the FEBS Journal

Local binding with globally distributed changes in a small protease inhibitor upon enzyme binding

Zoltán Gáspári[1], Borbála Szenthe[2], András Patthy[2], William M. Westler[3], László Gráf[2] and András Perczel[1]

1 Institute of Chemistry, Eötvös Loránd University, Budapest, Hungary
2 Institute of Biology, Eötvös Loránd University, Budapest, Hungary
3 National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, MA, USA

Keywords
enzyme–inhibitor complex; internal dynamics; NMR spectroscopy; pacifastin inhibitor family; SGCI

Correspondence
András Perczel, Eötvös Loránd University, Pázmány Péter sétány 1/A, Budapest, 1117, Hungary
E-mail: perczel@para.chem.elte.hu

(Received 2 January 2006, revised 6 February 2006, accepted 27 February 2006)

doi:10.1111/j.1742-4658.2006.05204.x

Complexation of the small serine protease inhibitor *Schistocerca gregaria* chymotrypsin inhibitor (SGCI), a member of the pacifastin inhibitor family, with bovine chymotrypsin was followed by NMR spectroscopy. ^{1}H–^{15}N correlation (HSQC) spectra of the inhibitor with increasing amounts of the enzyme reveal tight and specific binding in agreement with biochemical data. Unexpectedly, and unparalleled among canonical serine protease inhibitors, not only residues in the protease-binding loop of the inhibitor, but also some segments of it located spatially far from the substrate-binding cleft of the enzyme were affected by complexation. However, besides changes, some of the dynamical features of the free inhibitor are retained in the complex. Comparison of the free and complexed inhibitor structures revealed that most, but not all, of the observed chemical shift changes can be attributed to minor structural transitions. We suggest that the classical 'scaffold + binding loop' model of canonical inhibitors might not be fully valid for the inhibitor family studied. In our view, this feature allows for the emergence of both taxon-specific and nontaxon-specific inhibitors in this group of small proteins.

Schistocerca gregaria chymotrypsin inhibitor (SGCI) is a small, 35-residue protease inhibitor isolated from the desert locust, *Schistocerca gregaria* [1]. This molecule is a member of the pacifastin serine protease inhibitor family [2–4], the characteristic attributes of which are a well-defined secondary structure consisting of three antiparallel β sheets stabilized by three disulfide bridges [5–7], a reactive site located at the C-terminus and considerable heat stability [1,8]. In desert locust, SGCI is synthesized as part of a precursor molecule [9] that is cleaved to yield SGCI and also *Sch. gregaria* trypsin inhibitor (SGTI), a paralog of SGCI with surprising taxon specificity: this molecule is a selective inhibitor of arthropod trypsins over mammalian ones [10,11]. Recently, these two and several related inhibitors were shown to be involved in the solitary–gregarious transition of the desert locust [12,13] opening up possible new perspectives in the fight against African locust invasions.

The solution structure and internal dynamics of these two inhibitors have been determined at pH 3.0 [7,14] and it was found that, despite the similar fold, the two molecules exhibit remarkably different dynamics at multiple time scales, which was suggested to contribute to the differences in taxon specificity of SGCI and SGTI.

The specificity of the interaction of SGCI and SGTI with proteases can only be assessed by investigating the appropriate enzyme–inhibitor complexes. To date, the crystal structures of three such complexes have

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

been reported: the complex of the SGCI ortholog pars intercerebralis major peptide C (PMP-C) and a modified form of the SGTI ortholog pars intercerebralis major peptide D2 (PMP-D2) with bovine chymotrypsin [15] (PMP-C and PMP-D2, are isolated from the migratory locust *Locusta migratoria*) as well as the tight complex formed between SGTI and crayfish trypsin [11]. Detailed analysis of the interactions in the latter ('arthropod–arthropod') complex revealed the importance of an extended protease-binding site in SGTI unparalleled among canonical serine protease inhibitors [11]. Despite the known crystal structures, NMR spectroscopic measurements of the complexes are expected to yield important complementary information about the process of complex formation as well as the structural and dynamical changes of the inhibitors relative to the free state. The two possible approaches for NMR titration studies are to follow the induced changes in the isotope-labeled inhibitors using unlabeled protease or to monitor the changes in the protease using the reverse of the previous labeling scheme. The first approach proved fruitful in investigations of complexes of Kazal-type inhibitors [16,17] with proteases, and the second was shown to be feasible using selectively labeled trypsin variants and several inhibitors [18]. Detailed investigation of the internal dynamics of molecular partners in enzyme–substrate complexes in general has recently been shown to contribute to the understanding of enzymatic mechanisms [19].

In this study we report NMR titration experiments of labeled SGCI with bovine α-chymotrypsin and characterization of the complex formed including dynamical features. In addition, we also describe NMR measurements of free SGCI at pH 6.0, as this state is the starting point of the titration experiments. To interpret chemical shift changes upon titration and analyze SGCI conformation in the bound state, the crystal structure of the nearly identical ortholog PMP-C with bovine α-chymotrypsin (the same enzyme as in this study) is used.

Results

SGCI at near-neutral pH

All our previous measurements were carried out at pH 3.0 in order to suppress chemical-exchange phenomena, which are due to rapid exchange of amide protons with water. However, the natural pH of the inhibition is around pH 6, thus all titration measurements were performed in a buffered environment to ensure optimal pH. Because several resonances appear at different positions at low and near-neutral pH, resonance assignment of the free inhibitors before titration was necessary. Moreover, several resonances become unobservable or weak in the ^{1}H–^{15}N correlation (HSQC) spectra at near-neutral pH possibly indicating increased chemical exchange relative to the low-pH state (Fig. S1). The quality of the homo- and heteronuclear spectra allowed clear resonance assignment for most of the residues, but the relatively low number of NOE cross-peaks made high-precision structure determination unfeasible at pH 6.0 (Fig. 1).

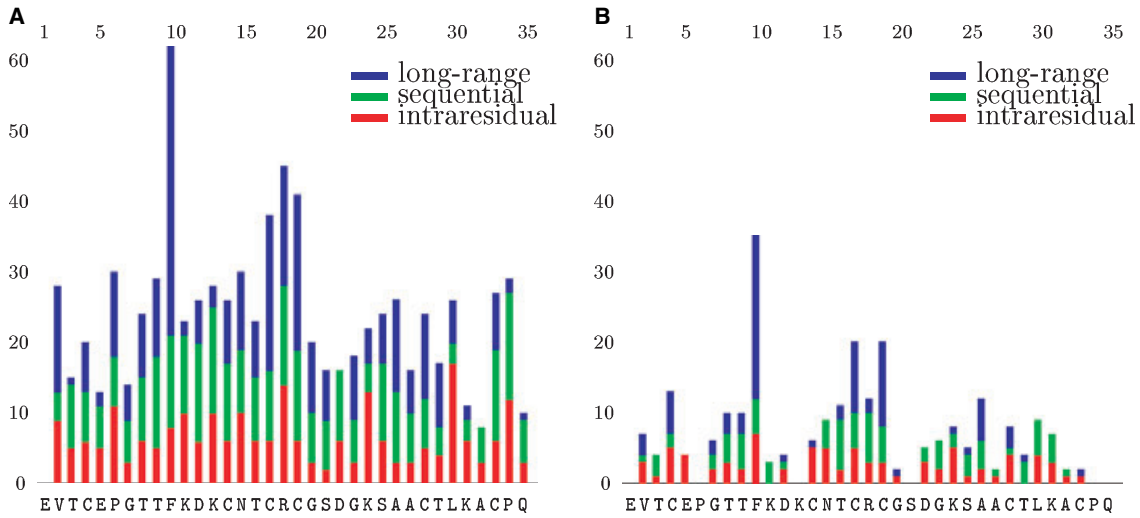

Fig. 1. Comparison of distance restraint distributions obtained for free SGCI at pH 3.0 (A) and pH 6.0 (B). Red, intraresidual restraints; green, sequential restraints; blue, long-range restraints. The total number of restraints is 526 (227 intraresidual, 149 sequential and 150 long-range) at pH 3.0 [7] and 163 (79, 37 and 47, respectively) at pH 6.0.

The collected distance restraints constitute a subset of those used for structure determination at low pH, thus, all the NOE cross-peaks observed at pH 6.0 are consistent with the published SGCI structure. The distribution of NOE-derived restraints is similar to that observed at pH 3.0 with a clear 'peak' at the hydrophobic 'core-forming' residue Phe10 (Fig. 1). Exploratory structure calculations yielded only a low-resolution structural model but confirmed the similarity of the backbone fold (data not shown). Random addition of restraints found only at pH 3.0 resulted in clear improvement of the structure, suggesting that the scarcity of NMR data is due to sample conditions (increased chemical exchange) rather than structural rearrangements.

Titration experiments

Step-by-step addition of the enzyme caused the emergence of a completely new set of resonances indicating slow exchange on the NMR time scale. The new resonance set can clearly be assigned to a single molecular species (see below). Upon complexation, several residues became unobservable in the HSQC spectra compared with the initial state. The linewidths of the peaks arising from the complex were greater than those of the uncomplexed inhibitor (linewidths for the complex were typically 25–30 Hz versus 16–19 Hz for the free inhibitor), consistent with an almost eightfold increased molecular mass of the complex over the free inhibitor (28.7 kDa for the complex versus 3.6 kDa for free SGCI).

Characterization of the complexed state

Intriguingly, amide resonances of residues in the canonical protease-binding loop (P3–P3′, Cys27–Cys32) [20] could not be identified and several clearly resolved peaks in the HSQC spectra escaped assignment. It is noteworthy that resonance assignment of the complexed state required the use of high-sensitivity spectrometers in order to gain sufficient signal-to-noise ratio in the triple-resonance experiments. The identified residues comprise a continuous segment from Gly7 to Lys24, i.e. the N-terminal and C-terminal parts of the molecule, including most of the third β strand and the full protease-binding site could not be unambiguously assigned.

Chemical shift changes upon complexation

Upon titration, the most striking feature of the emerging HSQC patterns was that almost all assigned resonances appeared in a new position compared with the uncomplexed state (free inhibitor at pH 6.0). This means that even residues far from the protease-binding site are greatly affected by complexation (Figs 2 and 3A,B). Interestingly, the least affected region is the loop between the second and the third β strands (Ser21–Ser25), which comprises the extended binding site in the related taxon-specific inhibitor SGTI. By contrast, residues in the first and second β strands (Thr9–Lys11 and Thr16–Cys19, respectively), being spatially far from the primary binding site, exhibit remarkable changes, Thr9 and Arg18 being the most prominent examples (Fig. 3B).

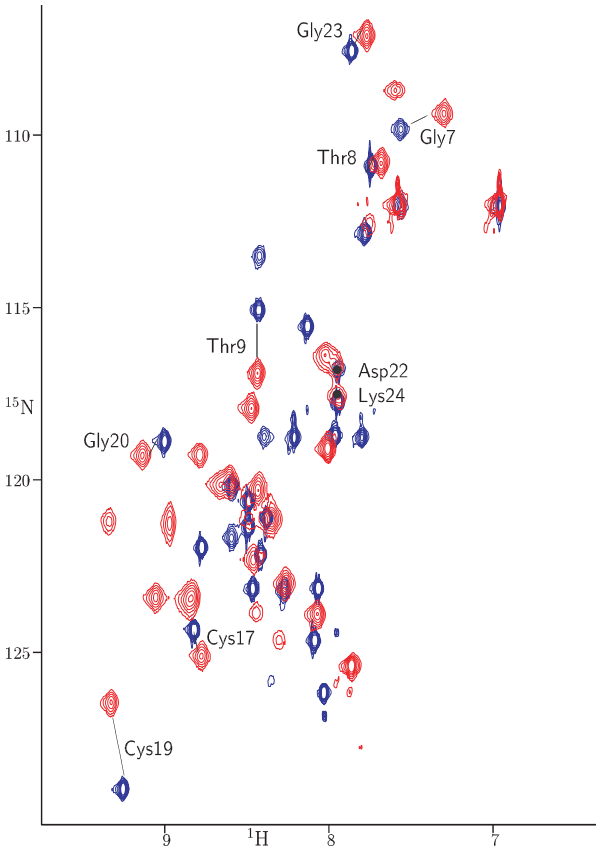

Fig. 2. Overlaid spectra of free (blue) and complexed (red) SGCI at pH 6.0 with some changed and virtually unchanged resonance peaks labeled. Figure generated with SYBYL [35].

Relaxation data

Relaxation parameters (T1, T2 and heteronuclear NOE) were measured for free and complexed SGCI at pH 6.0 and compared with the values obtained previously for free SGCI at pH 3.0 (Fig. 4). Relaxation rates for free SGCI at near-neutral pH are generally

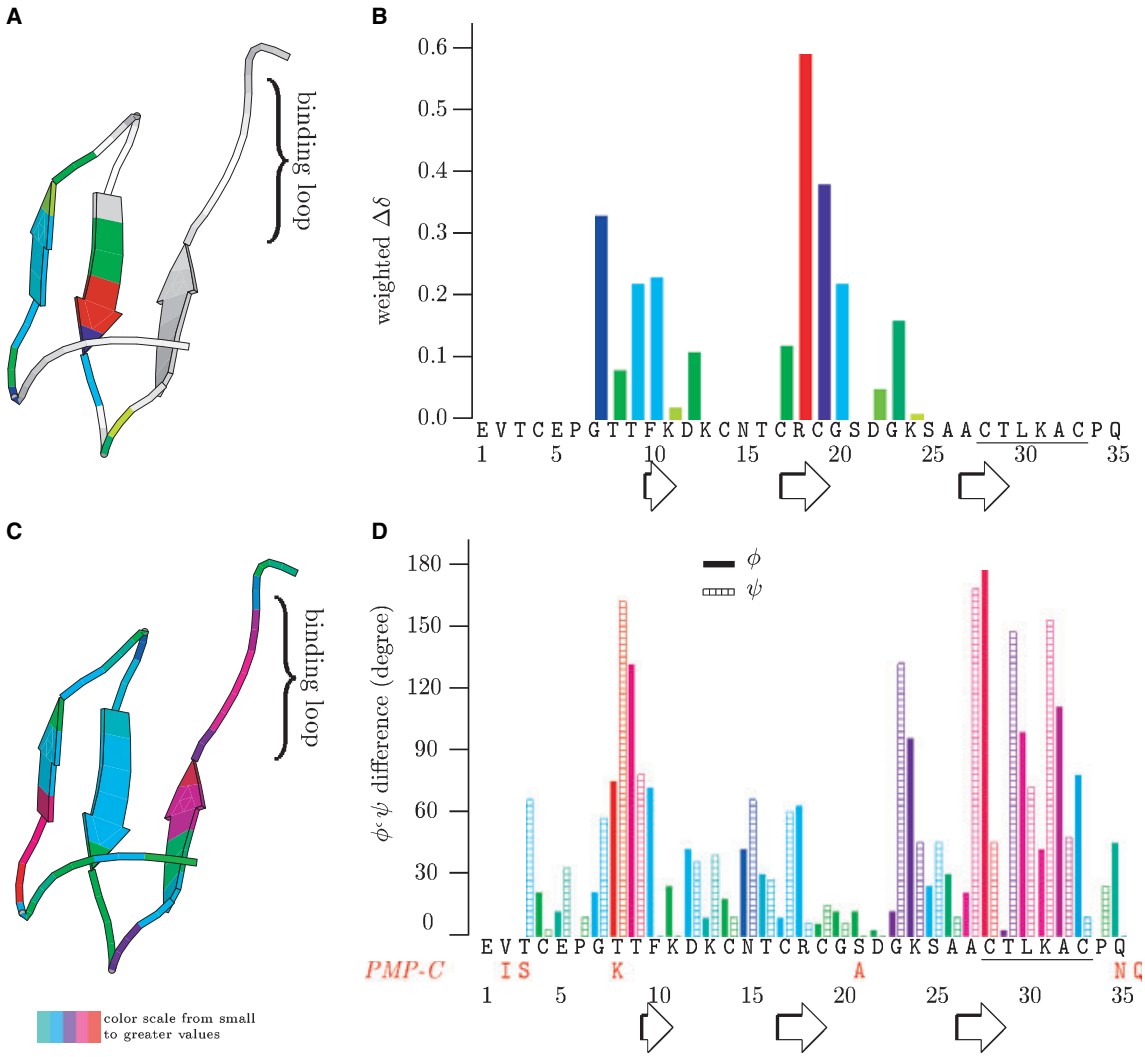

Fig. 3. (A,B) Chemical shift changes in SGCI upon complexation. Changes are indicated as weighted chemical shift differences ($\Delta\delta^1H + \Delta\delta^{15}N/6$ for glycines and $\Delta\delta^1H + \Delta\delta^{15}N/8$ for all other residues to compensate for the broader nitrogen chemical shift range) [42,43]. Residues in the structure (A) and bars (B) are color-coded according to the relative values of weighted Δδ. Position of the binding loop is indicated (underlined residues in B). Residues unambiguously assigned in both the free and complexed states are compared only. (C, D) Backbone torsion angle differences between the solution structure of SGCI and complexed PMP-C. Differences are calculated between the average values in the 10 deposited SGCI conformers (PDB ID 1KGM) and the averages of the 3 PMP-C structures in the asymmetric unit (PDB ID 1GL1). Residues (C) and bars (D) are color-coded according to the sum of ϕ and ψ differences. Note that as a residue has a single color in (C), columns for both dihedrals for each residue are colored the same irrespective of their contribution to the sum. As the conformations of different molecules are compared, amino acid substitutions in PMP-C relative to the SGCI sequence are indicated in (D). Cartoon structure representations (A) and (C) were prepared using MOLSCRIPT [44].

higher than those measured at low pH. In addition, rates show a greater deviation at pH 6.0, especially the spin-matrix rates (R1). Nevertheless, the general trend of the R2 rates is similar to that obtained at low pH (although individual values might differ). The calculated rotational correlation time ($\tau_c \approx 3$ ns) is close that calculated earlier from NMR at pH 3.0 ($\tau_c = 3.14$ ns) [14] supported by hydrodynamical calculations (2.81 ns).

For the complex, the R2 rates increase and the R1 rates decrease compared with free SGCI, in agreement with the almost eightfold increase in molecular mass [21]. The calculated correlation time is $\tau_c \approx 12$ ns, which is considerably smaller than that obtained from hydrodynamical calculations (16.4 ns, τ_c calculated for uncomplexed chymotrypsin is 13.9 ns). The discrepancy may be, at least in part, due to the insufficient sampling of relaxation parameters as data is available

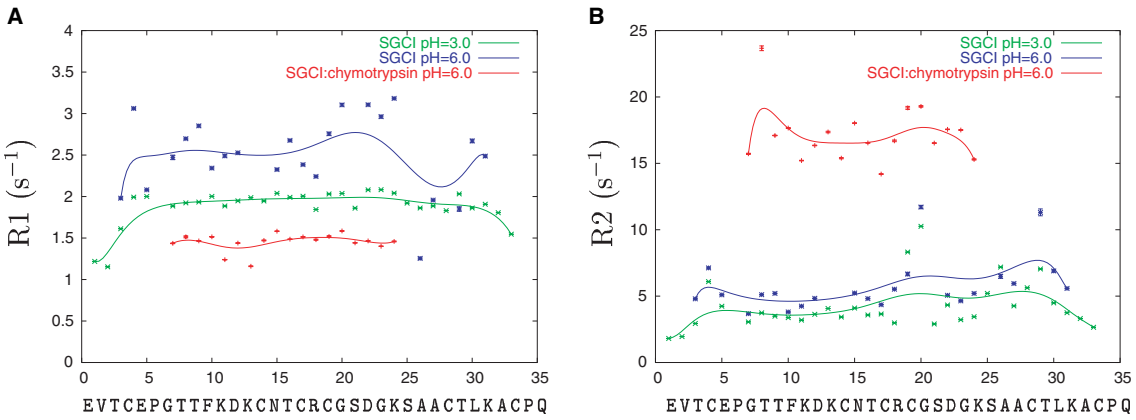

Fig. 4. R1 and R2 relaxation parameters of free SGCI at pH 3.0 and 6.0 as well as SGCI complexed with bovine chymotrypsin (green, blue and red points and lines, respectively). The lines are smoothed bezier curves intended only to guide the eye.

for only 18 of the 280 residues in the complex. Using τ_c of 16.4 ns to calculate model-free parameters yields better fit for most residues and chemical exchange (R_{ex}) should be considered for only six residues (Thr8, Thr9, Asn15, Thr16, Cys19, Gly20) compared with almost all residues when $\tau_c = 12$ ns was used. Because of the relative scarcity of underlying experimental data, these derived parameters can not be regarded as reliable and thus are not discussed further.

The trend of the R2 values can not be fully compared with those of the free states as the signals of several residues with above-average R2 values in free SGCI (Cys4, Ser25, Ala26, Ala27, Cys28, Thr29, Leu30) could not be assigned in the complex. However, R2 values for residues Cys19 and Thr20 are high (with T1/T2 nearly one standard deviation above the mean), which is also observed in the free states, especially for Thr20.

Discussion

SGCI structure at near-neutral pH

Changes in a HSQC spectrum induced by pH adjustment can generally occur for many reasons, the two most important being the changes in the exchange properties of amide protons with water and conformational rearrangements. The former is analyzed as the dynamics of the molecule is investigated at pH 6 in the free state. The data show that there are changes in the R2 rates although the general trend remains the same (correlation coefficient = 0.84). The similarity of the dynamics at low and near-neutral pH is most easily explained by assuming that conformational changes are negligible between the two states. Notably, changes in amide ^{1}H and ^{15}N shifts are, on average, about twice as small as for complexation (Fig. 5).

Structural information derived from NMR spectra recorded at near-neutral pH are consistent with the published SGCI structure, determined at pH 3.0. Structural calculations yielded ill-defined structures but with backbone fold clearly similar to the structure at low pH. Therefore, we argue that there are no significant structural changes upon elevating the pH but the scarcity of NOE data is due to increased chemical

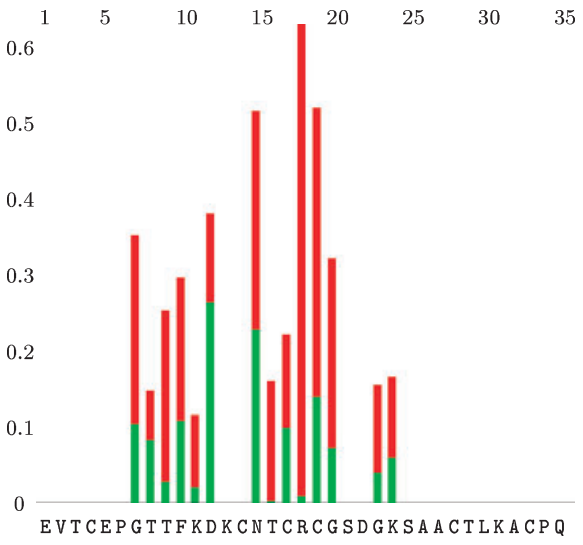

Fig. 5. Comparison of chemical shift changes of free SGCI upon pH change (green bars) and complexation (red bars). ^{1}H–^{15}N shifts for residues assignable in all three states are compared. Weights are calculated as for Fig. 3. On average, changes upon pH elevation are about twice as small as for complexation (average change 0.09 versus 0.21, respectively).

exchange. This is further supported by the observation that side-chain resonances are practically unaffected, including nonstandard shifts indicative of structural integrity (e.g. β protons of Cys17) and also chemical shift index data for the three states (Fig. S2). It should also be noted that the structure determined at low pH superimposes well with the complexed PMP-C structure and detailed investigation is needed to identify structural differences (see below). It is highly unlikely that there would be a significantly different third conformational state of free SGCI at pH 6.0 when these two are so close to each other. Nevertheless, our observations on chemical shift changes upon complexation are unaffected by the relevance of our arguments presented above (see below).

Interpretation of the titration experiments

The observed changes in the HSQC spectra of SGCI upon titration are interpreted as indicative of tight and selective binding. Tight binding is consistent with the emergence of a new set of resonances instead of a stepwise shift of peak positions. The specificity of the binding can be reasoned by the facts that: (a) the new resonance set is assignable to a single form of SGCI and no signs of other species are present in the spectra, (b) the protease-binding loop is affected by the binding (resonances for this part became unobservable), (c) aspecific binding is not expected to be tight, and (d) the crystal structure of the nearly identical PMP-C with bovine α-chymotrypsin reveals specific protease–inhibitor interaction in a system of this type. Biochemical evidence for tight binding is supported by measurements from independent laboratories (K_i values determined: SGCI–chymotrypsin, 6.2×10^{-12} mol.dm^{-3} [8]; SGCI–chymotrypsin, 3.0×10^{-10} mol.dm^{-3} [22]; PMP-C–chymotrypsin, 1.2×10^{-10} mol.dm^{-3} [23]).

We note that our observation that residues far from the binding site are affected upon complexation is independent of our speculations on the structure of SGCI at near-neutral pH. We compare only resonances clearly assignable in spectra recorded at both the start and endpoint of the titration experiments. Thus, although we argue that there are no significant structural changes in SGCI upon elevating the pH from 3.0 to 6.0 and use the structure determined at the former condition for comparison, the interpretation of chemical shift changes remains valid even if this assumption does not fully hold.

The most straightforward hypothesis based on our results is that no significant structural change occurs to SGCI on pH elevation but multiple regions are affected upon protease binding. This model is simpler than all the possible competing ones, e.g. assuming conformational rearrangement on pH elevation and a 'back-change' upon enzyme binding (chemical shift, NOE and mobility data do not support this and the close overall similarity of the free and bound conformations should be explained) or another scenario when the bound conformation would be 'preformed' during pH elevation (in this case, changes in the HSQC spectrum upon titration are hard to explain). Our proposed model is not affected by the fact that the crystal structure used for comparison is determined at pH 5.0 as it is reasonably close to the pH of our experiments and the effects of complexation are expected to be determinative compared with those of pH change. We note here that the observed spectral changes upon complexation were essentially the same in our exploratory titration experiments at pH 7.5 and 8.1, suggesting that the bound conformation is not influenced greatly by pH.

Comparison of the free and complexed inhibitors

As no structure of complexed SGCI is available, the X-ray coordinates of the PMP-C–chymotrypsin complex (PDB code 1GL1) [15] were used for comparison. This approach can be justified on the basis that PMP-C is the closest known homolog of SGCI [4] and there are only five substitutions beside a one-residue C-terminal extension in PMP-C relative to SGCI (Fig. 3D). Only two of the substitutions are not in the N- or C-terminal part. The enzymes used for complexation are the same, bovine α-chymotrypsin in both cases. Therefore, the published PMP-C–chymotrypsin structure [15] can reliably be regarded as being practically identical with the proposed SGCI–chymotrypsin complex, the molecular species present at our titration endpoint.

The structures were compared using two different methods, by backbone root mean square deviation (RMSD) values and using the backbone dihedral angles φ and ψ. Whereas the former is sensitive to conformational changes involving segments of several residues, the latter is able to detect smaller, residue-specific alterations which may average out to yield similar backbone conformation and small RMSDs. In addition, distances corresponding to the NMR restraints used for structure calculation of free SGCI [7] (available in PDB) were measured in the complexed PMP-C conformers, where appropriate (i.e. considering identical side chains only).

Backbone RMSD values were calculated for different regions of the inhibitors (Table 1) using two different approaches: first, models of complexed PMP-C

Table 1. Backbone RMSD values [Å] of free SGCI and complexed PMP-C. Values were calculated using the program MOLMOL [41] after fitting the molecules to the region considered. The representative conformers are model 5 for free SGCI (PDB ID 1KGM) and chain I for complexed PMP-C (PDB ID 1 GL1).

	Whole molecule (4–33)	Protease-binding loop (28–33)	β strands (9–11, 16–19, 26–28)	N-terminal region (3–6)	12–15 loop (12–15)	21–25 loop (21–25)
Free SGCI, 10 models	0.76 ± 0.17	0.78 ± 0.23	0.39 ± 0.11	0.22 ± 0.10	0.42 ± 0.21	0.47 ± 0.19
Free SGCI, 10 models + complexed PMP-C, three models (13 models altogether)	0.94 ± 0.25	0.86 ± 0.27	0.52 ± 0.19	0.30 ± 0.13	0.45 ± 0.20	0.50 ± 0.18
Representative models of free SGCI and complexed PMP-C (2 models altogether)	1.07	0.91	0.55	0.35	0.66	0.45

(three different conformers in the asymmetric unit of the structure 1GL1) were used to 'extend' the 10-conformer NMR ensemble of SGCI (1KGM) (Fig. 6A) yielding a 13-member 'ensemble' testing whether PMP-C would fit into the outcome of our structure calculations, and second, one conformer PMP-C (chain I in the 1 GL1 structure) was compared with the representative conformer of SGCI (model 5 in the deposited ensemble). The values obtained are not indicative of significant structural changes upon complexation. On the one hand, the RMSD intervals calculated including or excluding complexed PMP-C overlap (the ranges defined by the standard deviations have an intersection in all cases), indicating that complexed PMP-C structures fit well into the deposited 10-confomer ensemble of SGCI. On the other hand, although values calculated for the representative SGCI and PMP-C conformers are outside the RMSD interval calculated for free SGCI in four of the six cases shown (Table 1) there is a maximum deviation value of only 0.13 Å (residues 4–33). These differences are well within the range usually observed for solution and crystal structures of the same protein [24] and therefore can not be unambiguously attributed to effects of complexation.

Another comparison of the free and enzyme-bound structures can be made by comparing the backbone torsion angles in the two forms. The ϕ/ψ differences can easily be compared with the chemical shift changes of the amide NH groups (Fig. 3).The most affected regions in terms of backbone dihedral differences are the protease-binding loop and the N-terminal part of the first β strand. The alterations furthest from the binding site, in segment Gly7–Lys10, are reflected in the changes in the chemical shifts. Intriguingly, residues Arg18 and Cys19, the two with the greatest observed chemical shift changes do not undergo a conformational transition comparable with the greatest observed using either RMSD or ϕ/ψ analysis. Located in the second β strand, they are also reasonably far from the protease to exclude contact effects (Fig. 6B). Thus, there is no straightforward explanation for the chemical shift changes of these two residues. It should be noted that Cys19 and Gly20 exhibit high R2 values in free SGCI (the two highest at pH 3.0 and Gly20 the highest at pH 6.0) and also in the complex (Cys19 and Gly20 the third and second highest, respectively), suggesting that the corresponding region of the second β strand is subject to extensive motions on the μs/ms

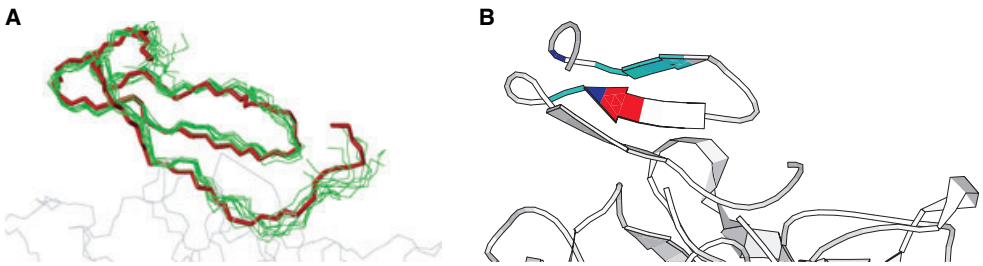

Fig. 6. (A) Comparison of free SGCI (PDB ID 1KGM, 10 conformers, thin green lines) and PMP-C (1PMC, thick red line) complexed to bovine chymotrypsin (1 GL1, thin gray line, only a part of it shown). Figure prepared using MOLMOL [41]. (B) Model of the SGCI–bovine chymotrypsin complex (only part of the protease is shown) residues with remarkable chemical shift changes in SGCI upon complexation (Arg18, Cys19, Gly20 as well as Gly7, Thr9 and Phe10) are colored (coloring scheme as for Fig. 3). Figure prepared using MOLSCRIPT [44].

time scale in all of the states investigated. Cys19 can be contrasted to Cys17, a residue exhibiting much smaller changes in chemical shifts despite undergoing minor conformational changes and being linked by a disulfide to Cys28 of the binding loop (Figs 2 and 3).

It is noteworthy that although RMSD analysis did not reveal significant structural alterations upon complexation, ϕ/ψ analysis shows differences as large as 179° (Cys28 ϕ) of backbone dihedrals in the two states (Fig. 3D, S3 and S4). The solution to this apparent contradiction lies in the relative direction of the occurring backbone dihedral rotations as they systematically compensate each other in neighboring residues resulting in a virtually unchanged main chain conformation (Fig. S3).

Analysis of NMR distance restraint violations in the complexed PMP-C structure supports the above findings. Only one backbone–backbone restraint is violated by > 0.1 Å in the complexed form, namely the one between Thr9 Hα and Cys19 Hα. This corresponds to a conformational change of Thr9 captured also by ϕ/ψ analysis. Although this particular restraint could not be derived from NOESY spectra recorded at near-neutral pH, the peak indicating spatial proximity of the γ2 methyl group of Thr9 and the amide proton of Cys19 is present at pH 3.0 and pH 6.0, and the corresponding restraint is violated in the complex structure lending support for the relevance of this conformational change.

Other violated restraints indicate changes in the protease-binding loop, the first β strand, and, not detected by the former two methods, a rotamerization of the Arg18 side chain. However, the guanidino group of this residue is pointing away from its amide NH in both conformations and can thus not be made responsible for the observed chemical shift changes in this region (Fig. 5).

The internal dynamics of the complexed inhibitor is also changed relative to the free state. The distribution of high R2 values, indicative of motions on the μs/ms time scale, is similar in free SGCI at both pH 3.0 and 6.0, affected residues mostly located in the third β strand and the loop connecting it to the second. Although some of these residues could not be assigned in the complex, it is noteworthy that in this state the residue with the highest R2 value is Thr8, indicating mobility changes in the first β strand upon complexation beside structural ones affecting the neighboring Thr9. However, as mentioned above, Cys19 and Gly20 are characterized by high R2 values in all three states investigated, suggesting that these residues exhibit similar dynamics in free and complexed SGCI, including significant motions on the μs/ms time scale. Although the significance of these motions is not yet clear, we note here that similarity of the dynamics of free and substrate-bound cyclophilin A was recently shown and there the correspondence with catalytic turnover was straightforward [19].

Implications for mechanism of inhibition

Canonical inhibitors are regarded as consisting of a 'scaffold' and a protease-binding loop which have highly similar conformations, even between unrelated molecules [25,26]. In most inhibitor families studied, the properties of the binding loop turned out to be sufficient to interpret even diverse biological activities of these proteins. NMR titration studies of Kazal-type inhibitors supported this view as only residues in the protease-binding loop and its spatial vicinity were affected upon complexation. Here we show that, for SGCI, a member of the pacifastin inhibitor family, complexation results in significant alterations even in regions far from the binding site. The observed changes differ from those reported for the taxon-specific subgroup of this inhibitor family, where, as judged by the crystal structures of the PMP-D2v–bovine chymotrypsin and the SGTI–crayfish trypsin complexes, an extended protease-binding site is responsible for the increased strength of the interaction [11,15,27]. In contrast to these inhibitors, SGCI displays only minor changes in the region corresponding to the 'extension' of the primary protease-binding site (Asp22, Gly23 and Lys24, Figs 2 and 3).

The fact that almost the whole molecule is affected by complexation may be due to the 'peptide-like' nature of SGCI: its small size and decreased rigidity on the ps/ns time scale (order parameters around 0.6) [14] place it between flexible peptides and larger proteins with well-defined structural cores, although undoubtedly closer to the latter group. This feature might explain that, although no remarkable structural changes occur in terms of backbone RMSD values, both ϕ/ψ dihedrals and chemical shifts of residues far from the interaction site are affected by complexation. We also suggest that the observed chemical shift changes of Cys19 and Gly20 and maybe also Arg18 can be attributed to the internal dynamics of SGCI. Two of these residues, Cys19 and Gly20 presumably retain some of their internal mobility-associated features in the bound state (see the R2 values in Fig. 4). This strengthens our previous suggestion that the different internal dynamics on the μs/ms time scale of SGCI and SGTI may play a role in taxon-specific inhibition [4,14].

Table 2. Inhibition constants of SGTI and modified SGCI on trypsin-like proteases. K_i values are given in mol.dm^{-3}. Values are from [8] and [10].

	Bovine trypsin	Crayfish trypsin
SGCI [L30R, K31M]	$5.0 \pm 0.3\ 10^{-12}$	$1.2 \pm 0.4\ 10^{-12}$
SGTI	$2.1 \pm 0.4\ 10^{-7}$	$1.4 \pm 0.4\ 10^{-12}$

Despite the availability of the crystal structure of the inhibitor complex, NMR spectroscopy provided valuable new information about the complexation process of SGCI. The observed chemical shift changes indicate that SGCI can not be easily described by the traditional 'scaffold + binding loop' concept of canonical inhibitors. This observation sheds new light at our previous results with SGCI model peptides [28,29], where the strength of inhibition was greatly dependent on the structure and dynamics of residues classified as 'scaffold'. Our findings indicate that inhibitors of the pacifastin family have a special design bringing together the dynamical features of peptides and structural organization, i.e. specific binding sites, of larger proteins.

Although taxon specificity of SGTI and SGCI can not be directly compared, as no data with arthropod chymotrypsins are yet available, K_i values of wild-type SGTI and modified SGCI clearly demonstrate the presence of this unparalleled specificity (Table 2).

Taxon specificity of SGTI was attributed to the presence of an extended protease-binding region. We showed for the related SGCI that even residues far from the primary enzyme binding site are affected by complexation. Thus, almost the whole molecule undergoes changes upon interaction with the protease, which corresponds to the concept of an 'extended binding site'. This organization might allow for the emergence of diverse inhibitor subgroups with and without taxon specificity in the pacifastin family.

Experimental procedures

Protein expression and purification

To obtain unlabeled, as well as isotope-labeled, SGCI, the SGTMCI-pET17b vector was used as described previously [14]. To obtain the double-labeled inhibitor, the SGTMCI precursor protein was expressed in BL21 DE3 pLysS cells (Novagen, Merck, Darmstadt, Germany). Cells were grown on 1 L minimal media containing 0.6% Na_2HPO_4 (Sigma, St. Louis, MO), 0.3% KH_2PO_4 (Sigma), 0.05% NaCl (Sigma), 0.1% $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Andover, MA, USA), and 0.2% U-[^{13}C] glucose (Cambridge Isotope Laboratories) at 37 °C. Cells were induced at $A_{600} = 1.0$ with a final isopropyl thio-β-D-galactoside (Sigma) concentration of 100 μg·mL^{-1} for 4 h at 37 °C. Protein isolation and purification was performed as described previously [14].

NMR measurements

Samples were dissolved in a buffer containing 10 mM Mes; 0.001% NaN_3; pH 6.0. Sample concentration was 0.76–1.72 mM ^{15}N, ^{13}C and ^{15}N SGCI were titrated in four steps to 98% saturation with unlabeled bovine α-chymotrypsin (purchased from Sigma). At each titration point, ^{1}H–^{15}N HSQC spectra were recorded on a Bruker DRX 500 spectrometer. For resonance assignment of the initial state (0% enzyme), homonuclear TOCSY and NOESY (typically 2048 data points and 512 increments) as well as 3D TOCSY–HSQC and NOESY–HSQC spectra (typically $1024 \times 100 \times 32$ data points in the direct and indirect ^{1}H and ^{15}N dimensions, respectively) were measured. To assign the complexed state, triple-resonance experiments (HNCA, HNCOCA, HNCACB, and COCACBNH, $1031 \times 64 \times 48$ data points in the ^{1}H, ^{15}N and ^{13}C dimensions, respectively) were collected on a Varian Inova 900 MHz NMR spectrometer and a Varian Inova 600 MHz NMR spectrometer equipped with a cryogenic probe. NMR relaxation parameters (T1, T2 and heteronuclear NOE) were measured at 500 MHz for the free and the complexed state at pH 6.0 using the pulse sequences described by Farrow *et al.* [30] with sensitivity enhancement [31,32].

Processing of NMR data was carried out with NMRPIPE using zero filling to the next power of 2 and shifted sinebell window functions in all dimensions. For the triple-resonance experiments, backward linear prediction was applied in the ^{13}C dimension. For spectral analysis, the programs XEASY [33], SPARKY [34] as well as the TRIAD module of SYBYL [35] were used. Linewidths were calculated using Gaussian fitting by SPARKY and taking the arithmetic average of the reported values in the ^{1}H and ^{15}N dimensions. Chemical shifts and relaxation parameters for free SGCI at pH = 6.0 and the SGCI–chymotrypsin complex were deposited in the BMRB database (http://www.bmrb.wisc.edu) [36] with Accession nos 6880 and 6881, respectively.

Exploratory structure calculations for free SGCI at pH = 6.0 were carried out as described previously [7]. Hα, Cα and Cβ chemical shift indices were calculated according to the procedures described by Wishart *et al* [37,38].

Fitting of relaxation and dynamical parameters

Fitting of R1 and R2 rates and calculating heteronuclear NOE values was carried out as described previously [14]. Peak volumes were obtained by careful integration of the central region of each peak using TRIAD. Fitting of dynamical parameters was performed using the program TENSOR2.0 [39]. Hydrodynamical calculations were done with the program HYDROPRO [40]. The structural model of the

SGCI–chymotrypsin complex used for input to both TENSOR2.0 and HYDROPRO was built by replacing PMP-C (chain I, see below) in the published PMP-C–chymotrypsin complex (PDB code 1GL1) [15] with the representative model (model 5) of the deposited SGCI solution structures (1KGM) [7] using MOLMOL [41].

Structural comparison of free and complexed SGCI

To monitor the structural changes in SGCI, the published structures of the free and complexed molecules were analyzed: free SGCI (1KGM) [7], free and the complex of the SGCI ortholog PMP-C with bovine chymotrypsin (1 GL1) [15]. The PMP-C–chymotrypsin complex is an excellent substitute for the SGCI–chymotrypsin one as both the primary and the three-dimensional structure (free SGCI and complexed PMP-C) of the two closely related inhibitors is highly similar [8,14], see Fig. 3D for differences in sequence. Chain I of the structure 1 GL1 was chosen as representative model for complexed PMP-C on the basis that it is more complete than chains J and K (lacks coordinates for only one residue opposed to two in the other chains) and has the lowest RMSD relative to the other two structures (0.44 ± 0.23 Å) Structural superpositions, RMSD and average torsion angle calculations were performed with the program MOLMOL [41].

Acknowledgements

This research was supported by grants from the Hungarian Scientific Research Fund (OTKA T046994, TS044730, TS49812 and T047154), Medichem 2 and ICGEB (Hun04-03). 900 MHz and 600 MHz NMR data were collected at the National Magnetic Resonance Facility at Madison, which is supported by grants P41-RR02301 from the NIH National Center for Research Resources and P-41G66326 from the NIH Institute of General Medical Sciences. The authors thank Antal Lopata, Chemicro Ltd and Tripos, Inc for their valuable help in obtaining and using SYBYL. The useful comments of the anonymous referees are acknowledged.

References

1 Hamdaoui A, Wataleb S, Devreese B, Chiou S-J, Vanden Broeck J, Van Beeumen J, De Loof A & Schoofs L (1998) Purification and characterization of a group of five novel peptide serine protease inhibitors from ovaries of the desert locust, *Schistocerca gregaria*. *FEBS Lett* **422**, 74–78.

2 Liang Z, Sottrup-Jensen L, Aspán A, Hall M & Söderhall K (1997) Pacifastin, a novel 155-kDa heterodimeric proteinase inhibitor containing a unique transferrin chain. *Proc Natl Acad Sci USA* **94**, 6682–6687.

3 Simonet G, Claeys I, Fransens V, De Loof A & Vanden Broeck J (2003) Genomics, evolution and biological functions of the pacifastin peptide family: a conserved serine protease inhibitor family in arthropods. *Peptides* **24**, 1633–1644.

4 Gáspári Z, Ortutay C & Perczel A (2004) A simple fold with variations: the pacifastin inhibitor family. *Bioinformatics* **20**, 448–451.

5 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**, 15397–15407.

6 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 protease inhibitors. *J Mol Biol* **258**, 158–171.

7 Gáspári Z, Patthy A, Gráf L & Perczel A (2002) Comparative structure analysis of proteinase inhibitors from the desert locust, *Schistocerca gregaria*. *Eur J Biochem* **269**, 527–537.

8 Malik Z, Amir S, Pál G, Buzás Zs, Várallyay É, Antal J, Szilágyi Z, Vékey K, Asbóth B, Patthy A *et al.* (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.

9 Vanden Broeck J, Chiou S-J, Schoofs L, Hamdaoui A, Vandenbussche F, Simonet G, Wataleb S & De Loof A (1998) Cloning of two cDNAs encoding three small serine protease inhibiting peptides from the desert locust *Schistocerca gregaria* and analysis of tissue-dependent and stage-dependent expression. *Eur J Biochem* **254**, 90–95.

10 Patthy A, Amir S, Malik Z, Bódi Á, Kardos J, Asbóth B & Gráf L (2002) Remarkable phylum selectivity of a *Schisotcerca gregaria* trypsin inhibitor: the possible role of enzyme–inhibitor flexibility. *Arch Biochem Biophys* **398**, 179–187.

11 Fodor K, Harmat V, Hetényi C, Kardos J, Antal J, Perczel A, Patthy A, Katona G & Gráf L (2005) Extended intermolecular interactions in a serine protease–canonical inhibitor complex account for strong and highly specific inhibition. *J Mol Biol* **350**, 156–169.

12 Simonet G, Claeys I, Bergelmans B, Van Soest S, De Loof A & Vanden Broeck J (2004) Transcript profiling of pacifastin-like peptide precursors in crowd- and isolated-reared desert locusts. *Biochem Biophys Res Commun* **317**, 565–569.

13 Simonet G, Bergelmans B, Proost P, Claeys I, Van Damme J, De Loof A & Vanden Broeck J (2005) Characterization of two novel pacifastin-like peptide precursor isoforms in the desert locust (*Schistocerca gregaria*): cDNA cloning, functional analysis and real-time RT-PCR gene expression studies. *Biochem J* **388**, 281–289.

14 Szenthe B, Gáspári Z, Nagy A, Perczel A & Gráf L (2004) Same fold with different mobility: backbone dynamics of small serine protease inhibitors from the desert locust, *Schistocerca gregaria*. *Biochemistry* **43**, 3376–3384.
15 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 rolesfor binding and selectivity. *J Biol Chem* **276**, 38893–38898.
16 Song J & Markley JL (2001) NMR chemical shift mapping of the binding site of a protein proteinase inhibitor: changes in the ^{1}H, ^{13}C and ^{15}N NMR chemical shifts of turkey ovomucoid third domain upon binding to bovine chymotrypsin A_{α}. *J Mol Recogn* **14**, 166–171.
17 Schlott B, Wöhnert J, Icke C, Hartmann M, Ramachandran R, Gührs K-H, Glusa E, Flemming J, Görlach M, Große F *et al.* (2002) Interaction of Kazal-type inhibitor domains with serine proteinases: biochemical and structural studies. *J Mol Biol* **318**, 533–546.
18 Peterson FC, Gordnon NC & Gettins PGW (2001) High-level bacterial expression and ^{15}N-alanine-labeling of bovine trypsin. Application to the study of trypsin-inhibitor complexes and trypsinogen activation by NMR spectroscopy. *Biochemistry* **40**, 6275–6283.
19 Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev W, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE & Kern D (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121.
20 Schechter I & Berger A (1967) On the size of the active site in proteases I. Papain. *Biochem Biophys Res Commun* **27**, 157–162.
21 Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
22 Boigegrain R-A, Pugnière M, Paroutaud P, Castro B & Brehélin M (2000) Low molecular weight serine protease inhibitors from insects are proteins with highly conserved sequences. *Insect Biochem Mol Biol* **30**, 145–152.
23 Biogegrain R-A, Mattras H, Brehélin M, Paroutaud P & Coletti-Previero M (1992) Insect immunity: two proteinase inhibitors from hemolymph of *Locusta migratoria*. *Biochem Biophys Res Commun* **189**, 790–793.
24 Chakravarty S, Wang L & Sanchez R (2005) Accuracy of structure-derived properties in simple comparative models of protein structures. *Nucleic Acids Res* **33**, 244–259.
25 LaskowskiM Jr & Qasim MA (2000) What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* **1477**, 324–337.
26 Bode W & Huber R (1992) Natural protein proteinase inhibitors and their interaction with proteases. *Eur J Biochem* **204**, 433–451.
27 Kellenberger C & Roussel A (2005) Structure–activity relationship within the serine protease inhibitors of the pacifastin family. *Protein Peptide Lett* **12**, 409–414.
28 Mucsi Z, Perczel A & Orosz G (2002) Engineering new peptidic inhibitors from a natural chymotrypsin inhibitor. *J Peptide Sci* **8**, 643–655.
29 Mucsi Z, Gáspári Z, Orosz G & Perczel A (2003) Structure-oriented rational design of chymotrypsin inhibitor models. *Protein Eng* **16**, 673–681.
30 Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD & Kay LE (1994) Backbone dynamics of free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984–6003.
31 Palmer AG, Cavanagh J, Wright PE & Rance M (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J Magn Reson* **93**, 151–170.
32 Kay LE, Keifer P & Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* **114**, 10663–10665.
33 Bartels C, Xia T-H, Billeter M, Güntert P & Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **5**, 1–10.
34 Goddard TD & Kneller GD (2005) *SPARKY 3*. University of California, San Fransisco.
35 Tripos Inc (2001). *SYBYL Molecular Modelling Package*, Version 6.7. Tripos, St Louis.
36 Seavey BR, Farr EA, Westler WM & Markley JL (1991) A relational database for sequence-specific protein NMR data. *J Biomol NMR* **1**, 217–236.
37 Wishart DS, Sykes BD & Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **34**, 1647–1651.
38 Wishart DS & Sykes BD (1994) The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR* **4**, 171–180.
39 Dosset P, Hus J-C, Blackledge M & Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* **16**, 23–28.
40 Garcia de la Torre J, Huertas ML & Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* **78**, 719–730.
41 Koradi R, Billeter M & Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51–55.

42 Shuker SB, Hajduk PJ, Meadows RP & Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.
43 Wishart DS, Sykes BD & Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* **222**, 311–333.
44 Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**, 946–950.

Supplementary material

The following supplementary material is available online:

Fig. S1. Comparison of $^{1}H-^{15}N$ HSQC spectra of free SGCI at pH 3.0 (green peaks) and pH 6.0 (blue peaks). Peaks corresponding to residues also labeled in Fig. 1 are labeled.

Fig. S2. Chemical shift index (CSI) values for free and complexed SGCI. (A) Hα CSI values for free SGCI at pH 3.0. (B) Hα CSI values for free SGCI at pH 6.0. (C) Cα (red bars) and Cβ (brown bars) CSI values for SGCI complexed with bovine α-chymotrypsin. The respective position of the three β strands is indicated by arrows. CSI values for all three species are in good agreement with the observed secondary structure (note that the sign characteristic for β strands is the opposite for Cα CSI values than the other CSI types in the figure).

Fig. S3. (A) Signed ϕ and φ differences for free SGCI and complexed PMP-C (average dihedral angles of ten models for SGCI and three conformers for PMP-C are compared). (B–D) Similar backbone fold despite different backbone dihedrals in free SGCI (blue) and complexed PMP-C (red): backbone heavy atoms for the whole molecules (B), the first β strand (C) and the protease-binding loop (D) are shown. (B–D) were prepared with RASMOL.

Fig. S4. Ramachandran plot of representative conformers of free SGCI (blue boxes) and complexed PMP-C (red boxes). Part of the lower region of the plot is duplicated at the top to yield a clearer view of some of the changes (e.g. residues 10, 15 and 25). Figure prepared with MOLMOL [41].

Fig. S5. Comparison of the side chain conformation of Arg18 in free SGCI (blue) and complexed PMP-C (red). Only heavy atoms and the amide proton of Arg18 in the representative models of SGCI and PMP-C are shown. The two molecules are fitted using backbone atoms of the region 4–33. Figure prepared with RASMOL.

This material is available as part of the online article from http://www.blackwell-synergy.com.