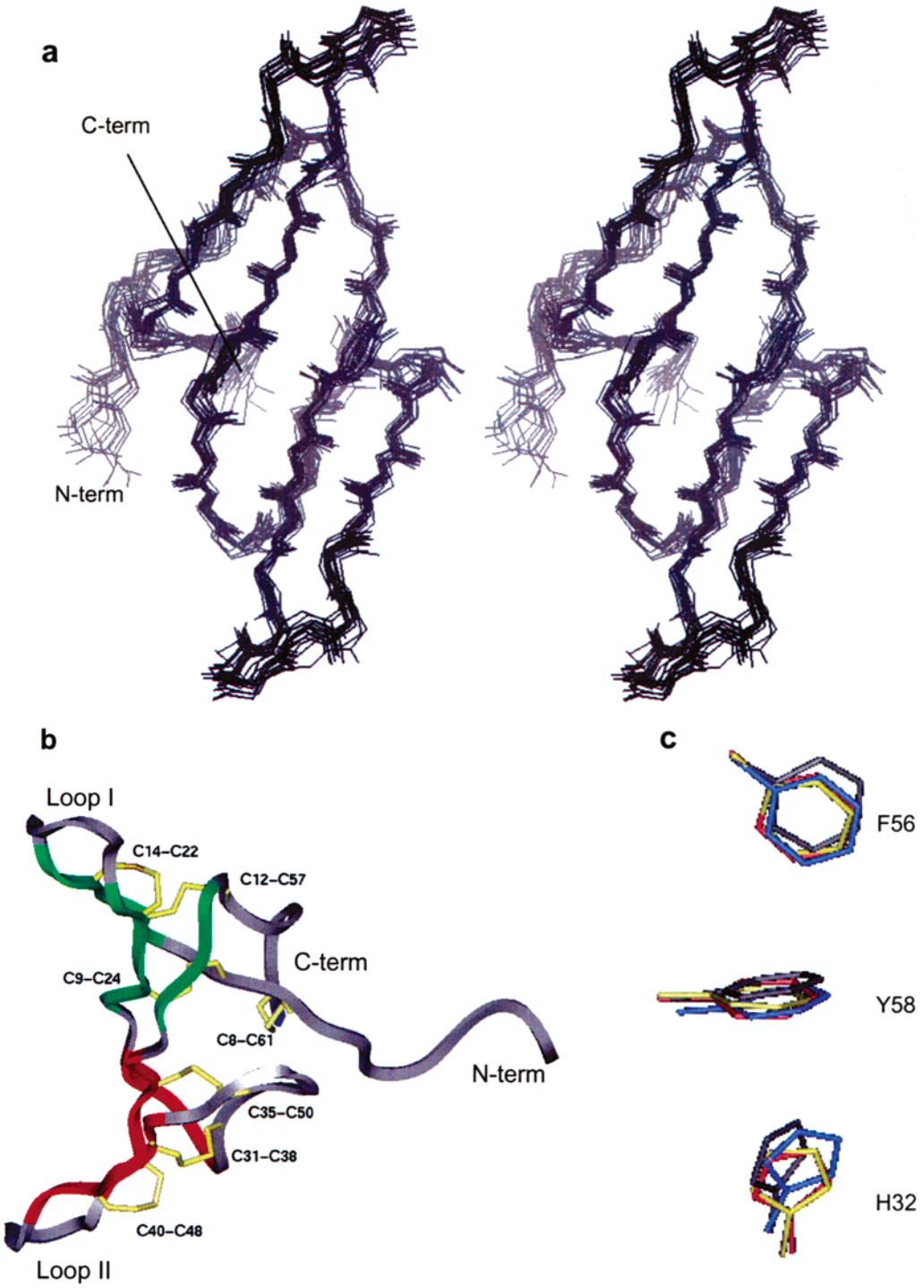

FIGURE 6: Stereoview of backbone superposition of the 15 final structures of MSTI. Superposition was performed in the range of residues 6−61 (a). Lowest-energy MSTI structure drawn in ribbon representation. Residues 11−25 and 53−55, corresponding to the first three-stranded β-sheet, are colored in green; residues 27−29 and 37−51, corresponding to the second three-stranded β-sheet, are colored in red. The active loops, the disulfide bridges and N and C-termini are indicated (b). Superposition of residues belonging to the aromatic cluster present in all BBIs of known structure (MSTI, gray; BBI-I, yellow; PI-II, red; PsTI-IVb, blue) (c).

tures, the side-chain oxygen of T41 was involved in H-bond with the amide proton or the side-chain oxidrile of S43 (P1′). Both Q47 and S43 amides have low-temperature coefficients but do not show a significant protection from exchange. A possible interchange of donors for Oγ T41 can be hypothesised, suggesting the presence of higher mobility in the second domain loop region with respect to the first one.

The lack of inter-domain disulfide bridges (all disulfide bridges are indeed localized either within a domain or between a residue of one domain and a residue at the border of the same domain) (see Figure 6) allows for flexibility of the spatial orientation of the two domains relative to each other, as reflected by the lack of NOEs between the two domains. These observations underline the double headed feature of this inhibitor, requiring mobility of one domain relative to the other for its precise adjustment during the simultaneous inhibition of two trypsin molecules (*2*).

Self-Association. Spectra of MSTI were performed at different protein concentrations ranging from 0.6 to 2.0 mM in order to characterize the self-association behavior. The

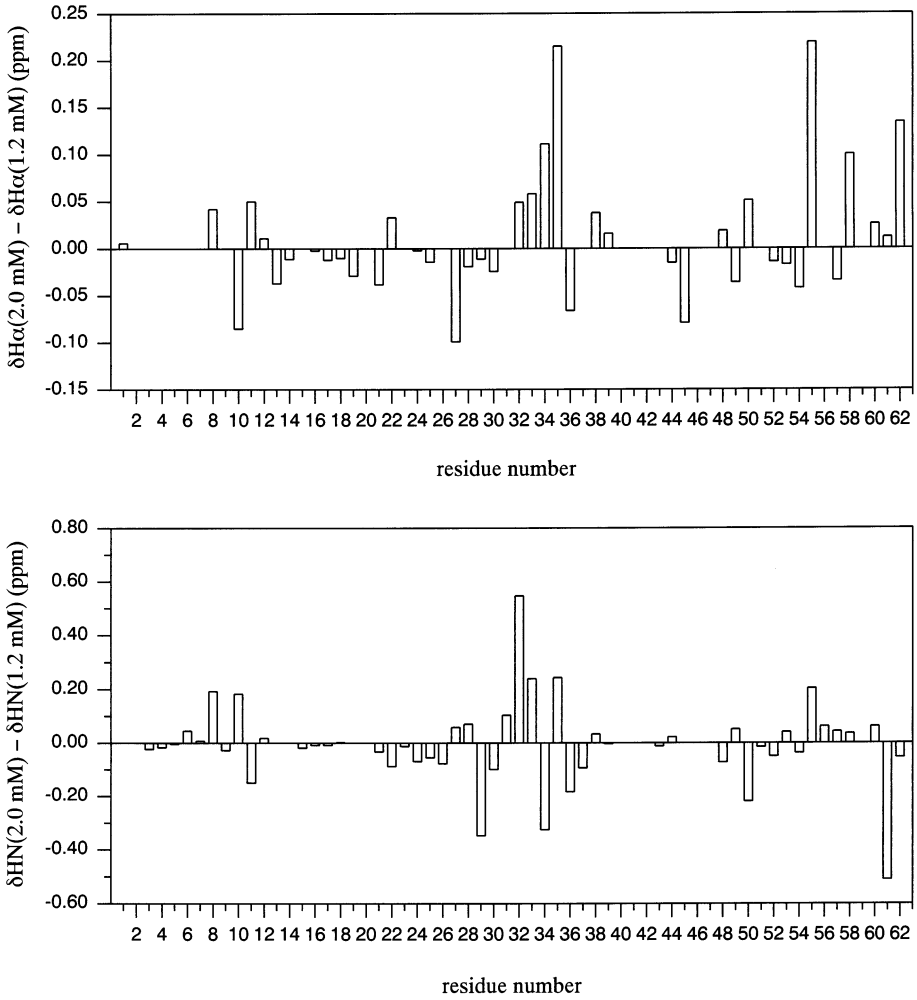

FIGURE 7: Hα (upper panel) and HN (lower panel) chemical shift differences measured for 1.2 and 2.0 mM MSTI samples.

assignment of the spectra at different protein concentrations was easily extended from that obtained for 1.2 mM sample. The comparison of Hα conformational shifts at 1.2 mM and 2.0 mM indicates that the secondary structure of the domains is preserved upon association.

At 2.0 mM protein concentration a few amide, α, and side-chain protons shifts were observed (Figure 7), clearly due to self-association. The residues showing amide shifts greater than 0.1 ppm and/or Hα or side-chain hydrogen shifts greater than 0.05 ppm are listed: C8, D10, F11, V27, R28, E29, K30, C31, H32, S33, A34, C35, K36, S37, C48, C50, D55, F56, C57, C61. It is clear that all residues belonging to the loop connecting the two domains are affected by aggregation together with some residues located in the N and C-terminal regions and in the C and E strands of the second domain.

DISCUSSION

Despite extensive studies on inhibitory and anticarcinogenic activity of BBIs, only few three-dimensional structures have been solved, namely, those of tracy soybean (*5*) (PI-II, PDB codes 1PI2), soybean (*7*, *8*) (BBI-I, PDB codes 2BBI and 1K9B), and pea seeds (*2*) (PsTI-IVb, PDB code 1PBI) inhibitors.

The topology of MSTI, as determined by ^{1}H NMR spectroscopy and molecular dynamics calculations, is common to all known Bowman Birk inhibitors and comprises two distinct domains each composed by a triple stranded β-sheet and the active site, located in a VIb type turn. The structural comparison of MSTI with all BBIs of known structure performed through the VAST search service (http://www.ncbi.nlm.nih.gov/Structure/VAST) indicated that PsTI-IVb has the highest structural similarity, exhibiting an RMSD for alignable regions of 1.3 Å. The RMSD with other BBIs was slightly higher and comprised between 1.5 and 1.7 Å. The presence of an unusual buried cluster of charged side-chains, located in the inter-domain region, has been identified in the X-ray structure of BBI-I (*8*, *44*). The cluster of charged side-chains is also present in MSTI, involving D26, R28, and H32 residues, while the same geometry cannot be established in PI-II and PsTI-IVb due to local sequence mutations at the level of R28.

Interestingly, three aromatic residues, highly conserved in all BBIs (Figure 1), namely, H32, F56, and Y58, form in MSTI an aromatic cluster showing a typical geometry in which pairs of aromatic residues are almost orthogonal to each other and with centroids at distances minor than 7 Å. This geometry is known to confer thermostability to many proteins (*45*), in agreement with the observation that MSTI shows high-temperature stability. The aromatic cluster is present in all the other BBIs of known structure (Figure 6c) conferring, together with the H-bond network and the seven disulfide bridges, a high stability to this class of inhibitors.

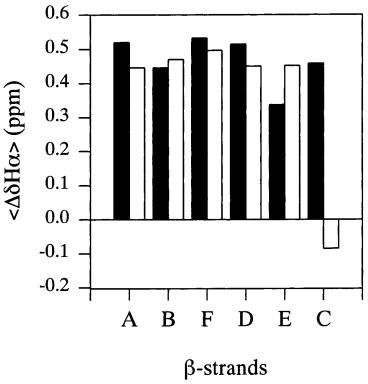

FIGURE 8: Comparison of average secondary chemical shifts of MSTI (black bars) and BBI-I (white bars). The results are reported for strands belonging to the first (A, B, and F) and the second domain (D, E, and C).

A comparison of NMR structural data can be performed between MSTI and BBI-I, sharing the highest sequence similarity and whose NMR solution structure has been previously determined (*46*). Hα secondary shifts show a similar behavior for the two proteins with only few exceptions attributed to the mutation of the residue itself or of its neighbors. It is, however, interesting to note that the main difference occurs at the level of the third strand of the second domain (C strand). The average Hα secondary shifts ($\langle \Delta H_\alpha \rangle$) within each strand were calculated for MSTI and BBI-I and reported in Figure 8. BBI-I C strand shows a negative $\langle \Delta H_\alpha \rangle$, suggesting that this strand has a very low stability. Indeed, the authors stated that the second domain (antichymotryptic) presents fewer stabilizing tertiary interactions with respect to the first one. The two β-sheets of MSTI display comparable stability on the basis of Hα secondary shifts, a few differences were, however, evidenced in the active loop regions based on the analysis of the 15 final structures. Our data indicate that in the first domain a hydrogen bond between the hydroxyl group of threonine, in P2 position of the active loop, and the amide proton of residue in P5′ position is formed in all the structures. In the second domain, this H-bond is present only in six structures out of 15 suggesting a higher mobility. Threonines in P2 position (T15 and T41) are highly conserved in BBIs, implying that these residues are functionally important for inhibitory properties.

It is reported that when BBIs interacts with the enzyme, the intramolecular hydrogen bond involving T15 Oγ stabilizes the active loop also favoring the correct positioning of T15 in order to maximize the hydrophobic interactions with H57 and L99 residues of trypsin (*11*). Further studies on amino acid variations in P2 position of cyclic peptides, encompassing the reactive site loops, demonstrated that a threonine confers both the lowest dissociation constant and the slowest hydrolysis rate (*47*), as expected for a good inhibitor. All these data underline the utmost importance of P2 threonine, whose side chain can provide both hydrophobic interactions via its methyl group and H-bonds via the hydroxyl group.

Protease recognition mechanism is triggered by the distribution of hydrophilic and hydrophobic residues in the

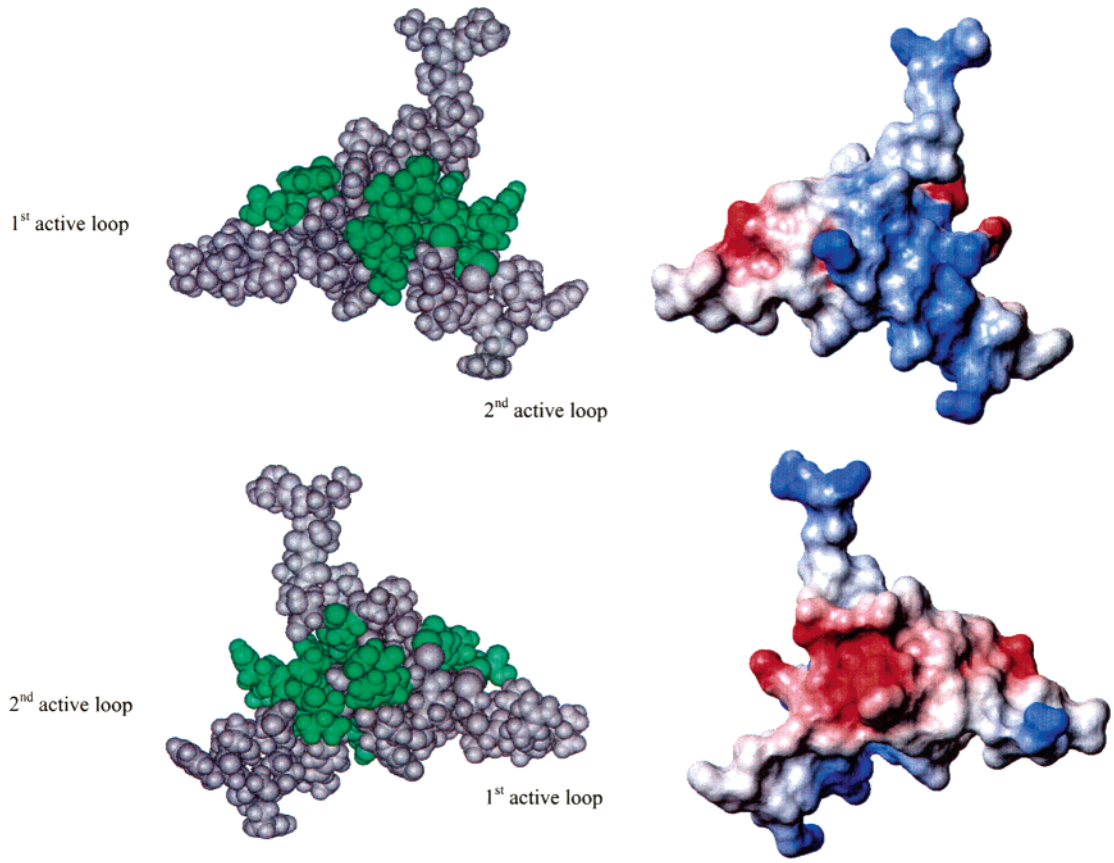

FIGURE 9: Self-association interfaces and electrostatic potential surfaces of MSTI. Two views of MSTI are offered in the lower and upper panel (one rotated by 180 degrees with respect to the other about *z*-axis). Residues affected by self-association are colored in green on CPK representation of the molecule (left panel); electrostatic potential surfaces were calculated by MOLMOL program (*39*) (right panel). The location of the first and second active loop in the two views are indicated.

domain regions. Hydrophilic bridges and conformational rigidity are characteristic of narrow specificity against trypsin, while high hydrophobicity and significant conformational mobility, typical of antichymotryptic domains, confer a broad inhibition specificity (*11*). Indeed, it is reported that BBI-I second domain is able to interact either with chymotrypsin or with trypsin (*4*). Polar and charged side chains are predominant in both domains of MSTI as expected for trypsin inhibition specificity, anyway the second domain presents a higher hydrophobicity. As an example the proline residue, in P4 position of the first domain, is replaced in the second domain by the longer hydrophobic side chain of L39. The same change takes place in BBI-I, where P4 position of the first domain (anti-tryptic) is a short hydrophobic side-chain A13, replaced in the anti-chymotryptic by an isoleucine (I40). This observation suggests a low specificity of MSTI second domain for trypsin, supported as well by the structural results pointing at its lower conformational rigidity. In agreement with our hypothesis, titration studies have shown a stoichiometric ratio MSTI:trypsin of 1:1.5. Moreover, the same K_d value of 1×10^{-10} M was measured for MSTI (*25*), presenting two antitryptic domains, and for BBI-I, presenting one antitryptic and one antichymotriptic domain (*11*).

Preliminary investigations have shown that MSTI has cytotoxic activity, reducing of about 20% the cellular survival of both MCF7 and HeLa, and improves the cisplatin effect on a few tumoral cell lines. (*26*). Since MSTI presents two antitryptic domains, these data indicate that antichymotryptic activity is not a strict requirement for antitumoral effect, as often reported in the literature. In addition it has been shown that inhibitory activity and antitumoral activity are not correlated. Indeed, after cleavage of BBI-I and linearization of the two resulting peptides, carrying the chymotrypsin and trypsin inhibitory sites, respectively, both fragments, lacking inhibitory activity, act radio-protectively (*48*). In agreement with these results, a recent analysis of a few nona-peptides, reproducing the second active loop of BBI-I, suggested that radio-protective and inhibitory properties are inherent to different residues (*49*). In this line we propose that the less rigid second domain of MSTI might bear some relevance relative to the recognition mechanism and to the antitumoral activity.

We further analyzed self-association behavior of MSTI collecting NMR data at different protein concentrations. The residues affected by self-association are localized in two well defined regions on opposite faces of the molecule (Figure 9, left panel). The calculation of the electrostatic potential surfaces shows that these two faces present an opposite potential (Figure 9, right panel), and we suggest that electrostatic interactions between them may play an important role in MSTI association. Residues involved in MSTI self-association, located at the border of the domains and in C and E strands (second domain), are highly conserved in BBIs from different seeds suggesting a functional relevance of the two superficial patches. The presence of self-association has also been reported for PsTI-IVb (*2*), PI-II (*5*) and BBI-I (*7*), but the structural data for the dimer have been reported only for PsTI-IVb. Its dimer interface comprises an extensive hydrogen-bond network together with hydrophobic contacts among residues belonging to both domains and to the longer C-terminal tail**.** It is important to stress that different residues are involved in self-association of MSTI and PsTI-IVb. Residues involved in PsTI-IVb self-association are not conserved in BBI family, at variance with those of MSTI. From these data we infer that residues of MSTI affected by self-association may represent those regions of the protein capable of interacting with other enzymes or macromolecules, which could be of the highest importance in antitumoral activity.

The identification of association/interaction interfaces will be further investigated to correlate the association properties of MSTI with its antitumoral activity.

ACKNOWLEDGMENT

We gratefully acknowledge the technical support offered by the Large Scale Facility of Firenze (CERM) for the acquisition of NMR spectra at 700 and 800 MHz. The authors thank A. Lanza and R. Iori for helpful discussions and F. Greco and G. Zannoni for technical assistance.

SUPPORTING INFORMATION AVAILABLE

List of proton chemical shifts for MSTI at 27 °C and pH 5.6. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Laskowski, M., Jr., and Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593−626.
2. Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J., and Brunie, S. (1999) *J. Mol. Biol. 285*, 1195−207.
3. Odani, S., and Ikenaka, T. (1978) *J. Biochem. 83*, 747−53.
4. Jensen, B., Unger, K. K., Uebe, J., Gey, M., Kim, Y. M., and Flecker, P. (1996) *J. Biochem. Biophys. Methods 33*, 171−85.
5. Chen, P., Rose, J., Love, R., Wei, C. H., and Wang, B. C. (1992) *J. Biol. Chem. 267*, 1990−1994.
6. Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S., and Ikenazka, T. (1993) *J. Mol. Biol. 234*, 722−734.
7. Werner, M. H., and Wemmer, D. E. (1992) *Biochemistry 31*, 999−1010.
8. Voss, R. H., Ermler, U., Essen, L. O., Wenzl, G., Kim, Y. M., and Flecker, P. (1996) *Eur. J. Biochem. 242*, 122−31.
9. Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S., and Takahashi, K. (1986) *J. Biochem. 100*, 1637−1646.
10. Lin, G., Bode, W., Huber, R., Chi, C., and Engh, R. A. (1993) *Eur. J. Biochem. 212*, 549−555.
11. Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., and Flecker, P. (2000) *J. Mol. Biol. 298*, 477−491.
12. Luckett, S., Garcia, R. S., Barker, J. J., Konarev, A. V., Shewry, P. R., Clarke, A. R., and Brady, R. L. (1999) *J. Mol. Biol. 290*, 525−33.
13. Brauer, A. B., Kelly, G., McBride, J. D., Cooke, R. M., Matthews, S. J., and Leatherbarrow, R. J. (2001) *J. Mol. Biol. 306*, 799−807.
14. Brauer, A. B. E., Domingo, G. J., Cooke, R. M., Matthews, S. J., and Leatherbarrow, R. J. (2002) *Biochemistry 41*, 10608−10615.
15. Kennedy, A. R. (1998) *Am. J. Clin. Nutr. 68*, 1406S−1412S.
16. Miyata, S., Koshikawa, N., Yasumitsu, H., and Miyazaki, K. (2000) *J. Biol. Chem. 275*, 4592−4598.
17. Sorsa, T., Salo, T., Koivunen, E., Tyynelä, J., Konttinen, Y. T., Bergmann, U., Tuuttila, A., Niemi, E., Teronen, O., Heikkilä, P., Tschesche, H., Leinonen, J., Osman, S., and Stenman, U. H. (1997) *J. Biol. Chem. 272*, 21067−21074.
18. Miyata, S., Miyagi, Y., Koshikawa, N., Nagashima, Y., Kato, Y., Yasumitsu, H., Hirahara, F., Misugi, K., and Miyazaki, K. (1998) *Clin. Exp. Metastasis 16*, 613−622.
19. Ware, J. H., Wan, X. S., Rubin, H., Schechter, N. M., and Kennedy, A. R. (1997) *Arch. Biochem. Biophys. 344*, 133−138.
20. Ware, J. H., Wan, X. S., and Kennedy, A. R. (1999) *Nutr. Cancer 33*, 174−177.
21. Billings, P. C., St. Clair, W., Owen, A. J., and Kennedy, A. R. (1998) *Cancer Res. 48*, 1798−17802.

22. Dittmann, K. H., Gueven, N., Mayer, C., Ohneseit, P., Zell, R., Begg, A. C., and Rodemann, H. P. (1998) *Radiat. Res. 150*, 648−655.
23. Dittmann, K. H., Gueven, N., Mayer, C., and Rodemann, H. P. (1998) *Int. J. Radiat. Biol. 74*, 225−230.
24. St. Clair, W. H., Billings, P. C., and Kennedy, A. R. (1990) *Cancer Lett. 52*, 145−152.
25. Ceciliani, F., Tava, A., Iori, R., Mortarino, M., Odoardi, M., and Ronchi, S. (1997) *Phytochemistry 44*, 393−398.
26. Lanza, A., Tava, A., Zetta, L., Di Francesco, S., Robustelli della Cuna, F. S., and Robustelli della Cuna G. (2001) *Proc. XXVII Congresso Nazionale Oncolog.*, 151.
27. Kennedy, A. R. (1998) *Am. J. Clin. Nutr. 68*, 1406−1412.
28. Kennedy, A. R., and Wan, X. S. (2002) *Prostate 50*, 125−133.
29. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
30. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 66*, 355−360.
31. Sklenar, V., Piotto, M., Leppik, R., and Sudek, V. (1993) *J. Magn. Reson., Ser A 102*, 241−245.
32. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
33. Hwang, T. L., and Shaka, A. J. J. (1995) *J. Magn. Resn. ser A 112*, 275−279.
34. Bartels, C., Xia, T., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biom. NMR 5*, 1−10.
35. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
36. Wagner, G., and Wuthrich, K. (1982) *J. Mol. Biol. 160*, 343−361.
37. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
38. Kim, Y., and Prestegard, J. H. (1989) *J. Magn. Reson. 84*, 9−13.
39. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J. Mol. Graph. 14*, 51−5, 29−32.
40. Mumenthaler, C., Guntert, P., Braun, W., and Wuthrich, K. (1997) *J. Biomol. NMR 10*, 351−362.
41. Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167−339.
42. Baxter, N. J., and Williamson, M. P. (1997) *J. Biomol. NMR 9*, 359−369.
43. Apostoluk, W., and Otlewski, J. (1998) *Proteins 32*, 459−474.
44. Bode, W., Greyling, H. J., Huber, R., Otlewski, J., and Wilusz, T. (1989) *FEBS Lett. 242*, 285−292.
45. Burley, S. K., and Petsko, G. A. (1985) *Science 229*, 23−28.
46. Werner, M. H., and Wemmer, D. E. (1991) *Biochemistry 30*, 3356−3364.
47. McBride, J. D., Brauer, A. B., Nievo, M., and Leatherbarrow, R. J. (1998) *J. Mol. Biol. 282*, 447−458.
48. Gueven, N., Dittmann, K., Mayer, C., and Rodemann, H. P. (1998) *Cancer Lett. 125*, 77−82.
49. Dittmann, K. H., Gueven, N., Mayer, C., and Rodemann, H. P. (2001) *Protein Eng. 14*, 157−160.

BI020576W