Eur. J. Biochem. **269**, 527–537 (2002) © FEBS 2002

Comparative structure analysis of proteinase inhibitors from the desert locust, *Schistocerca gregaria*

Zoltán Gáspári[1], András Patthy[2], László Gráf[3] and András Perczel[1]

[1]*Department of Organic Chemistry, Eötvös L University, Budapest, Hungary;* [2]*Agricultural Biotechnology Center, Gödöllõ, Hungary;* [3]*Department of Biochemistry, Eötvös L. University, Budapest, Hungary*

The solution structure of three small serine proteinase inhibitors, two natural and one engineered protein, SGCI (*Schistocerca gregaria* chymotrypsin inhibitor), SGCI[L30R, K31M] and SGTI (*Schistocerca gregaria* trypsin inhibitor), were determined by homonuclear NMR-spectroscopy. The molecules exhibit different specificities towards target proteinases, where SGCI is a good chymotrypsin inhibitor, its mutant is a potent trypsin inhibitor, and SGTI inhibits both proteinases weakly. Interestingly, SGTI is a much better inhibitor of insect proteinases than of the mammalian ones used in common assays. All three molecules have a similar fold composed from three antiparallel β-pleated sheets with three disulfide bridges. The proteinase binding loop has a somewhat distinct geometry in all three peptides. Moreover, the stabilization of the structure is different in SGCI and SGTI. Proton–deuterium exchange experiments are indicative of a highly rigid core in SGTI but not in SGCI. We suggest that the observed structural properties play a significant role in the specificity of these inhibitors.

Keywords: serine proteinase inhibitor; specificity; NMR structure; flexibility.

Despite of the enormous work dedicated to understanding the structure, function and specificity of canonical serine proteinase inhibitors (reviewed in [1]), there are still unanswered questions awaiting detailed investigation concerning the precise mechanism of action of these molecules. All known canonical inhibitors share at least one common structural motif, the proteinase binding loop, although their overall fold displays no similarity in different families [2]. This loop contains the scissile peptide bond to be hydrolysed by the target enzyme. Cleavage occurs between residues labeled as P1 and P1′ [3]. Unlike in substrates, the cleaved form of reversible inhibitors remains associated with the enzyme in a manner that allows re-formation of the scissile bond. This is suggested to be a key point in the mechanism of canonical serine proteinase inhibition [4].

The major target enzyme of canonical inhibitors, e.g. trypsin or chymotrypsin, is determined primarily by the amino acid at position P1, which fits into the substrate binding site of the proteinase. The side chains of the neighbouring residues, especially from P3 to P3′ can affect this interaction by making additional interactions with the enzyme (for a comprehensive study on chymotrypsin, see [5]).

In this paper, we focus on inhibitors of the 'grasshopper family', as recently named by Laskowski [1]. A small chymotrypsin-inhibiting peptide, isolated from the desert locust *Schistocerca gregaria,* SGCI (*Schistocerca gregaria* chymotrypsin inhibitor) could be converted into a potent trypsin inhibitor by engineering both residues P1 and P1′ [6]. Replacing only Leu30 with Arg at position P1 yields a weaker and less specific inhibitor. On the other hand, a related inhibitor, SGTI (*Schistocerca gregaria* chymotrypsin inhibitor) sharing 45% sequence identity with SGCI, is inherently a much less effective inhibitor of both trypsin and chymotrypsin [6] (Fig. 1). A study on the related *Locusta* peptides revealed similar results [7]. Furthermore, Gráf and colleagues [8] found that SGTI inhibits trypsins isolated from arthropods (crayfish and shrimp) much better than the bovine enzyme used for routine assays. This is further supported by results for PMP-D2 cited in [9]. Prior to our work, structures of two inhibitors of the grasshopper family were determined by NMR spectroscopy [10,11]. Both molecules (PMP-C, pars intercerebralis major peptide C and PMP-D2, pars intercerebralis major peptide D2, orthologous in this order to SGCI and SGTI) were isolated from the migratory locust *Locusta migratoria*. These small proteins (35 and 36 residues, respectively) contain a previously unknown fold among the small proteinase inhibitors. The molecules contain three twisted antiparallel β strands and three disulfide bridges (Fig. 1). The P1 residue is located near the C-terminus of the polypeptide chain between two disulfide bonds that stabilize the binding loop. The hydrophobic core is organized differently in the two molecules either by a Phe of the first or by a Trp of the third β strand. The natural form of PMP-C is fucosylated at Thr9. As revealed by an NMR study [12], the fucosylation contributes to the structural stabilization of the peptide, although the nonglycosylated form is fully active [7]. The internal dynamics of PMP-D2 was extensively studied using both theoretical and experimental methods [13]. The conclusion of these investigations was that there are slow

Correspondence to A. Perczel, Department of Organic Chemistry, Eötvös L. University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary. Fax: + 36 1 372 2620, Tel.: + 36 1 209 0555/1653, E-mail: perczel@para.chem.elte.hu

Abbreviations: SGCI, *Schistorerca gregaria* chymotrypsin inhibitor; SGTI, *Schistorerca gregaria* trypsin inhibitor; PMP-C, pars intercerebralis major peptide C; PMP-D2, pars intercerebralis major peptide D2.

(Received 2 August 2001, revised 14 November 2001, accepted 16 November 2001)

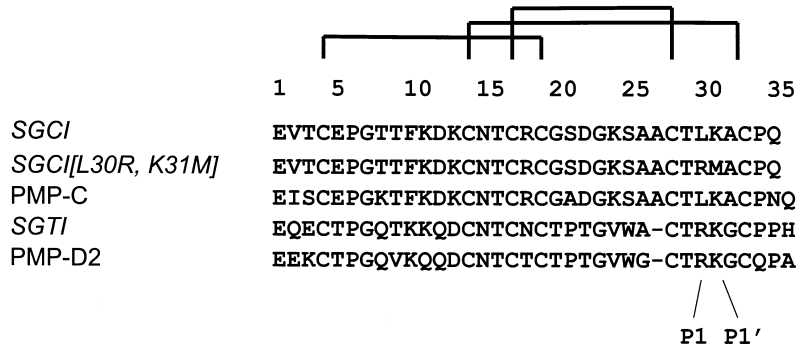

Fig. 1. Sequence alignment of SGCI, SGCI[L30R, K31M] and SGTI with members of the grasshopper inhibitor family with known structure. The conserved disulfide bonding pattern and the position of the P1 and P1′ residues is indicated. PMP-C and PMP-D2 are peptides from the migratory locust *Locusta migratoria*.

conformational motions in PMP-D2. Both peptides were shown to inhibit mammalian N-type Ca^{2+}-channels [14], providing the basis for comparing the structure of these inhibitors with ω-conotoxin GVIA [10]. Recently, Roussel *et al*. reported the crystal structure of PMP-C and PMP-D2 complexed to chymotrypsin [9].

The fold of PMP-C and PMP-D2 is in many ways similar to those of small insect toxins, defensins and some other proteinase inhibitors [15–17]. The common feature is the presence of three disulfide bridges and three β strands. However, the topology of the disulfide bonds is different, the most common arrangement being pairs 1–4, 2–5 and 3–6 (*abcabc*, according to the notation in [17]) in contrast to the pattern 1–4, 2–6, 3–5 (*abcacb*) observed in PMP-C and PMP-D2. The disulfide pattern of these inhibitors does not match the criteria established for the 'cystine knot' motif (e.g [18]). This structural motif consisting of three β strands connected by three disulfide bridges is believed to be very stable. The observed rigidity of the PMP-D2 structure [10] is consistent with these expectations.

Our goal was to determine the solution structure of SGCI and SGTI, as well as an engineered variant of SGCI that effectively inhibits trypsin (SGCI[L30R, K31M]) [6]. Homonuclear NMR as well as CD measurements were carried out, addressing the question of structural organization and stability of the peptides. The increasing number of proteinase inhibitor structures solved by NMR allows us to compare the structural organization of the investigated peptides with those described in the literature in recent years. In this paper, we attempt to interpret our results with respect to the specificity of these peptides towards proteinases.

MATERIALS AND METHODS

Sample preparation

All peptides were synthesized by solid phase synthesis on an ABI 431 A peptide synthesizer using Fmoc strategy. The purification protocol was identical to that described previously [6]. The identity of these peptides with the natural ones was confirmed by HPLC coelution analysis [6].

NMR spectroscopy

1D and 2D homonuclear NMR spectra were typically recorded at 500 MHz, at 297–303 K and pH ≈ 3. In the direct dimension the number of acquired data points was typically 2048, the number of increments was 512. Data processing was carried out by applying zero filling up to 4 K and a Kaiser window function in the t_1 and a shifted sinebell (80°) window function in the t_2 dimension. Mixing time for TOCSY experiments was 50 ms, and for NOESY measurements 80–125 ms. Spectra were referenced to internal sodium 3-(trimethyl silyl)propane-1-sulfate (DSS). Resonance assignment was carried out with the TRIAD module of the program SYBYL [19], running on SGI Octane R10000 workstations, according to the procedure described by Redfield [20] and originally established by Wüthrich [21]. Chemical shift indices for proton resonances were calculated as described by Wishart *et al*. [22].

Proton–deuterium exchange experiments were carried out for SGCI and SGTI. Samples lyophilized from H_2O were dissolved in pure (+99.9%) ^2H_2O and 1D spectra of 16 K complex points were recorded at various times after starting the experiments (at 3.5, 6.5, 12.5, 22, 30, 45, 60, 90, 120, 180, and 11280 (SGCI) and 2.5, 5, 7.5, 15, 30, 75, 100, 120 and 10980 (SGTI) min, respectively). Although the experiments were carried out in ^2H_2O, to eliminate residual water signals, a WATERGATE-type pulse sequence was applied. All data were acquired at 292 K.

Distance restraints and structure calculations

Structure calculations were solely based on NOE-derived distance restraints. Restrained distances were obtained from integration of the NOESY cross-peaks using the program module TRIAD. According to the integral values, restraints were classified into three categories corresponding to distance ranges of 1.8–2.5, 1.8–3.5 and 1.8–5.0 Å, respectively. The 1.8-Å value was chosen as a good approximation of the sum of the Van der Waals radii of two adjacent but nonoverlapping hydrogen atoms.

Structure calculations were performed with the program X-PLOR 3.851 using standard simulated annealing protocols [23,24]. High-temperature calculations were run at 1000 K for 6000 steps, with the number of cooling steps set to 3000 and the NOE scale parameter set to 50. In the course of calculations, disulfide bonds were introduced using the DISU patch available in X-PLOR.

For each molecule, an ensemble of 100 structures was calculated with X-PLOR. Accepted structures (no NOE violation greater than 0.3 Å and good geometry) were selected using the 'accept.inp' protocol. In this paper, we describe the 10 best structures for each molecule, so that structural features can be compared. Angular order parameters for the selected structures were calculated as described by Hyberts *et al*. [26] using the program MOLMOL [27].

CD spectroscopy

CD spectra for each molecule were acquired in water containing 0, 25, 50 and 75% trifluorethanol and pure trifluorethanol (SGCI and SGTI). In the case of SGCI[L30R, K31M], data were recorded in water containing 50 and 87.5% trifluorethanol and in pure trifluorethanol. Measurements were performed on a Jobin Yvon VI spectrometer at 25 °C.

RESULTS

Resonance assignment

For SGCI and its mutant, only resonances belonging to the N-terminal Glu could not be resolved. In the case of SGTI, signals of the N-terminal residue and the second of the two consecutive prolines (Pro34) escaped identification. Stereospecific assignments were not made at this stage, although where it was possible, prochiral protons were distinguished according to their position in the precalculated structures.

Comparing the NMR chemical shift values for the protons of SGCI and SGCI[L30R, K31M], the difference is the largest for the amide proton resonances (see below). Both the H^{α} and the side-chain protons can be found at rather similar chemical shift values in the two molecules. The position-specific chemical shift bias of amide protons and its significance in light of the resulting structures will be discussed below.

All chemical shift values will be deposited in the BioMagResBank database (http://www.bmrb.wisc.edu/).

NMR distance restraints

The number of NMR distance restraints obtained from NOESY spectra is summarized in Table 1 (see also Fig. 2A) It should be noted that in the case of SGCI and SGCI[L30R, K31M] the total number of restraints is over 500; this well exceeds the number of distance restraints used for the structure determination of the homologous inhibitor PMP-C [11].

Structure calculations

For structure calculation, only NMR distance restraints were used (see Materials and methods). After several exploratory runs, some stereospecific assignments could be made and subsequently ambiguous NOE peaks were identified.

The aromatic side-chain protons of Phe10 in SGCI and in SGCI[L30R, K31M] were observed at the same chemical shift value, possibly indicative of a rapidly flipping side chain [20]. Therefore, instead of two H^{δ} and two H^{ε} protons we used NMR pseudoatoms with properly modified distance restraints for structure calculations (we have obtained essentially the same structure for SGCI in calculations where a rigid Phe ring was assumed and the protons were treated separately).

In SGTI, NOEs observed between Lys10 and Trp25 were indicative of either spin diffusion or multiple conformers of one or both side chains. The corresponding distance restraints were all included but with a larger tolerance (6.5 Å).

For each molecule, a family of 100 structures was calculated with X-PLOR. In this paper, we describe and compare the 10 best structures for each molecule (RCSB PDB codes 1 kgm, 1 kio and 1 kjo for SGCI, SGCI [L30R, K31M] and SGTI, respectively). The corresponding data are summarized in Table 2.

Disulfide bridges

All three molecules contain three disulfide bridges. From sequence data it is clear that this is a conserved feature of this inhibitor family (Fig. 1), thus the disulfide pattern was expected to be essentially the same in all molecules. Indeed, our NOE data confirm that the following pairs are formed: Cys4–Cys19, Cys17–Cys28(27) and Cys14–Cys33(32) (sequence numbering in brackets account for SGTI throughout the text). H^{α}–H^{β} and H^{β}–H^{β} NOEs bearing significant information about the disulfide pairings [28] are observed between Cys14 and Cys33, as well as Cys17 and Cys28(27), respectively, in all three molecules. For SGCI and its variant, we cannot find indicative NOE peaks in the case of the Cys4–Cys19 bridge; however, the two other disulfides could be unambiguously identified. The observed NOE contacts between Phe10 and Cys4 as well as between Phe10 and Cys19 are consistent with the existence of the Cys4–Cys19 bridge. For SGTI, the Cys17–Cys27 bond was difficult to demonstrate unambiguously due to spectral overlap of both the amide and the α proton resonances of these two residues.

Overall fold of the molecules

All three inhibitors (SGCI, SGCI[L30R, K31M] and SGTI) share a common fold typical for the grasshopper inhibitor family (Fig. 3). The fold is defined by a secondary structure composed of a three-stranded antiparallel β sheet with three disulfide bridges. All three molecules have an elongated shape with the N- and C-terminus located at the opposite ends. The backbone of the N-terminal residues adopts an extended conformation and runs parallel with the loop connecting the second and third β strands, perpendicular to the strands, respectively. The cysteine residue closest to the N-terminus, Cys4, is connected to Cys19, which lies in the second β strand. The P1 and P1′ residues are located near

Table 1. Number of NMR-based distance restraints used for structure calculations of SGCI, SGCI[L30R, K31M] and SGTI. The corresponding values for the homologuous molecules PMP-C [10] and PMP-D2 [11] are also given.

	SGCI	SGCI[L30R, K31M]	PMP-C	SGTI	PMP-D2
Intraresidual	227	214	20	138	10
Sequential	149	142	66	81	93
Long-range	150	170	78	123	119
Total	526	526	164	322	222

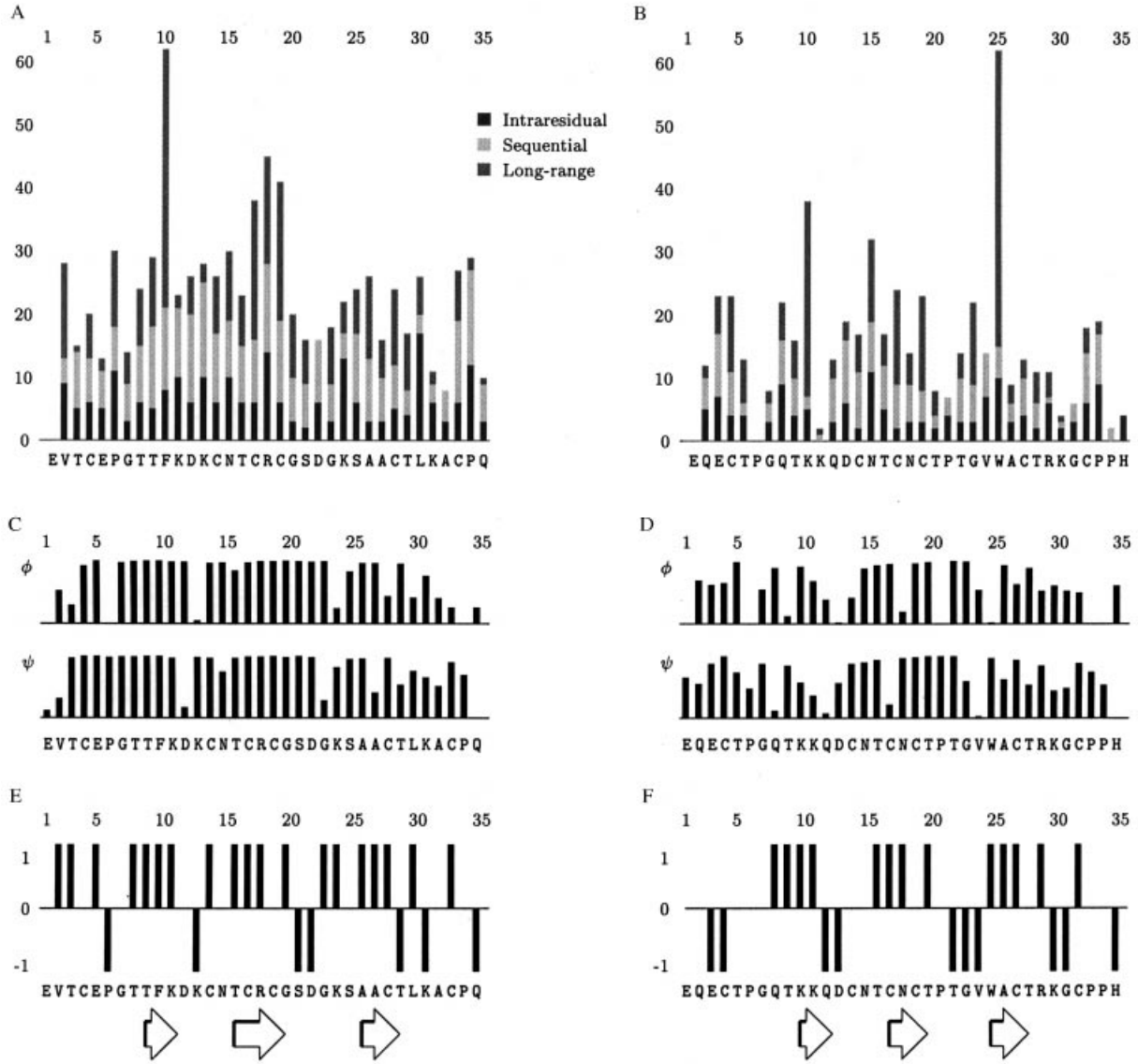

Fig. 2. Number of intraresidual, sequential, and long-range (including medium-range) NOE distance restraints per residue for SGCI (A) and SGTI (B); angular order parameters [26] of ϕ and ψ torsion angles of SGCI (C) and SGTI (D) and chemical shift indices [22] for SGCI (E) and SGTI (F). Positions of the β strands are indicated by arrows.

the C-terminus between two disulfide bonds involving the P3 and P3′ residues [Cys28(27) and Cys33(32)] and their pairs (Cys17 and Cys14).

Rmsd data for the superimposed structures of SGCI and its variant are summarized in Table 2. The best rmsd values can be observed when only the β-pleated regions, residues 9–11, 16–19 and 26–28, respectively, are compared (backbone rmsd: 0.39 ± 0.11 Å). In this view the two terminal regions and the loops interconnecting the strands show a somewhat disordered structure compared to the β strands. For SGCI[L30R, K31M], however, the binding loop region is less well defined than in SGCI and can be superimposed with an rmsd value greater than 1 Å.

For SGTI, both the rmsd values (Table 2) and the angular order parameters (Fig. 2D) indicate a less well defined overall structure than for SGCI. This can be attributed to the lower number of NMR distance restraints obtained for this molecule. In our view, this fact (i.e. the lower number of distance restraints) should not be regarded as indication of a more flexible structure of SGTI.

Detailed structural features

In SGCI and its mutant, the three β strands comprise residues 9–10, 16–19, and 26–28, respectively, in good agreement with the chemical shift indices (Fig. 2E) and NOE data. The case is similar for SGTI (Fig. 2F). As also suggested from the NOE pattern, residues 5–8 form a type II β turn [21] in all three molecules.

All three inhibitors possess a structural core reminiscent of the hydrophobic core of larger proteins. In SGCI and its variant, the core is organized around the aromatic side chain of Phe10. Residues giving side-chain NOEs to the aromatic ring comprise Val2, Cys4, Thr8, Asp12, Cys17, Cys19, and Ala26, respectively. The aromatic ring is almost completely buried.

Table 2. Rmsd values of the 10 best conformers of SGCI, SGCI[L30R, K31M] and SGTI. rmsd values are given in Å.

	Backbone	Heavy
SGCI		
4–33	0.76 ± 0.17	1.35 ± 0.23
4–28	0.55 ± 0.11	1.18 ± 0.20
9–11, 16–19, 26–28[a]	0.39 ± 0.11	0.94 ± 0.19
28–33[b]	0.78 ± 0.23	1.35 ± 0.33
SGCI[L30R, K31M]		
4–33	1.03 ± 0.31	2.11 ± 0.51
4–28	0.68 ± 0.14	1.43 ± 0.29
9–11, 16–19, 26–28[a]	0.51 ± 0.12	1,14 ± 0.23
28–33[b]	1.22 ± 0.48	3.04 ± 1.07
SGTI		
4–32	1.22 ± 0.25	1.93 ± 0.36
4–27	0.94 ± 0.20	1.56 ± 0.27
9–11, 16–19, 25–27[a]	0.67 ± 0.14	1,31 ± 0.28
27–32[b]	1.02 ± 0.43	2.07 ± 0.64

[a] β sheet regions; [b] binding loop region.

In SGTI, the key residue is Trp25 with its side-chain indole ring. Trp25 is located in the third β strand, in contrast to Phe10 of SGCI, which lies in the first one. The ring is involved in an 'intimate' interaction with the side chain of Lys10. The unusual chemical shift values of the side-chain protons of Lys10 are consistent with the vicinity of the aromatic ring. Other residues giving NOE peaks to the indole ring are Glu3, Cys4, Gln8, Thr9, Cys17 and Cys19. The side chain of Trp reaches to the α proton of Thr9 (two NOEs observed), thus bridging all three β strands, respectively.

The loop interconnecting strands 2 and 3, comprising residues 20–25 (called the 20–25 loop below) is located near the N-terminus of the molecules. In addition to the adjacent Cys4–Cys19 disulfide, a number of NOE peaks are observed, e.g. betweenVal2 and Lys24, Ser25 and Ala26, Cys4 and Lys24, Pro6 and Gly23 in SGCI, Glu3 and Gly23, Cys4 and Gly23, Thr5 and Gly23 in SGTI. Although this loop is one residue longer in SGCI than in SGTI, its trace (the α-carbons) of this region can be superimposed well in the two structures except for Lys24 in SGCI. It should be noted, however, that the orientation of the interacting N-terminal region is different in the two molecules (Fig. 5).

Conformation of the binding loop

The proteinase-binding loop is located near the C-terminus of the molecules flanked by two disulfide bridges. The corresponding NOE pattern as well as ϕ and ψ values are shown in Tables 3 and 4. Values for SGCI deviate from the canonical values and also from values observed for PMP-C [11]. However, the trace (α-carbons) of SGCI and that of the canonical inhibitor turkey ovomucoid third domain (RCSB PDB accession no. 1cho [29]), and the Bowman–Birk inhibitor (RCSB PDB accession no. 1bbi [30]), can be superimposed well (Fig. 4).

NOE peaks observed between the 11–16 loop and the binding loop are different for each observed molecule. (Table 3). It is remarkable that the SGCI variant [L30R, K31M] displays a different NOE pattern compared to SGCI. In the case of the former molecule, the number of NOEs is greater, though most of the 'extra' peaks can be classified as weak. The binding loop of SGCI[L30R, K31M] is disordered compared to the corresponding segment in SGCI. However, this is not the only difference between the binding loops of the two molecules: in SGCI[L30R, K31M], the average conformation of the loop is clearly altered in comparison to the wild-type SGCI (Fig. 5A).

In SGTI, we have observed a slightly different organization of the binding loop (Fig. 5B). This region is less well defined than that in SGCI. This is probably due to the lower number of distance restraints obtained for SGTI. The NOE peaks between residues of the binding loop and that of the 11–16 loop are also different from that of SGCI and its variant (Table 3).

Qualitative evaluation of H–D exchange data

Proton–deuterium exchange experiments can provide valuable information about the hydrogen-bonding pattern and internal dynamics of the inhibitors [21]. In SGCI, the amide protons of Gly7, Thr8, Phe10 Thr16, Cys17, Cys19 and Gly20 are exchanging slowly. In SGTI, clearly distinguishable residues with slowly exchanging amide protons are Thr16, Cys19, Thr20, Val24. Ala26 and Cys27. These data together with the calculated structures are indicative of a number of intersheet hydrogen bonds in both of the molecules. The exchanging properties of the amide protons of SGCI[L30R, K31M] are very similar to that of SGCI.

Both the amide proton of Lys13 of SGCI and of SGCI[L30R, K31M] variant and Asp13 of SGTI are quicklyexchanging, the corresponding signal can be hardly observed even after 3 min of incubation in ^2H_2O (see Materials and methods), indicating that this residue is exposed to solvent.

This experiment revealed striking difference in the structural flexibility of SGCI and SGTI. There are amide protons of SGTI that are not exchanging even after several days of incubation, while SGCI does not show such behavior. Residues displaying extreme resistance to H–D exchange comprise Thr16, Cys17, Asn18, Cys19, Thr20, Val24 and Cys27, respectively. Four of the listed residues are located in the second β strand, indicating the presence of a rigid structural core of this molecule.

CD spectroscopy

The recorded CD spectra of the molecules show little change upon altering the ratio of trifluorethanol, suggesting a compact structure for these systems, with resistance to change of the chemical environment. The spectra of SGCI and SGCI[L30R, K31M] are typical for a peptide with high β sheet content. In contrast to this, SGTI was found to show an uncommon CD spectra containing a large negative peak at 202 nm and a positive maximum at 227 nm. A highly similar spectrum was recorded for PMP-D2 and was explained on the basis of the presence of the Lys10–Trp25 interaction [10].

DISCUSSION

Structural comparison of SGCI and SGTI

Analyzing the individual NMR structures of SGCI and SGTI, it is found that they exhibit different structural

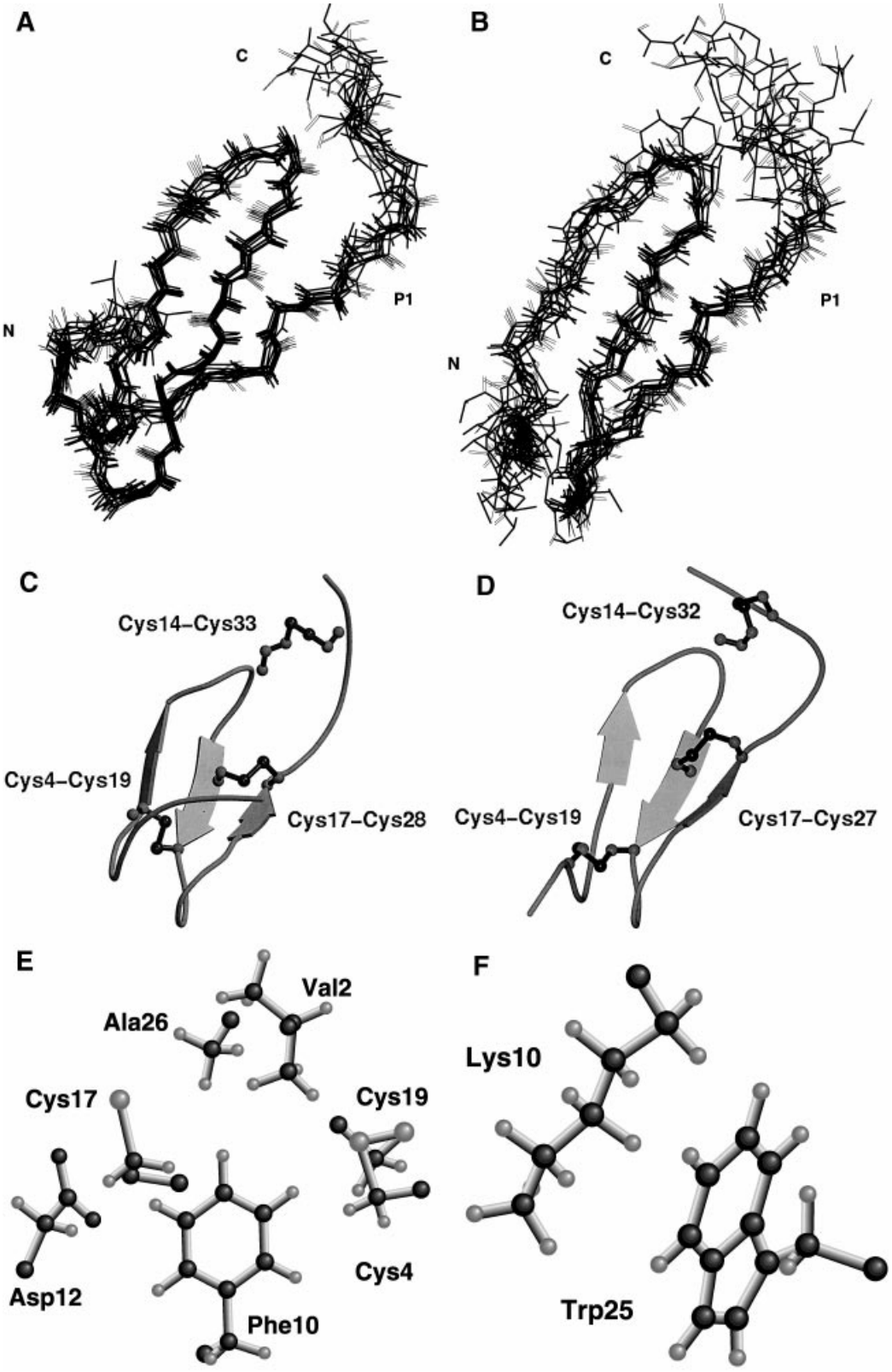

Fig. 3. Backbone of 10 superimposed structures of SGCI (A) and 10 superimposed structures of SGTI (B). Schematic representation of the secondary structure elements and the disulfide bonding pattern of SGCI (C) and SGTI (D). Residues giving sidechain NOEs to Phe10 in SGCI (E) and the Lys10–Trp25 interaction of SGTI (F). (A), (B), (E) and (F) were prepared with SYBYL [19], (C) and (D) with MOLSCRIPT [38] and RASTER3D [39].

Table 3. Observed NOEs between the binding loop (residues 28–33 or 27–32) and the 14–16 region of SGCI, SGCI[L30R, K31M] and SGTI. Notation of NOEs is as follows: first atoms correspond to residues in columns, second atoms to residues in rows. The number of NOEs, where it is greater than 1, is showed in parentheses.

	28	29	30	31	32
SGCI		**Thr**	**Leu**	**Lys**	**Ala**
Cys14					
Asn15		Hβ-HN (2), Hα-Hγ	Hβ-Hα (2)	Hδ-HN (2)	
Thr16		HN-HN, Hβ-HN, Hγ-HN, Hβ-Hγ, HN-Hγ			
SGCI[L30K, K31M]		**Thr**	**Arg**	**Met**	**Ala**
Cys14					HN-HN
Asn15		Hβ-HN (2)	Hβ-Hα (2), Hβ-HN (2), Hδ-HN (2)	Hδ-HN (2), Hβ-HN (2), Hδ-Hβ (4)	Hδ-HN (2), Hβ-HN (2)
Thr16		HN-HN, Hβ-HN, Hγ-HN, Hβ-Hγ, HN-Hγ			
SGTI	**Thr**	**Arg**	**Lys**	**Ala**	
Cys14					
Asn15	Hβ-HN (2), Hα-HN	Hβ-Hα (1), Hδ-Hβ, Hδ-Hα	Hδ-HN		
Thr16	HN-HN, Hβ-HN				

Table 4. φ and ψ torsion angles in the binding loop in the molecules SGCI, SGCI[l30R, K31R] and SGTI. For comparison, the corresponding values of PMP-C [11] and chymotrypsin inhibitor 2 [37], and six torsion angles in one of the most well-defined regions of SGCI (Cys17-Cys19) are also given.

	P3		P2		P1	
Molecule	φ	ψ	φ	ψ	φ	ψ
SGCI	49.5 ± 85.3	117.2 ± 15.7	− 70.6 ± 21.1	15.6 ± 62.6	− 11.8 ± 71.7	118.2 ± 45.3
SGCI[L30K, K31M]	55.9 ± 109.9	140.6 ± 6.5	− 56.7 ± 13.4	− 51.4 ± 31.7	− 150.5 ± 87.7	− 175.3 ± 99.4
SGTI	54.3 ± 70.5	123.4 ± 22.0	− 102.0 ± 29.8	142.0 ± 73.9	− 116.7 ± 86.1	126.2 ± 36.4
PMP-C	−127 ± 4	−175.5 ± 3	−141 ± 3	163 ± 3	−110 ± 6	167 ± 5
CI2	−115.1 ± 14.1	−122.8 ± 56.5	−84.9 ± 9.1	122.8 ± 46.0	93.2 ± 8.8	54.6 ± 86.8

	P1′		P2′		P3′	
Molecule	φ	ψ	φ	ψ	φ	ψ
SGCI	−123.5 ± 62.4	3.7 ± 52.6	152.9 ± 72.1	−179.4 ± 68.2	174.6 ± 86.4	126.4 ± 28.1
SGCI[L30K, K31M]	−140.4 ± 77.8	−13.9. ± 79.8	179.3 ± 63.4	114.5 ± 60.7	−114.5 ± 60.7	127.4 ± 34.2
SGTI	−173.3 ± 65.2.	−24.1 ± 79.6	163.7 ± 65.8	151.8 ± 76.6	−123.6 ± 87.8	125.9 ± 30.1
PMP-C	−165 ± 3	182 ± 8	−48 ± 9	151 ± 3	−130 ± 13	117 ± 19
CI2	−78.0 ± 8.6	−166.2 ± 37.3	−123.0 ± 19.3	87.0 ± 27.4	−93.4 ± 10.3	65.4 ± 11.1

	Cys17		Arg18		Cys19	
Molecule	φ	ψ	φ	ψ	φ	ψ
SGCI	−135.5 ± 16.9	106.0 ± 11.1	−68.7 ± 9.4	123.1 ± 11.3	−66.9 ± 12.2	1198 ± 7.2

features. As noted for PMP-C and PMP-D2, the two types of inhibitors share a common backbone fold stabilized in a molecule-specific manner. The aromatic residues are located in different β strands and display a different network of interaction. The different organization of the hydrophobic core together with the one-residue deletion from the 20–25 loop results in a similar overall fold with clearly distinctive structural features (Fig. 5B). Furthermore, the results of the H–D exchange experiment suggest a significant difference in the internal flexibility of the molecules. The rigid nature of the second β strand is consistent with the presence of the indole ring of Trp25 bridging strands 1, 2 and 3. It should be

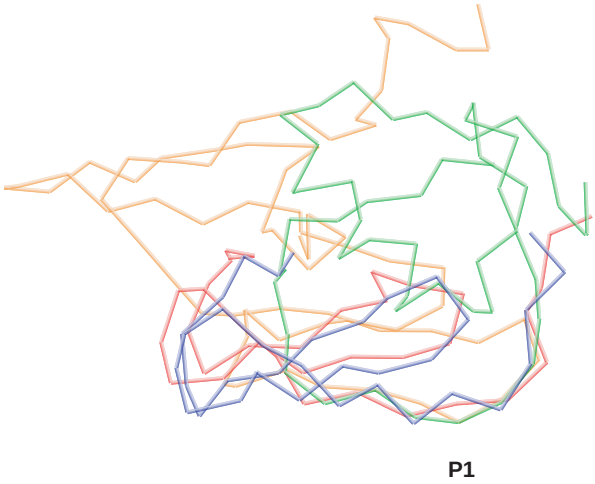

Fig. 4. Comparison of the Hα-trace of the binding loop of SGCI (blue), PMP-C (red), the turkey ovomucoid third domain inhibitor [29] (green) and the Bowman-Birk inhibitor [30] (orange). Structures are superimposed between residues P3 and P3′. This figure was prepared with SYBYL [19].

noted here that the natural form of SGCI is fucosylated on Thr9 and the homologous protein PMP-C has an altered internal dynamics upon fucosylation [12].

The most intriguing difference between SGCI and SGTI is the differing conformation of their binding loops. The corresponding NOE pattern of SGTI is different from that of SGCI as well as from that recorded for SGCI[L30R, K31M]. The number of observed NOEs for this region is lowest for SGTI (in agreement with the lowest number of total distance restraints for SGTI). The average conformation of the loop is somewhat different from that of the corresponding region in SGCI.

The data presented here concerning the difference in the internal flexibility of the two molecules are in agreement with FT-IR experiments (carried out at different temperatures) reported previously [8].

Effect of the [L30R, K31M] mutation on the structure of SGCI

The observed chemical shift values are highly similar for SGCI and SGCI[L30R, K31M], with the NH protons showing the greatest bias (Fig. 6). For 18 residues out of the 30 that could be compared (the N-terminal Glu, the prolines and the mutated P1–P1′ residues are excluded), only few exhibit greater bias than 0.3 p.p.m., located in the 11–16 loop and and in the 20–25 loop, as well as in the P2′ and P3′ positions of the binding loop. The residues with the greatest $\Delta\delta_{NH}$ values are Thr16 and, more interestingly, Gln35. The majority of the residues affected, including Thr16 and the other residues in the 11–16 loop, are in structural vicinity of the engineered P1–P1′ positions. For the observed $\Delta\delta_{NH}$ of Gln35 we have no plausible explanation.

In the observed NOE pattern of the two molecules, there are generally only a few differences, the majority of which do not reflect significant conformational alterations. However, in the structural vicinity of the mutated positions, the NOE pattern is clearly different: the number of NOE peaks observed between the 11–16 loop and the binding loop is greater for SGCI[L30R, K31M] than for SGCI (Table 3).

Thus, the overall fold of SGCI and its mutant is highly similar but the conformation of the binding loop is slightly different in the two molecules (Fig. 5A), and this part of the molecule seems less well defined in the variant molecule than in SGCI (see the rmsd values in Table 2).

Locusta and *Schistocerca* inhibitors

Previously published structures of the grasshopper inhibitor family comprise PMP-C and PMP-D2, orthologous to SGCI and SGTI, respectively. The overall fold of the *Locusta* and *Schistocerca* inhibitors is essentially the same, as expected from sequence similarity (Fig. 1). However, comparing the 36 structures of PMP-C available from the RSCB PDB (accession no. 1pmc) and the 10 structures of SGCI reported here (Fig. 5C), there are some small structural differences. The most important is the different orientation of the binding loop with respect to the rest of the molecule, although both the 11–16 loop and the binding loop can be fitted individually (backbone rmsd of the average structures is 1.02 Å for residues 28–33 and 0.55 Å for residues 11–16). This bias may be attributed either to the different approaches (the lack of incorporation of dihedral angle constraints but using more NOE distance restraints in this study) or slightly different experimental conditions.

Comparison of the structures of PMP-D2 and SGTI cannot be discussed in detail because there is no deposited structure of PMP-D2 in the RCSB PDB. The chemical shift values reported for PMP-D2 [10] are consistent with our results. The Glu2–Lys10 salt bridge reported for PMP-D2 [11] is likely to be present in SGTI; however, without pH titration experiments, we were unable to show evidence for the Lys10–Asp13 bridge.

Structure, mechanism and specificity of the inhibitors

As reported previously, SGCI, a potent inhibitor of chymotrypsin, can be converted into a powerful trypsin inhibitor simply by replacing the P1 and P1′ residues. On the other hand, SGTI is only a moderate inhibitor of both mammalian trypsin and chymotrypsin [6]. However, as observed recently [8], SGTI turns out to be a much better inhibitor of arthropodal trypsins than of the bovine enzyme used for routine assays. These observations raise the question of the universal nature of serine proteinase structure among animals and particularly, mechanism of inhibition by natural protein inhibitors.

In the light of the solution structures reported here, we attempt to address the question of taxon-specific inhibitory action from the inhibitor side. The binding loops of each of the molecules described in this paper exhibit a different amino-acid sequence, average conformation and degree of stabilization by interactions with the core of the molecule. Although we only have indirect evidence concerning the flexibility of the binding loop of the inhibitors, we suggest that the mobility of this region plays a key role in the mechanism of inhibition. An increasing amount of structural data indicates that a flexible proteinase binding loop is not an uncommon feature of proteinase inhibitors. The reported solution structure of the *Cucurbita maxima*

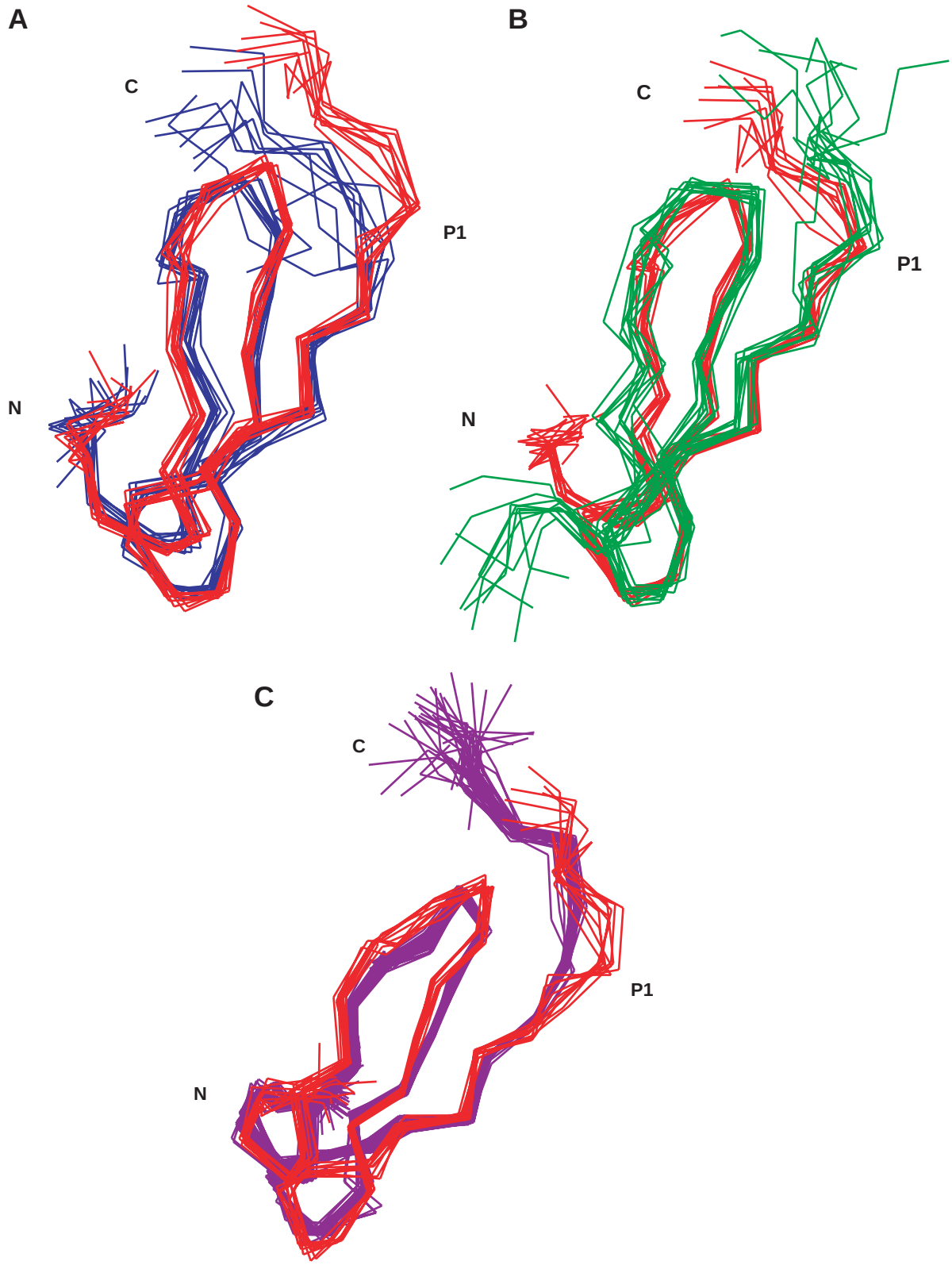

Fig. 5. Comparison of SGCI (10 structures, red lines) and of SGCI[L30R, K31M] (10 structures, blue lines; A), SGCI and SGTI (10 structures, gray lines; B) as well as SGCI and PMP-C (36 structures, purple lines; C).

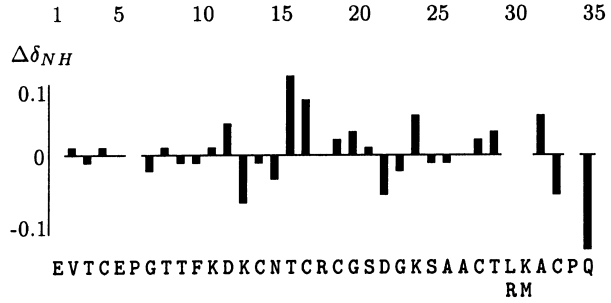

Fig. 6. Comparison of the NH chemical shift values of SGCI and SGCI[L30R, K31M]. The $\Delta\delta_{NH}$ values are indicated relative to the chemical shift values of SGCI.

trypsin inhibitor V (CMTI-V) [31] and R-elafin [32], as well as experimental and theoretical studies of backbone dynamics of a recombinant form of CMTI-V [33,34], CMTI-III [34] and chymotrypsin inhibitor 2 [35,36] provide the evidence for various serine proteinase inhibitors. Thus, our results indicating the somewhat increased flexibility of the proteinase-binding loop compared to the rest of the molecules agree well with the findings obtained from recently determined NMR structures of other proteinase inhibitors.

Based on the three structures reported in this paper we suggest a simple model explaining the inhibitory properties of the peptides. We believe that the following two key points should be considered: first, the organization of the binding loop is slightly different in the 'SGCI-type' molecules (SGCI and SGCI[L30R, K31M]) and SGTI, and secondly that the overall flexibility of SGTI is restricted in comparison to the other two peptides.

While this work was in progress, the crystal structures of PMP-C and PMP-D2 complexed to chymotrypsyn were solved [9]. Phylum selectivity of the inhibitor PMP-D2, similarly to SGTI, has also been demonstrated. The conformation of the binding loop of PMP-C has a somewhat altered conformation in the complex compared to that found in solution. Moreover, PMP-C and PMP-D2 show different interaction with the enzyme: while PMP-C establishes contacts only with its binding loop region, PMP-D2 makes additional contacts with residues in the 20–25 loop, respectively. The authors suppose that the latter interaction is responsible for the phylum selectivity of PMP-D2. To quantitavely address the question of the role of internal flexibility of the inhibitors will need ^{15}N relaxation experiments.

CONCLUSION

We have determined the solution structure of SGCI, SGCI[L30R, K31M] and SGTI by NMR spectroscopy. Our results indicate that the proteinase binding loop of these peptides is more flexible than the rest of the molecules. We have observed a significant difference in the internal dynamics of SGCI and SGTI revealed by H–D exchange experiments. We suggest that the differences between the molecules concerning the conformation and equally importantly, the flexibility of the binding loop can, at least in part, explain the intriguing species specificity of these small serine proteinase inhibitors.

ACKNOWLEDGEMENTS

Grants of the Hungarian Scientific Research Foundation (OTKA T032486 and T030841) are acknowledged. The authors also thank Antal Lopata, Chemicro Ltd. and Tripos, Inc. for their valuable help with obtaining and using SYBYL, Prof Iain D. Campbell for the opportunity offered to us to record the preliminary NMR experiments, Imre Jákli for his efforts dedicated to establishing the computational background of this study, Dr Sándor Pongor for providing us their paper prior publication, and Bence Asbóth and Gregory A. Chass for their help with preparing the manuscript. The useful suggestions of the anonymous referees are also welcome.

REFERENCES

1. Laskowski, M. & Qasim, M.A. (2000) What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme–substrate complexes? *Biochim. Biophys. Acta* **1477**, 324–337.
2. Bode, W. & Huber, R. (1991) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.
3. Schechter, I. & Berger, A. (1967) On the size of the active site in proteases I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162.
4. Longstaff, C., Campbell, A.F. & Fersht, A.R. (1990) Recombinant chymotrypsin inhibitor 2: expression, kinetic analysis of inhibition with alpha-chymotrypsin and wild-type and mutant subtilisin BPN′, and protein engineering to investigate inhibitory specificity and mechanism. *Biochemistry* **29**, 7339–7347.
5. Schellenberger, V., Braune, K., Hofmann, M.-J. & Jakube, H.-D. (1991) The specificity of chymotrypsin: a statistical analysis of hydrolysis data. *Eur. J. Biochem.* **199**, 623–636.
6. Malik, Z., Amir, S., Pál, G., Buzás, Zs. Várallyay, É., Antal, J., Szilágyi, Z.-, Vékey, K., Asbóth, B., Patthy, A. & Gráf, L. (1999) Proteinase inhibitors from desert locust, Schistocerca gregaria: engineering of both P1 and P1′ residues converts a potent chymotrypsin inhibitor to a potent trypsin inhibitor. *Biochim. Biophys. Acta.* **1434**, 143–150.
7. Kellenberger, C., Boudier, C., Bermudez, I., Bieth, J.G., Luu, B. & Hietter, H. (1995) Serine proteinase inhibition by insect peptides containing a cysteine knot and a triple-stranded β-sheet. *J. Biol. Chem.* **270**, 25514–25519.
8. Patthy, A., Amir, S., Malik, Z., Bódi, Á., Kardos, J., Asbóth, B. & Gráf, L. Remarkable phylum selectivity of a *Schisotcerca gregaria* trypsin inhibitor: the possible role of enzyme–inhibitor flexibility. *Arch. Biochem. Biophys.* in press.
9. Roussel, A., Mathieu, M., Dobbhs, A., Luu, B., Cambillau, C. & Kellenberger, C. (2001) Complexation of two proteic insect inhibitors to chymotrypsin's active site suggests decoupled roles for binding and selectivity. *J. Biol. Chem.* **276**, 38893–38898.
10. Mer, G., Kellenberger, C., Koehl, P., Stote, R., Sorokine, O., Van Dorsselaer, A., Luu, B., Hietter, H. & Lefèvre, J.-F. (1994) Solution structure of PMP-D2, a 35-residue peptide isolated from the insect *Locusta migratoria. Biochemistry* **33**, 154397–115407.
11. Mer, G., Hietter, H., Kellenberger, C., Renatus, M., Luu, B. & Lefèvre, J.-F. (1996) Solution structure of PMP-C: a new fold in the group of small serine proteinase inhibitors. *J. Mol Biol.* **258**, 158–171.
12. Mer, G., Hietter, H. & Lefèvre, J.-F. (1996) Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated proteinase inhibitor. *Nat. Struct. Biol.* **3**, 45–53.
13. Mer, G., Dejaegere, A., Stote, R., Kiefer, B. & Lefèvre, J.-F. (1996) Structural dynamics of PMP-D2: an experimental and theoretical study. *J. Phys. Chem.* **100**, 2667–2674.
14. Scott, R.H., Gorton, V.J., Harding, L., Patel, D., Pacey, S., Kellenberger, C., Hietter, H. & Bermudez, I. (1997) Inhibition of neuronal high voltage-activated calcium channels by insect pep-

tides: a comparison with the actions of ω-conotoxin GVIA. *Neuropharmacology* **36**, 195–208.
15. Bontems, F., Roumestand, C., Gilquin, B., Ménez, A. & Toma, F. (1991) Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. *Science* **254**, 1521–1523.
16. Pallaghy, P.K., Nielsen, K., Craik, D.J. & Norton, R.S. (1994) A common structural motif incorporating a cystine knot and a triple-stranded beta-sheet and inhibitory polypeptides. *Protein Sci.* **3**, 1833–1839.
17. Carugo, O., Lu, S., Luo, J., Gu, X., Liang, S., Strobl, S. & Pongor, S. (2001) Structural analysis of the amaranth alpha-amylase inhibitor: classification within the knottin fold superfamily and analysis of its functional flexibility. *Protein Eng.* **14**, 629–637.
18. Craik, J.D., Daly, N.L. & Waine, C. (2001) The cystine knot motif in toxins and implications for drug design. *Toxicon* **39**, 43–60.
19. Tripos Inc. (2001) *SYBYL Molecular Modelling Package*, Version 6.6. and 6.7. Tripos, Inc, St Louis, MO.
20. Redfield, C. (1993) Resonance assignment strategies for small proteins. In *NMR of Macromolecules* (Roberts, G.C.K., ed.), pp. 71–99. Oxford University Press, Oxford.
21. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
22. Wishart, D.S., Sykes, B.D. & Richards, F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651.
23. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett.* **239**, 129–136.
24. Nilges, M., Kuszewski, J. & Brunger, A.T. (1991) In: *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J.C., eds). Plenum Press, New York.
25. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bournem, P.E. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242.
26. Hyberts, S.G., Goldberg, M.S., Havel, T.F. & Wagner, G. (1992) The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.
27. ETH Zürich (2000) *MOLMOL-Molecule Analysis and Molecule Display*, Version 2K.1. Institut für Molekularbiologie und Biophysik, ETH Zürich, Spectrospin AG, Faellanden, Switzerland & Koradi, R.
28. Klaus, W., Broger, C., Gerber, P. & Senn, H. (1993) Determination of the disulfide bonding pattern in proteins by local and global analysis of nuclear magnetic resonance data. *J. Mol. Biol.* **232**, 897–906.
29. Fujinaga, M., Sielecki, A.R., Read, R.J., Ardelt, W., Laskowski, M. Jr & James, M.N.G. (1987) Crystal and molecular structures of the complex of alpha-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 angstroms resolution. *J. Mol. Biol.* **195**, 397–418.
30. Werner, M.H. & Wemmer, D. (1992) Three-dimensional structure of soybean trypsin (slash) chymotrypsin bowman-birk in solution. *Biochemistry* **31**, 999–1010.
31. Cai, M., Gong, Y., Kao, J.L.-F., K. & Krishnamoorthi, R. (1995) Three-dimensional solution structure of *Cucurbita maxima* trypsin inhibitor-V determined by NMR spectroscopy. *Biochemistry* **34**, 5201–5211.
32. Francart, C., Dauchez, M., Alix, A.J.P. & Lippens, G. (1997) Solution structure of R-elafin, a specific inhibitor of elastase. *J. Mol. Biol.* **268**, 666–677.
33. Liu, J., Prakash, O., Cai, M., Gong, Y., Huang, Y., Wen, L., Wen, J.J., Huang, J.-K. & Krishnamoorthi, R. (1996) Solution structure and backbone dynamics of recombinant Cucurbita maxima trypsin inhibitor-V determined by NMR spectroscopy. *Biochemistry* **35**, 1516–1524.
34. Liu, J., Gong, Y., Prakash, O., Wen, L., Lee, I., Huang, J.-K. & Krishnamoorthi, R. (1998) NMR studies of internal dynamics of serine proteinase protein inhibitors: binding region mobilities of intact and reactive-site hydrolized Cucurbita maxima trypsin inhibitor (CMTI)-II of the squash family and comparison with those of counterparts of CMTI of the potato I family. *Protein Sci.* **7**, 132–141.
35. Li, A. & Daggett, V. (1995) Investigation of the solution structure of chymotrypsin inhibitor 2 using molecular dynamics: comparison to X-ray crystallographic and NMR data. *Protein Eng.* **8**, 1117–1128.
36. Shaw, G.L., Davis, B., Keeler, J. & Fersht, A.R. (1995) Backbone dynamics of chymotrypsin inhibitor 2: effect of breaking the active site bond and its implications for the mechanism of inhibition of serine proteinases. *Biochemistry* **34**, 2225–2233.
37. Ludvigsen, S., Shen, H., Kjaer, M., Madsen, J.C. & Poulsen, F.M. (1991) Refinement of the three-dimensional solution structure of barley serine proteinase inhibitor 2 and comparison with the structures in crystals. *J. Mol. Biol.* **222**, 621–635.
38. Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
39. Merritt, E.A. & Bacon, D.J. (1997) Raster3D Photorealistic Molecular Graphics. *Methods Enzymol* **277**, 505–524.

SUPPLEMENTARY MATERIAL

The following material is available from http://www.ejbiochem.com

Table S1. Chemical shift values of SGCI (A), SGCI[L30R, K31M] (B) and SGTI (C) recorded in H_2O:D_2O (9 : 1) at 297 K.

Figure S1. Proton-deuterium exchange experiments of SGCI (A) and SGTI (B).

Figure S2. Circular dichroism spectra of the molecules in solvents containing different amount of trifluorethanol. SGCI (A), SGCI[L30R, K31M] (B) and SGTI (C).