doi:10.1016/S0022-2836(03)00245-6 *J. Mol. Biol.* (2003) **328**, 205–219

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Homologous Proteins with Different Folds: The Three-dimensional Structures of Domains 1 and 6 of the Multiple Kazal-type Inhibitor LEKTI

Thomas Lauber[1], Axel Schulz[2], Kristian Schweimer[1], Knut Adermann[2] and Ute C. Marx[1]*

[1]*Lehrstuhl für Biopolymere Universität Bayreuth Universitätstrasse 30 D-95440 Bayreuth, Germany*

[2]*IPF PharmaCeuticals GmbH Feodor-Lynen-Strasse 31 30625 Hannover, Germany*

**Corresponding author*

We have determined the solution structures of recombinant domain 1 and native domain 6 of the multi-domain Kazal-type serine proteinase inhibitor LEKTI using multi-dimensional NMR spectroscopy. While two of the 15 potential inhibitory LEKTI domains contain three disulfide bonds typical of Kazal-type inhibitors, the remaining 13 domains have only two of these disulfide bridges. Therefore, they may represent a novel type of serine proteinase inhibitor. The first and the sixth LEKTI domain, which have been isolated from human blood ultrafiltrate, belong to this group. In spite of sharing the same disulfide pattern and a sequence identity of about 35% from the first to the fourth cysteine, the two proteins show different structures in this region. The three-dimensional structure of domain 6 consists of two helices and a β-hairpin structure, and closely resembles the three-dimensional fold of classical Kazal-type serine proteinase inhibitors including the inhibitory binding loop. Domain 6 has been shown to be an efficient, but non-permanent serine proteinase inhibitor. The backbone geometry of its canonical loop is not as well defined as the remaining structural elements, providing a possible explanation for its non-permanent inhibitory activity. We conclude that domain 6 belongs to a subfamily of classical Kazal-type inhibitors, as the third disulfide bond and a third β-strand are missing. The three-dimensional structure of domain 1 shows three helices and a β-hairpin, but the central part of the structure differs remarkably from that of domain 6. The sequence adopting hairpin structure in domain 6 exhibits helical conformation in domain 1, and none of the residues within the putative P3 to P3′ stretch features backbone angles that resemble those of the canonical loop of known proteinase inhibitors. No proteinase has been found to be inhibited by domain 1. We conclude that domain 1 adopts a new protein fold and is no canonical serine proteinase inhibitor.

Keywords: LEKTI; SPINK5; NMR structure; proteinase inhibitor; Kazal-type

Abbreviations used: BUSI IIA, bull seminal plasma acrosin inhibitor IIA; CD, circular dichroism; DIPSI, decoupling in the presence of scalar interactions; DQF-COSY, double quantum filtered correlation spectroscopy; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; ESI-MS, electrospray ionization mass spectrometry; GARP, globally optimized alternating-phase rectangular pulses; HF, hemofiltrate; HNHA, 3D heteronuclear ($^{1}H-^{15}N-^{1}H$) shift correlation double resonance NMR experiment; HSQC, heteronuclear single quantum coherence NMR experiment; IC_{50}, inhibitory concentration causing 50% enzyme inactivation; LEKTI, lympho-epithelial Kazal-type-related inhibitor; NOE, nuclear Overhauser effect, also used for NOESY cross-peak; NOESY, NOE spectroscopy; rHF6478, recombinant LEKTI domain 1; RP-HPLC, reversed-phase high-performance liquid chromatography; SPINK5, serine protease inhibitor, Kazal-type 5, gene encoding LEKTI; SVD, singular value decomposition; TOCSY, total correlation spectroscopy; WATERGATE, water suppression by gradient-tailored excitation; PDB, Protein Data Bank.

E-mail address of the corresponding author: ute.marx@uni-bayreuth.de

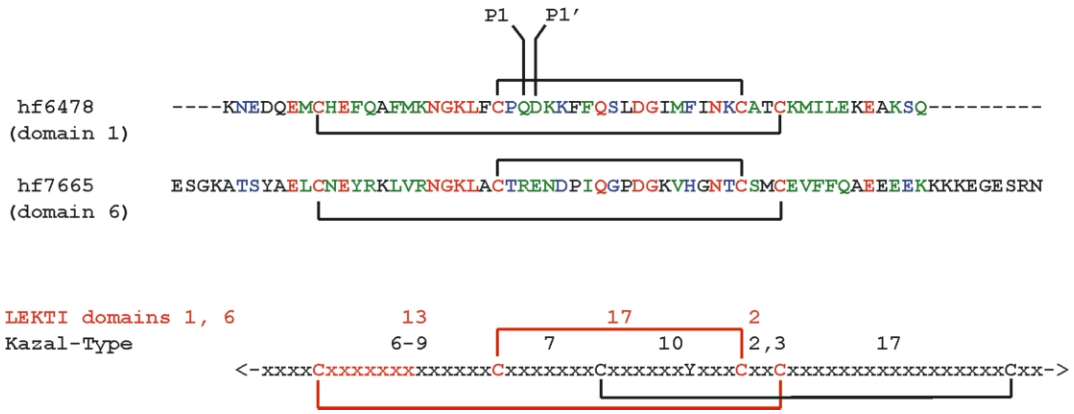

Figure 1. Top: primary structure alignment[68] of LEKTI domain 1 (residues 23 to 77) and domain 6 (residues 356 to 423), as found in human hemofiltrate. Identical residues are displayed in red, conserved and semi-conserved substitutions are colored green and blue, respectively. The 1–4, 2–3 disulfide bond pattern and the putative P1–P1′ sites of both LEKTI domains are indicated. Bottom: schematic comparison of LEKTI domains 1 and 6 with a typical Kazal-type inhibitor indicating conserved disulfide bonds and the number of amino acids spacing the cysteine residues.

Introduction

Proteinases are essential for the regulation of most physiological processes, but can also be potentially dangerous to their proteogenic environment, and therefore, must be carefully controlled. Once a proteinase is activated by liberation from its zymogen, the *in vivo* activity is controlled by specific endogenous proteinase inhibitors. Many pathological effects originate from misregulated endogenous proteinases or proteinases encoded by bacteria, viruses, or parasites, underlining the important function of the respective inhibitors.[1] In the case of serine proteinase inhibitors at least 18 non-homologous families can be distinguished according to their sequences, their disulfide pattern and their three-dimensional structures.[2]

Two previously unknown polypeptides of 55 and 68 amino acid residues (HF6478, 55 amino acid residues, and HF7665, 68 amino acid residues, *h*emofiltrate peptides with molecular masses of 6478 and 7665 Da, respectively) have been isolated from human blood ultrafiltrate, and were shown to represent domain 1 and domain 6 of a large precursor protein termed lympho-epithelial Kazal-type-related inhibitor (LEKTI), which contains 15 potential serine proteinase inhibitory domains.[3] Two out of these domains resemble typical Kazal-type serine proteinase inhibitors, as deduced from their primary structure and characteristic pattern of six cysteine residues. The other 13 domains share partial homology to Kazal-type inhibitors, but lack one of the three conserved Kazal-typical disulfide bridges. Thus, it has been suggested that they may represent a novel type of serine proteinase inhibitor. In addition, all 15 LEKTI domains contain a longer sequence stretch between the first two cysteine residues compared to typical Kazal-type inhibitors (13 and 12, respectively, instead of six to nine residues; Figure 1).

No proteinase has so far been found to be inhibited by domain 1, whereas domain 6 has been shown to inhibit trypsin efficiently, but temporarily with an apparent IC_{50} of approximately 150 nM.[3–5] Since the LEKTI gene (SPINK5) is mainly expressed in mucous epithelia and associated glands as well as in lymphoid organs, the digestive enzyme trypsin is unlikely to be a physiological target proteinase of LEKTI or its domains.

A possible role of LEKTI in the regulation of T-lymphocyte differentiation or antimicrobial protection is suspected, but the target proteinases of LEKTI still remain to be determined.[5] Interestingly, it appears that defects in SPINK5 are directly correlated with the congenital skin disorder disease Netherton Syndrome. Those defects are various mutations leading to premature termination codons,[6] generation of altered splice products,[7] and point mutations resulting in exchanges of single amino acid residues.[8] Especially, an E420K mutation (numbering according to full-length LEKTI) shows a significant association with atopic dermatitis and a weak association with asthma as demonstrated with two independent panels of families. Additionally, the point mutations N368S and D386N were identified, all three located in domain 6 of LEKTI, with the latter amino acid exchange being part of the putative canonical loop.[8,9] These results demonstrate, that significant genotype–phenotype correlations exist, but the actual physiological function of LEKTI or its domains still remains unclear. Komatsu *et al.*[10] localized mRNA encoding for LEKTI in the uppermost epidermis of normal human skin, and they determined elevated stratum corneum hydrolytic activity in Netherton Syndrome patients suggesting a regulation of desquamation by LEKTI. This is a first hint for a natural target proteinase of LEKTI or its domains.[10]

Since LEKTI domains 1 and 6 circulate in human blood, they represent naturally occurring processed forms of LEKTI. Despite their sequence identity of 35% within the region from the first to the fourth cysteine and their identical cysteine

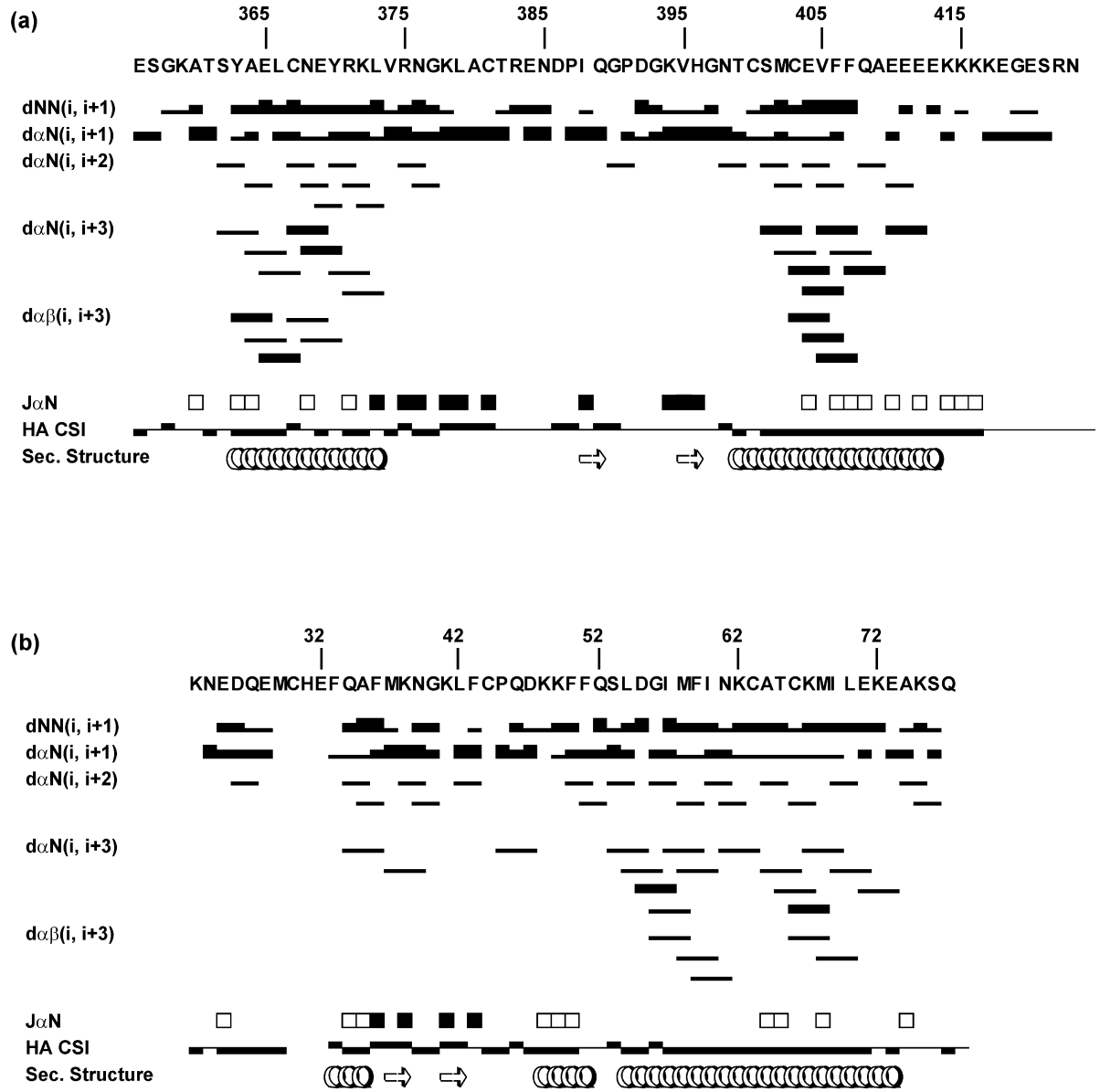

Figure 2. Secondary structure of domain 6 (a) and domain 1 (b). Summary of sequential and medium-range NOEs and chemical shift indices (HA CSI). The relative strength of the NOEs, classified as weak, medium and strong from cross-peak intensities of 2D-NOESY spectra (domain 6, domain 1) and the ^{15}N-NOESY-HSQC spectrum (domain 1), is indicated by the strength of the horizontal bars. Positive and negative chemical shift indices are typical of extended and helical structure, as denoted by rectangles above and below the axis. Values of $^{3}J(H^{N},H^{\alpha}) < 6$ Hz and >8 Hz, typical of helical and extended structure, are shown as white and black squares, respectively. The secondary structures of both proteins as derived from structure calculations are depicted in the bottom line.

connectivity pattern[3,11] (Figure 1), they exhibit different activities against trypsin and other proteinases. These features and the fact that some of the point mutations, which seem to be related to Netherton Syndrome, are located within the sequence of domain 6, make them both interesting target molecules for structure determination. Here we report the three-dimensional structures of recombinant domain 1 and native domain 6 of LEKTI. Despite their significant sequence homology the proteins show different overall folds.

Results

NMR spectroscopy

One-dimensional ^{1}H NMR spectra at different temperatures revealed an optimal resonance dispersion at 298 K for both proteins. Therefore, all NMR experiments were carried out at this temperature. The spin systems of all amino acid residues of domain 6 except for the COOH-terminal Asn were assigned sequence-specifically according to standard procedures[12] using data

Table 1. Structural statistics

	Domain 1	Domain 6
Experimental restraints used for the structure calculation		
Intraresidual NOEs	102	68
Interresidual NOEs	421	616
Sequential	203	279
Medium-range	162	207
Long-range	56	130
Dihedral angle restraints	14	24
Molecular dynamics simulation statistics		
Energies (kcal/mol)		
Total	79.71 ± 1.43	90.21 ± 1.77
Bond lengths	2.01 ± 0.17	1.94 ± 0.14
Bond angles	62.06 ± 0.79	72.93 ± 0.98
Improper angles	7.01 ± 0.21	8.28 ± 0.28
van-der-Waals repulsion	4.31 ± 0.53	2.63 ± 0.56
Distance restrains	4.31 ± 0.41	4.42 ± 0.68
dihedral angle restraints	0.00053 ± 0.0012	0.01 ± 0.02
RMSDs from ideal distances (Å)		
Bond lengths	0.0015 ± 0.00006	0.0013 ± 0.00005
Distance restraints	0.013 ± 0.0006	0.011 ± 0.0009
RMSDs from ideal angels (°)		
Bond angels	0.49 ± 0.0032	0.50 ± 0.0038
Dihedral angel restraints	0.0038 ± 0.007	0.05 ± 0.07

Atomic RMSDs from the average structure (Å)

	Domain 1		Domain 6	
	Back bone	Heavy atoms	Back bone	Heavy atoms
Overall	1.80	2.45	2.77	3.26
Structured region[a]	0.57	0.99	0.62	1.13
Regular secondary structure[b]	0.47	0.83	0.43	0.82

Except for the experimental restraints all values are average values over the 21 accepted structures in the form average value ± standard deviation.
[a] Residues 29–74 for domain 1, residues 363–413 for domain 6.
[b] Residues 34–38, 40–41, 48–52, 54–74 for domain 1, residues 363–371, 388–413 for domain 6.

from homonuclear 2D spectra. An almost complete sequence-specific assignment of the backbone as well as side-chain resonances of domain 1 was obtained from heteronuclear and homonuclear spectra except for the spin systems of Cys30, His31, and Glu32. The spin systems of all leucine residues (Leu42, 54, and 70) were assigned from homonuclear spectra. All Xxx-proline peptide bonds of both domains are in *trans* conformation, as evidenced by unambiguous Hα(i) Pro-Hδ($i+1$) nuclear Overhauser effect (NOE) signals in the NOE spectroscopy (NOESY) spectra and the fact that Hα(i) Pro-Hα($i+1$) connectivities were absent.[12]

Secondary structures of LEKTI domains 1 and 6

^{1}Hα chemical shifts

^{1}Hα chemical shifts of both peptides were analyzed by the chemical shift index strategy[13,14] using the random coil values given by Wishart *et al.*[15] For all spin systems followed by a Pro residue, corrected random coil values were used for the analysis.[15] Only deviations of more than 0.1 ppm from the random coil values were taken into account for secondary structure determination, represented by +1 and −1 values (Figure 2(a) and (b), next to last line, respectively). Two potential helical regions were obtained for domain 6: from Tyr363 to at least Leu366, and from Thr399 to Lys416. Furthermore, indication for extended structure is found from Lys378 to Cys381 of domain 6 (Figure 2(a)). For domain 1 upfield shifted ^{1}Hα resonances suggest helical regions from Glu25 to at least Met29, from Asp47 to Phe50, and from Leu54 to Glu71. A short stretch with indications for extended structure can be deduced for Phe36 to Lys38 of domain 1 (Figure 2(b)).

Coupling constants

For domain 6 a total of 49 $^3J(H^N,H^\alpha)$ coupling constants was determined from a high digital resolution double quantum filtered correlated spectroscopy (DQF-COSY) spectrum, with 24 coupling constants unambiguously transformable into ϕ-angle constraints. A total of 14 helix-typical coupling constants (ϕ-angle: $-60(\pm 20)°$) were identified within the two putative helices. Around

Leu373 to Leu379 and Ile388 to His396 another ten coupling constants characteristic for extended structure (ϕ-angle: $-120(\pm 40)°$) were found (Figure 2(a)).

A total of 20 $^3J(H^N,H^\alpha)$ coupling constants were determined from the HNHA spectrum of ^{15}N-labeled recombinant domain 1, with ten helix-typical coupling constants in the NH_2-terminal region up to Ala35, in the central region from Lys48 to Phe50 and in the COOH-terminal region around Ala64 to Ala74. Four coupling constants characteristic for extended structure are located in the area from Phe36 to Phe43 (Figure 2(b)).

Medium range NOEs

The patterns of sequential and medium range NOE connectivities for domain 6 and domain 1 corroborate the results from chemical shift data and coupling constants (Figure 2). For domain 6, a high density of $d_{\alpha N}(i,i+3)$ and $d_{\alpha\beta}(i,i+3)$ NOEs, which are characteristic for helical structure, was detected from Ser362 to Leu373 and Thr399 to Glu413 (Figure 2(a)). Similar NOEs were found for domain 1 in the sequence stretch from Gln52 to Ala74, and a few isolated $(i,i+3)$ NOEs were determined in the central region of the protein (Figure 2(b)). These data together with the ^{1}Hα chemical shifts and the coupling constants suggest a COOH-terminal helix for both proteins and an NH_2-terminal one for domain 6.

Structure calculations

Structure calculation of domain 6 was based on a total number of 684 distance constraints comprising 68 non-trivial intraresidual, 279 sequential, 207 medium range, and 130 long range NOE contacts, as well as 24 dihedral angle constraints (Table 1). For domain 1, the respective number of distance constraints is 523, comprising 102 intraresidual, 203 sequential, 162 medium range, and 56 long range NOEs together with 14 dihedral angle restraints (Table 1). The distance constraints of domain 6 were derived from homonuclear NOESY spectra. For domain 1, a ^{15}N-NOESY-HSQC and homonuclear NOESY spectra were used, each at 298 K. *Trans* conformation of all Xxx-Pro peptide bonds was considered. Structure calculations were performed iteratively, including only unambiguous distance restraints in the first rounds. A set of 60 structures for both domains was calculated in each round of structure calculations. The solution conformation of domains 6 and 1 (Figure 3) is represented by a family of 21 structures each, that was chosen from the final rounds of calculations by the criterion of the lowest overall energy. The overall energies and RMSD values from ideal geometry of both proteins are similar, and small deviations from ideal bond lengths and bond angles reflect a good covalent geometry (Table 1). The structures have no distance violations greater than 0.13 Å and no dihedral angle violations greater than five degrees. The best fit superimposition of the structures (Figure 3), as well as the corresponding RMSD values (Figure 4, Table 1) demonstrate well-defined tertiary structures for LEKTI domains 1 and 6. Both proteins, however, exhibit exposed and flexible termini, which might relate to post-translational processing of the complete 15-domain precursor protein LEKTI or parts of it.

Tertiary structures

LEKTI domain 6

A best-fit backbone superimposition of the 21 structures of lowest overall energy of domain 6 together with one representative structure indicating regular secondary structural elements are given in Figure 3(a) and (b). The RMSD value of the 21 converging structures is 2.77 Å for the backbone heavy atoms and 3.26 Å for all heavy atoms. Omission of the unstructured terminal residues resulted in lower RMSD values of 0.62 Å and 1.13 Å, respectively, for residues 363 to 413 (Table 1). The three-dimensional structure of domain 6 consists of two helices and a β-hairpin structure, with their relative almost orthogonal orientation defined by the two disulfide bonds and a distinct hydrophobic core. The NH_2-terminal helix extends from Tyr363 to Arg371, with the second half of the helix having 3_{10} helix character in some structures of the calculated family. The COOH-terminal helix represents a stable α-helix from Thr399 to at least Glu413, in some cases followed by a short 3_{10} helix. The end of the COOH-terminal helix is not well defined as a result of missing NOE data due to severe spectral overlap for this region. The analyses of ^{1}Hα chemical shifts, as well as helix-typical coupling constants imply an extension of the COOH-terminal helix up to residue Lys416. Most structures show a β-hairpin from Ile388 to His396, as β-sheet-typical backbone/backbone NOEs were present, e.g. a HαHα NOE between Gln389 and Val395. The experimentally determined secondary structure of domain 6 is summarized in the lower part of Figure 2(a). The hydrophobic core, which is indicated by a high number of long range NOE connectivities, mainly comprises the residues Tyr370, Leu379, Phe406, and Phe407, and is additionally stabilized by the two disulfide bonds. The well-defined secondary structure, the disulfide bonds, and the hydrophobic core act as a supporting scaffold for the proteinase binding loop with the side-chains of the P2 to P3′ residues (numbering according to the nomenclature of Schechter & Berger[16]), i.e. Thr382 to Asp386, being exposed. As deduced from RMSD values per residue (Figure 4(a)), the exposed binding loop and the preceding residues exhibit higher local disorder than residues located in regular secondary structure. From Figure 3(a) a second backbone conformation for the binding loop residues can be deduced for three out of the

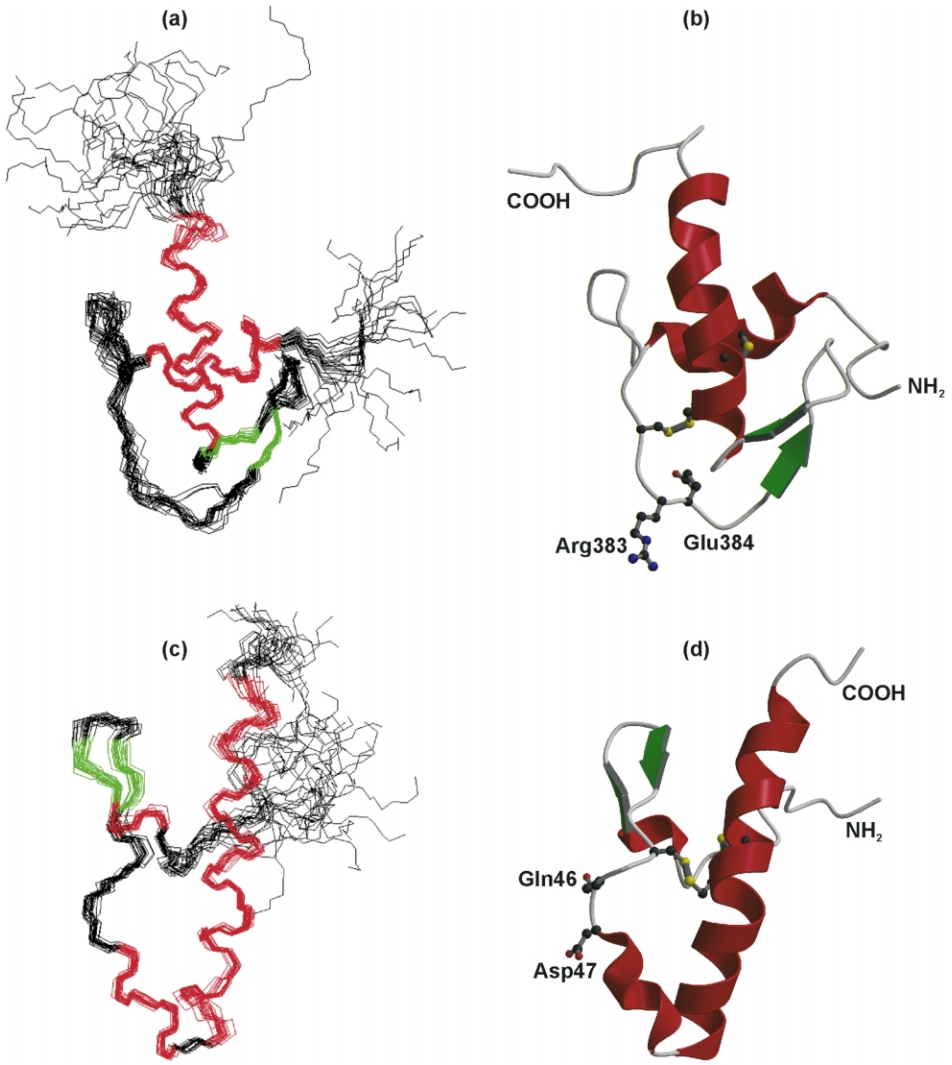

Figure 3. (a) and (c) Best-fit backbone superimposition of the 21 lowest overall energy structures of domain 6 (a) and domain1 (c) (generated with MOLMOL[63]). Helical elements are displayed in red, β-sheet structure is colored in green. (b) and (d) Schematic drawing of a representative structure of domain 6 (b) and domain 1 (d), indicating regular secondary structural elements (same color code as in (a) and (c)). The heavy side-chain atoms of the P1 and P1′ residues of domain 6 (Arg383 and Glu384) and the corresponding residues of domain 1 (Gln46 and Asp47) are shown (generated with MOLSCRIPT[65] and Raster3D[66,67]).

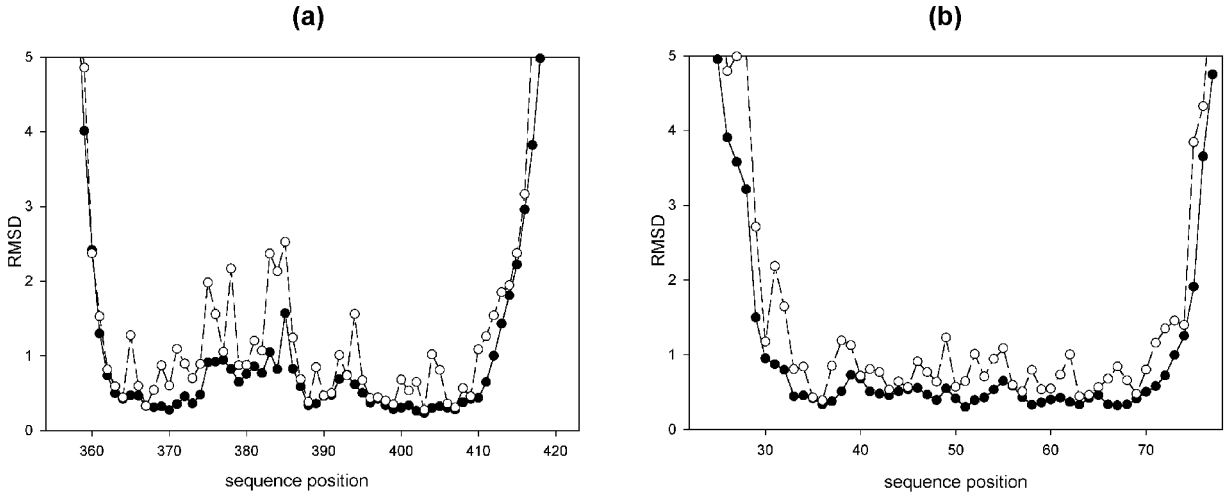

Figure 4. Atomic RMSD values from the average structures of domain 6 (a) and domain 1 (b). The average structures were calculated over the residues 363 to 413 for domain 6, and residues 29 to 74 for domain 1. Values for the backbone heavy atoms and all heavy atoms are displayed as filled and open circles, respectively.

Table 2. ϕ and ψ angles of P3 to P3′ residues of domain 6 and domain 1 in comparison to characteristical backbone angle values of the canonical binding loop of proteinase inhibitors[69]

Loop position	Canonical loop		Domain 6		Domain 1	
P3	ϕ	−140– −120	CYS 381	−136.9 ± 26.7	CYS 44	−51.9 ± 3.8
	ψ	140– 170		179.2 ± 14.9		−46.5 ± 2.5
P2	ϕ	−100– −60	THR 382	−105.6 ± 19.8	PRO 45	−62.8 ± 1.6
	ψ	139– 180		−176.5 ± 8.4		79.9 ± 4.5
P1	ϕ	−120– −95	ARG 383	−84.5 ± 16.3	GLN 46	−118 − 6 ± 8.7
	ψ	10– 50		21.2 ± 38.9		−79.5 ± 40.8
P1′	ϕ	−100– −60	GLU 384	−52.9 ± 61.7	ASP 47	−135.0 ± 43.7
	ψ	139– 180		150.3 ± 13.9		68.9 ± 7.0
P2′	ϕ	−140– −99	ASN 385	−110.7 ± 21.5	LYS 48	−61.8 ± 5.9
	ψ	70– 120		59.2 ± 6.6		−8.7 ± 5.1
P3′	ϕ	−140– −99	ASP 386	−142.1 ± 60.6	LYS 49	−41.7 ± 0.9
	ψ	70– 120		78.8 ± 7.9		−26.9 ± 0.7

21 structures. Omission of these structures for the calculation of the RMSD values from the average structure (Figure 4(a)) only improves the backbone and side-chain RMSD values of residue Asn385 and the side-chain RMSD value of Glu384. But nevertheless, the RMSD values for residues Arg375 to Asp386 are still increased compared to the remaining structured regions. Closer examination of the ϕ and ψ angles of the P3 to P3′ residues reveals that the second conformation can mainly be ascribed to different ϕ angles of Glu384 and Asp386 compared to the remaining 18 structures. All $^3J(H^N,H^\alpha)$ couplings found for residues 381 to 386 are in agreement with the canonical loop geometry (only $^3J(H^N,H^\alpha)$ of Glu384 could not be determined due to spectral overlap): The $^3J(H^N,H^\alpha)$ coupling constant for Cys381 (P3′) unambiguously indicates extended structure. The values for the P2, P1, P2′ and P3′ sites could not unambiguously be transformed into ϕ angle restraints but are compatible with the backbone angles shown in Table 2.

LEKTI domain 1

A best-fit backbone superimposition of the 21 structures of lowest overall energy of domain 1 together with one representative structure indicating regular secondary structural elements are given in Figure 3(c) and (d). The RMSD value of the 21 converging structures is 1.80 Å for the backbone heavy atoms and 2.45 Å for all heavy atoms. Omission of the unstructured terminal residues resulted in lower RMSD values of 0.57 Å and 0.99 Å, respectively, for residues 29 to 74 (Table 1). Common features of the experimentally determined structures of domain 1 are three helices and a β-hairpin. The first helix extends from Gln34 to Phe36, directly followed by the β-hairpin motif from Met37 to Leu42. For few structures, helical structure is identified for residues Glu25 to Glu28. Due to the unassigned spin systems of Cys30 to Glu32, it is not clear if the NH_2-terminal helix extends from Glu25 to Phe36. A 3_{10} helix extends from Lys48 to Gln52, followed by a kink and a regular α-helix from Leu54 to Ala74. The disulfide bonds and a hydrophobic core, mainly comprising the residues Phe33, Phe36, Leu42, Phe59, and Leu70, define the relative orientation of the secondary structural elements and the overall shape of domain 1, which is dominated by the long COOH-terminal helix (Figure 3(c) and (d)).

Discussion

The structure of LEKTI domain 6 resembles that of classical Kazal-type inhibitors

The order and arrangement of the secondary structural elements, as well as the overall structure of LEKTI domain 6 closely resembles the three-dimensional fold of classical Kazal-type serine proteinase inhibitors, apart from the fact that the third disulfide bond and a third β-strand connecting the COOH terminus with the core region of the protein are missing. The remaining structural elements, i.e. the β-hairpin and the first half of the COOH-terminal helix, however, align well with the respective structural elements of classical Kazal-type inhibitors (C^α RMSD ~1 Å; Figure 5(a)). Also, the putative binding loop of domain 6 resembles that of classical inhibitors and the backbone angles of the corresponding residues are within the typical range for canonical inhibitory loops (Table 2). The backbone geometry of the canonical loop and the preceding residues (Arg375-Pro387), however, is not as well defined as the remaining structural elements of domain 6 (Figures 3(a) and 4(a)). This disorder might be due to a lack of experimental restraints within this region. Assuming the disorder of this region being an intrinsic feature of domain 6, however, would provide an explanation for its partial substrate-like behavior, and thus, for its non-permanent inhibitory activity: During a trypsin-inhibitory assay, cleavage of the scissile peptide bond (P1–P1′) occurs over the time, reducing the overall inhibitor concentration, and therefore, leading to an increasing trypsin activity.[3] The lack of the third disulfide bond and of the third β-strand, as

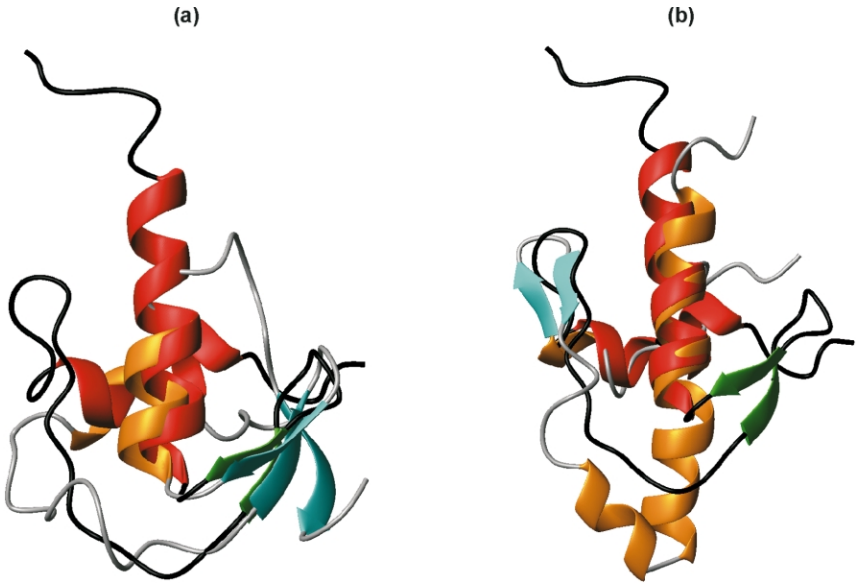

Figure 5. Schematic drawings of the best-fit superimpositions of matching structural elements of domain 6 and BUSI IIA (2bus.PDB) (a) and of domains 1 and 6 (b). For BUSI IIA and domain 1 helical elements are displayed in orange, β-sheet structure is colored in cyan (for domain 6 same color code as in Figure 3).

well as the additional amino acid residues between the first two cysteine residues, might contribute to an intrinsic flexibility of the canonical loop region compared to classical Kazal-type inhibitors. This assumption is supported by the fact that domain 15 of LEKTI containing three disulfide bonds and lacking one residue between the first two cysteine residues compared to domain 6 is a permanent inhibitor of trypsin, not showing a substrate-like behavior.[4,5]

The sequence stretch between the first two cysteine residues consists of four to seven additional residues compared to classical Kazal-type serine proteinase inhibitors. Analyzing Protein Data Bank (PDB) structures of Kazal-type inhibitors, the usually six to nine residues between the first two cysteine residues do not exhibit regular secondary structure. Only in the case of bull seminal plasma acrosin inhibitor IIA (BUSI IIA, PDB code: 2bus[17]) with nine residues in this sequence stretch, a 3_{10} helix for the four residues directly following the first cysteine was identified. Domain 6 also exhibits helical structure in this region, followed by a turn-like loop stabilized by hydrophobic side-chain interactions of Val374 and Leu379. These structural features lead to a shortening of the distance between the first two cysteine residues, thus matching the orientation of the disulfide bridges of Kazal-type inhibitors (Figure 5(a)). A best fit superimposition of the C^{α} atom positions of the cysteine residues of domain 6 with the corresponding atoms of Kazal-type inhibitors results in a C^{α} RMSD value of $\sim$0.7 Å (PDB codes: 2bus: 0.6 Å, 2ovo:[18] 0.7 Å).

Up to now, 18 different serine proteinase inhibitor families have been described.[2] This classification is based on sequence similarities, disulfide patterns, location of the active site, and on topological similarity of the three-dimensional overall structures. All families share the canonical conformation of the inhibitory binding loop, but each family has its own specific three-dimensional fold.[2,19] A DALI protein structure comparison search[20]† with the experimentally determined structure of domain 6 reveals similarity to Kazal-type proteinase inhibitors such as the third domain of the ovomucoid inhibitor. From this result and from our structural studies we conclude, that domain 6 and maybe the other domains of LEKTI that contain only two disulfide bridges instead of three compared to typical Kazal-type inhibitors belong to a subfamily of classical Kazal-type inhibitors, with the exception of domain 1. This conclusion is based on the conserved disulfide connectivity pattern for the remaining two bridges, on the order and arrangement of the secondary structural elements, namely the β-hairpin structure and the COOH-terminal helix, as well as the backbone structure of the exposed binding loop (Figure 5(a)). This is further stressed by the experimental identification of the P1–P1′ site of domain 6, resembling that of classical Kazal-type inhibitors (P. Kreutzmann, personal communication). Differences from classical Kazal-type inhibitors are the additional amino acid residues between the first two cysteine residues, the elongation of the COOH-terminal helix by one or two helix turns and in particular the absence of the third β-strand, which is connected to the other two strands by the third disulfide bridge in classical Kazal-type inhibitors (Figure 5(a)).

Point mutations of domain 6 and their location in the structure

Three point mutations in SPINK5 leading to single amino acid exchanges have been identified in patients with atopy, atopic dermatitis, and/or Netherton syndrome, all being located in LEKTI

† http://www.ebi.ac.uk/dali/

domain 6. These mutation are namely N368S, D386N, and E420K,[8,9] with the latter showing a significant genotype–phenotype correlation.[8] As only persons with atopy, atopic dermatitis, and/or Netherton syndrome have been examined, it is not clear if the observed polymorphisms for N368 and D386 are associated with these diseases or if they also can be found in healthy persons.

Glu420 resides in the unstructured COOH terminus of domain 6. Thus, a particular structure/activity relationship cannot be deduced, but an influence of this substitution on the correct processing of LEKTI and the release of the single inhibitory domain 6 is likely. The point mutation D386N might have an influence on the inhibitory activity and/or on the specificity of domain 6, as this mutation affects the P3′ site. In ten out of the 15 LEKTI domains an aspartate residue is found in the P3′ position. Domains 10 and 8, however, have an asparagine residue in this position, making an influence on proteinase specificity more likely than an influence on inhibitory activity. The point mutation N368S is located in the α-helical region around the first cysteine residue of domain 6 and is involved in side-chain interactions with Glu365, as well as Arg371. As this residue is not very conserved among the LEKTI domains, an influence of the mutation on the helical structure as well as on the overall assembly or stability of the three-dimensional structure can only be suspected.

Domain 1 is no canonical serine proteinase inhibitor and adopts a new protein fold

The three-dimensional structure of domain 1 shows three helices and a β-hairpin structure. Except for the central part of the COOH-terminal helix and the conserved disulfide pattern, domain 1 does not show similarity to classical Kazal-type inhibitors. In contrast to the β-hairpin structure that follows the binding loop in classical Kazal-type inhibitors as well as in domain 6, domain 1 exhibits helical conformation in this region, and only one of the residues within the putative P3 to P3′ stretch features backbone angles that resemble those of the canonical loop of proteinase inhibitors, in agreement with the experimentally observed $^{3}J(H^{N},H^{\alpha})$ coupling constants (Figures 2(b) and 5(b); Table 2). Up to now, no proteinase has been found to be inhibited by domain 1. We do not ascribe this fact to the unusual residue in the putative P1 position, i.e. Gln46. In fact, at least ten of the over 470 known Kazal-type inhibitors possess a glutamine residue at the P1 position.†[2] Usually, the P1 residue of a particular serine proteinase inhibitor is identical to the corresponding substrate residue, i.e. arginine or lysine for trypsin inhibitors. However a couple of exceptions are known, and the P1 position is very tolerant to variations.[19] For example, in turkey ovomucoid third domain the P1 leucine has been replaced by any of the 19 residual amino acids, and inhibitor/proteinase complexes have been characterized. In all cases the same residue position served as P1 residue, independent of the particular amino acid.[2,21–23] A similar study was conducted for bovine pancreatic trypsin inhibitor (BPTI), where crystal structures of ten P1 variants of BPTI in complex with trypsin have been determined. All ten mutants use residue 15 as P1 residue, even the mutant with glycine at that position.[24] These results demonstrate that the location of the P1 residue is not determined by its side-chain character, but rather by the canonical backbone geometry of the binding loop.[2] Taken together, we ascribe the missing serine proteinase inhibitory activity of domain 1 to the backbone conformation of the assumed binding loop that does not fulfill the requirements for being a canonical loop, which is a requisite for proteinase inhibition by the standard mechanism.[2] We conclude that domain 1 is no canonical serine proteinase inhibitor. A non-canonically inhibition mechanism, however, cannot totally be excluded, as a few non-canonically binding serine proteinase inhibitors have also been identified.[25,26]

The experimentally determined three-dimensional structure of domain 1 has been classified as a new protein fold by the SCOP database (structural classification of proteins‡), and no hit was found applying a DALI protein structure comparison search.[20] Furthermore, a pairwise comparison of the experimentally determined structures of domains 1 and 6 using the DALI server did not result in significant similarity. The experimentally determined structure of domain 1, thus, exhibits a new protein fold.

Differences between the structures of LEKTI domains 1 and 6

Up to now, only a few examples are known for homologous proteins with different folds, for example structural variations within the RING finger domain family.[27] In contrast, quite often similar folds are found for proteins that exhibit only very low to no sequence similarities. An example are the β-defensins: The members of this large mammalian peptide family, which are small cationic and cysteine-rich antimicrobial peptides, exhibit only low sequence similarity, but share a considerable similarity on the level of secondary and tertiary structure, suggesting that the fold is mainly stabilized by the presence of three conserved disulfide bonds.[28] Another example is a study on BPTI, where almost 50% of all residues have been substituted by alanine. Despite those drastic changes in the sequence, the mutants folded into a native-like three-dimensional structure,

† http://www.chem.purdue.edu/LASKOWSKI/qtable.html

‡ http://scop.mrc-lmb.cam.ac.uk/scop/

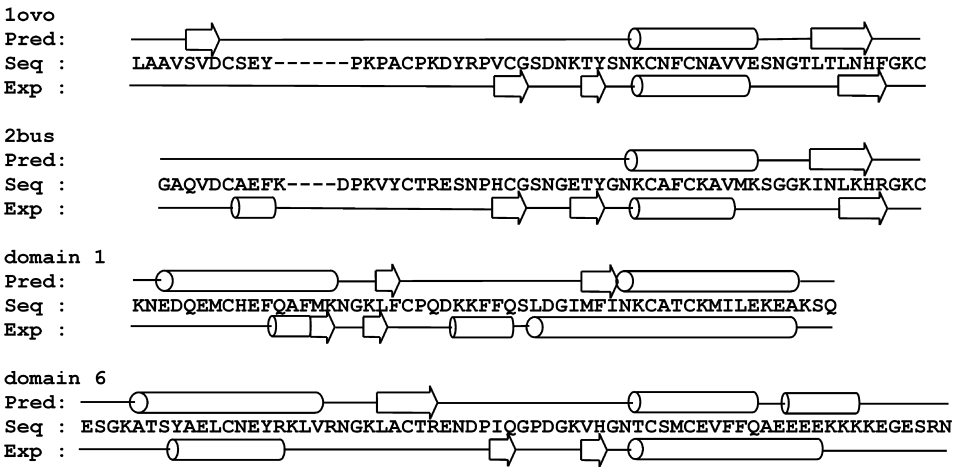

Figure 6. Comparison of predicted (Pred)[31] and experimental (Exp) secondary structure for Japanese quail ovomucoid third domain (1ovo), BUSI IIA (2bus), and LEKTI domains 1 and 6. The sequences (Seq) are aligned with respect to conserved cysteine residues. Helical elements and β-sheet structure are displayed as cylinders and arrows, respectively.

even though only one of the originally three disulfide bonds remained intact.[29]

Despite their significant sequence similarity and their conserved disulfide pattern, the experimentally determined structures of domains 1 and 6 exhibit different folds. The C^{α} atom RMSD between the well-structured region of the average structure of domain 1, i.e. residues 29 to 74, and the corresponding residues of domain 6 (366 to 411) is ~6.2 Å. Also, considering only the 37 residues from the first to the fourth cysteine, the C^{α} atom RMSD is still ~6 Å. In contrast, omitting the central parts of the proteins and considering only the residues 29 to 44 and 62 to 74 of domain 1 and the corresponding residues of domain 6 (366 to 381 and 399 to 411), the fit improves substantially, resulting in an RMSD value of ~1.8 Å (Figure 5(b)). Therefore, similar structural features of both proteins are the second half of the long COOH-terminal helix and partially the NH_2-terminal helix. In addition, the backbone course of residues 37 to 42 of domain 1 resembles the corresponding sequence of domain 6 (residues 374 to 379), but only domain 1 shows a regular β-hairpin structure in this region, as only for this domain a β-sheet typical inter-strand HαHα NOE was present, also in ^2H_2O solution. For domain 6, the interactions within this loop region are dominated by side-chain/side-chain interactions, whereas backbone/backbone interactions dominate the NOE pattern of domain 1. This applies particularly to Val374 of domain 6 and the corresponding Met37 of domain 1, which show significantly different secondary ^{1}Hα chemical shifts.

The most striking differences between the structures of domains 1 and 6 are located around the putative binding loops and the following 12 amino acid residues. The P3 to P3′ residues of domain 6 largely match the characteristic ϕ and ψ angles of canonical inhibitory loops (Table 2) and show mainly extended conformation, whereas the corresponding residues of domain 1 exhibit turn-like and helical secondary structure. Furthermore, domain 1 exhibits helical structure where domain 6 shows a β-hairpin in the corresponding sequence stretch, and the COOH-terminal helix of domain 1 thus starts eight residues NH_2-terminally of the according helix of classical Kazal-type inhibitors and that of domain 6 (Figure 5(b)). These structural differences are sufficient to classify domains 1 and 6 into different structural folds using tools like the DALI server and the SCOP database. Domain 6 shows typical features of Kazal-type proteinase inhibitors, including the arrangement of secondary structural elements and the geometry of its canonical binding loop. This classification is also confirmed by its biological activity. In contrast, domain 1 does not show a canonical binding loop anywhere in the protein and its secondary and tertiary structure do not resemble hitherto known protein folds.

In order to find an explanation for these findings we carried out a similarity search, as well as secondary and tertiary structure predictions and compared the results with our experimentally determined structures. Based on their sequences, domains 1 and 6 are generally considered to be homologous and to adopt similar folds by programs like BLAST and 3D-PSSM. Using the BLASTP2 similarity search† for LEKTI domain 1, one of the first similar sequences not belonging to LEKTI itself was that of BUSI IIA, which is a typical member of the Kazal-type proteinase inhibitor family. Further Kazal-type inhibitors were found with lower score values beside other proteins with various functions. Compared to domain 1, the BLASTP2 search for domain 6 resulted in a slightly higher score value for BUSI IIA, and all other hits, with only two exceptions,

† http://blast.wustl.edu

belonged to serine proteinase inhibitors of the Kazal-type. Using 3D-PSSM,[30] a program that is able to recognize remote protein sequence homologues as well as structural relationships, a fold similar to the BUSI IIA structure (PDB code: 2bus) was predicted for domain 1 with a certainty of 50%. Also, domain 6 was predicted to adopt a BUSI IIA-like fold with the same certainty. Interestingly, domain 6 was also predicted to adopt a domain 1-like fold (PDB code: 1hdl) with a certainty of 70%. Thus, using 3D-PSSM, domain 6 is predicted to be structurally more homologous to domain 1 than to classical Kazal-type inhibitors (for comparison see Figure 5(a) and (b)). These findings are likely due to two steps in the structure prediction procedure: (1) the significant sequence similarity of both domains (Figure 1), and (2) the secondary structure prediction used by 3D-PSSM, i.e. PSIpred†,[31] resulting in similar predictions for both domains (Figure 6). Comparison of the predicted and experimentally determined secondary structures of domains 1 and 6, as well as the classical Kazal-type inhibitors BUSI IIA and the Japanese quail ovomucoid third domain (PDB code: 1ovo)[70] reveals that for domains 1 and 6 two helices in similar sequence stretches are predicted, and that the COOH-terminal helix is also predicted for the two classical Kazal-type inhibitors (Figure 6). These helices largely match the experimentally determined ones in all cases. In contrast, for none of the examples the Kazal-typical first two β-strands have been predicted, reflecting the general difficulties in prediction of short β-strands. The tertiary structures of the two classical Kazal-type inhibitors are predicted correctly by 3D-PSSM due to a high sequence identity to the Kazal-type inhibitors used in the database. From the similarities of their sequences as well as of the predicted secondary structures it is not unexpected that 3D-PSSM also predicts similar tertiary structures for domains 1 and 6. For domain 1, however, a higher similarity between predicted and experimental secondary structure can be obtained by the use of other prediction methods‡, resulting in a consensus prediction that closely matches the experimental data for all helical elements of this domain, including the central 3_{10} helix (data not shown). Therefore, the results of theoretical structure prediction methods should generally be interpreted with caution, as the accuracy of the prediction of tertiary structure is usually associated with that of secondary structure.

We suspect the key residues for the switch between helical and β-hairpin structure found in the central parts of the structures of LEKTI domains 1 and 6 are located within the central 3_{10} helix of domain 1. Over 95% of all known classical Kazal-type inhibitors exhibit a Proline residue in the P4′ position (position Phe50 of domain 1) followed by a β-branched amino acid residue, mainly valine, in about 90% of the cases (Phe51 in domain 1). The only exceptions are the ovomucoid inhibitors, which often possess a leucine residue in this position, and BUSI IIA as one of the rare examples where a histidine residue follows the conserved proline residue. Ten of the 15 LEKTI domains exhibit a valine residue, and four domains, including domain 6, exhibit an isoleucine residue in this position. All LEKTI domains except domains 1 and 2 feature the proline residue in the P4′ position. The helix breaker proline followed by a β-branched amino acid residue seems to be associated with an extended structure in the region directly following the canonical loop of Kazal-type proteinase inhibitors. Whether the presence of these two residues is a commensurate condition for β-sheet structure in the central part of the protein or if further amino acid exchanges are necessary is currently under investigation in our laboratory.

Materials and Methods

Source of peptide material

Isolation of domain 6

Native LEKTI domain 6 was isolated from a peptide library generated from human blood ultrafiltrate as described.[32] The isolation was guided by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry and included multiple steps of chromatographic purification. Primary structure, molecular mass, and homogeneity of native LEKTI domain 6 were verified by sequencing, electrospray ionization mass spectrometry, and capillary zone electrophoresis. Disulfide bonds were determined by sequence analysis following tryptic cleavage of the peptide[3] (P. Kreutzmann, personal communication).

Peptide synthesis of domain 1

Non-labeled domain 1 was synthesized applying solid-phase technology on a preloaded TentaGel R Trt-Gln-Fmoc resin (Rapp Polymere, Tübingen, Germany). Fmoc-protected amino acids were purchased from Orpegen (Heidelberg, Germany) or Novabiochem (Bad Soden, Germany) with the following side-chain protections: Asn(Trt), Asp(OtBu), Gln(Trt), Glu(OtBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Cys(Acm), and Cys(Trt). Acylations were carried out with TBTU/HOBt activation in *N*-methyl-2-pyrrolidinone. For selective introduction of the disulfide bonds, Cys30 and Cys66 were Acm-protected, and Cys44 and Cys63 were Trt-protected. After cleavage of crude peptide from the resin, disulfide bonds were subsequently introduced by air oxidation followed by iodine treatment, and the product was purified using preparative HPLC. Peptide purity was checked by analytical RP-HPLC (Vydac C18), capillary zone electrophoresis, and ESI-MS (Perkin–Elmer, Sciex API 100).

† http://bioinf.cs.ucl.ac.uk/psipred/
‡ http://npsa-pbil.ibcp.fr/cgi-bin/align_clustalw.pl

Expression and purification of domain 1

A detailed description of the procedure has been published recently.[11] In brief, the gene fragment coding for domain 1 (HF6478) was combined with a sequence encoding a factor Xa cleavage site at its 5′ end, using standard PCR techniques,[33] and was cloned into the vector pET-32a (Novagen, Madison, WI). The resulting vector construct pET-32a-Xa-*hf6478* produces domain 1 with an NH_2-terminal *Escherichia coli* thioredoxin fusion (Trx-tag, Novagen, Madison, WI) connected by a 43-amino acid residue linker containing six histidine residues and a factor Xa cleavage site. As host strain for expression an *E. coli* $trxB^-/gor522^-$ double mutant (*E. coli* Origami (DE3); Novagen, Madison, WI) was used, allowing disulfide formation in the oxidative cytoplasm of this strain.[34,35] The soluble fraction of cell lysate was used for further purification, using a one-step Ni^{2+}-chelating-affinity chromatography.[36] Trx-tag and linker region were removed by factor Xa cleavage. Expression of recombinant domain 1 (rHF6478) as Trx-fusion protein in *E. coli* Origami (DE3) resulted in a homogeneous product with the correct disulfide bonds. For uniform (>95%) ^{15}N-labeling domain 1 was isolated and purified from *E. coli* cultures grown in M9 minimal medium enriched with $^{15}NH_4Cl$, using the same purification protocol. As the expression strain lacks the capability to synthesize leucine, unlabeled leucine was added to the medium according to the manufacturers instructions (Novagen, Madison, WI).

NMR spectroscopy and data processing

Domain 6

The sample of native domain 6 contained approximately 2 mM protein in 0.5 ml of $H_2O/^2H_2O$ (9:1, v/v, pH 4.0) or ^2H_2O (99.994 at.%). All two-dimensional NMR experiments were carried out at 298 K on a commercial Bruker DRX600 spectrometer equipped with triple resonance $^1H/^{13}C/^{15}N$ probes and pulsed field gradient capabilities. For resonance assignment DQF-COSY, TOCSY (80 ms mixing time), and NOESY (150 and 200 ms mixing time) spectra were performed using standard techniques.[37] Water suppression was accomplished by excitation sculpting[38] for NOESY and z-filtered TOCSY spectra,[39] and by coherence selection with magic angle gradient for the DQF-COSY spectrum.[40] All experiments were acquired with 4096 data points in t_2 and 720 to 1024 data points in t_1, and a sweep width of 6613.8 Hz in both dimensions. Quadrature detection was used in both dimensions with the time-proportional phase incrementation (TPPI) technique in t_1.[41] Spectra data were multiplied with a squared sinebell window function shifted by $\pi/2$, $\pi/3$, or $\pi/4$ prior to Fourier transformation. Sixth-order baseline and phase correction were used. Data processing was performed using the NDee software package (SpinUp Inc., Dortmund, Germany) on Sun and DEC workstations.

Sequence-specific resonance assignment from the two-dimensional spectra was carried out using standard methods.[12] All distance restraints for structure calculation were estimated from cross-peak intensities of the NOESY spectra. $^3J(H^N,H^\alpha)$ coupling constants were obtained from line-shape analysis of the anti-phase cross-signal splitting in a high digital resolution DQF-COSY spectrum using a Lorentzian function for peak fitting. The chemical shifts are in reference to DSS as an external standard.

Domain 1

For homonuclear two-dimensional NMR spectra of synthetic (2 mM in $H_2O/^2H_2O$ (9:1, v/v), pH 4.0) and non-labeled recombinant domain 1 (1.2 mM in $H_2O/^2H_2O$ (9:1, v/v), pH 4.5) the same standard techniques as for domain 6 were used, i.e. COSY, NOESY, and TOCSY spectra.

All NMR experiments described below were performed on a sample containing ^{15}N-labeled recombinant domain 1 at a final concentration of 1.5 mM in $H_2O/^2H_2O$ (9:1, v/v), pH 4.5 at 298 K on a Bruker DRX600 spectrometer equipped with triple resonance $^1H/^{13}C/^{15}N$ probes and pulsed field gradient capabilities. Quadrature detection in the indirectly detected dimensions was obtained by the States-TPPI method[41,42] or by the echo–antiecho method[43] if coherence selection with gradients was employed.[44,45] The HSQC step in the ^{15}N-TOCSY-HSQC consists of a sensitivity enhanced HSQC with gradient coherence selection for water suppression. The ^{15}N-NOESY-HSQC[46] and the HNHA[47] were recorded with a watergate and a water flipback scheme. For proton TOCSY mixing the DIPSI2rc sequence,[39] and for heteronuclear decoupling during acquisition the GARP pulse sequence[48] was applied.

All data were processed using in-house written software and visualized using NMRView 4.1.0[49] and the NDee software package (SpinUp Inc., Dortmund, Germany) on Sun and DEC workstations. Data processing consists typically of SVD-linear prediction[50] with root-reflection[51] in the heteronuclear dimension, apodization with $\pi/2$ or $\pi/3$ shifted squared sinebells, one zero-filling in all dimensions, and Fourier transformation. Finally baseline correction in the acquisition dimension was performed using a model free algorithm.[52] The proton chemical shifts were referenced to external DSS at 0.0 ppm. The chemical shifts of the ^{15}N resonances were referenced indirectly using the Ξ ratios of the zero-point frequencies at 298 K: 0.101329118 for $^{15}N/^1H$.[53]

Sequence-specific resonance assignment was carried out using characteristic NOE cross-peaks from ^{15}N-NOESY-HSQC and ^{15}N-HMQC-NOESY-HSQC spectra. Intraresidual proton resonances were assigned using the ^{15}N-TOCSY-HSQC spectrum. $^1H\alpha$ chemical shifts were taken from HNHA and ^{15}N-TOCSY-HSQC spectra.

The distance restraints for structure calculation were taken from the ^{15}N-NOESY-HSQC spectrum for NOEs involving amide protons and from the homonuclear NOESY spectra for NOEs between aliphatic and aromatic protons. $^3J(H^N,H^\alpha)$ coupling constants were measured from cross-peak to diagonal peak intensity ratios in the HNHA spectrum, corrected by a factor of 1.05.[54,55]

Structure calculation

The total number of non-trivial unambiguous NOESY cross-peaks used for structure calculation was 523 for domain 1 and 684 for domain 6 (Table 1). These cross-peaks were divided into three groups according to their relative intensities and converted into upper distance constraints: strong, 2.7 Å, medium, 3.5 Å and weak, 5.0 Å. $^3J(H^N,H^\alpha)$ were converted into ϕ-angles according to the Karplus equation allowing the following deviations from derived angle: $^3J(H^N,H^\alpha)$ coupling constants <6.0 Hz were translated to ϕ-angle constraints of $-60(\pm 20)$ degrees, coupling constants >8.0 Hz were translated to ϕ-angle constraints of $-120(\pm 40)°$. Disulfide bonds were taken into account by one NOE

distance restraint (dss = 2.02(±0.10) Å) for each disulfide bond.

All structures were calculated using X-PLOR 3.851[56] and a modified *ab initio* simulated annealing protocol.[57] The structure calculation strategy is similar to those described previously, i.e. a three-stage simulated annealing protocol,[58] and included floating assignments of prochiral groups,[59] a conformational database potential term,[60] which was modified in the case of domain 6,[61] and a reduced presentation for non-bonded interactions for part of the calculation[58] to increase efficiency. The conformational search phase (80 ps of molecular dynamics at 2000 K) was followed by a refinement in which the system was cooled to 1000 K within 60 ps, accompanied by increasing of the force constant for non-bonded interactions and angle energy constants for the diastereospecifically unassigned groups to their final values. Applying these high force constants the system was cooled from 1000 to 100 K within 30 ps. To approach the energy minimum, 1200 steps of energy minimization were performed, the final 1000 steps without conformational database potential.

An iterative approach using several rounds of structure calculations with subsequent distance analysis was used to solve ambiguities in the NOE cross-peak assignments. Dihedral angle constraints were introduced in later rounds of the structure calculation. Each round of structure calculation started from templates with randomized backbone torsion angles. Using the modified database term for structure calculation of domain 1 as well, results in virtually the same overall structure.

For each protein, a family of 60 structures was calculated and 21 structures were selected, respectively, with the criterion of lowest overall energy for further characterization. Geometry of the structures as well as elements of secondary structure were analyzed using PROCHECK[62] and MOLMOL.[63] RASMOL,[64] MOLSCRIPT,[65] Raster3D,[66,67] and MOLMOL[63] were used for visualization of the structure data.

Data Bank accession numbers

The coordinates and NMR restraints have been deposited to the Protein Data Bank (codes: 1hdl for domain 1, and 1h0z for domain 6), chemical shift values and coupling constants have been deposited to the BioMagResBank (accession numbers: 4910 for domain 1, and 5551 for domain 6).

Acknowledgements

We thank Professor Paul Rösch (Universität Bayreuth) for providing excellent laboratory and NMR facilities. We further thank Ludger Ständker for providing native LEKTI domain 6. The study was supported by the Deutsche Forschungsgemeinschaft (DFG) (project MA2317-1). U.C.M. acknowledges support by the "Bayerischer Habilitationsförderpreis".

References

1. Seife, C. (1997). Blunting nature's Swiss army knife. *Science*, **277**, 1602–1603.
2. Laskowski, M., Jr & Qasim, M. A. (2000). What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta*, **1477**, 324–337.
3. Mägert, H. J., Ständker, L., Kreutzmann, P., Zucht, H. D., Reinecke, M., Sommerhoff, C. P. *et al.* (1999). LEKTI, a novel 15-domain type of human serine proteinase inhibitor. *J. Biol. Chem.* **274**, 21499–21502.
4. Mägert, H. J., Kreutzmann, P., Drögemüller, K., Ständker, L., Adermann, K., Walden, M. *et al.* (2002). The 15-domain serine proteinase inhibitor LEKTI: biochemical properties, genomic organization, and pathophysiological role. *Eur. J. Med. Res.* **7**, 49–56.
5. Mägert, H. J., Kreutzmann, P., Ständker, L., Walden, M., Drögemüller, K. & Forssmann, W. G. (2002). LEKTI: a multidomain serine proteinase inhibitor with pathophysiological relevance. *Int. J. Biochem. Cell Biol.* **34**, 573–576.
6. Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A. D. *et al.* (2000). Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nature Genet.* **25**, 141–142.
7. Sprecher, E., Chavanas, S., DiGiovanna, J. J., Amin, S., Nielsen, K., Prendiville, J. S. *et al.* (2001). The spectrum of pathogenic mutations in SPINK5 in 19 families with Netherton syndrome: implications for mutation detection and first case of prenatal diagnosis. *Invest. Dermatol.* **117**, 179–187.
8. Walley, A. J., Chavanas, S., Moffatt, M. F., Esnouf, R. M., Ubhi, B., Lawrence, R. *et al.* (2001). Gene polymorphism in Netherton and common atopic disease. *Nature Genet.* **29**, 175–178.
9. Bitoun, E., Chavanas, S., Irvine, A. D., Lonie, L., Bodemer, C., Paradisi, M. *et al.* (2002). Netherton syndrome: disease expression and spectrum of SPINK5 mutations in 21 families. *J. Invest Dermatol.* **118**, 352–361.
10. Komatsu, N., Takata, M., Qtsuki, N., Ohka, R., Amano, O., Takehara, K. & Saijoh, K. (2002). Elevated stratum corneum hydrolytic activity in Netherton syndrome suggests an inhibitory regultion of desquamation by SPINK5-derived peptides. *J. Invest. Dermat.* **118**, 436–443.
11. Lauber, T., Marx, U. C., Schulz, A., Kreutzmann, P., Rösch, P. & Hoffmann, S. (2001). Accurate disulfide formation in Escherichia coli: overexpression and characterization of the first domain (HF6478) of the multiple Kazal-type inhibitor LEKTI. *Protein Expr. Purif.* **22**, 108–112.
12. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
13. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31**, 1647–1651.
14. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333.
15. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.
16. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. *Biochem. Biophys. Res. Commun.* **27**, 157–162.

17. Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315.
18. Bode, W., Epp, O., Huber, R., Laskowski, M., Jr & Ardelt, W. (1985). The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur. J. Biochem.* **147**, 387–395.
19. Laskowski, M., Jr & Kato, I. (1980). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593–626.
20. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
21. Huang, K., Lu, W., Anderson, S., Laskowski, M., Jr & James, M. N. (1995). Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci.* **4**, 1985–1997.
22. Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L. *et al.* (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441–461.
23. Fujinaga, M., Huang, K., Bateman, K. S. & James, M. N. (1998). Computational analysis of the binding of P1 variants of domain 3 of turkey ovomucoid inhibitor to *Streptomyces griseus* protease B. *J. Mol. Biol.* **284**, 1683–1694.
24. Helland, R., Otlewski, J., Sundheim, O., Dadlez, M. & Smalas, A. O. (1999). The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. *J. Mol. Biol.* **287**, 923–942.
25. Bode, W. & Huber, R. (2000). Structural basis of the endoproteinase–protein inhibitor interaction. *Biochim. Biophys. Acta*, **1477**, 241–252.
26. van de Locht, A., Stubbs, M. T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W. & Huber, R. (1996). The ornithodorin-thrombin crystal structure, a key to the TAP enigma? *EMBO J.* **15**, 6011–6017.
27. Borden, K. L. & Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence–structure family. *Curr. Opin. Struct. Biol.* **6**, 395–401.
28. Bauer, F., Schweimer, K., Klüver, E., Conejo-Garcia, J. R., Forssmann, W. G., Rösch, P. *et al.* (2001). Structure determination of human and murine beta-defensins reveals structural conservation in the absence of significant sequence similarity. *Protein Sci.* **10**, 2470–2479.
29. Kuroda, Y. & Kim, P. S. (2000). Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. *J. Mol. Biol.* **298**, 493–501.
30. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 501–522.
31. Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202.
32. Schulz-Knappe, P., Schrader, M., Ständker, L., Richter, R., Hess, R., Jürgens, M. & Forssmann, W. G. (1997). Peptide bank generated by large-scale preparation of circulating human peptides. *J. Chromatog., sect. A*, **776**, 125–132.
33. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1999). *Unit 15.1* Current Protocols in Molecular Biology, Wiley, New York.
34. Derman, A. I., Prinz, W. A., Belin, D. & Beckwith, J. (1993). Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, **262**, 1744–1747.
35. Prinz, W. A., Aslund, F., Holmgren, A. & Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667.
36. Porath, J. (1992). Immobilized metal ion affinity chromatography. *Protein Expr. Purif.* **3**, 263–281.
37. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. III & Skelton, N. J. (1996). *Protein NMR Spectroscopy,* Academic Press, San Diego.
38. Hwang, T. L. & Shaka, A. J. (1995). Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson., ser. A*, **112**, 275–279.
39. Cavanagh, J. & Rance, M. (1992). Suppression of cross-relaxation effects in TOCSY spectra *via* a modified DIPSI-2 mixing sequence. *J. Magn. Reson.* **96**, 670–678.
40. Vanzijl, P. C. M., Johnson, M. O., Mori, S. & Hurd, R. E. (1995). Magic-angle-gradient double-quantum-filtered COSY. *J. Magn. Reson., ser. A*, **113**, 265–270.
41. States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. *J. Magn. Reson.* **48**, 286–292.
42. Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Applications to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.
43. Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665.
44. Schleucher, J., Schwendinger, M. G., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J. *et al.* (1994). A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR*, **4**, 301–306.
45. Sattler, M., Schwendinger, M. G., Schleucher, J. & Griesinger, C. (1995). Novel strategies for the sensitivity enhancement in heteronuclear multidimensional NMR experiments employing pulsed field gradients. *J. Biomol. NMR*, **5**, 11–22.
46. Talluri, S. & Wagner, G. (1996). An optimized 3D NOESY-HSQC. *J. Magn. Reson., ser. B*, **112**, 200–205.
47. Zhang, W., Smithgall, T. & Gmeiner, W. H. (1997). Three-dimensional structure of the Hck SH2 domain in solution. *J. Biomol. NMR*, **10**, 263–272.
48. Shaka, A., Barker, P. B. & Freeman, R. (1985). Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Reson.* **64**, 547–552.
49. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
50. Barkhuijsen, H., de Beer, R., Bovee, W. M. M. J. & van Ormondt, D. (1985). Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure. *J. Magn. Reson.* **61**, 465–481.
51. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C*, 2nd edit., Cambridge University Press, Cambridge.

52. Friedrich, M. S. (1995). A model-free algorithm for the removal of baseline artifacts. *J. Biomol. NMR*, **5**, 147–153.
53. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E. *et al.* (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
54. Vuister, G. W. & Bax, A. (1993). Quantitative *J* correlation: a new approach for measuring homonuclear three-bound $J(H^{N}–H^{\alpha})$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
55. Düx, P., Whitehead, B., Boelens, R., Kaptein, R. & Vuister, G. W. (1997). Measurements of $^{15}N–^{1}H$ couplings constants in uniformly ^{15}N-labeled proteins: application to the photoactive yellow protein. *J. Biomol. NMR*, **10**, 301–306.
56. Brünger, A. T. (1993). *X-PLOR Version 3.1: A system for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
57. Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M. & Pastore, A. (1995). Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* **14**, 3572–3584.
58. Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins: Struct. Funct. Genet.* **17**, 297–309.
59. Folmer, R. H., Hilbers, C. W., Konings, R. N. & Nilges, M. (1997). Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with *J*-coupling data. *J. Biomol. NMR*, **9**, 245–258.
60. Kuszewski, J. & Clore, G. M. (2000). Sources of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force. *J. Magn. Reson.* **146**, 249–254.
61. Neudecker, P., Sticht, H. & Rösch, P. (2001). Improving the efficiency of the Gaussian conformational database potential for the refinement of protein and nucleic acid structures. *J. Biomol. NMR*, **21**, 373–375.
62. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
63. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 32–51.
64. Sayle, R., (1995). RasMol V2.6 molecular visualisation programme. Glaxo Wellcome Research and Development Stevevage, Hertfordshire, UK.
65. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
66. Bacon, D. & Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220.
67. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0, a program for photorealistic molecular graphics. *Acta Crystallog.* D**50**, 869–873.
68. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
69. Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.
70. Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I. & Laskowski, M., Jr (1982). Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J. Mol. Biol.* **158**, 515–537.

Edited by M. F. Summers

(Received 18 October 2002; received in revised form 10 February 2003; accepted 10 February 2003)