PROTEINS: Structure, Function, and Genetics 32:459–474 (1998)

Variability of the Canonical Loop Conformations in Serine Proteinases Inhibitors and Other Proteins

Włodzimierz Apostoluk and Jacek Otlewski*
Institute of Biochemistry and Molecular Biology, University of Wrocław, Wrocław, Poland

ABSTRACT **Canonical loops of protein inhibitors of serine proteinases occur in proteins having completely different folds. In this article, conformations of the loops have been analyzed for inhibitors belonging to 10 structurally different families. Using deviation in C_α-C_α distances as a criterion for loop similarity, we found that the P3-P3′ segment defines most properly the length of the loop. When conformational differences among loops of individual inhibitors were compared using root mean square deviation (rmsd) in atomic coordinates for all main chain atoms (Δr method) and rmsd operating in main chain torsion angles (Δt method), differences of up to 2.1 Å and 72.3°, respectively, were observed. Such large values indicate significant conformational differences among individual loops. Nevertheless, the overall geometry of the inhibitor-proteinase interaction is very well preserved, as judged from the similarity of C_α-C_α distances between C_α of catalytic Ser and C_α of P3-P3′ residues in various enzyme-inhibitor complexes. The mode of interaction is very well preserved both in the chymotrypsin and subtilisin families, as distances calculated for subtilisin-inhibitor complexes are almost always within the range of those for chymotrypsin-inhibitor complexes. Complex formation leads to conformational changes of up to 160° for χ_1 angle. Side chains of residue P1 and P2′ adopt in different complexes a similar orientation (χ_1 angle = −60° and −180°, respectively). To check whether the canonical conformation can be found among non–proteinase-inhibitor Brookhaven Protein Data Bank structures, two selection criteria—the allowed main chain dihedral angles and C_α-C_α distances for the P3-P3′ segment—were applied to all these structures. This procedure detected 132 unique hexapeptide segments in 121 structurally and functionally unrelated proteins. Partial preferences for certain amino acids occurring at particular positions in these hexapeptides could be noted. Further restriction of this set to hexapeptides with a highly exposed P1 residue side chain resulted in 40 segments. The possibility of complexes formation between these segments and serine proteinases was ruled out in molecular modeling due to steric clashes. Several structural features that determine the canonical conformation of the loop both in inhibitors and in other proteins can be distinguished. They include main chain hydrogen bonds both within the P3-P3′ segment and with the scaffold region, P3-P4 and P3′-P4′ hydrophobic interactions, and finally either hydrophobic or polar interactions involving the P1′ residue. Proteins 32:459–474, 1998.** © 1998 Wiley-Liss, Inc.

Key words: protein inhibitors; serine proteinases; protein loop; canonical conformation

INTRODUCTION

Serine proteinases and their natural protein inhibitors are among the most intensively studied models of protein-protein recognition.[1–4] Protein inhibitors do not form a single class; instead, they can be grouped into about 18 different families.[5,6] The inhibitor binds to the enzyme through an exposed and convex proteinase binding loop, which is highly complementary to the active site. Binding loops of different inhibitors adopt a similar, so-called canonical conformation.[3] The loop serves as a rather simple recognition motif, which depending on the amino

Abbreviations: AB I, adzuki bean inhibitor; ALTI, *Ascaris lumbricoides* trypsin inhibitor; ASA, accessible surface area; BPTI, basic pancreatic trypsin inhibitor (Kunitz); BUSI IIA, bull seminal plasma proteinase inhibitor IIA; CDR, complementarity determining region; CI-2, chymotrypsin inhibitor 2 from barley; CMTI I, *Cucurbita maxima* trypsin inhibitor I; ETI, *Erythrina caffra* trypsin inhibitor; HLE, human leukocyte elastase; hPSTI, human pancreatic secretory trypsin inhibitor (Kazal); MBTI, mung bean trypsin inhibitor; MCTI A, *Momordica charantia* trypsin inhibitor A; OMJPQ3, Japanese quail ovomucoid third domain; OMSVP3, silver pheasant ovomucoid third domain; OMTKY3, turkey ovomucoid third domain; PCI 1, polypeptide chymotrypsin inhibitor 1; PDB, Brookhaven Protein Data Bank; PI A-II, tracy soybean seeds Bowman-Birk proteinase inhibitor; pPSTI, porcine pancreatic secretory trypsin inhibitor (Kazal); rmsd, root mean square deviation; SGPB, *Streptomyces griseus* proteinase B; SSI, *Streptomyces* subtilisin inhibitor; STCI, soybean trypsin/chymotrypsin inhibitor; STI, soybean trypsin inhibitor (Kunitz).

Grant sponsor: Polish Committee for Scientific Research; Grant number: 6 PO4B 002 10.

*Correspondence to: Jacek Otlewski, Institute of Biochemistry and Molecular Biology, University of Wrocław, Tamka 2, PL-50–137 Wrocław, Poland. E-mail: otlewski@bf.uni.wroc.pl

Received 9 December 1997; Accepted 28 April 1998

acid sequence, is able to block strongly and specifically the active site of a particular enzyme. The largest effects on the association energy almost always involve the P1 residue (numbering system of Schechter and Berger[7]), located in the center of the loop and specifically recognized in the S1 specificity pocket of the proteinase.[8,9] Amino acid residues that precede or follow this segment (e.g., P6 or P4′) and residues from a sequentially remote region, called the secondary contact region, can also influence the association energy. The proteinase binding loop represents a unique model for studying many different aspects of protein-protein recognition and protein conformation.

First, the loop forms a linear functional epitope of six residues, which can be recognized by many serine proteinases of different specificities. The mode of recognition is always very similar and resembles that of an ideal substrate: a short antiparallel β-sheet formed between the P3-P1 residues and the 214–216 (the chymotrypsinogen numbering system is used throughout this article) segment of the enzyme. Other very important features include a short (about 2.7 Å) contact between the P1 carbonyl carbon and the catalytic serine residue and two hydrogen bonds between carbonyl oxygen of P1 and Gly193/Ser195 amides. All the hydrogen bonds mentioned above and shape complementarity of the interacting areas ensure very similar recognition between different proteinases and inhibitors. Free energies of individual interactions between residues forming the loop and the proteinase are very often found to be additive.[10,11] This introduces the possibility of installing multiple substitutions in the binding loop to maximize the association energy and specificity toward a particular enzyme.

Second, conformations of the binding loops of inhibitors representing different families are similar. Thus, inhibitors are an excellent example of a convergent evolution: the same conformation of the binding loop appeared several times during evolution independently in different groups of proteins. Global structures of proteins belonging to different inhibitor families comprise α-helical proteins, β-sheet proteins, α/β-proteins, and different folds of small disulfide-rich proteins.[3] Because amino acid sequences of the loops are also very different even within a particular inhibitor family,[6] the question arises as to what structural factors determine the canonical conformation of the binding loop. The problem becomes even more intriguing if one keeps in mind that identical pentapeptide[12] or hexapeptide[13] sequences can adopt completely different conformations in different proteins. Shorter loops in proteins (turns of three or four residues, Rose et al.[14]) and even some pentapeptide conformations[15] show a strong propensity for certain amino acids to occur at specific positions. No such rules have been described for the proteinase binding loop conformation.

Third, it has been postulated that short- and medium-sized loops (e.g., β-hairpin loops, Ω-loops, CDR loops) are constrained to a limited number of conformations and should be considered as secondary structure elements.[16–22] Because the loop occurs in very different protein folds, the following question arises: is the loop a very rare structural motif occurring only in canonical inhibitors or can it be considered as a more popular secondary structure element?

In this article, we have surveyed the binding loops of protein inhibitors for which three-dimensional structures are available. This helped us to define the range of binding loop conformations. The conformations in free and proteinase-complexed forms were compared to gain better understanding of stereochemical rules that govern enzyme-inhibitor recognition. Next, we searched Brookhaven Protein Data Bank (PDB) for hexapeptide sequences that are compatible with the range of binding loop conformations to see whether such conformations occur more frequently in proteins and, eventually, might protect these proteins against proteolysis. Finally, the analysis of binding loops and similar conformations found in proteins allowed us to better understand structural determinants of the canonical conformation.

MATERIALS AND METHODS

All distances and angles were calculated using PROMOTIF software.[23] For the accessible surface area (ASA) calculations, Lee and Richards'[24] method was used, implemented in NACCESS program.[25] Van der Waals' radii of Chothia[26] were used. Relative ASA values were calculated and compared with ASA of the respective residue in an extended Ala-Aaa-Ala tripeptide. We searched all protein structures available in the PDB January 1997 release. Δr values were calculated for the main chain atoms of the superimposed hexapeptides with the SYBYL package (Tripos Inc., St. Louis, MO). For Δt calculations, we used our own program based on the algorithm of Karpen et al.[27] Stereo plots were generated with the SYBYL package.

Molecular docking of various protein-protein complexes was performed using "O" software.[28] Hexapeptide protein fragments similar in conformation to the canonical conformation of proteinase inhibitors were superimposed on the conformation of the proteinase binding loop, as it exists in the particular inhibitor-proteinase complex for which the three-dimensional structure is known. Next, the inhibitor scaffolding was "removed" and "substituted" with the remaining part of the protein.

RESULTS

Definition of the Proteinase Binding Loop

Table I shows all inhibitor structures used in this study. They belong to 10 structurally distinct inhibi-

TABLE I. Protein Inhibitors of Serine Proteinases and Their Complexes With Proteinases Used in This Study

Family	Inhibitor	Enzyme in complex	PDB	Resolution (Å)
Kazal	hPSTI(K18Y,I19E,D21R,N29D)	—	1HPT	2.3
	OMJPQ3	—	1OVO	1.9
	OMSVP3	—	2OVO	1.5
	BUSI IIA	—	2BUS	NMR
	hPSTI(K18Y,I19E,D21R,N29D)	α-Chymotrypsinogen	1CGI	2.3
	hPSTI(K18L,I19E,D21R,N29D)	α-Chymotrypsinogen	1CGJ	2.3
	pPSTI	Trypsinogen	1TGS	1.8
	OMTKY3	α-Chymotrypsin	1CHO	1.8
	OMTKY3	HLE	1PPF	1.8
	OMTKY3	SGPB	1SGR	1.8
Potato 1	CI-2	—	2CI2	2.0
	Eglin C	α-Chymotrypsin	1ACB	2.0
	Eglin C	Subtilisin Carlsberg	1CSE	1.2
	Eglin C(L45R)	Subtilisin NOVO	1SBN	2.1
	Eglin C(R53K)	Subtilisin NOVO	1SIB	2.4
	N-acetyl Eglin C	Subtilisin Carlsberg	2SEC	1.8
	CI-2	Subtilisin NOVO	2SNI	2.1
	Eglin C	Thermitase	2TEC	2.0
BPTI	BPTI crystal form II	—	5PTI	1.0
	BPTI crystal form III	—	6PTI	1.7
	BPTI	Kallikrein A	2KAI	2.5
	BPTI	Trypsin	2PTC	1.9
	BPTI	Trypsinogen	2TGP	1.9
	BPTI(K15R)	Trypsinogen	4TPI	2.2
Bowman-Birk	STCI	—	1BBI	NMR
	PI A-II	—	1PI2	2.5
	MBTI	Trypsin	1SMF	2.1
	AB I	Trypsin	1TAB	2.3
Squash	CMTI I	—	1CTI	NMR
	MCTI A	Trypsin	1MCT	1.6
	CMTI I	Trypsin	1PPE	2.0
SSI	SSI	—	2SSI	2.6
	SSI	Subtilisin NOVO	2SIC	1.8
	SSI(M73K)	Subtilisin NOVO	3SIC	1.8
	SSI(M70G, M73K)	Subtilisin NOVO	5SIC	2.2
STI	ETI	—	1TIE	2.5
Potato 2	PCI 1	SGPB	4SGB	2.1
Ecotin	Ecotin	—	1ECY	2.2
	Ecotin(A86H)	Trypsin	1SLU	1.8
Ascaris	ALTI	—	1ATB	NMR

tor families. Generally, we considered all high-resolution crystal structures (resolution below 2.0 Å) deposited in PDB. In a few cases, medium-resolution crystal and NMR solution structures were also included to make possible comparisons between free and enzyme-complexed conformations. Table II lists amino acid sequences of the inhibitor binding loops representing different families.

The first aim of this study was to define the range of conformations which can be adopted by the proteinase binding loop in different free inhibitor structures. Because we did not know a priori which particular inhibitor loop defines the most representative conformation and how long is the segment which defines the loop, first we calculated C_α-C_α distances for the P6-P6′ region of all free and complexed inhibitor structures included in Table I (Table IIIA). It is clear that C_α-C_α distances within the P3-P3′ segment of free inhibitors exhibit relatively small deviations compared with longer sequences (the P3 C_α-P3′ C_α distance difference does not exceed 3.05 Å), whereas the deviations are significantly larger for any other segment of similar or even shorter length (e.g., P1-P4′, 7.47 Å; P2-P4′, 9.27 Å; P4-P2′, 8.38 Å). Because the P3-P3′ distance shows little variability, it seems that it defines the convex shape of the loop and is crucial for its fitting to the concave active site of the enzyme. The distance deviations become significantly smaller (P3 C_α-P3′ C_α, 1.44 Å) when one specific inhibitor family (Bowman-Birk inhibitors) is eliminated from the analysis (Table IIIB). However, we decided to include all inhibitors, as all are known to inhibit serine proteinases. Moreover, from our further analysis it appeared that Bowman-Birk in-

TABLE II. Alignment of Amino Acid Sequences and Relative Solvent Accessible Surface Areas (in Parentheses) of the P3-P3′ Segments in Proteinase Binding Loops of Free Inhibitors*

Inhibitor	P3	P2	P1	P1′	P2′	P3′
hPSTI	CYS (23.5)	THR (58.1)	TYR (103.5)	GLU (50.1)	TYR (73.7)	ARG (70.1)
OMSVP3	CYS (28.7)	THR (70.7)	MET (99.1)	GLU (47.3)	TYR (63.9)	ARG (70.8)
CI-2	VAL (40.4)	THR (73.0)	MET (102.7)	GLU (52.6)	TYR (72.9)	ARG (48.8)
BPTI	PRO (47.3)	CYS (37.7)	LYS (91.5)	ALA (43.5)	ARG (79.2)	ILE (35.2)
STCI 1	CYS (58.3)	THR (53.3)	LYS (100.2)	SER (46.1)	ASN (101.2)	PRO (75.7)
STCI 2	CYS (46.6)	ALA (43.5)	LEU (100.9)	SER (50.5)	TYR (101.0)	PRO (72.9)
PI A-II 1	CYS (46.6)	THR (53.4)	ARG (103.6)	SER (51.3)	MET (91.5)	PRO (66.4)
PI A-II 2	CYS (48.4)	THR (50.6)	ARG (96.1)	SER (47.0)	GLN (91.6)	PRO (53.3)
CMTI I	CYS (23.4)	PRO (60.4)	ARG (92.7)	ILE (58.7)	LEU (77.2)	MET (43.3)
SSI	CYS (26.4)	PRO (70.2)	MET (106.4)	VAL (47.5)	TYR (83.4)	ASP (56.5)
Ecotin	SER (36.1)	THR (57.5)	MET (88.5)	MET (72.3)	ALA (80.6)	CYS (33.3)
ALTI	CYS (50.0)	THR (74.6)	ARG (101.2)	GLU (73.9)	CYS (48.3)	LYS (60.5)

*The solvent accessible surface areas (ASA) were calculated according to Lee and Richards[24] using van der Waals' radii of Chothia.[26] The relative ASA values were calculated as the ratio of the side chain surface area found in the three-dimensional structure to ASA of the side chain in an extended Ala-Aaa-Ala tripeptide.

TABLE III. (A) C_α–C_α Distances (Å) of the P6-P6′ Segments in Inhibitors*

	P1′	P2′	P3′	P4′	P5′	P6′
P1	3.755 (0.159)	5.851 (1.483)	7.650 (2.271)	5.572 (7.466)	6.867 (8.217)	8.455 (9.514)
	3.705 (0.164)	*6.292 (0.822)*	*7.840 (2.148)*	*5.621 (7.957)*	*7.746 (6.981)*	*9.390 (8.208)*
P2	5.287 (1.415)	7.787 (2.154)	8.587 (2.764)	5.127 (9.270)	5.057 (10.996)	6.014 (12.358)
	5.518 (0.708)	*8.603 (0.904)*	*8.498 (2.726)*	*5.025 (9.994)*	*5.290 (10.526)*	*6.395 (11.871)*
P3	6.624 (2.889)	7.815 (5.034)	9.725 (3.047)	5.895 (9.767)	5.364 (10.789)	4.141 (13.789)
	8.373 (1.077)	*10.426 (1.373)*	*9.513 (3.359)*	*5.666 (10.557)*	*5.286 (11.739)*	*3.944 (13.786)*
P4	7.658 (5.587)	8.147 (8.384)	8.973 (6.854)	6.263 (12.581)	8.330 (10.967)	5.438 (14.869)
	9.822 (2.817)	*10.987 (4.128)*	*11.226 (5.237)*	*9.936 (9.882)*	*8.738 (11.890)*	*5.887 (14.684)*
P5	9.193 (6.762)	8.968 (9.616)	9.167 (9.663)	7.565 (12.324)	9.600 (10.685)	6.776 (15.884)
	9.310 (6.878)	*9.173 (9.108)*	*9.176 (10.047)*	*9.901 (11.212)*	*9.620 (11.690)*	*6.054 (16.229)*
P6	11.838 (7.307)	11.925 (10.028)	10.482 (10.319)	8.642 (14.060)	10.220 (11.790)	10.514 (13.221)
	12.909 (5.661)	*12.594 (7.702)*	*11.852 (9.589)*	*11.722 (11.472)*	*10.219 (12.198)*	*9.362 (14.041)*
	(B) Data as in Table IIIA but Distances for Bowman-Birk Inhibitors Have Been Excluded					
P1	3.755 (0.112)	5.851 (1.284)	7.650 (2.271)	9.059 (3.979)	11.390 (3.694)	13.078 (4.891)
	3.705 (0.164)	*6.292 (0.822)*	*8.910 (1.078)*	*10.935 (2.643)*	*13.623 (1.104)*	*11.988 (5.610)*
P2	5.287 (1.415)	7.787 (2.154)	10.113 (1.238)	9.049 (5.348)	10.604 (5.449)	12.991 (5.381)
	5.518 (0.708)	*8.612 (0.895)*	*9.929 (1.295)*	*12.305 (2.714)*	*13.985 (1.831)*	*11.615 (6.651)*
P3	6.624 (2.889)	7.815 (5.034)	11.336 (1.436)	9.445 (6.217)	11.037 (5.116)	12.787 (5.143)
	8.373 (1.077)	*10.694 (1.105)*	*11.350 (1.522)*	*13.000 (3.223)*	*13.773 (3.252)*	*12.212 (5.518)*
P4	7.658 (5.587)	8.147 (8.384)	8.973 (6.854)	6.263 (12.581)	8.562 (10.735)	10.643 (9.664)
	9.822 (2.817)	*10.987 (4.128)*	*11.226 (5.237)*	*12.859 (6.959)*	*12.563 (8.065)*	*14.259 (6.312)*
P5	9.193 (6.762)	8.968 (9.616)	9.167 (9.663)	7.565 (12.324)	9.600 (10.685)	11.270 (11.390)
	9.310 (6.878)	*9.173 (9.108)*	*9.176 (10.047)*	*9.901 (11.212)*	*9.620 (11.690)*	*11.110 (11.173)*
P6	11.838 (7.307)	11.925 (10.028)	10.482 (9.981)	8.642 (14.060)	10.220 (11.790)	10.795 (12.940)
	12.909 (5.661)	*12.594 (7.702)*	*11.852 (9.589)*	*11.722 (11.472)*	*10.219 (12.198)*	*10.554 (12.849)*

*Distances were calculated for free inhibitor structures (upper entry) and complexed inhibitors (lower entry, in italics) listed in Table I. Each entry contains two numbers: minimum distance and, given in parentheses, the largest deviation from the minimum distance. The sum of these two values gives the maximum distance.

hibitors are indeed rather typical. Some of the distances (particularly P3-P1′, P3-P2′ and P2-P1′, P2-P2′) are significantly more restricted when calculated only for the complexed state of the inhibitor loop (Table IIIA, B). This is in agreement with the restriction of ϕ and ψ angles accompanying complexation (see below).

In the next step of the analysis, we defined the binding loop conformation in terms of the range of main chain ϕ and ψ angles of the P3-P3′ binding loop segment in different inhibitors. The range of loop conformations in the proteinase-complexed state is rather restricted, particularly for the P1-P2′ segment (Fig. 1). In contrast, in free inhibitors ϕ and ψ angles of the corresponding residues can adopt quite different values, often over a range of 100–120°. This shows again that loops of free inhibitors can differ significantly in their conformations.

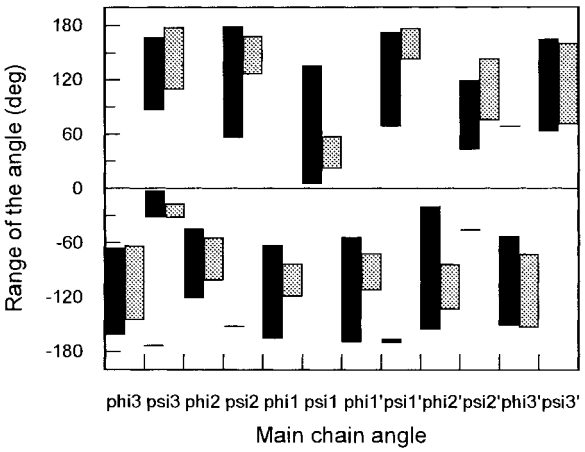

Fig. 1. Ranges of the ϕ and ψ angles that are adopted by the main chain of the P3-P3′ region in free inhibitors (black bars) and in their complexes with proteinases (gray bars).

Differences Among Individual Inhibitor Loops

Table IV presents pairwise calculations of root mean square deviation (rmsd) operating in atomic coordinates for all main chain atoms (Δr method) and in main chain torsion angles[27] (Δt method) for P3-P3′ loop segments. The calculations were performed both for free and complexed inhibitors. Correlation of individual Δr and Δt values calculated for free inhibitors is shown in Figure 2. The correlation is poor (correlation coefficient r = 0.515) and therefore shows that comparisons made jointly for distances and angles are necessary as criteria of conformational similarity. Based on the values from Table IV, average values of Δr and Δt for a particular loop were calculated. This facilitates comparison of any binding loop to all other inhibitory loops. Based on the Δr criterion, the loop of Kazal inhibitor in a free state exhibits the lowest average value of Δr = 0.954 Å and thus seems to be the most representative conformation. This particular loop exhibits Δr = 0.309 Å with respect to the homologous ovomucoid third domain and Δr in the range of 0.54–1.39 Å to loops representing other inhibitor families. The largest Δr difference between two free loops representing different families approaches 2.1 Å (*Ascaris* inhibitor ALTI versus the second binding loop of Bowman-Birk inhibitor STCI 2). However, when the Δt criterion is applied, the first binding loop of PI A-II inhibitor appears to be the most typical (average Δt = 35.9°). In this case, the range of Δt values is 24.6–52.6°. The largest Δt value observed is 72.3° (BPTI versus ecotin). By the Δt criterion, the binding loop of BPTI (probably the best-known proteinase inhibitor) differs the most from all other inhibitor loops (average Δt = 57.2°). Examples of binding loop pairs most similar and most different by Δt and Δr criteria are given in Figure 3. All of these loops are capable of binding strongly to the active site of a serine proteinase. In conclusion to this part of the analysis, differences among individual loops are too large to warrant the consideration of the canonical conformation as an element of protein secondary structure.

Pairways comparisons of individual loops in free versus complexed state (the diagonal of Table IV) shows that complexation leads to conformational changes of canonical loops (Δr in the range 0.228–1.018 Å; Δt in the range 12.2–34.3°). These changes vary for different inhibitors; BPTI shows insignificant conformational change upon complexation with trypsin, whereas for ecotin Δr and Δt values approach the upper limits.

Enzyme-Inhibitor Docking

Because inhibitor loops exhibit significant differences, we examined whether docking of these loops into the active sites of different enzymes is similar. To test this, we calculated distances from C_α of catalytic Ser195 (or Ser221 in the case of subtilisin) to C_α atoms of P3 to P3′ residues for different proteinase-inhibitor pairs (Table V). It appears that these distances are very similar (the differences do not generally exceed 1 Å), which indicates a geometrically very similar mode of interaction. The distances to C_α of P3-P1 segment exhibit generally smaller distance differences than to the P1′-P3′ segment. This might be due to the conserved antiparallel β-sheet interactions between the P3-P1 and the 214–216 (in subtilisin 125–127) region of the proteinase. There are no similarly conserved inhibitor-enzyme interactions at the prime site of the loop. It should be stressed that the proteinases included in Table V belong to two structurally distinct families: chymotrypsin and subtilisin. In both families the geometrical mode of interaction is very well preserved, as distances calculated for subtilisin family-inhibitor complexes are almost always within the range of those for chymotrypsin family-inhibitor complexes.

Changes in the Side Chain Conformation on Complex Formation

The analysis of changes in main chain torsion angles of the binding loop accompanying complex formation has already been described.[29] Therefore, we focus here on local changes of the individual side chains manifested by changes of side chain torsion angles. Figure 4 A–F shows the changes of χ_1 angle for the P3-P3′ residues. Several conclusions can be drawn from this plot. Side chains of residues P1 and P2′ adopt a similar orientation in the complex (χ_1 angle = −60° and −180°, respectively). The most dramatic transition of 160° occurs at the P1 χ_1 angle of CI-2 inhibitor upon complexation with subtilisin NOVO. The χ_1 angle transition at P2′ can reach 120°. However, for a number of loops, no conformational reorientation occurs at P1 and P2′. At positions P3

TABLE IV. Δr and Δt Values for the P3-P3′ Loop Segments of Free and Complexed Inhibitors*

Inhibitor	hPSTI	OMSVP3 (OMTKY3)	CI-2	BPTI	STCI 1	STCI 2	PI A-II 1	PI A-II 2	AB I	CMTI I	SSI	Ecotin	ALTI	PCI 1
hPSTI	13.0	0.309	0.541	0.633	1.329	1.213	1.226	1.321	—	0.718	0.619	1.191	1.392	—
	0.468	*0.249*	*0.518*	*0.556*	—	—	—	—	*1.172*	*0.480*	0.282	*0.999*	—	*0.817*
OMSVP3	15.7	14.3	0.564	0.598	1.376	1.242	1.280	1.390	—	0.809	0.616	1.152	1.550	—
(OMTKY3)	*10.0*	0.313	*0.403*	*0.574*	—	—	—	—	*1.301*	*0.538*	*0.321*	*0.927*	—	*0.734*
CI-2	20.5	25.6	17.0	0.729	1.564	1.451	1.452	1.537	—	0.696	0.784	1.326	1.324	—
	17.1	*16.0*	0.400	*0.685*	—	—	—	—	*1.487*	*0.544*	*0.533*	*0.716*	—	*0.560*
BPTI	53.4	52.4	54.9	12.2	1.384	1.265	1.301	1.343	—	0.982	0.696	1.398	1.591	—
	55.9	57.4	*52.1*	0.228	—	—	—	—	*1.280*	*0.688*	*0.583*	*1.085*	—	*0.888*
STCI 1	53.8	56.9	46.1	60.8		0.571	0.651	0.568	—	1.309	1.137	1.755	2.012	—
	—	—	—	—	—		—	—	—	—	—	—	—	—
STCI 2	52.7	53.4	43.4	58.5	18.1		0.745	0.712	—	1.317	0.991	1.582	2.100	—
	—	—	—	—	—		—	—	—	—	—	—	—	—
PI A-II 1	38.4	37.8	28.0	52.6	29.0	27.6		0.444	—	1.363	1.188	1.840	1.956	—
	—	—	—	—	—	—		—	—	—	—	—	—	—
PI A-II 2	44.2	48.7	35.7	60.9	17.3	20.5	24.6		—	1.347	1.155	1.964	2.042	—
	—	—	—	—	—	—	—		—	—	—	—	—	—
AB I	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	33.3	*33.7*	*24.7*	*48.5*	—	—	—	—		*1.231*	*1.161*	1.702	—	*1.558*
CMTI I	43.2	47.7	38.5	45.4	35.1	37.3	35.6	31.6	—	28.6	0.632	1.520	1.429	—
	21.9	*21.0*	*19.2*	*46.1*	—	—	—	—	*34.2*	0.645	*0.448*	*0.913*	—	*0.805*
SSI	27.2	32.3	32.5	55.2	43.1	39.9	40.8	34.8	—	34.3	18.8	1.369	1.809	—
	9.3	*9.6*	*14.8*	*55.5*	—	—	—	—	*33.0*	*20.2*	0.447	*1.026*	—	*0.861*
Ecotin	41.8	44.7	45.1	72.3	54.6	55.9	46.3	45.6	—	51.3	45.1	34.3	1.852	—
	33.0	*31.9*	*20.1*	*56.6*	—	—	—	—	*23.2*	*26.8*	*30.3*	1.018	—	*0.465*
ALTI	45.1	45.0	39.9	62.8	49.2	52.6	34.7	43.7	—	45.4	55.1	49.8		—
	—	—	—	—	—	—	—	—	—	—	—	—		—
PCI 1	—	—	—	—	—	—	—	—	—	—	—	—	—	
	30.3	*28.9*	*17.7*	*57.2*	—	—	—	—	*20.9*	*25.7*	*29.1*	*17.9*	—	
average Δt	39.6	41.8	37.3	57.2	42.2	41.8	35.9	37.1	—	40.5	40.0	50.2	47.6	—
	26.3	*26.1*	22.7	*53.7*	—	—	—	—	*31.4*	*26.9*	*25.2*	*30.0*	—	*28.5*
average Δr	0.954	0.990	1.088	1.084	1.241	1.199	1.222	1.257	—	1.102	1.000	1.541	1.732	—
	0.634	*0.631*	*0.681*	*0.792*	—	—	—	—	*1.361*	*0.706*	*0.652*	*0.979*	—	*0.836*

*Δr values are given in the upper right and Δt value in the lower left triangle of the matrix. The values were calculated for free inhibitor structures (upper entry) and complexed inhibitors (lower entry, in italics). The diagonal shows Δt (upper entry) and Δr (lower entry) values for comparison of free versus complexed state of the same inhibitor. The average Δr and Δt values are given below the matrix. The largest Δr and Δt values in the matrix are boxed; the smallest ones are underlined. The smallest average Δr and Δt values are double underlined. The minimum values are for representatives of different inhibitor families, therefore, the value for hPSTI and OMSVP3 was not considered.

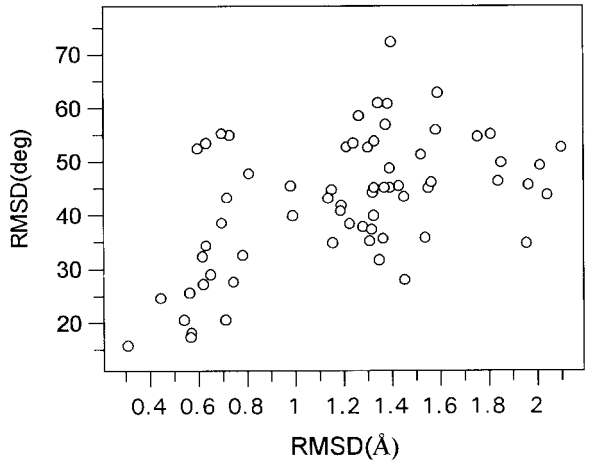

Fig. 2. Correlation of Δr with Δt values. All Δr and Δt values that compare conformations of the P3-P3′ free loops reported in Table IV are included in the plot (Pearson-product moment correlation, r = 0.515; slope, s = 13.905).

and P2, no changes of χ_1 angle are observed. This is because the P3 position is often involved either in disulfide bond formation or is occupied by a Pro. Similarly, only minor adjustments are observed at P2. Larger reorientation, up to 120°, occurs at positions P1′ and P3′.

Due to a small number of longer side chains in the P3-P3′ segment, more limited data are available for χ_2. It is clear, however, that χ_2 angle of different side chains (Tyr, Lys, Arg) at P1 adopts a value of approximately 180° in complexes with enzymes of chymotrypsin family, whereas for Met at this position in two subtilisin complexes χ_2 is −60° (Fig. 5A). At P3′, regardless of the proteinase and side chain (Arg, Ile, Met, and Asp) involved, the χ_2 angle is invariably ±180° (Fig. 5B).

Occurrence of the Canonical Conformation in Protein Structures

Because the canonical loop in free inhibitors shows a considerable range of conformations, we wanted to see whether it is limited to proteinase inhibitors or, perhaps, is a more popular structural motif. We checked how many protein fragments in PDB (4308 protein structures) fulfill our binding loop conformation criteria. First, we defined the range of binding loop conformations in terms of the P3-P3′ main chain dihedral angles as shown in Figure 1 for free and complexed inhibitors adding extra 5° to lower and upper limits. As a second criterion, we used all nine P3-P3′ C_α-C_α distances with ranges as defined in Table IIIA (e.g., for P3-P3′ the range is 9.5–12.9 Å).

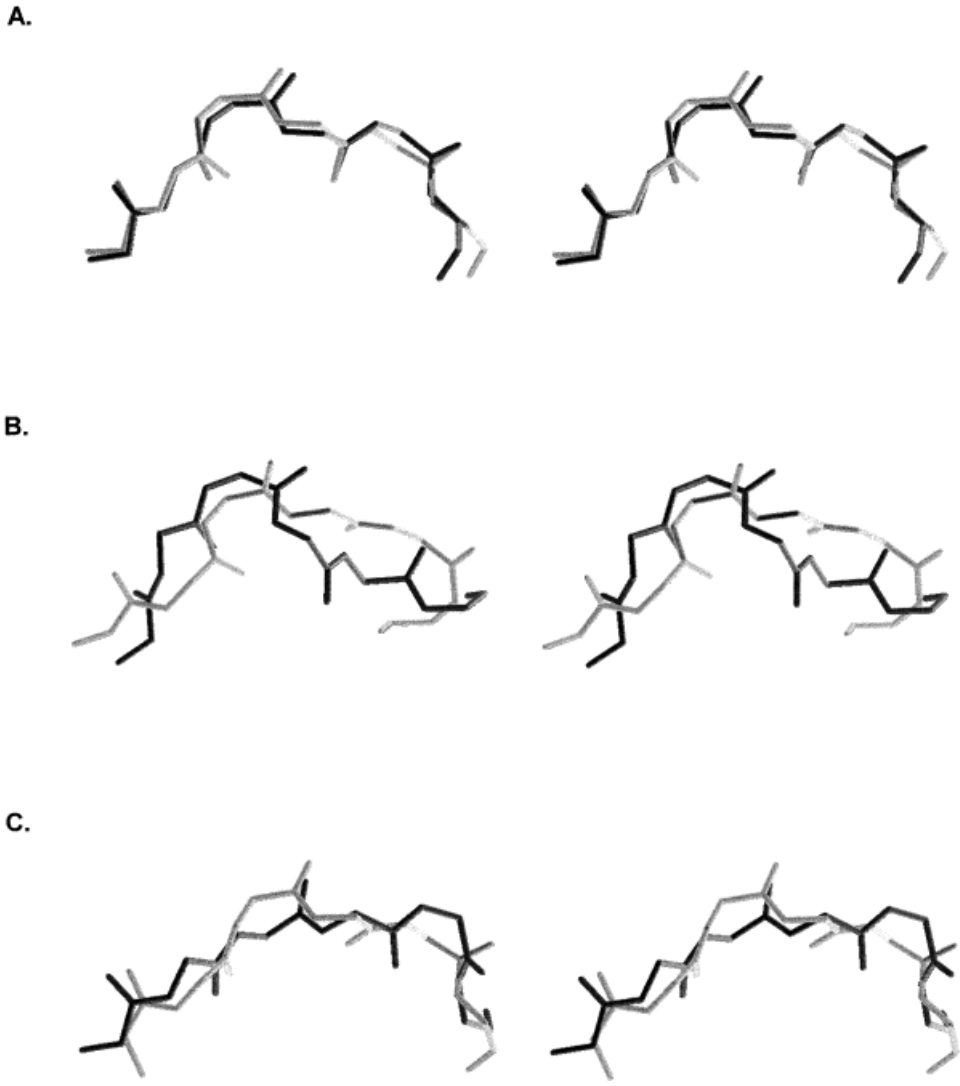

Fig. 3. Superposition of the main chain conformations of the most similar and most dissimilar proteinase binding loops found in structures of free inhibitors. **A:** The P3-P3′ loops of hPSTI and CI-2 that are most similar by Δr and Δt criteria (Δr = 0.541 Å; Δt = 20.5°). **B:** The canonical P3-P3′ loops of ALTI and second domain of STCI that are most different by Δr criterion (Δr = 2.1 Å). **C:** The canonical loops of ecotin and BPTI that are most different by Δt criterion (Δt = 72.3°).

TABLE V. Intermolecular Distances (Å) Between C_α of Ser 195 (Chymotrypsin Family) or Ser 221 (Subtilisin Family) and C_α Atoms of the P6-P6′ Binding Loop in Different Inhibitor Complexes*

Residue	P6	P5	P4	P3	P2	P1	P1′	P2′	P3′	P4′	P5′	P6′
MIN	15.890	13.259	12.476	10.033	6.866	4.510	4.873	7.866	10.299	9.112	10.943	13.296
Δ	3.900	3.796	1.251	0.880	0.497	0.368	0.609	1.065	1.656	6.133	6.467	7.524

*Data for complexes reported in Table I are included. The minimum values of distances are in the first row; the largest deviations are in the second row.

Both criteria, when applied jointly to PDB protein structures, provided 235 hexapeptide sequences (160 unique) in 149 different proteins. From this set, we discarded 28 NMR solution structures in cases when only one particular substructure fulfilled the criteria. The final set contained 161 hexapeptides (132 unique) from 121 proteins.

Table VI lists the occurrences of amino acids at particular positions of the loop. Not a single position contains exclusively one particular amino acid. However, several conclusions can be drawn. First, Cys occurs only twice at P3, which is much less than found in inhibitors. The Kazal, potato 2, Bowman-Birk, squash, SSI, and *Ascaris* inhibitor families

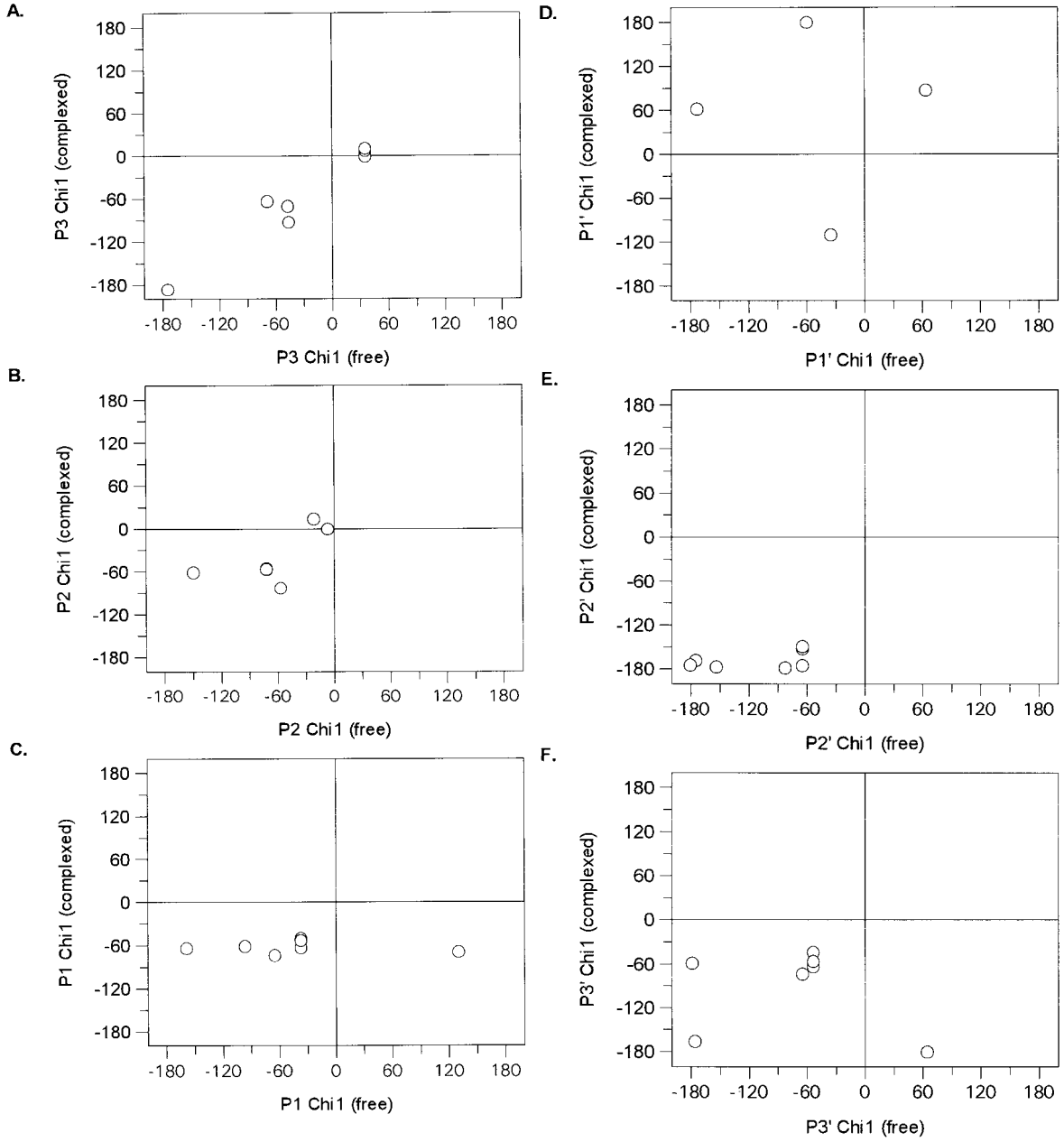

Fig. 4. Changes of the χ_1 angle during complex formation. The graph is based on inhibitors for which X-ray structures of free and complexed forms were available, with the exception of the high-quality NMR structure of free CMTI I. **A:** P3, **B:** P2, **C:** P1, **D:** P1′, **E:** P2′, and **F:** P3′.

contain cystine at position P3, often in combination with Pro at P2 (Tables II and IX). In the hexapeptide set Cys(P3)-Pro(P2) sequence occurs once. However, in this particular case, Cys residue does not form a disulfide bond. Another combination that occurs in BPTI family is Pro(P3)-Cys(P2). This sequence does not occur in the set of protein hexapeptides. Second, there is a relatively strong tendency for Pro to occur at positions P2, P1, P2′, or P3′ but never at P1′. Pro is found often at P2 in various families of proteinase inhibitors. Next, there is a rather strong tendency in the set for β-branched side chain of Ile, Val, or Thr to occur at P1′ position (35% of all cases). Another striking tendency is a very rare occurrence of Trp and (to a lesser extent) Met side chains at any position of the loop. Trp is the only amino acid that is not found at P1 position. The side chains that are most frequently observed at P1 are Asp, Asn, Pro, Gly, Lys, or Glu. These side chains are charged, strongly polar, or of unique conformational propensities (Gly and Pro).

More than half of the hexapeptides (75 of 132) are similar (Δt lower than 30°) to at least one of the

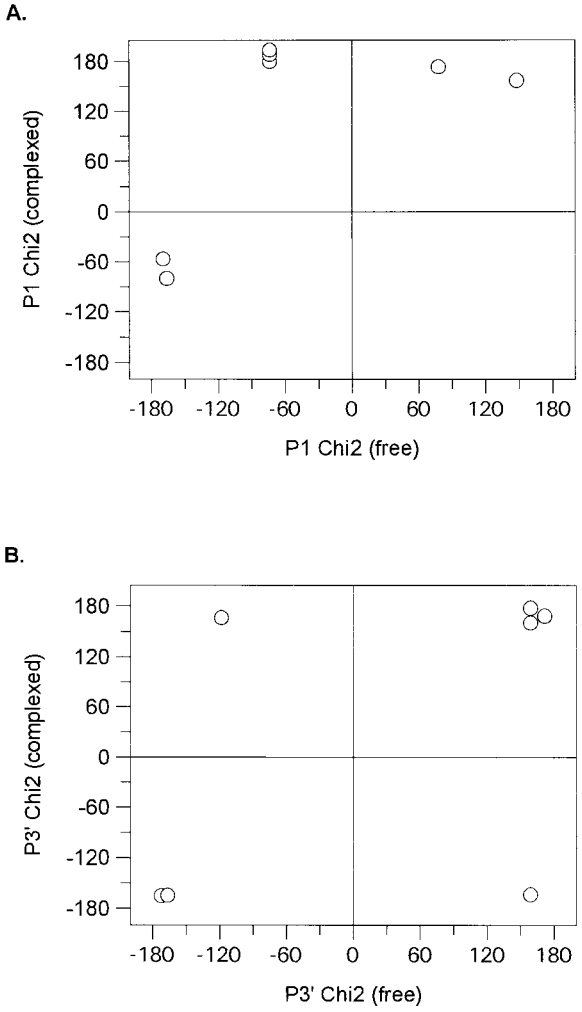

Fig. 5. Changes of the χ_2 angle of the P1 (**A**) and P3′ (**B**) residues during complex formation. The graph is based on inhibitors for which X-ray structures of free and complexed forms were available, with the exception of the high-quality NMR structure of free CMTI I.

TABLE VI. Frequencies of Occurrence of 20 Amino Acids in the Positions Equivalent to P3-P3′ Hexapeptides That Fulfill Conformational Criteria of the Canonical Proteinase Binding Loop*

	P3	P2	P1	P1′	P2′	P3′
ALA	12/2	8/2	9/4	10/3	12/6	12/3
ARG	8/2	3/1	3/1	4/2	6/1	3/1
ASN	5/2	4/0	14/6	5/2	6/1	6/1
ASP	7/5	13/3	15/4	8/1	8/3	8/3
CYS	2/1	4/1	3/0	3/0	1/0	2/0
GLN	6/0	1/0	5/0	5/1	5/0	3/0
GLU	6/4	8/3	11/4	8/2	14/6	4/1
GLY	6/1	5/1	13/2	8/3	1/0	2/1
HIS	5/2	2/0	3/2	7/0	2/1	6/4
ILE	12/4	9/2	1/0	13/7	7/1	10/3
LEU	12/1	11/2	6/0	4/1	3/1	18/8
LYS	13/5	4/1	11/5	8/3	7/2	8/3
MET	3/0	4/1	2/0	3/0	2/0	1/0
PHE	9/2	4/2	2/0	2/2	5/1	4/1
PRO	5/3	18/9	15/4	0/0	14/4	13/2
SER	5/1	12/3	8/5	4/1	8/2	4/0
THR	5/1	8/4	6/1	15/5	12/5	6/3
TRP	2/1	0/0	0/0	1/0	2/0	2/0
TYR	4/2	7/2	2/0	5/2	6/2	5/1
VAL	5/1	7/3	3/2	19/5	11/4	15/5

*First value is the frequency of a given residue found in all 132 PDB unique hexapeptides that fulfill the criteria of the canonical loop. The second entry is the frequency found in the 40 hexapeptides with highly (>70%) exposed side chain of the P1 equivalent residue.

proteinase binding loops listed in Table IV. In the most striking cases the similarity of the conformations is very high. For example, $\Delta t = 14.6°$ was found for the comparison between the Phe144-Tyr149 fragment of lipoxygenase-1 (PDB code 2SBL) and the binding loop of CI-2 inhibitor. This value is lower than those for all possible comparisons between two proteinase binding loops in uncomplexed state (Table IV). Another observation is that in many cases, loops found in protein structures are highly similar (Δt below 30°) to several (up to five) different inhibitor loops presented in Table IV. In several proteins two loops of unique sequences were found in one polypeptide chain. Dimer of pertussis toxin contains four such loops. Another interesting case is the Gly142-Asn143-Thr144-Lys145-Ser146-Ser147 segment of trypsinogen. Lys145 is well known as the autolytic cleavage site and serves as the P1 residue for trypsin recognition. According to our analysis, if the nicksite were at Thr144 (relative ASA = 20.5%), this loop would resemble substrate conformation to a higher degree, because the canonical conformation mimics that of an ideal substrate. However, even the hexapeptide with Lys145 at P1 resembles the BPTI loop (rmsd for backbone atoms 1.35 Å), as noted by Hubbard et al.[29]

A prominent feature of the proteinase binding loops is a very high solvent exposure of the P1 side chain (relative ASA greater than 90%) (Table II). This greatly facilitates the penetration of the P1 side chain into the S1 pocket of the proteinase during complex formation. To further search for conformations similar to the inhibitor loops, we restricted the set to hexapeptides in which P1 residues exhibited high solvent accessibility. In 40 of 132 loops of unique sequences, the relative ASA of P1 residue was higher than 70%. These loops are listed in Table VII. They represent 39 proteins; in ecotin, two other loops were found in addition to the canonical loop (P1 Met84) that serves to inhibit serine proteinases. With the exceptions for Pro, Val, Gly ,and Ala (total 12), the side chains acceptable at P1 are either charged Arg, Asp, Glu, His, Lys (total 16) or polar Asn, Ser, Thr (total 12).

TABLE VII. The Values of Δr and Δt, Hexapeptide Segment Amino Acid Sequences, and the Accessible Surface Areas of the P1 Equivalent Residue in the 40 Hexapeptides That Fulfill Conformational Criteria of the Canonical Proteinase Binding Loop*

PDB code	Sequence no.	Δt (degree)		Δr (Å)		P3	P2	P1	P1′	P2′	P3′
1avd	A39	36.9	BPTI	0.989	BPTI	A	T	S (94.2)	N	E	I
1cdg	553	31.7	PI A-II 1	1.051	hPSTI	I	P	A (90.7)	V	A	G
1cgt	552	33.0	PI A-II 2	1.054	hPSTI	I	P	S (89.1)	V	A	A
1ecy	92	23.3	CMTI I	0.954	ecotin	K	E	K (74.7)	K	F	V
1ecy	132	29.1	PI A-II 2	0.797	CI-2	A	E	E (91.2)	K	I	D
1gd1	Q52	26.7	PI A-II 2	0.759	SSI	R	L	D (101.1)	A	E	V
1gff	2 152	26.5	CMTI I	0.713	CMTI I	W	T	A (77.0)	S	T	I
1ggt	A542	33.5	BPTI	0.959	BPTI	N	S	H (74.5)	N	R	Y
1gtm	C286	49.7	BPTI	1.055	PI A-II 2	E	L	E (81.9)	V	D	V
1har	2	23.6	PI A-II 1	0.849	CI-2	I	S	P (72.4)	I	E	T
1hge	A171	31.8	BPTI	1.065	BPTI	N	D	N (91.0)	F	D	K
1hnv	B315	50.7	BPTI	1.813	ecotin	H	G	V (74.2)	Y	Y	D
1hum	A15	37.6	ALTI/PI A-II 1	0.972	CI-2	Y	T	A (91.5)	R	K	L
1ibg	H127	39.7	BPTI	0.899	CMTI I	G	C	G (106.2)	D	T	T
1ige	B210	26.3	PI A-II 2	0.896	CI-2	K	T	S (99.8)	G	P	R
1ijs	P473	50.9	OMSVP3	1.027	PI A-II 1	E	F	D (102.5)	T	D	L
1lla	603	52.8	PI A-II 2	1.428	PI A-II 2	P	I	H (80.4)	T	E	H
1lpp	328	27.2	PI A-II 2	0.743	CI-2	Y	A	N (75.1)	I	P	V
1ltd	A94	40.2	STCI 1	1.480	SSI	C	P	P (73.4)	Y	A	P
1lts	A135	21.1	CI-2	0.601	CI-2	I	D	E (72.3)	R	L	H
1obr	56	31.7	CMTI I	0.972	hPSTI	D	E	N (96.8)	E	P	E
1pcz	A52	20.0	CMTI I	1.121	CMTI I	D	P	K (78.4)	V	A	L
1prt	A182	28.4	hPSTI	0.783	SSI	R	S	V (97.1)	A	S	I
1svb	331	25.0	PI A-II 1	1.185	ecotin	T	F	S (74.2)	G	T	K
1tad	A285	32.0	BPTI	1.092	ecotin	D	Y	N (98.6)	G	P	N
1tbp	B108	22.0	CMTI I	1.096	CI-2	E	P	K (83.0)	T	T	A
1tgh	A293	20.8	CMTI I	1.040	CI-2	K	P	R (75.0)	I	V	L
1ton	123	32.9	PI A-II 1	1.199	BPTI	L	P	T (83.1)	K	E	P
1trz	B1	24.2	PI A-II 2	0.796	SSI	F	V	N (85.6)	Q	H	L
1ukz	39	22.9	BPTI	0.795	CI-2	D	Y	S (72.3)	F	V	H
1vok	A157	18.0	CMTI I	1.117	CI-2	V	P	K (71.7)	I	V	L
1ytb	A199	<u>16.8</u>	CMTI I	1.053	CI-2	K	P	K (84.9)	I	V	L
2cmd	48	30.7	BPTI	0.863	PI A-II 2	H	I	P (99.9)	T	A	V
2ctx	7	29.2	BPTI	0.805	BPTI	P	D	G (94.0)	T	S	K
2ebn	7	21.8	CI-2	<u>0.506</u>	CI-2	K	A	N (87.1)	I	K	L
2por	219	27.6	PI A-II 2	1.086	PI A-II 2	P	V	A (107.9)	A	A	A
2reb	34	23.9	PI A-II 2	0.771	CMTI I	S	M	D (108.7)	V	E	T
3est	60	28.9	BPTI	0.932	BPTI	D	R	E (90.2)	L	T	F
821p	28	27.1	SSI	0.952	SSI	F	V	D (82.1)	E	Y	D
8dfr	126	40.3	BPTI	1.077	hPSTI	E	K	P (75.1)	I	N	H

*The amino acid sequence numbering as described in PDB. The sequence number refers to the first position of the hexapeptide. The reported Δt and Δr are the smallest values obtained from an rmsd comparison to any of the 12 inhibitory loops included in Table IV (the lowest values of Δt and Δr are underlined). The inhibitor name is shown. The relative ASA values are given in parentheses next to the one-letter codes for the P1 equivalent residue.

Molecular Docking of Protein-Serine Proteinase Complexes

We performed a simple docking procedure to further check whether any of the 39 proteins could inhibit a serine proteinase. The hexapeptide protein fragments similar in conformation to the canonical loops of proteinase inhibitors were superimposed on the conformation of the proteinase binding loop, as it exists in a particular inhibitor-proteinase complex for which the three-dimensional structure has been determined. We chose inhibitors to which a particular hexapeptide exhibited the highest resemblance, based on the Δt criterion. For example, the 7–12 hexapeptide of cobratoxin (2CTX) is most similar to the canonical loop of BPTI (Table VII); therefore, the 7–12 fragment was superimposed on the BPTI 13–18 binding loop, as seen in the BPTI-trypsin complex. Next, the inhibitor scaffold was "removed" and "substituted" with the remaining part of the protein. We did not try to match the P1 residue specifically with a particular proteinase, as in our hexapeptide set there were many examples of P1 residues (Asn, Gly,

His, Pro, Ser, Thr) for which no highly specific serine proteinases are known. Rather, our aim was to see whether there are steric conflicts between a particular protein and a serine proteinase. When all 39 proteins (40 loops) were checked for their potential inhibitory properties, it appeared that steric conflicts did exist in virtually all cases. We suppose that they would efficiently prevent complex formation between any of the 39 proteins and serine proteinases tested.

Structural Determinants of the Canonical Conformation

Are there any clear structural determinants of the canonical conformation of the loop? In analogy to the situation described for medium-size loops by Tramontano et al.,[30] we grouped these determinants to a rather extensive network of hydrogen bonds and to interactions involving three particular loop residues (P1′, P3, and P3′).

In particular, we analyzed the hydrogen bonds both in inhibitors and in the 40 hexapeptide segments in which main chain atoms of P3-P3′ segment are involved. Generally, the system of hydrogen bonds is different for representatives of different inhibitor families (Table VIII). The most intensive system occurs in BPTI (nine hydrogen bonds), ETI, and hPSTI (six hydrogen bonds) inhibitors. In the case of BPTI, this might partly reflect a high quality of the neutron structure. In ALTI and ecotin there are only 1 and 0 such bonds, respectively. In the 40 protein hexapeptides this system is somewhat better developed (data not shown), which might result from their closer contacts with protein scaffold (lower protrusion of the hexapeptides compared to the inhibitor loops). Some hydrogen bonds in the inhibitors are partly conserved. For example, the P2 O . . . N P1′ hydrogen bond exists in 3 of 13 inhibitors studied. Similarly, it occurs in 14 of 40 protein hexapeptide segments. Main chain atoms of P3 and P1 are involved in a lower number of hydrogen bonds than the remaining main chain atoms of P3-P3′ segment (e.g., P2, P1′, and P3′). Further, in a number of the inhibitor families, the binding loop conformation is maintained by a system of hydrogen bonds to the inhibitor scaffolding. In OMSVP3 and CI-2 inhibitors carbonyl oxygens of P2 and P1′ are involved in hydrogen bonds to side chain of Asn33 and Arg65, respectively. In BPTI and CMTI I, similar interactions are mediated through water molecules to Gly12 and to the side chain of Cys20, respectively. A similar network of hydrogen bonds stretching from the loop to the protein scaffold can be also found in the 40 hexapeptide segments. This system, however, is not regular, and characteristic features could not be easily extracted.

The second major determinant of the loop conformation are the interactions involving side chains of P3, P3′, and P1′ residues. There is a very strong tendency of the P3 side chain (found in 12 of the total of 13 inhibitor loops) to make hydrophobic contacts to P4 residue. Similarly, the side chain of P3′ almost always (in 12 of 13 cases) interacts with P4′. The situation observed in protein hexapeptides is very similar. In 40 cases, hydrophobic P3-P4 interactions were found in 36 hexapeptides, and P3′-P4′ interactions were found in 32 hexapeptides. In many cases there were often additional hydrophobic interactions involving P3 and P3′ residues.

The side chain of P1′ residue is engaged in either hydrophobic or polar interactions. This residue, therefore, can be considered the loop anchoring residue. Four hydrophobic and 5 polar P1′ side chains occur in 13 inhibitor loops (Table VIII). The remaining four inhibitors do not exhibit any anchoring residue features, as their P1′ does not form either clear hydrophobic or hydrogen bonding interactions. Among them there is, however, squash inhibitor CMTI I. In case of squash inhibitors, the P1′ residue is exclusively an Ile in 39 unique amino acid sequences. Replacement of this Ile by Leu leads to significant disordering of the loop and to a large drop in the energy of association with trypsin.[31] Hydrophobic interactions that involve P1′ are formed with atoms outside the binding loop (Table VIII). Hydrogen bonding interactions involving the P1′ side chain are in three of five cases to the side chain of P2 residue and also to the side chain or main chain of P3′ (Table VIII).

When the anchoring interactions of P1′ residue of the inhibitors are compared with those in the 40 protein hexapeptides, a clear preference for their hydrophobic property can be noted. Among these 40 loops, there are 18 with P1′ side chain participating in either hydrophobic (β-branched side chains 7 Ile, 4 Val, 1 Thr) or polar interactions. The hydrophobic interactions are formed predominantly to protein regions well outside the loop; in only four cases they involve the P4′ residue.

Polar interactions of P1′ residue were found in only six hexapeptide loops. In three cases they also involved β-branched side chain of Thr; in the remaining three loops, Arg, Lys, and Glu side chains. They were directed to scaffold residues or, in two cases, to the P3′ main chain or side chain.

Sequential Motifs

Finally, we analyzed all available inhibitor sequences for occurrence of conserved sequential motifs within P3-P3′ loop (Table IX). Four clear patterns could be distinguished. They comprise 91% of the total number of inhibitor sequences reported in Table IX. Representatives of six different families show Thr(P2) motif, whereas inhibitors belonging to five families exhibit Cys(P3)-Pro(P2) sequential motif. BPTI, hirustasin, and elafin families are characterized by Cys(P2) motif, whereas STI family is the only one with Pro at P3. Some other positions are

TABLE VIII. Patterns of Weak Interactions (Hydrogen Bonds and van der Waals Interactions) Stabilizing the Canonical Conformation of the Inhibitors*

Inhibitor/ interaction	Main chain hydrogen bonds												Anchoring residue (P1′)		van der Waals contacts		
	N P3	O P3	N P2	O P2	N P1	O P1	N P1′	O P1′	N P2′	O P2′	N P3′	O P3′	Polar	Apolar	P3	P2′	P3′
hPSTI				$Asn33_{sc}$		H_2O	O^{γ} P2	$Asn33_{sc}$	H_2O		$Pro32_{mc}$		$P3'_{sc}/H_2O$ (O^{γ} P2)			Pro22	Pro22
CI-2				N P1′			O P2	$Arg65_{sc}$			$Gly83_{mc}$	$Arg65_{mc}$	$Arg65_{sc}$		Leu51/Ile56/Arg67	Gly83	Ile63
BPTI		H_2O	H_2O ($Thr11_{mc}$/$Cys38_{mc}$)	H_2O ($Gly12_{mc}$)	H_2O		$Gly36_{mc}$	H_2O		H_2O	$Tyr35_{mc}$	$Tyr35_{mc}$		Gly36	Gly12	Val34	Gly37
SSI				N P1′			O P2							Asn99	Met70/Cys101	Pro77	Lys18/Pro77
OMSVP3			H_2O	$Asn33_{sc}$				H_2O/$Asn33_{sc}$		H_2O	$Gly32_{mc}$		O^{γ} P2/H_2O ($Cys35_{mc}$)		Ala15	Pro22	Pro22
STCI 1			$Gln21_{mc}$	$Gln21_{mc}$										Pro22	Ala13		Pro20
STCI 2			$Gln48_{mc}$	N P1′			O P2							Ala47	Ile40		Leu29/Ala47
PI A-II 1			$Gln20_{mc}$										O P3′/O^{γ} P2 ($Gln20_{mc}$)		Met12		Pro19
PI A-II 2								H_2O ($Cys20_{sc}$)/ H_2O ($Cys28_{mc}$)							Met38		Gly45
CMTI I				H_2O ($Cys20_{sc}$)							H_2O ($Cys28_{mc}$)	$Cys28_{mc}$			Val2/Cys20		Glu9
Ecotin															Val81		Pro88
ALTI											$Gly16_{mc}$				Cys18/Pro28	Cys15	Gly16/Pro35
ETI	H_2O ($Asp71_{sc}$/$Lys72_{mc}$)	H_2O/H_2O ($Tyr113_{sc}$)		$Ser60_{sc}$					H_2O	$Asn12_{mc}$		O^{γ} P1′	O P3′		Ser60/Tyr113		Val1/Val9/Ile67

*To show the common features, the left part of the table presents only the main chain (and not the side chain) interactions of the P3-P3′ segment. The right side combines the polar (hydrogen bonds) and van der Waals interactions involving the P1′ side chain and van der Waals interactions of the P3, P2′, and P3′ side chains. These interactions are to main chain (mc), side chain (sc), and water (H_2O) atoms. A slash is used in the cases when two or three interactions are present. Parentheses are used when the hydrogen bond was main chain, side chain, or water mediated.

TABLE IX. Most Typical Amino Acid Sequences of the Canonical Binding Loop Occurring in Different Families of Proteinase Inhibitors*

Motif	Family	No. of sequences	P3	P2	P1	P1′	P2′	P3′
Cys(P3)-Pro(P2)			Cys	Pro	Aaa	Aaa	Aaa	Aaa
	Kazal	24/156	Cys	Pro	Arg (11/24) Lys (6/24)	Asp (9/24) Ile (8/24)	Tyr (18/24) Phe (4/24)	Asn (9/24) Asp (6/24)
	Squash	48/48	Cys	Pro	Arg (34/48) Lys (6/48)	Ile	Leu (32/48) Tyr (8/48) Trp (8/48)	Met (46/48) Lys (2/48)
	SSI	10/34	Cys	Pro	Met (6/10) Lys (3/10)	Val (8/10) Ile (1/10) His (1/10)	Tyr (9/10) Phe (1/10)	Asp
	Potato 2	21/22	Cys	Pro	Leu (9/21) Arg (9/21)	Asn (17/21) Phe (2/21) Tyr (2/21)	Cys	Asp
	Ascaris	1/6	Cys	Pro	Leu	Met	Cys	Arg
Thr(P2)			Aaa	Thr	Aaa	Aaa	Aaa	Aaa
	Kazal	108/156	Cys	Thr	Leu (40/108) Met (31/108)	Glu (90/108) Asp (9/108)	Tyr (78/108) Asp (7/108)	Arg (51/108) Met (20/108)
	Potato 1	48/53	Val (47/48) Ala (1/48)	Thr	Met (18/48) Lys (8/48)	Asp (37/48) Glu (11/48)	Tyr (23/48) Phe (12/48)	Arg (42/48) Lys (3/48)
	Bowman-Birk 1st domain	59/63	Cys	Thr	Lys (39/59) Arg (14/59)	Ser (58/59) Met (1/59)	Ile (29/59) Met (14/59)	Pro
	2nd domain	45/60	Cys (44/45) Ser (1/45)	Thr	Arg (30/45) Phe (8/45)	Ser	Met (17/45) Ile (15/45)	Pro
	SSI	19/34	Cys	Thr	Arg (10/19) Lys (9/19)	Glu (15/19) Gln (4/19)	Tyr (11/19) Trp (7/19)	Ala (7/19) Asn (5/19)
	Ecotin	5/5	Ser	Thr	Met	Met	Ala (4/5) His (1/5)	Cys
	Ascaris	2/6	Cys	Thr	Arg	Glu	Cys (1/2) Thr (1/2)	Lys
Cys(P2)			Aaa	Cys	Aaa	Aaa	Aaa	Aaa
	BPTI	131/131	Pro (124/131) Ile (3/131) Leu (3/131)	Cys	Lys (73/131) Arg (40/131)	Ala (116/131) Gly (10/131)	Arg (61/131) Met (30/131)	Ile (98/131) Met (18/131)
	Hirustasin	31/31	Arg (13/31) Asn (12/31)	Cys	Arg (26/31) Ala (2/31)	Lys (12/31) Val (7/31)	Thr (12/31) His (9/31)	Cys
	Elafin	11/13	Arg (9/11) Gln (2/11)	Cys	Leu (4/11) Ala (3/11)	Met (9/11) Leu (2/11)	Leu (5/11) Val (3/11)	Asn
Pro(P3)			Pro	Aaa	Aaa	Aaa	Aaa	Aaa
	STI	37/37	Pro	Ile (27/37) Val (7/37)	Arg (24/37) Thr (11/37)	Ile (36/37) Phe (1/37)	Thr (14/37) Ser (8/37)	Ser (21/37) Pro (16/37)

*Amino acid sequences of homologous inhibitors representing 10 families were taken from GenBank, PDB, SwissProt, and PIR databases and analyzed for occurrence of similar sequential features. The four most popular sequential motifs are included in the table. Aaa denotes any amino acid. In "No. of sequences," first entry denotes how often a particular sequential motif occurs in the total number of family sequences. The second entry is the total number of family sequences considered. For a given position of a particular family, only two most popular amino acids are given (first entry) of a total number of sequences fulfilling particular sequential motif (second entry).

further strongly restricted in particular families (e.g., squash inhibitors have exclusively Ile at P1′; *Ascaris* and potato 2 families have Cys at P2′; only Asp is found at P3′ in potato 2 and SSI families; first and second domains of Bowman-Birk inhibitors have Pro at P3′ and Ser at P1′ [with single exceptions]). Cys and/or Pro either at P3 or P2 seem to rigidify the loop conformation and protect it against irreversible proteolysis. Thr at P2 donates a hydrogen bond to the side chain of P1′ across the P1-P1′ peptide bond in different inhibitor structures belonging to the Thr(P2) group. Interestingly, in all cases studied this hydrogen bond is significantly shortened upon complex formation, which should lead to a higher association energy. The Thr(P2) motif is, in all but one family (potato 1), associated with the presence of Cys at P3, which leads to rigidification of the loop similarly as in the Cys(P3)-Pro(P2) motif. However,

because Cys at P3 is not absolutely necessary to achieve canonical conformation, we prefer to call this motif Thr(P2) rather than Cys(P3)-Thr(P2).

Many other positions in the sequences of the binding loops are also severely restricted to just a few amino acids, which can be easily noted in Table IX. In this respect, canonical loops are similar to loops existing in noninhibitor proteins (see above), although the sequential motifs used to generate canonical conformation are generally different for inhibitor and noninhibitor proteins.

DISCUSSION

The binding loop of protein inhibitors is usually considered as a highly conserved structural motif responsible for tight binding of serine proteinases.[1,3,5] Using C_α-C_α distances as the criterion of structural similarity, we found that P3-P3′ segment defines most properly the length of the loop in different inhibitor families. The conformational differences among loops of different inhibitors in free (uncomplexed) state were compared using rmsd in atomic coordinates for all main chain atoms (Δr method). As a second method we used rmsd calculations operating in the main chain torsion angles[27] (Δt method). The two methods should give generally similar results, although the torsion angles method is more sensitive to peptide plane rotation (which is not readily identified by the rmsd in atomic coordinates); the Δr method is very useful for short protein segments. When Δr and Δt values calculated for free inhibitor loops were compared, no significant correlation could be noted (Pearson correlation coefficient, r = 0.515). Therefore, both methods combined were used to compare the canonical conformations of the binding loop in the whole data set. Despite similar shapes of the loop and similar mode of docking in the proteinase active site, severe differences up to 2.1 Å (rmsd in atomic coordinates) and 72.3° (rmsd in torsional angles) were observed among individual loops in their free state (Table IV). Karpen et al.[27] suggested a threshold Δt value of approximately 25–35° as differentiating between similar and dissimilar structures. Distribution of pairwise rmsd operating in dihedral angles shows that in the case of free inhibitors, most of the pairs exhibit values larger than 35° (Fig. 6A). This points to large conformational differences between individual loops. Unger et al.[32] classified hexapeptide segments as similar if rmsd between their C_α atoms was below 1 Å. Distribution of pairwise rmsd operating in distances is distinctly bimodal (Fig. 6B). The first maximum (rmsd 0.3–0.7 Å) comprises inhibitors belonging to the same family and also some other closely similar conformations, which can be evaluated in Table IV. The second maximum (1.1–1.9 Å) covers the majority of the pairs. In summary, Δr and Δt values are large and show that canonical loops of free inhibitors cannot be considered as a novel secondary structure element.

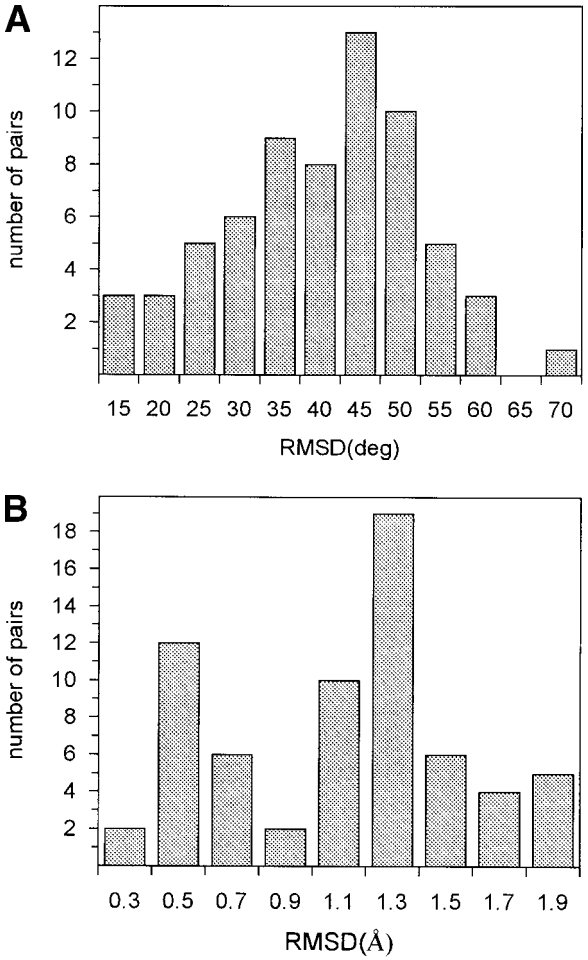

Fig. 6. **A:** Distribution of Δt values measured for each pair of the canonical loops of free inhibitors reported in Table IV. **B:** Distribution of Δr values measured for each pair of the canonical loops of free inhibitors (Table IV).

When Δr and Δt values are calculated for complexed loops, it can be easily seen that they are 20–35% lower than for free inhibitors (Table IV). However, although complexation leads to a more uniform loop conformation, genuine differences among individual loops remain, expressed in respective Δr and Δt values. The effect of complexation is also observed in changes of the side chain χ_1 and χ_2 angles up to 160° (Fig. 4 A–F). The C_α-C_α distances calculated from catalytic serine residue to P4-P3′ of the inhibitor show little (often below 1Å) variability in different complexes, including enzymes from both chymotrypsin and subtilisin families (Table V). Many of these distances, particularly those of low variability, could be used together with fixed side chain χ_1 angle of P1 and P2′ residues for fast computer docking of different proteinase-inhibitor pairs.

The proteinase binding loop can adopt a quite broad range of different conformations, which occur

in many different inhibitor scaffolds and serve to inactivate different proteinases. Is this conformation limited to proteinase inhibitors or does it occur more frequently in protein structures? Unger et al.[32] pointed out that conformations of six amino acid residue oligopeptides derived from PDB do not vary in a continuous manner but are concentrated in specific conformational clusters. In other words, protein structures can be fairly well approximated if reconstructed from such hexameric building blocks. The list of 30 most common hexameric building blocks does not include the proteinase binding loop (which, according to our definition, is also hexameric).[32] This suggests that the loop conformation occurs rather rarely in protein structures (according to Table III of Unger et al.,[32] less than $68/12{,}973 \times 100\% = 0.5\%$).

Search of PDB data base for hexapeptide segments similar to the canonical loop provided 132 unique hexapeptide segments in 121 structurally and functionally unrelated proteins which obey conformational criteria of the canonical conformation. Proteins that contain these loops do not exhibit any specific functional or structural properties; rather, the set of 121 proteins looks as a randomly generated one. To the best of our knowledge, no information is available on antiproteolytic activity for any of these proteins. Partial restrictions regarding the character of the amino acid at particular positions of the loop could be observed. Within this set of protein loops, 40 hexapeptides (Table VII) fulfill two basic structural requirements to inhibit any serine proteinase. First, their conformation resembles that of the canonical binding loop. Second, the third position (equivalent to the P1 position in inhibitors) is always highly (above 70%) exposed to the solvent and capable of penetrating the S1 binding pocket of a serine proteinase.

Is it possible that any of these protein loops might exhibit inhibitory properties against serine proteinase(s)? A positive answer might suggest that these proteins have a built-in mechanism preventing proteolysis. A successful example of the introduction of a proteinase binding loop into another protein scaffold has already been described,[33] suggesting that proteins can be protected against proteolysis in this way. Our docking experiments show that the major reason preventing complex formation are steric conflicts due to slight protrusion of the hexapeptide conformation from the protein scaffold. This is in sharp contrast to the situation observed in canonical inhibitor-serine proteinase complexes that are formed according to the lock-and-key principle.[3] In the case of protein inhibitors, the most numerous contacts are donated from the binding loop and no steric conflicts are observed, which results from pronounced protrusion of the loop. It should be stressed that in many cases the interactions and distances between the hexapeptide part of the protein and the proteinase appear to be well-adjusted, similar to the inhibitor-proteinase complexes. This indicates that by themselves, the hexapeptide segments fulfill the inhibition criteria.

Our modeling experiments should be considered as preliminary. The recently determined structure of the E192Q thrombin mutant with BPTI[34] reveals limits of the rigid body docking procedure. The favorable complex formation with a submicromolar affinity constant is possible due to significant rearrangements of several loops occluding thrombin's active site which could not be anticipated from modeling experiments.[35] Therefore, highly unpredictable conformational changes can accompany complex formation, indicating that our modeling experiments cannot rule out the formation of stable inhibitory complexes. Thus, although improvements in the docking procedure are clearly possible, at this stage we predict that inhibition involving the canonical loops is restricted to protein inhibitors and that other proteins that contain fragments resembling the canonical inhibitory loops will not show antiproteinase activity.

It is generally accepted that short loops in proteins are constrained to a limited number of conformations and can be rather well described. Loops comprising six residues are already long enough to adopt predictable conformations, as observed for example for CDR loops of antibodies.[19,36] This also appears to be true from our analysis of the canonical conformation of the loop, which can adopt a rather wide range of similar conformations. Are there any sequential constraints in the inhibitor loops and in the hexapeptide sequences extracted from PDB that ensure this conformation?

When sequences of the inhibitor loops are compared among different families, they show a considerable level of variability (Table II). Similarly, the amino acid sequences of the binding loops from orthologous inhibitors (most intensively studied for ovomucoid third domains)[37] are also very different and, in some cases (in analogy to CDR loops of antibodies), hypervariable. It appears that the canonical conformation is very tolerant of multiple amino acid substitutions. Nevertheless, 91% of inhibitor sequences analyzed can be grouped into four major sequential motifs (Table IX).

Compared with inhibitors, the amino acid sequences of hexapeptides are generally different and do not show similar sequential motifs. Of the 40 hexapeptides reported in Table VII, only 9 can be assigned to any of the inhibitor sequential motifs. However, in analogy to the sequential motifs, there are clear, although not very strong propensities for certain amino acids to occupy a particular position of the hexapeptide. This shows that in noninhibitory proteins, the loop conformation is achieved in a sequentially different way.

Despite sequential variability, several structural features that determine the canonical conformation of the loop both in inhibitors and in other proteins can be distinguished. They include main chain hydrogen bonds both within the P3-P3′ segment and with the scaffold region, P3-P4 and P3′-P4′ hydrophobic interactions, and either hydrophobic or polar interactions involving P1′ residue. In the case of hexapeptides, however, the system of hydrogen bonds is somewhat better developed and the P1′ side chain is usually apolar.

ACKNOWLEDGMENTS

J.O. thanks Professors Michael Laskowski, Jr. and Wolfram Bode for numerous discussions as well as Dr. Michal Dadlez for critical reading of the manuscript.

REFERENCES

1. Read, R., James, M.N.G. Introduction to the proteinase inhibitors: X-ray crystallography. In: "Proteinase Inhibitors." Barrett, A.J., Salvesen, G. (eds.). Amsterdam: Elsevier, 1986:301–336.
2. Janin, J., Chothia, C. The structure of protein-protein recognition sites. J. Biol. Chem. 265:16027–16030, 1990.
3. Bode, W., Huber, R. Natural protein proteinase inhibitors and their interaction with proteinases. Eur. J. Biochem. 204:433–451, 1992.
4. Jones, S., Thornton, J.M. Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA 93:13–20, 1996.
5. Laskowski, M., Jr., Kato, I. Protein inhibitors of proteinases. Annu. Rev. Biochem. 49:593–626, 1980.
6. Laskowski, M., Jr. Protein inhibitors of serine proteinases—Mechanism and classification. In: "Nutritional and Toxicological Significance of Enzyme Inhibitors." Friedman, M. (ed.). New York: Plenum, 1986:1–17.
7. Schechter, P., Berger, A. On the size of the active site in proteases. I. Papain. Biochem. Biophys. Res. Commun. 27:157–162, 1967.
8. Krystek, S., Stouch, T., Novotny, J. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. J. Mol. Biol. 234:661–679, 1993.
9. Lu, W., Apostol, I., Qasim, M.A., et al. Binding of amino acid side chains to S_1 cavities of serine proteinases. J. Mol. Biol. 266:441–461, 1997.
10. Wells, J.A. Additivity of mutational effects in proteins. Biochemistry 29:8509–8517, 1990.
11. Qasim, M.A., Ganz, P.J., Saunders, C.W., Bateman, K.S., James, M.N.G., Laskowski, M., Jr. Interscaffolding additivity. Association of P_1 variants of eglin c and of turkey ovomucoid third domain with serine proteinases. Biochemistry 36:1598–1607, 1997.
12. Kabsch, W., Sander, C. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. Proc. Natl. Acad. Sci. USA 81:1075–1078, 1984.
13. Cohen, B.I., Presnell, S.R., Cohen, F.E. Origins of structural diversity within sequentially identical hexapeptides. Protein Sci. 2:2134–2145, 1993.
14. Rose, G.D., Gierasch, L.M., Smith, J.A. Turns in peptides and proteins. Adv. Protein Chem. 37:1–109, 1985.
15. Karpen, M.E., Neet, K.E., de Haseth, P.L. A common pentapeptide conformation occurs in viral acid proteases and other proteins. J. Mol. Biol. 216:201–206, 1990.
16. Leszczynski, J.F., Rose, G.D. Loops in globular proteins: A novel category of secondary structure. Science 234:849–855, 1986.
17. Thornton, J.M., Sibanda, B.L., Edwards, M.S., Barlow, D.J. Analysis, design and modification of loop regions in proteins. BioEssays 8:63–69, 1988.
18. Sibanda, B.L., Blundell, T.L., Thornton, J.M. Conformation of β-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. J. Mol. Biol. 206: 759–777, 1989.
19. Martin, A.C.R., Thornton, J.M. Structural families of homologous proteins: Automatic classification, modelling and application to antibodies. J. Mol. Biol. 263:800–815, 1996.
20. Oliva, B., Bates, P.A., Querol, E., Aviles, F.X., Sternberg, M.J.E. An automated classification of the structure of protein loops. J. Mol. Biol. 266:814–830, 1997.
21. Rufino, S.D., Donate, L.E., Canard, L.H.J., Blundell, T.L. Predicting the conformational class of short and medium size loops connecting regular secondary structures: Application to comparative modelling. J. Mol. Biol. 267:352–367, 1997.
22. van Vlijmen, H.W.T., Karplus, M. PDB-based protein loop prediction: Parameters for selection and methods for optimization. J. Mol. Biol. 267:975–1001, 1997.
23. Hutchinson, E.G., Thornton, J.M. PROMOTIF—A program to identify and analyze structural motifs in proteins. Protein Sci. 5:212–220, 1996.
24. Lee, B.K., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
25. Hubbard, S.J., Thornton, J.M. NACCESS computer program. Department of Biochemistry & Molecular Biology, University College, London, 1993.
26. Chothia, C. The nature of the accessible and buried surface in proteins. J. Mol. Biol. 105:1–14, 1976.
27. Karpen, M.E., de Haseth, P.L., Neet, K.E. Comparing short protein substructures by a method based on backbone torsion angles. Proteins 6:155–167, 1989.
28. Jones, T.A., Zou, J-Y, Cowan, S.W., Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A. 47:110–119, 1991.
29. Hubbard, S.J., Campbell, S.F., Thornton, J.M. Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. J. Mol. Biol. 220:507–530, 1991.
30. Tramontano, A., Chothia, C., Lesk, A.M. Structural determinants of the conformation of medium-sized loops in proteins. Proteins 6:382–394, 1989.
31. Nielsen, K.J., Alewood, D., Andrews, J., Kent, S.B.H., Craik, D.J. An ^{1}H NMR determination of the three dimensional structures of mirror image forms of a Leu-5 variant of the trypsin inhibitor from *Ecballium elaterium* (EETI II). Protein Sci. 3:291–302, 1994.
32. Unger, R., Harel, D., Wherland, S., Sussman, J.L. A 3D building blocks approach to analyzing and predicting structure of proteins. Proteins 5:355–373, 1989.
33. Wolfson, A.J., Kanaoka, M., Lau, F.T.K., Ringe, D. Insertion of an elastase-binding loop into interleukin-1β. Protein Eng. 4:313–317, 1991.
34. van de Locht, A., Bode, W., Huber, R., et al. The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. EMBO J. 16:2977–2984, 1997
35. Bode, W., Turk, D., Karshikov, A. The refined 1.9-Å X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human α-thrombin: Structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. Protein Sci. 1:426–471, 1992.
36. MacCallum, R.M., Martin, A.C.R., Thornton, J.M. Antibody-antigen interactions: Contact analysis and binding site topography. J. Mol. Biol. 262:732–745, 1996.
37. Laskowski, M., Jr., Kato, I., Ardelt, W., et al. Ovomucoid third domains from 100 avian species. Isolation, sequence and hypervariability of enzyme-inhibitor contact residues. Biochemistry 26:202–221, 1987.