NMR Structure of the Cathelin-like Domain of the Protegrin-3 Precursor†

Yinshan Yang,‡ Jean Frédéric Sanchez,‡ Marie-Paule Strub,‡ Bernhard Brutscher,§ and André Aumelas*,‡

Centre de Biochimie Structurale, UMR 5048 CNRS-UM1/UMR 554 INSERM-UM1, Université Montpellier 1, Faculté de Pharmacie, 15 avenue Charles Flahault, 34093 Montpellier Cedex 5, France, and Institut de Biologie Structurale, Laboratoire de RMN, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France

Received November 7, 2002; Revised Manuscript Received February 24, 2003

ABSTRACT: In mammals, numerous precursors of antibacterial peptides with unrelated sequences share a similar prosequence of 94−114 residues, termed the cathelin-like domain. The cathelin-like domain of protegrin-3 (ProS) was overexpressed in *Escherichia coli* and uniformly labeled with ^{15}N or ^{15}N and ^{13}C, and its three-dimensional structure was determined by heteronuclear NMR at pH 6.2. Under these conditions and due to the cis−trans isomerization of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds, the ProS structure was found to adopt four almost equally populated conformations in slow exchange on the NMR chemical shift time scale. The ProS structure consists of an N-terminal α-helix ($Y^{34}-N^{48}$) cradled by a four-stranded antiparallel β-sheet (β1, $N^{53}-L^{60}$; β2, $K^{74}-P^{86}$; β3, $V^{104}-V^{111}$; and β4, $I^{122}-C^{124}$). The solution structure of ProS, which is monomeric, allowed us to determine the structure of the L1 and L2 loops, which are too mobile in the crystal structure. The regions common to the solution and X-ray structures were found to be very similar. Finally, since the overall fold of ProS is very similar to that of cystatins despite a low degree of sequence identity, the ProS solution structure was compared to the solution and X-ray structures of the chicken cystatin. This comparison revealed that the structures of the L1 and L2 loops as well as that of the appending domain are quite different in the two proteins. These differences are mainly due to the high proline residue content (10%) which disorganizes the hydrogen bond network of a part of the ProS β-sheet in contrast to that of the chicken cystatin structure.

In mammals, cathelicidin antimicrobial peptides (CAMPs) are initially synthesized as a proform that is usually inactive (*1*−*9*). However, unprocessed rabbit cathelicidin, notably, proCAP18, in concert with other host defense molecules, can contribute to antimicrobial host defense (*10*). This peptide family includes ∼15 structurally diverse antimicrobial peptides, whereas their precursors share a highly conserved prosequence of 94−114 residues (*8*, *11*). Moreover, it has been shown that to release the active antimicrobial peptides, they must be processed in activated neutrophils (*7*, *12*−*15*). It has been shown that the LL37 antimicrobial peptide is cleaved from hCAP-18 by proteinase-3 (*16*). The fact that these CAMPs of unrelated sequences share this highly conserved proregion suggests a similar role for these prosequences, which has not yet been clearly identified. Nevertheless, a monocyte chemoattractant activity was proposed for ProBac5 and a cysteine protease inhibitory activity for proBac7 (*10*, *17*). Although unrelated to the CAMP, the proform of the human α-defensin HNP-1 was also found to be inactive. In this case, it has been proposed that electrostatic interactions between the antibacterial peptide, positively charged, and the negatively charged groups of the prosequence could be responsible for such an inactivation (*18*).

Protegrins (PG-1−5), which are antibacterial peptides of 16−18 residues either isolated or identified from cDNA of porcine leucocytes, belong to the CAMP family (*1*, *11*, *19*). They are characterized by two intramolecular disulfide bridges and a high arginine content. The structure of PG-1 consists of an amphipathic β-sheet stabilized by the two disulfide bonds (*20*, *21*). Protegrins are initially synthesized as precursors of 149 residues (147 residues for PG-2) with three regions: a signal peptide (residues 1−29), a prosequence (residues 30−130 here termed cathelin-like domain[2] or ProS), and the PG sequence (residues 131−148) which is followed by an amidation consensus sequence (*11*, *22*). It has been shown that the proform of protegrin-1 (residues

† This research was supported by a grant from the Ministère de l'Education Nationale, de la Recherche et de la Technologie and the Fondation pour la Recherche Médicale (JFS).

* To whom correspondence should be addressed: Centre de Biochimie Structurale, UMR 5048 CNRS-UM1/UMR 554 INSERM-UM1, Faculté de Pharmacie, 15 avenue Charles Flahault, 34093 Montpellier Cedex 5, France. Telephone: +33 467 043 432. Fax: +33 467 529 623. E-mail: aumelas@cbs.cnrs.fr.

‡ Université Montpellier 1.

§ Laboratoire de RMN.

[1] Abbreviations: CAMP, cathelicidin antimicrobial peptide; CSI, chemical shift index; 1D, 2D, and 3D, one-, two-, and three-dimensional, respectively; DQF-COSY, 2D double-quantum-filtered correlation spectroscopy; DSS, sodium 4,4-dimethyl-4-silapentane-1-sulfonate; hCAP-18, human cationic antimicrobial peptide; HSQC, heteronuclear single-quantum coherence; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D nuclear Overhauser effect spectroscopy; PG, protegrin; ProBac5 and ProBac7, proforms of bactenecin-5 and bactenecin-7, respectively; proCAP18, proform of the cationic antimicrobial peptide; ProS, prosequence of the protegrin; rmsd, root-mean-square deviation; TPPI, time-proportional phase incrementation; TOCSY, total correlation spectroscopy.

[2] In previous papers, we called the prosequence of protegrins (ProS) the "cathelicidin motif" whereas other labs called it the "cathelin-like domain". To standardize, the "cathelin-like domain" terminology was used in this paper.

10.1021/bi027133c CCC: $25.00
Published on Web 04/03/2003

30−148) has no antibacterial activity and that the V^{130}−R^{131} amide bond has to be cleaved by an elastase-like enzyme to yield the active protegrin (*14*). The main features of the ProS sequence are the presence of two conserved disulfide bonds as well as its high proline residue content (8−10%).

While it is well-established that the PG antimicrobial activity is produced by cytoplasmic membrane disruption and that the overall structure (amphipathicity, charge, and disulfide bridges) is of primary importance, the role of the cathelin-like domain is not yet well-known. This domain is widespread, and to better understand its involvement in the maturation process of cathelicidin peptides, we set out to determine the three-dimensional (3D) solution structure of the cathelin-like domain of the protegrin-3 precursor using multidimensional heteronuclear NMR spectroscopy.

In this paper, we report on the identification of two X−Pro amide bonds that are able to adopt both the cis and trans conformations, thus giving rise to four conformers which were fully assigned. The structure of the all-trans and all-cis conformers of ProS complements the X-ray structure in which two loops were either lacking or poorly defined (*23*). Finally, the solution structure of ProS, which displays the cystatin fold, was compared with that of the chicken cystatin. The presence of numerous proline residues in ProS, which break the β-sheet hydrogen bond network, explains the major structural differences in the vicinity of the loops which contrasts with the high-level structural similarity of their core.

MATERIALS AND METHODS

Expression and Purification. Cloning, expression, and purification of ProS were described elsewhere (*24*). Uniform ^{15}N or ^{15}N and ^{13}C labeling was obtained by growing the cells in the HCDM1 minimal medium (*25*) containing $^{15}NH_4Cl$ (99.5%) or $^{15}NH_4Cl$ and [$^{13}C_6$]glucose (99%), respectively, as sole the nitrogen and carbon sources.

All NMR samples of ProS were prepared in 50 mM sodium phosphate, 50 mM NaCl, and 0.56 mM benzamidine. First, two 500 μL samples of unlabeled ProS (2 mM) were prepared in a 95:5 H_2O/D_2O mixture and in 99.98% D_2O. Then, three 300 μL samples containing 1.5−2.0 mM labeled ProS were prepared in a Shigemi tube (Shigemi Co., Ltd., Tokyo, Japan): one with the uniformly ^{15}N-labeled protein in a 95:5 H_2O/D_2O mixture and two with the uniformly ^{15}N- and ^{13}C-labeled sample in a 95:5 H_2O/D_2O mixture and in 99.98% D_2O, respectively. The pH was adjusted to 6.2 by addition of DCl or NaOD. pH values were uncorrected for the deuterium isotope effect.

NMR Spectroscopy. NMR experiments were performed both on a Bruker AMX 600 spectrometer and on a Varian INOVA 800 spectrometer equipped with a triple-resonance (^1H, ^{15}N, and ^{13}C) probe and shielded triple-axis pulsed-field gradients. The sample temperature was set to 27 °C. Quadrature detection in the indirect dimensions of the multidimensional experiments was achieved by the echo/antiecho detection scheme (*26*) for ^{15}N, and by the TPPI−States method (*27*) for ^1H and ^{13}C. The carrier frequencies were set to 4.75 ppm for ^1H, to 118.6 ppm for ^{15}N, to 35.5 ppm for $^{13}C_{aliph}$, to 174.5 ppm for $^{13}C'$, and to 55.5 ppm for $^{13}C_\alpha$. All chemical shifts were referenced with respect to DSS for ^1H and ^{13}C, and liquid NH_3 for ^{15}N, according to the IUPAC recommendations (*28*).

With the unlabeled samples, data sets of two-dimensional (2D) experiments (DQF-COSY, z-TOCSY, and NOESY) spectra were acquired in the phase-sensitive mode using the States−TPPI method and using a time domain data size of 512 t_1 × 2048 t_2 (800 t_1 × 2048 t_2 for DQF-COSY) complex points and 64 transients per complex t_1 increments. 2D data sets were recorded at 37 and 27 °C. Except for the DQF-COSY spectra (where low-power irradiation was used), the water resonance was suppressed by the WATERGATE method. z-TOCSY spectra were obtained with a mixing time of 80 ms and NOESY spectra with a mixing time of 150 ms.

With the ^{15}N and ^{13}C doubly labeled sample, 2D $^1H-^{13}C$ HSQC and $^1H-^{15}N$ HSQC correlation spectra were recorded with 512 (^1H) × 256 (^{13}C) and 512 (^1H) × 200 (^{15}N) complex points, respectively, for spectral widths of 12.5 (^1H), 29.6 (^{15}N), and 60 ppm (^{13}C).

For sequential backbone assignment, a set of triple-resonance experiments (*29*) comprising 3D HNCA, 3D HNCO, and 3D CBCA(CO)NH experiments was performed. The pulse sequences included standard features such as sensitivity-enhanced quadrature detection (*26*) in the ^{15}N dimension, pulsed field gradient coherence pathway selection (*30*), and water flip back pulses to align the water signal along the *z*-axis prior to detection (*31*). The spectral widths were set for ^1H to 12.5 ppm, for ^{15}N to 29.6 ppm, for $^{13}C'$ to 20 ppm, for $^{13}C_\alpha$ to 30 ppm, and for $^{13}C_{\alpha/\beta}$ to 66 ppm. The time domain data size was 512 (^1H) × 67 (^{15}N) × 128 ($^{13}C_\alpha$) complex points for HNCA using four scans per (t_1, t_2) increment, 512 (^1H) × 67 (^{15}N) × 120 ($^{13}C'$) and two scans for HNCO, and 512 (^1H) × 57 (^{15}N) × 79 ($^{13}C_{\alpha/\beta}$) and 12 scans for CBCA(CO)NH.

^1H and ^{13}C side chain assignments were made on the basis of 3D HC(C)H-TOCSY and 3D ^{13}C-edited NOESY-HSQC spectra recorded on the sample dissolved in D_2O. The ^1H-NOESY mixing time was set to 100 ms. The ^{13}C-TOCSY mixing time was set to 12 ms using a DIPSI-2 composite pulse train applied at a field strength of 10 kHz [$(\gamma B_1)/(2\pi)$] (*32*). Adiabatic broadband ^{13}C decoupling STUD+ (*33*) was used during detection. Data sets of 512 (^1H) × 110 (^{13}C) × 110 (^1H) complex points were recorded for both experiments for a spectral width of 12.5 ppm (^1H) × 60 ppm (^{13}C) × 10 ppm (^1H) and four scans per (t_1, t_2) increment.

The ^1H, ^{13}C, and ^{15}N chemical shifts of ProS are available as Supporting Information.

Slowly exchanging amide protons were identified from successive $^1H-^{15}N$ HSQC spectra with a duration of 25 min recorded at 27 °C following the solubilization of ProS from 30 min to 3 months. Between experiments, the sample was kept at 4 °C.

Structure Calculations. The intensities of NOESY cross-peaks were measured from the NOESY-($^1H-^{15}N$)HSQC and NOESY-($^1H-^{13}C$)HSQC spectra recorded with a mixing time of 100 ms and were divided into five classes, according to their intensities. Very strong, strong, medium, weak, and very weak NOEs were then converted to distance constraints of 1.8−2.4, 1.8−2.8, 1.8−3.6, 1.8−4.4, and 1.8−5.0 Å, respectively. For equivalent protons or non-stereospecifically assigned protons, pseudoatoms were introduced. When obtained, the χ_1 and χ_2 dihedral angle restraints were derived from the combined analysis of DQF-COSY data recorded

in D_2O and intra- and inter-residue NOEs. No hydrogen bond restraint was used.

To calculate the 3D structures of the all-trans and all-cis conformers of ProS, the distance and dihedral angle restraints were used as input in the standard distance geometry (DG)/simulated annealing (SA) refinement and energy minimization protocol using X-PLOR 3.8 (*34*). In the first stage of the calculation, an initial ensemble of 60 structures was generated from a template structure with randomized backbone dihedral angles ϕ and ψ and extended side chains, using a DG protocol followed by restrained SA and refinement (*35*). Analysis of the obtained structures and their comparison with the NMR data allowed us to identify more NOE restraints which were introduced into the subsequent calculation. After a number of these processes, for the all-trans conformer, 1227 NOE-derived distance restraints (142 medium-range and 346 long-range) and 48 dihedral restraints (36 ϕ and 12 χ_1) were used as final input data. Finally, 15 structures with a minimum number of restrained violations were submitted to 5000 cycles of restrained Powell energy minimization.

By using 1229 NOE-derived distance restraints and 46 dihedral restraints, a similar structure calculation was carried out for the all-cis conformer.

Structure Analysis. Calculations of rmsds and visual display were performed with INSIGHT 97 (Molecular Simulations Inc., San Diego, CA). Hydrogen bonds were considered present if the distance between heavy atoms was less than 3.5 Å and the donor−hydrogen−acceptor angle was greater than 120°. The Ramachandran ϕ and ψ plot analysis was performed with PROCHECK (*36*). The limits of the secondary structure elements and the van der Waals surfaces were determined with STRIDE (*37*).

The structure of the all-trans conformer of ProS was compared with the X-ray structures of the ProS monomer (PDB entry 1kwi) (*23*) and of the chicken cystatin (PDB entry 1cew) (*38*) as well as the solution structure of the chicken cystatin (PDB entry 1a67) (*39*, *40*).

The coordinates of the energy-minimized all-trans and all-cis conformers of ProS have been deposited in the Protein Data Bank as entries 1n5p and 1n5h, respectively.

RESULTS

Resonance Assignments and Identification of Two Cis−Trans Amide Bonds. NMR spectra of ProS were obtained at 27 °C, with a 1.5−2.0 mM sample at pH 6.2. As illustrated in Figure 1, the $^1H-^{15}N$ HSQC spectrum exhibited more cross-peaks than expected, suggesting that the structure of ProS could adopt several conformers in slow exchange on the ^1H and ^{15}N chemical shift time scale. The sequential assignment of the backbone atoms (H_N, ^{15}N, and $^{13}C_\alpha$) and the assignment of $^{13}C_\beta$ and $^3C'$ were obtained from the combined analysis of $^1H-^{15}N$ HSQC and HNCA, HNCO, and CBCA(CO)NH spectra recorded at 600 and 800 MHz, respectively. A total of 99% of ^1H_N (excluding the 10 prolines), 98% of ^{15}N (excluding the 10 prolines), 100% of $^1H_\alpha$, 100% of $^{13}C_\alpha$, and 87% of C′ resonances were assigned.

The assignment of the side chain resonances, including the 10 prolines, was obtained in part from the NOESY-($^1H-^{15}N$)-HSQC and complemented with HC(C)H-TOCSY and ^{13}C-edited NOESY-HSQC spectra recorded in D_2O. Eight of the 10 amide bonds involving proline residues were found to adopt the trans conformation. In contrast, both the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds adopt the cis and trans conformations in slow exchange on the NMR time scale (*41*); this explains the presence of the four conformers. As shown in Figure 2, the typical $d_{\alpha D118-\alpha P119}$ and $d_{\alpha D118-\delta\delta' P119}$ sequential NOEs allowed us to characterize both the cis and trans conformations for the $D^{118}-P^{119}$ amide bond (*42*). For the $R^{87}-P^{88}$ amide bond, the cis conformation was clearly characterized from the strong $d_{\alpha R87-\alpha P88}$ NOE, while weak $d_{\alpha R87-\delta\delta' P88}$ NOEs were observed for the trans conformation, suggesting $d_{\alpha-\delta\delta'}$ distances larger than those observed for a typical trans conformation. Since these conformers are in slow exchange, four chemical shifts were observed for residues which experience different environments in the four conformers. In the $^1H-^{15}N$ HSQC spectrum, this is clearly the case for several residues mainly located in the $R^{40}-R^{56}$ segment (R^{40}, A^{41}, V^{42}, D^{43}, R^{44}, L^{45}, N^{46}, E^{47}, S^{49}, Y^{55}, and R^{56}). For the $R^{87}-E^{94}$ segment, due to the overlapping of two cross-peaks, only three cross-peaks of the four (R^{87}, T^{89}, R^{90}, Q^{91}, and E^{94}) were observed (Figure 1), while for others, only two resonances (C^{107}, I^{122}, V^{111}, and D^{114}) were observed. Residues characterized by a unique cross-peak (L^{58} and E^{100}) are those which have a similar environment in all conformers. By a detailed analysis of the 2D and 3D NMR data, except for residues in the four conformers which really give identical chemical shifts, the almost full assignment was obtained for each conformer. This allowed us to further determine the structural differences induced by the cis−trans isomerization of the $D^{118}-P^{119}$ and $R^{87}-P^{88}$ amide bonds. The ^1H, ^{15}N, and ^{13}C chemical shifts of the four conformers are provided as Supporting Information. The details of the NMR experiments and chemical shifts for the four conformers were deposited in the BioMagResBank database (http://www.bmrb.wisc.edu) under accession number 5688.

Secondary Structure Determination. The elements of secondary structure of ProS were identified first from the analysis of the chemical shift deviations measured for H_α, $^{13}C_\alpha$, and $^{13}C'$ nuclei (CSI) (*43*) and then on the basis of sequential, medium-range, and long-range NOEs involving HN, H_α, and H_β protons. These data are summarized in Figure 3A. Because of the negative chemical shift deviations for H_α and the positive chemical shift deviations for C_α and C′ nuclei, the presence of a large helical structure could be inferred for the N-terminal part of ProS (Figure 3B). Moreover, strong d_{NN}, weak $d_{\alpha N}$, $d_{\alpha N(i,i+3)}$, $d_{\alpha\beta(i,i+3)}$, and $d_{\alpha\beta(i,i+4)}$ successive NOEs characterized this α-helical structure, spanning residues $S^{33}-S^{50}$. In contrast, positive chemical shift deviations observed for H_α and negative chemical shift deviations for C_α and C′, along with strong $d_{\alpha N}$ and weak d_{NN} NOEs, allowed us to identify four β-strands located approximately in fragments $N^{53}-Q^{62}$, $V^{76}-C^{85}$, $R^{103}-T^{112}$, and $I^{122}-N^{125}$. Moreover, in the diagonal plot presentation of NOEs displayed in Figure 3C, the distribution of long-distance constraints clearly shows three clusters perpendicular to the diagonal. Such a distribution is indicative of three interfaces between four antiparallel β-strands supported by several interstrand $d_{\alpha\alpha}$ NOEs.

Slowly Exchanging Amide Protons. The hydrogen−deuterium exchange of amide protons was followed by recording the $^1H-^{15}N$ HSQC spectra. After 30 min, 5 days, and 3 months, 36, 17, and 13 cross-peaks were still observed,

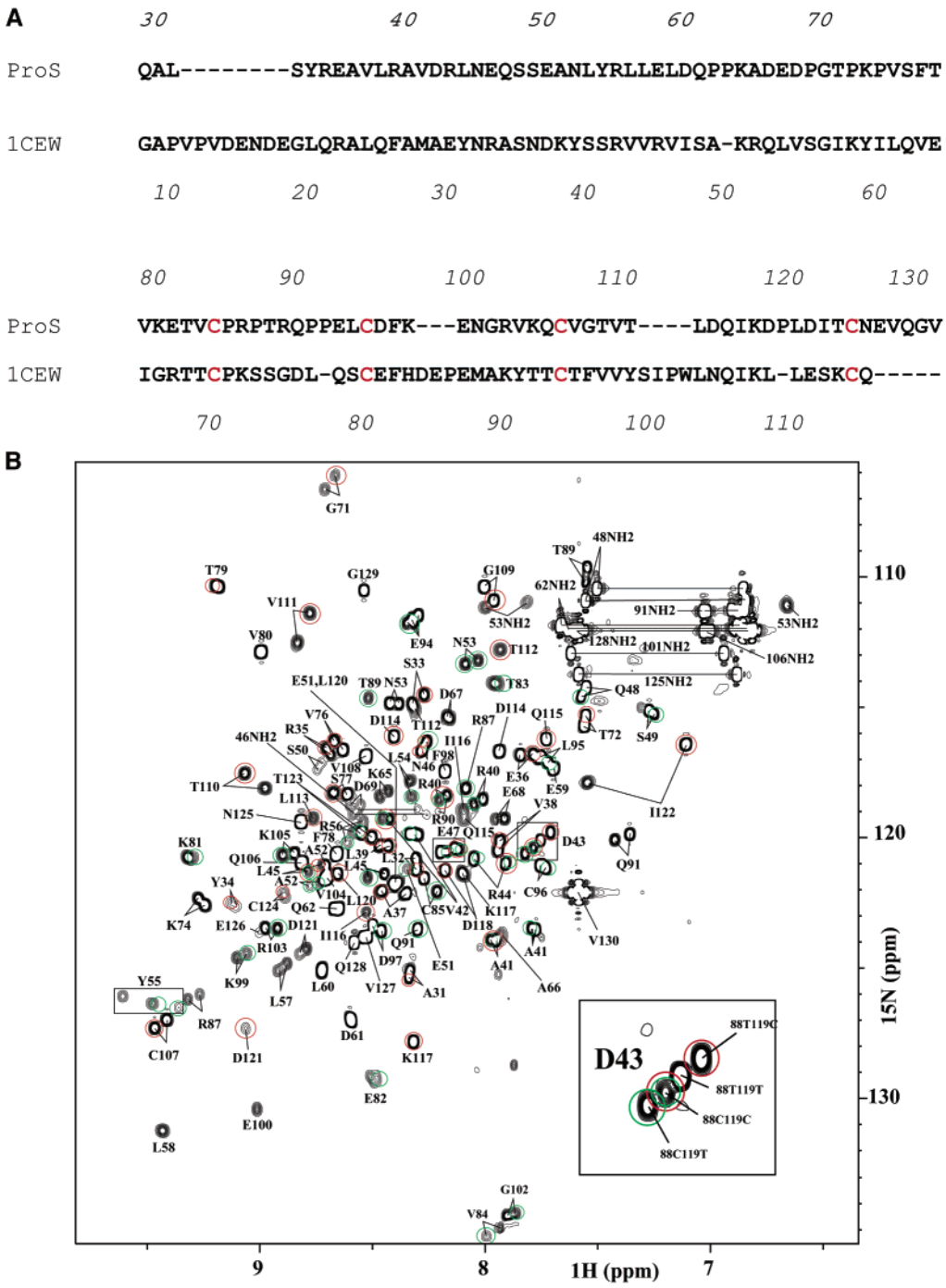

FIGURE 1: Sequence of ProS and $^{1}H-^{15}N$ HSQC spectrum of ProS showing the complete assignment (pH 6.2 and 27 °C). (A) Alignment of ProS and chicken cystatin (1cew) sequences (20% identical). For the two proteins, their own numbering was used. Notice the alignment of the four cysteines displayed in red which are involved in a similar disulfide bonding pattern ($C^{85}-C^{96}$ and $C^{107}-C^{124}$ for ProS and $C^{71}-C^{81}$ and $C^{95}-C^{115}$ for chicken cystatin). (B) Assignment of the $^{1}H-^{15}N$ HSQC resonances of ProS. The inset shows the cross-peaks of the four conformers observed for the D^{43} residue due to the cis−trans isomerization of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds. The cross-peak intensity suggests that the four conformers are roughly equally populated. The V^{84} and G^{102} cross-peaks are folded back. Red and green circles distinguish cross-peaks belonging to $D^{118}-P^{119}$ and $R^{87}-P^{88}$ cis conformers, respectively. The cross-peaks are labeled with the amino acid type and position in the protegrin-3 precursor.

respectively (Figure 3B). These slowly exchanging amide protons suggest that several amide bonds are either involved in strong hydrogen bonds or buried. Interestingly, most of these amide protons belong to four main fragments of the sequence spanning residues $A^{37}-N^{46}$ of the helix and residues $Y^{55}-E^{59}$, $T^{79}-V^{84}$, and $R^{103}-V^{111}$ of the β-strands.

Three-Dimensional Structure Determination. The combination of the cis−trans isomerization of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds leads to four conformers: $transP^{88}-transP^{119}$, $transP^{88}-cisP^{119}$, $cisP^{88}-transP^{119}$, and $cisP^{88}-cisP^{119}$. Therefore, four conformer calculations must be taken into account.

The analysis of the H_N, H_α, and ^{15}N chemical shifts for the different conformers showed that the effects due to the $R^{87}-P^{88}$ amide bond involved mainly the chemical shifts of the residues in fragments $R^{87}-Q^{91}$ and to a lesser extent $L^{39}-L^{45}$ and $L^{53}-Y^{55}$. A similar analysis showed that for the isomerization of the $D^{118}-P^{119}$ amide bond, the chemical

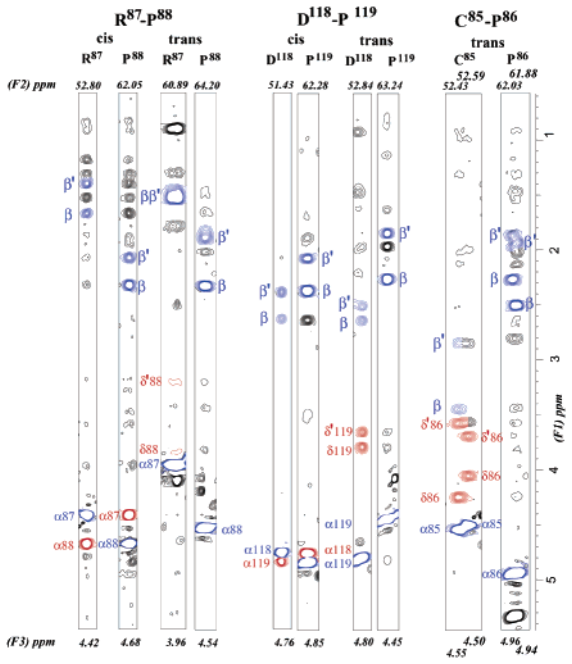

FIGURE 2: Strips of the ^{13}C-edited NOESY-HSQC spectrum of ProS recorded in D_2O showing the assignment of the cis and trans conformers due to $R^{87}-P^{88}$ or $D^{118}-P^{119}$ amide bond isomerization. The cross-peaks involving the trans $C^{85}-P^{86}$ amide bond are shown for comparison. The inter-residue NOE cross-peaks are displayed in red and the intraresidue cross-peaks in blue.

shifts of fragments $T^{112}-I^{122}$ and $R^{40}-L^{45}$ were strongly and weakly affected, respectively (see Figure 5A). This chemical shift analysis suggests that the main structural alterations are located in the vicinity of the two amide bonds, although on the basis of chemical shift variations some minor changes were detected for the helix. Preliminary calculations showed that the isomerization of the two amide bonds induced localized conformational alterations. Therefore, the structure calculation was only carried out for the two extreme conformers, transP^{88}−transP^{119} and cisP^{88}−cisP^{119}, later termed the all-trans and all-cis conformers, respectively.

In the case of the all-trans conformer, a total of 1227 distance constraints derived from NOE experiments recorded in H_2O and D_2O were used for structure calculations. Figure 4A shows for the all-trans conformer the distribution of these interproton distance restraints all along the sequence. The structure of the all-cis conformer was obtained by using practically the same number of NOE-derived distance restraints (a total of 1229) similarly distributed.

A set of 60 structures was calculated for both the all-trans and the all-cis conformers. The PROCHECK analysis of the 15 best structures showed that all ϕ and φ angles were localized to allowed (96.6%) and generously allowed (3.4%) areas of the Ramachandran plot. The statistics are summarized in Table 1.

Figure 4B shows the family of the 15 best all-trans structures superimposed by using the backbone atoms of residues 32−126. The mean value of the rmsd per residue as a function of the sequence indicates that the global 3D structure is well-defined except for the N- and C-terminal parts (Figure 4C).

The structure of the all-trans conformer consists of an N-terminal α-helix ($Y^{34}-Q^{48}$) cradled by four antiparallel

Table 1: Structural Statistics[a] of the All-Trans and All-Cis Conformers of ProS

	all-trans	all-cis
distance restraints		
intraresidue ($i-j=0$)	290	290
sequential ($i-j=1$)	449	448
medium-range ($i-j\leq 5$)	142	145
long-range ($i-j>5$)	346	346
total	1227	1229
dihedral angle restraints		
ϕ	36	37
χ_1	12	9
total	48	46
mean rmsd from experimental restraints		
NOE (Å)	0.0579 ± 0.0027	0.0548 ± 0.0031
dihedrals (deg)	0.464 ± 0.188	0.405 ± 0.146
mean rmsd from idealized covalent geometry		
bonds (Å)	0.00357 ± 0.00008	0.00342 ± 0.00008
angles (deg)	0.616 ± 0.007	0.611 ± 0.008
impropers (deg)	0.416 ± 0.011	0.413 ± 0.007
mean energy (kJ/mol)		
E_{noe}	61.46 ± 5.75	55.11 ± 6.10
E_{cdih}	0.182 ± 0.137	0.132 ± 0.086
E_{vdW}	−249.72 ± 14.06	−261.53 ± 11.92
E_{bond}	20.99 ± 0.92	19.33 ± 0.95
$E_{improper}$	21.56 ± 1.17	21.22 ± 0.67
E_{angle}	173.49 ± 4.01	170.97 ± 4.61
E_{total}	−442.67 ± 25.82	−460.88 ± 32.83
pairwise atomic rmsd difference (Å)		
backbone atoms (32−126)	1.13 ± 0.20	0.89 ± 0.14

[a] For these calculations, the XPLOR all-hydrogen force fields topoallhdg and parallhdg all were used. The final minimization of the structures was carried out with force constants of 15 kcal mol^{-1} Å^{-2} and 50 kcal mol^{-1} rad^{-2} for the NOE and dihedral angle potentials, respectively. None of the structures exhibited distance violations of >0.3 Å or dihedral violations of >5°. The program PROCHECK was used to assess the overall quality of the structures.

β-strands, β1 ($N^{53}-L^{60}$), β2 ($K^{74}-P^{86}$),[3] β3 ($V^{104}-V^{111}$), and β4 ($I^{122}-C^{124}$). These four β-strands are connected by three loops, L1 ($D^{61}-P^{73}$), L2 ($T^{112}-D^{121}$), and a third longer one ($T^{83}-G^{102}$, including the C-terminal part of β2), termed the appending domain (see below). This domain, which joins the β2 and β3 strands, contains the $C^{85}-C^{96}$ disulfide bridge and three proline residues (P^{88}, P^{92}, and P^{93}). Its well-defined structure consists of a β-turn for the $P^{86}RPT^{89}$ sequence, which contains the $R^{87}-P^{88}$ amide bond that is able to adopt both the cis and trans conformations, and a 3_{10}-helical turn ($P^{92}PEL^{95}$), including proline residues P^{92} and P^{93}.

Comparison of the All-Trans and All-Cis Conformers of ProS. The structures of the all-trans and the all-cis conformers display the same elements of secondary structure. Their superimposition highlights their similar global fold and reveals that the structure of the two conformers differ mainly in the neighbor of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds. This is supported both by the analysis of chemical shift alterations and by the local rmsd per residue measured for the two structures (Figure 5A,B). Indeed, except for the N- and C-terminal parts, the analysis of the rmsd per residue plot as a function of the sequence shows higher values for the two regions spanning residues $P^{86}-R^{90}$ and $Q^{115}-D^{121}$ than for the other parts of the molecule. Moreover, this

[3] The limits of the β2 strand were determined with STRIDE. On the basis of the interstrand hydrogen bonds, the β2 strand would span residues $K^{74}-E^{82}$.

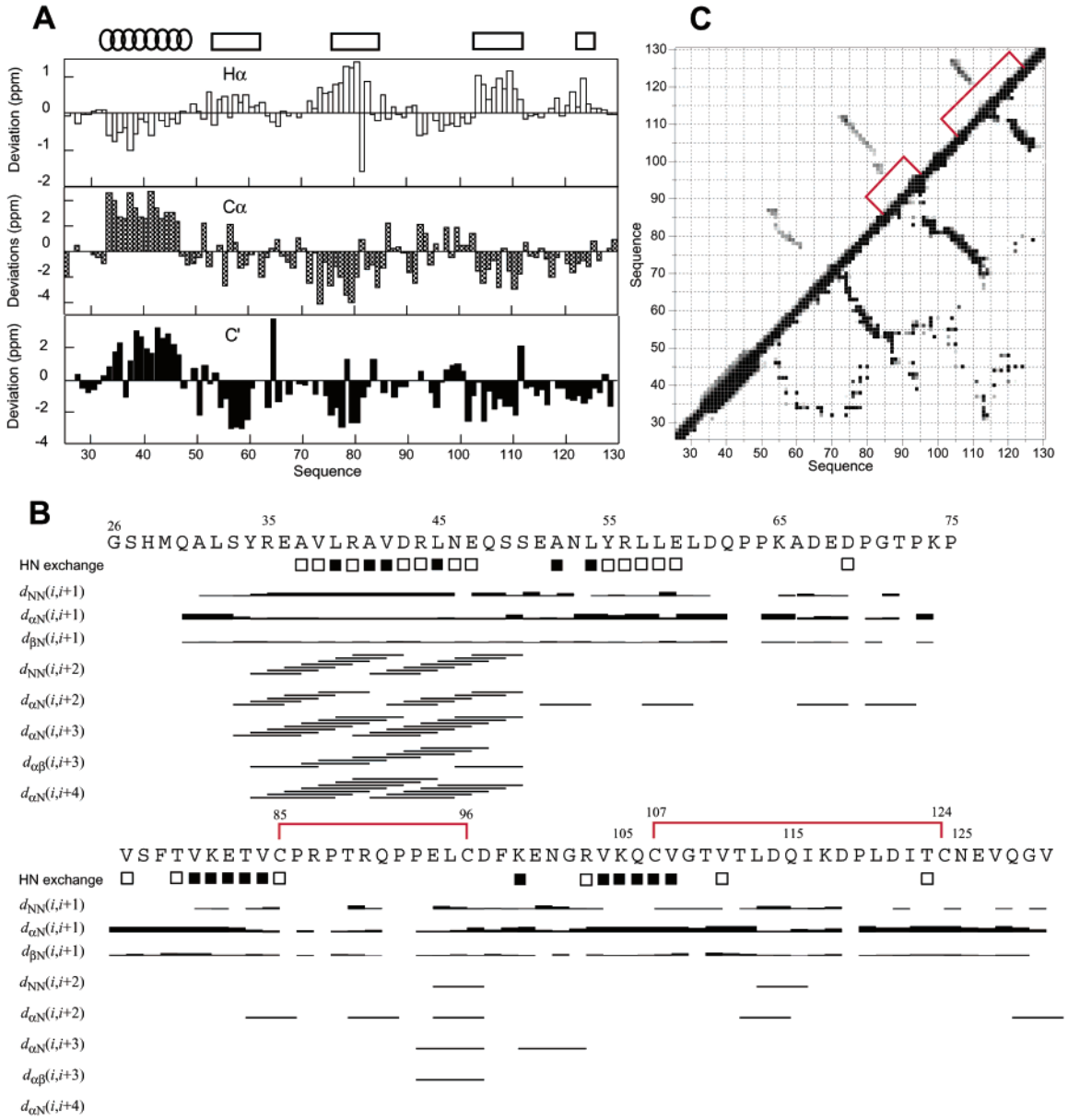

FIGURE 3: (A) Chemical shift deviations from random coil chemical shifts of H_α, C_α, and C′ resonances as a function of the sequence. Notice that the mean value of the chemical shifts of the four conformers was used for the calculation of deviations. The random coil chemical shifts are those of Wishart and Case (*43*). The locations of helical and β-strand structures are displayed above as stacked circles and rectangles, respectively. (B) Summary of the sequential and medium-range NOEs for the all-trans conformer of ProS. The relative intensity of NOEs is represented by the thickness of the bars. Slowly and intermediate-exchanging protons are identified by filled and empty squares, respectively. The two disulfide bonds are displayed in red. (C) Diagonal plot of the NOEs observed for the all-trans conformer of ProS. Backbone and side chains NOEs are displayed in the upper and lower parts, respectively. Short and long distances are displayed in black and gray squares, respectively.

analysis indicates that the extent of alterations due to the cis−trans isomerization is more localized for the $R^{87}-P^{88}$ amide bond isomerization, which is in a β-turn, than for the $D^{118}-P^{119}$ amide bond, which is in the L2 loop upstream from the β4 strand. The superimposition of these two areas of the two all-trans and all-cis conformers is displayed in panels C and D of Figure 5.

DISCUSSION

The initial length of the cathelin-like domain of the protegrin precursor is 101 residues. The overexpression of ProS in *E. coli* contains four additional amino acids in its N-terminus that were generated by subcloning into the *Nde*I and *Bam*HI sites of pET-15b as previously described (*24*). The ProS sequence is particularly poor in aromatic residues (4%; Y^{34}, Y^{55}, F^{78}, and F^{98}). This contrasts with the high content of proline, valine, and leucine residues (10% for each). Furthermore, the $C^{85}-C^{96}$ and $C^{107}-C^{124}$ disulfide bonds are conserved in all the cathelin-like domains (*24*).

Because of the presence of four conformers at pH 6.2, it initially appeared to be difficult to carry out the NMR structural study of ProS. However, despite the risk of the cleavage of the acid labile DP amide bonds ($D^{69}-P^{70}$ and $D^{118}-P^{119}$) (*3*, *44*), an attempt to perform this study was carried out at pH 3.0, since one conformer is significantly favored at this pH. It was then identified as being the $transP^{88}-transP^{119}$ conformer. Nevertheless, as previously discussed, because of the presence of the numerous proline

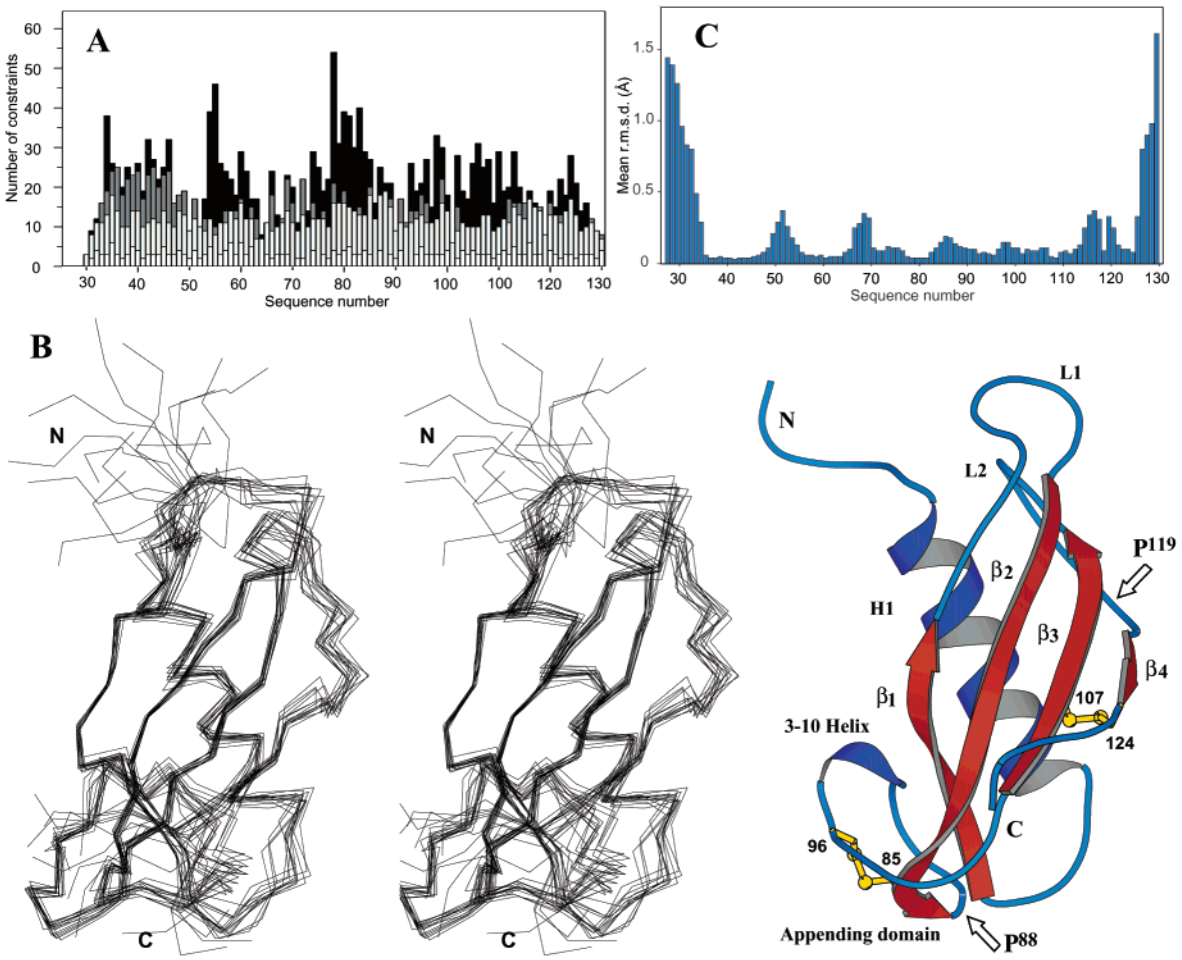

FIGURE 4: Distribution of NOE-derived restraints, calculated structures, and rmsds per residue for the all-trans conformers of ProS. (A) Number and distribution of the NOE-derived interproton distance restraints per residue as a function of the sequence. Intraresidue and short-, medium-, and long-range distance restraints are displayed from light gray to black, respectively. (B) Stereoview of the backbone of the global fold of a family of 15 ProS structures. The backbone atoms of residues 32−126 were used for the superimposition. Elements of secondary structure are labeled on the structure prepared with Molscript (*52*). The locations of P^{88} and P^{119} residues are shown by arrows. Disulfide bonds are shown in yellow. (C) Mean value for the rmsd per residue calculated for the backbone of the 15 best structures.

residues, only 80% of the assignment was obtained for this major conformer, and the minor ones could not be characterized (*24*).

Therefore, despite the presence of the four conformers, by using the ^{15}N and ^{13}C doubly labeled sample we performed the study at pH 6.2 for two main reasons. The first one was because it was interesting to know why the protein adopts four conformers. Moreover, given the fact that they are almost equally populated, it could be expected that each one could be characterized. The second reason takes into account the fact that electrostatic interactions were hypothesized to be essential in the interaction between the antibacterial peptide, usually positively charged, with the acidic side chains of the prosequence. This hypothesis was proposed in the case of the human α-defensin HNP-1 (*18*). Therefore, experimental conditions of pH in which the carboxyl groups are fully ionized, supposed to better mimic biological conditions, appeared to be more relevant to our study.

Fold and Structure of ProS. On the basis of the analysis of chemical shift deviations, elements of secondary structure were approximately located and confirmed for the all-trans and the all-cis calculated conformers. Structure calculations obtained from NMR-derived constraints showed that the ProS global fold consists of an N-terminal helix wrapped by a antiparallel four-stranded β-sheet linked by the L1 (D^{61}−V^{76}) and L2 (T^{112}−D^{121}) loops and the appending domain (T^{83}−G^{102}). The C^{107}−C^{124} and C^{85}−C^{96} disulfide bonds provide stabilization to the β-sheet and to the appending domain, respectively. The β-sheet is stabilized by a network of six, eight, and five hydrogen bonds at the $\beta1$−$\beta2$, $\beta2$−$\beta3$, and $\beta3$−$\beta4$ interfaces, respectively. This is supported by the H−D exchange study, which showed that most of the slowly exchanging amide protons belong to the helix and to the β-sheet, and therefore highlights the stability of these two structural elements which, whatever the conformer, constitute the hydrophobic core of the ProS structure (see below). Nevertheless, the large hydrogen bond network and the hydrophobic core are not sufficient to maintain the integrity of the 3D structure of ProS. Indeed, the DTT treatment of ProS induced alterations in the ^1H NMR spectrum, thus suggesting that the disulfide bonds are essential to maintenance of the global fold of ProS (data not shown).

For the structure of the all-trans conformer, on the basis of the last hydrogen bond between the $\beta1$ and $\beta2$ strands (D^{61} HN−S^{77} O), the L1 loop spans residues Q^{62}−V^{76} and includes five prolines (P^{63}, P^{64}, P^{70}, P^{73}, and P^{75}). Its structure

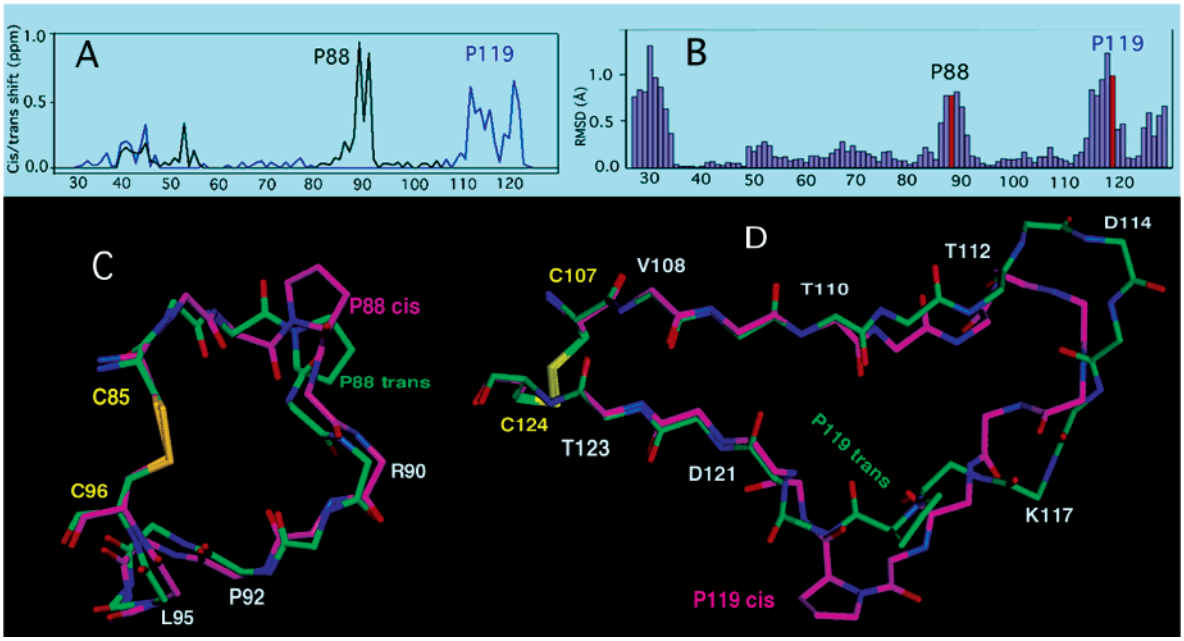

FIGURE 5: Comparison of the chemical shift of the cis and trans conformers as well as of the structures of the all-trans and all-cis conformers of ProS. (A) Chemical shift differences of amide resonances between the cis and trans conformers due to the isomerization of the $R^{87}-P^{88}$ (black line) and $D^{118}-P^{119}$ (blue line) amide bonds, respectively. For proline residues, the chemical shift differences measured for the δ proton were used. (B) Mean rmsd values per residue (backbone atoms) obtained with a three-residue smoothing window all along the sequence. (C and D) Superimposition of the backbone of the cis (purple) and trans (green) conformers (C) around the $R^{87}-P^{88}$ amide bond and (D) around the $D^{118}-P^{119}$ amide bond. Disulfide bonds are shown in yellow.

consists of a large macrocycle which starts with a well-defined bulge ($P^{63}-A^{66}$) identified by NOEs between the Y^{34} aromatic ring and the P^{63}, A^{66}, and D^{67} side chains. P^{70} is located in the middle of the L1 loop, whereas P^{73} and P^{75} belong to the C-terminal part of the loop. Four of these five prolines are solvent-exposed, whereas P^{63} is relatively buried.

Similarly, the last hydrogen bond between the β3 and β4 strands (T^{110} HN$-D^{121}$ O) delimits the L2 loop, which spans residues $V^{111}-D^{121}$ and consequently includes the $D^{118}-P^{119}$ amide bond that is able to adopt the cis and trans conformations. This loop consists of a large macrocycle whose structure was defined by several NOEs involving residues $T^{112}-Q^{115}$, $T^{112}-P^{119}$, and $T^{112}-D^{118}$. In addition, NOEs between residues of the L2 loop with those of the helix ($D^{114}-A^{31}$, $V^{111}-Y^{34}$, and $L^{113}-A^{37}$) and with those of the β2 strand ($L^{113}-K^{74}$, $L^{113}-P^{70}$, $T^{112}-P^{73}$, and $T^{110}-P^{75}$) were essential to positioning the L2 loop with regard to the helix and the L1 loop, respectively. Unlike the β-strands which lie approximately in the same plane to give rise to the β-sheet, the L1 and L2 loops are located in two different planes. NOEs involving the L^{113} side chain were particularly useful in defining the L1$-$L2 interface. The L2 loop was found to overlap with the middle of the helix, whereas the L1 loop is close to the N-terminus of the helix.

The ProS hydrophobic core consists of the interface between the helix and the concave side of the β-sheet. It involves Y^{34}, V^{38}, L^{39}, V^{42}, and L^{45} side chains of the helix and L^{54}, Y^{55}, L^{57}, L^{60}, P^{63}, A^{66}, V^{76}, F^{78}, V^{80}, C^{107}, V^{111}, L^{113}, L^{120}, I^{122}, and C^{124} side chains of the β-sheet.

The chemical shifts of E^{82} H_α (2.79 and 2.72 ppm) and C^{107} $H_{\beta\beta'}$ (1.84 and 1.02 ppm) showed two unusual values since they are particularly upfield shifted. The structure shows that the E^{82} H_α experiences the shielding area of the Y^{55} aromatic cycle since it is located 3.5 Å from the cycle. With regard to the C^{107} $\beta\beta'$ protons, they are located in the shielding area of the F^{78} ring with a mean distance of 3.0 Å. In contrast, the ^{1}H and ^{15}N chemical shifts of the N^{46} amide side chain are downfield shifted (8.56$-$8.29 and 116.7$-$116.4 ppm, respectively). Indeed, the structure shows that the N^{46} side chain is buried in a hydrophobic cluster (V^{42}, L^{45}, L^{54}, Y^{55}, L^{57}, and P^{93}) and involved in two hydrogen bonds with the Y^{55} backbone, thus explaining the observed downfield shift.

The resonances of the hydroxyl protons of threonine are rarely observed due to their exchange with water. Interestingly, that of T^{83} was observed at 4.58 and 4.60 ppm, suggesting that it is particularly buried and/or involved in a hydrogen bond. The structure shows that it is both surrounded by hydrophobic residues (L^{54}, Y^{55}, R^{56}, V^{84}, P^{93}, C^{85}, C^{96}, and F^{98}) and involved in a hydrogen bond with the oxygen atom of the L^{54} residue (T^{83} O$\gamma-L^{54}$ O distance of 2.9 Å) and is therefore in slow exchange with the solvent.

Structure Alterations Induced by the Isomerization of $R^{87}-P^{88}$ and $D^{118}-P^{119}$ Amide Bonds. Whatever the conformer, it appeared that the chemical shifts of some residues were not affected by the isomerization of the $D^{118}-P^{119}$ and $R^{87}-P^{88}$ amide bonds and displayed a unique resonance. This was the case for L^{58} and E^{100}, located at the surface of the molecule. In contrast, chemical shifts of some residues were strongly affected by the isomerization of the two amide bonds. For example, around the P^{88} residue (H_α at 4.54 and 4.68 ppm), the chemical shifts of R^{87} (HN at 9.27 ppm, H_α at 3.96 and 8.05 ppm, and H_α at 4.42 ppm) and T^{89} (HN at 7.55 ppm, H_α at 4.10 and 8.48 ppm, and H_α at 4.57 ppm) residues are very different (Figure 1). These chemical shifts are mainly representative of the trans and cis conformers of the $R^{87}-P^{88}$ amide bond, respectively. A similar observation was made for the P^{119} amide bond, where the main chemical shifts differences were measured for D^{118} (HN at 8.09 ppm, H_α at 4.80 and 8.13 ppm, and H_α at 4.76 ppm), P^{119} (H_α at

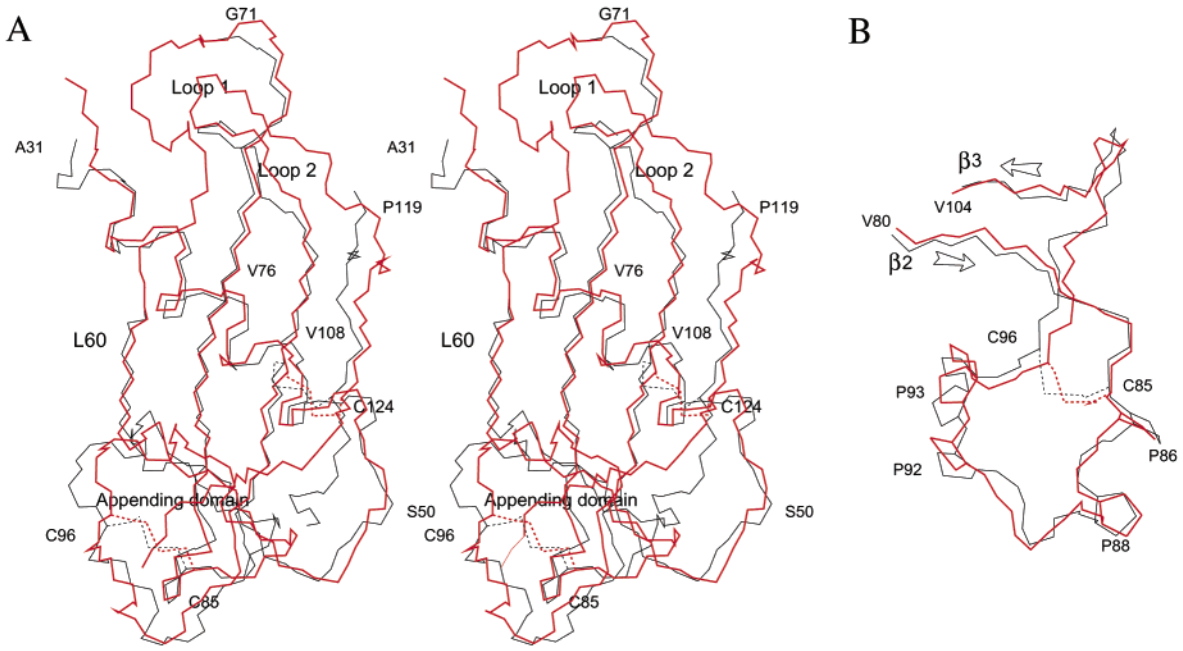

FIGURE 6: Comparison of the backbone structure of ProS in solution (all-trans conformer in red) and in the crystal (1kwi). Disulfide bonds are displayed as dashed lines. (A) Stereoview showing the global fold of ProS. The appending domain was excluded from the superimposition, which used 59 residues (rmsd of 0.938 Å). The locations of some residues are indicated. (B) Superimposition of the appending domain (residues 81−103, rmsd of 1.002 Å), showing the location of the proline residues.

4.45 and 4.85 ppm), L^{120} (HN at 8.40 ppm, H_α at 4.34 and 8.62 ppm, and H_α at 4.24 ppm), D^{121} (HN at 8.79 ppm, H_α at 4.83 and 9.06 ppm, and H_α at 4.90 ppm), and I^{122} (HN at 7.57 ppm, H_α at 4.72 and 7.10 ppm, and H_α at 4.77 ppm) residues. These chemical shifts are mainly representative of the trans and cis conformers of the $D^{118}-P^{119}$ amide bond. Residues $R^{40}-Y^{55}$ and $R^{87}-E^{94}$ are roughly located between the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds. Therefore, they experience conformational alterations from the two amide bonds, leading to four chemical shifts representative of the four different environments in the four conformers. The four cross-peaks are clearly observed for the $R^{40}-R^{56}$ sequence, indicating that this sequence experiences the cis−trans isomerization of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds.

Comparison of the all-trans and all-cis conformers of ProS showed that the main structural changes are located in the vicinity of the two $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds and that they do not involve the global fold. An rmsd of 0.616 Å was measured for the superimposition of the 72 residues of the core (33−85, 96−110, and 122−125). As shown in Figure 5, the main conformational changes are localized in the L2 loop (rmsd of 1.47 Å for the superimposition of residues 107−124). Indeed, the cis−trans alterations are extended all along the L2 loop ($V^{111}-D^{121}$). The L^{120} δCH_3−R40 H_α NOE which is strong and weak for the cis and trans conformers, respectively, and the NOEs between T^{112} and K^{117} side chains, only observed for the all-cis conformer, characterize some of the structural alterations of the L2 loop due to isomerization of the $D^{118}-P^{119}$ amide bond (Figure 5). In the case of the $R^{87}-P^{88}$ amide bond, which belongs to a β-turn structure, the induced conformational changes are more localized than they are for the $D^{118}-P^{119}$ amide bond and in the appending domain (rmsd of 1.19 Å for the superimposition of residues 86−96). Nevertheless, the isomerization of the $R^{87}-P^{88}$ and $D^{118}-P^{119}$ amide bonds induced some minor alterations detected by the analysis of chemical shifts for $R^{40}-Y^{55}$ and $A^{37}-E^{46}$, respectively. They could result from a small displacement of the β-sheet structure with regard to the helix upon isomerization of the two amide bonds.

Comparison of the Solution and Crystalline Structures of ProS. ProS was crystallized under acidic conditions (pH 3.8), and two types of crystals were obtained: one type with the monomeric form, for which the electronic density for the L1 and L2 loops was not observed, and another type in which the domain-swapped dimer was observed (*23*).

In solution at pH 6.2, on the basis of the line width, ProS was assumed to be monomeric. Moreover, critical NOEs involving the Y^{34} aromatic ring and P^{63} and A^{66} side chains are inconsistent with the domain-swapped dimer as determined in the crystal (*23*). Indeed, these distances are ∼6−7 and ∼10−11 Å. Nevertheless, it should be noticed that if the ProS sample is stable around neutral pH for 3−4 months, under acidic conditions, broad signals suggesting dimerization were observed.

Interestingly, the solution structure allowed us to determine the structures of the L1 and L2 loops, which were not defined in the X-ray structure of the monomer. Therefore, the structure of the $D^{118}-P^{119}$ amide bond, which belongs to the L2 loop, was not observed. In contrast, the well-defined electronic density of the $R^{87}-P^{88}$ amide bond characterized its trans conformation in the crystal structure. Therefore, the all-trans conformers and the crystal structures could be compared.

The superimposition of the all-trans and X-ray structures is displayed in Figure 6. The global fold of the two structures is very similar with an rmsd value of 0.938 Å for the superimposition of backbone atoms of residues 34−61, 72−85, 101−113, and 121−124 (59 residues), excluding the appending domain, and of 1.181 Å for the superimposition of residues 34−61, 72−113, and 121−124 (74 residues), including the appending domain. An rmsd of 1.002 Å was

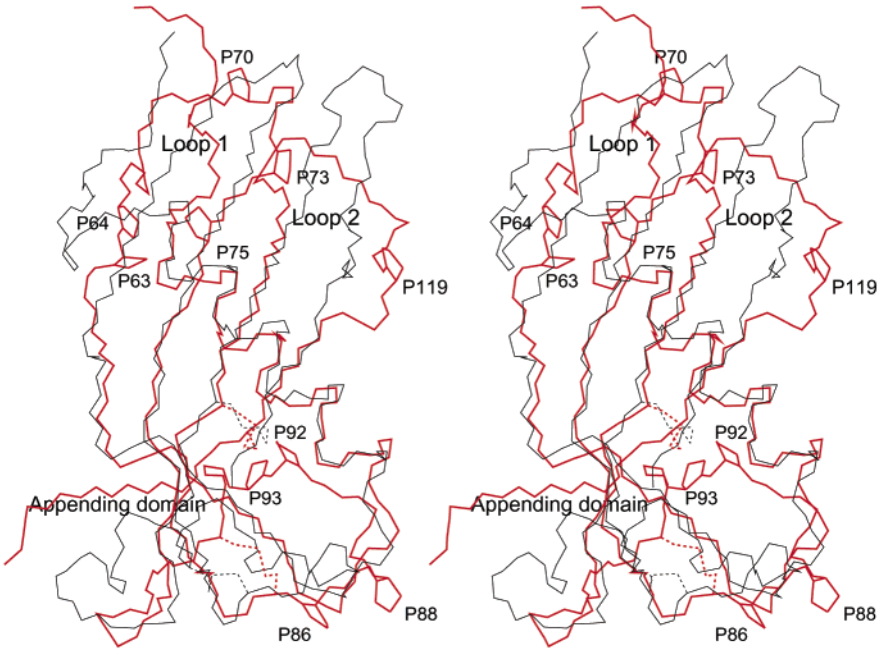

FIGURE 7: Stereoview of the superimposition of ProS (the all-trans conformer) and the chicken cystatin (1cew) structures displayed in red and black, respectively. For the global fold superimposition, 44 residues (rmsd of 1.014 Å) were used. The superimposition shows the similar fold and the β-sheet differences. The appending domain structures of ProS and cystatin are significantly different. Notice the similar location of the disulfide bonds displayed as dashed lines in the two structures. Proline residues of the ProS structure are labeled.

measured for the superimposition of residues 81−103 (23 residues), corresponding to the appending domain. These rmsd values indicate that the solution and the X-ray structure of the ProS monomer are very similar.

Can the ProS Structure Explain the pH-Mediated Conformational Changes Previously Observed? To explain the pH-mediated conformational changes, observed in the pH range for the ionization of carboxyl groups, the location of acidic residues in the close environment of the R^{87}−P^{88} and D^{118}−P^{119} amide bonds was examined. The E^{47} side chain is located in the vicinity of the R^{87}−P^{88} amide bond, whereas the D^{43} side chain is farther away. In contrast, the E^{51} side chain is oriented in the opposite direction. The D^{118} and D^{121} side chains are located in the close environment of the D^{118}−P^{119} amide bond. Nevertheless, it is difficult to conclude that for these two amide bonds the ionization change of their closest acidic side chains is the only explanation for the conformational change or to suggest a mechanism responsible for this. Furthermore, although more distant, some other acidic side chains are possibly involved in salt bridges that are thought to contribute to the stability of the structure (D^{118}−R^{40}, E^{51}−K^{105}, E^{47}−R^{87}, and to a lesser extent E^{82}−R^{103} and E^{94}−R^{56}). We therefore cannot exclude their indirect contribution in the equilibrium among the four conformers. To address this point, the mutation of either P^{88}, P^{119}, or both could be an interesting approach.

Comparison of the ProS Structure with That of the Chicken Cystatin. The structure of the chicken cystatin was first established by X-ray (*38*) and then by NMR (*39*, *40*). Its fold is basically characterized by a helix cradled by a five-stranded β-sheet and is the typical representative of the cystatin fold family. The cystatin family fold, which belongs to the cystatin/monellin superfamily (http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.e.ca.b.html), contains phytocystatin (oryzacystatin-I, PDB entry 1eqk) (*45*), chicken cystatin (PDB entry 1cew) (*38*−*40*), cystatin A (stefin A, PDB entries 1dvc and 1dvd) (*46*), cystatin B (stefin B, PDB entry 1stf) (*47*), and cystatin C (PDB entry 1g96) (*48*). These structures which share the same fold differ in the presence of the appending domain as well as in the presence of two disulfide bonds (*40*, *48*).

The all-trans ProS structure was compared to both the solution and X-ray structures of chicken cystatin (Figure 7). Whatever chicken cystatin structure was considered, it appeared that ProS and the chicken cystatin structures share a similar global fold. An rmsd of 1.014 Å (1cew) was measured for the superimposition of the backbone atoms of 44 residues (15 residues in the helix and 29 in the β-sheet), excluding the appending domain. In contrast, for a similar superimposition with its solution structure (1a67) a higher value (2.573 Å) was measured. Nevertheless, as shown in Figure 7, the cystatin and ProS structures differ significantly both in their L1 and L2 loops and in their appending domain structures.

The β-strands are longer in cystatin, resulting in a more extended β-sheet structure stabilized by a regular network of 30 hydrogen bonds, as opposed to 19 for ProS. Consequently, the β2 and β3 strands are more tightly linked together in the cystatin structure than they are in ProS, and the L1 and L2 loops are very short in chicken cystatin.

This difference in the β-sheet structure can be easily explained by a sequence analysis. The high content of proline residues (P^{63}, P^{64}, P^{70}, P^{73}, P^{75}, and P^{119}) of ProS significantly decreases the number of interstrand hydrogen bonds when compared to that of the chicken cystatin β-sheet which contains only one proline residue (P^{103}), located in the L2 loop. The lack of hydrogen bonding in ProS results in larger loops at the expense of the β-sheet structure.

The ProS and chicken cystatin appending domains were also compared. They are approximately similar in size and contain two cysteines involved in a disulfide bond generating macrocycles of 12 and 11 residues, respectively. It appears

that the appending domain of ProS is closer to the helix than that of the chicken cystatin structures (1cew and 1a67). The appending domain superimposition shows that they roughly have a similar global fold, although the 1cew appending domain displays a higher helical content, possibly due to the crystal packing (*40*). However, their detailed structures are significantly different, with rmsds of 3.954 (1cew) and 3.983 Å (1a67) for the backbone superimposition focused around the disulfide bridge ($K^{81}-P^{86}$ and $L^{95}-R^{103}$ for ProS with $G^{67}-P^{72}$ and $S^{80}-E^{88}$ for cystatin) since their loops have different sizes. Their macrocycle sequences are unrelated since they only share the two cysteines and a proline residue (P^{87} in ProS and P^{72} in the chicken cystatin). The sequence of the appending domain of ProS is also characterized by a high proline content (P^{86}, P^{87}, P^{92}, and P^{93}) compared with only one (P^{72}) in cystatin.

Despite a low level of sequence identity between it and cystatins (<20%), the ProS structure shares the cystatin fold (Figure 1). Moreover, like the chicken cystatin, ProS displays an equivalent appending domain as well as two equivalent disulfide bonds similarly located. In contrast, the N-terminal β-strand is lacking. From this comparison, it is clear that the ProS structure, and therefore the cathelin-like domain structures, belongs to the cystatin fold family. Moreover, although the biological role of the cathelin-like domain has not yet been clearly identified, it appears that the cystatin fold is able to host diverse biological functions from protease inhibition (cystatins) to sweetness (brazzein, a sweet protein) and here to a possible involvement in the processing of antibacterial peptides in mammals. This suggests that the cystatin fold could be a common ancestral scaffold which acquired various biological functions during evolution.

CONCLUSION

The solution structure of ProS, in which two amide bonds are able to adopt the cis and trans conformations, has been determined by double-isotope labeling (^{15}N and ^{13}C) and heteronuclear NMR at pH 6.2. To our knowledge, this is the first example of protein for which all four conformers have been assigned. The solution structures of the all-trans and all-cis conformers of ProS were determined. The structure of the cathelin-like domain of protegrins was found to display the cystatin fold despite a low level of sequence similarity. This structural study complements previously published solution (*24*) and X-ray studies (*23*) and allowed us to characterize the four conformers for the monomeric form of ProS. In contrast, the crystal structure revealed the ability of the ProS structure to give rise to a domain-swapped dimer which appears to be a feature of the cystatins (*48*−*50*). However, it seems that if such swapped dimers are physiologically relevant for some proteins, this is not the case for others (*51*).

The determination of the ProS structure is a first step in the study of the processing of antibacterial peptides in mammals. It is interesting to note that the P^{88} and P^{119} residues are only found in pig precursors of antimicrobial peptides, whatever their structure (β-sheet or helical). Consequently, it is difficult to establish a relevant relationship between the structure of the antimicrobial peptide and the role, if any, of the cis−trans isomerization in their maturation process. To gain insights into the interactions between the protegrin-3 and its cathelin-like domain, a structural study of the holoprotein is in progress to validate the previously proposed model (*23*). This structure of this model which displays an accessible elastase cleavage site consists of a two-stranded extension of the ProS β-sheet by the protegrin. Such a structure could help in the understanding of the processing of cathelicidin antimicrobial peptides that are essential host-defense molecules of innate immunity in mammals. Moreover, some aspects of this structural study could be useful in the design of new antibacterial peptides of therapeutic interest.

ACKNOWLEDGMENT

We thank Christian Dumas, Tomas Ganz, and Robert Lehrer for helpful discussions. We are indebted to Dr. S. L. Salhi for the editorial revision of the manuscript.

SUPPORTING INFORMATION AVAILABLE

^{1}H, ^{13}C, and ^{15}N chemical shifts of ProS conformers. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Zhao, C., Liu, L., and Lehrer, R. I. (1994) *FEBS Lett. 346*, 285−288.
2. Zanetti, M., Gennaro, R., and Romeo, D. (1995) *FEBS Lett. 374*, 1−5.
3. Storici, P., Tossi, A., Lenarcic, B., and Romeo, D. (1996) *Eur. J. Biochem. 238*, 769−776.
4. Scocchi, M., Wang, S., and Zanetti, M. (1997) *FEBS Lett. 417*, 311−315.
5. Wang, Y., Agerberth, B., and Johansson, J. (1998) *J. Protein Chem. 17*, 522−523.
6. Zanetti, M., Gennaro, R., Scocchi, M., and Skerlavaj, B. (2000) *Adv. Exp. Med. Biol. 479*, 203−218.
7. Gennaro, R., Scocchi, M., Merluzzi, L., and Zanetti, M. (1998) *Biochim. Biophys. Acta 1425*, 361−368.
8. Ramanathan, B., Davis, E. G., Ross, C. R., and Blecha, F. (2002) *Microbes Infect. 4*, 361−372.
9. Zaiou, M., and Gallo, R. L. (2002) *J. Mol. Med. 80*, 549−561.
10. Zarember, K. A., Katz, S. S., Tack, B. F., Doukhan, L., Weiss, J., and Elsbach, P. (2002) *Infect. Immun. 70*, 569−576.
11. Zhao, C., Ganz, T., and Lehrer, R. I. (1995) *FEBS Lett. 368*, 197−202.
12. Gudmundsson, G. H., Agerberth, B., Odeberg, J., Bergman, T., Olsson, B., and Salcedo, R. (1996) *Eur. J. Biochem. 238*, 325−332.
13. Zanetti, M., Litteri, L., Griffiths, G., Gennaro, R., and Romeo, D. (1991) *J. Immunol. 146*, 4295−4300.
14. Panyutich, A., Shi, J., Boutz, P. L., Zhao, C., and Ganz, T. (1997) *Infect. Immun. 65*, 978−985.
15. Gennaro, R., and Zanetti, M. (2000) *Biopolymers 55*, 31−49.
16. Sorensen, O. E., Follin, P., Johnsen, A. H., Calafat, J., Tjabringa, G. S., Hiemstra, P. S., and Borregaard, N. (2001) *Blood 97*, 3951−3959.
17. Verbanac, D., Zanetti, M., and Romeo, D. (1993) *FEBS Lett. 317*, 255−258.
18. Valore, E. V., Martin, E., Harwig, S. S., and Ganz, T. (1996) *J. Clin. Invest. 97*, 1624−1629.
19. Kokryakov, V. N., Harwig, S. S., Panyutich, E. A., Shevchenko, A. A., Aleshina, G. M., Shamova, O. V., Korneva, H. A., and Lehrer, R. I. (1993) *FEBS Lett. 327*, 231−236.
20. Aumelas, A., Mangoni, M., Roumestand, C., Chiche, L., Despaux, E., Grassy, G., Calas, B., and Chavanieu, A. (1996) *Eur. J. Biochem. 237*, 575−583.
21. Fahrner, R. L., Dieckmann, T., Harwig, S. S., Lehrer, R. I., Eisenberg, D., and Feigon, J. (1996) *Chem. Biol. 3*, 543−550.
22. Mirgorodskaya, O. A., Shevchenko, A. A., Abdalla, K. O., Chernushevich, I. V., Egorov, T. A., Musoliamov, A. X., Kokryakov, V. N., and Shamova, O. V. (1993) *FEBS Lett. 330*, 339−342.

23. Sanchez, J. F., Hoh, F., Strub, M. P., Aumelas, A., and Dumas, C. (2002) *Structure 10*, 1363−1370.
24. Sanchez, J. F., Wojcik, F., Yang, Y. S., Strub, M. P., Strub, J. M., Van Dorsselaer, A., Martin, M., Lehrer, R., Ganz, T., Chavanieu, A., Calas, B., and Aumelas, A. (2002) *Biochemistry 41*, 21−30.
25. Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W., and Wingfield, P. T. (1995) *Current Protocols in Protein Science*, Vol. 1, John Wiley & Sons, New York.
26. Palmer, A. G. I., Cavanagh, J., Wright, P. E., and Rance, M. (1991) *J. Magn. Reson. 93*, 151−170.
27. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson. 85*, 393−399.
28. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wuthrich, K. (1998) *Eur. J. Biochem. 256*, 1−15.
29. Bax, A., and Grzesiek, S. (1993) Methodological Advances in Protein NMR, in *NMR of Proteins* (Clore, G. M., and Gronenborn, E., Ed.) pp 33−52, Macmillan, London.
30. Bax, A., and Pochapsky, S. (1992) *J. Magn. Reson. 99*, 638−644.
31. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.
32. Shaka, A. J., Lee, C. J., and Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
33. Skinner, T. E., and Bendall, M. R. (1997) *J. Magn. Reson. 124*, 474−478.
34. Rice, L. M., and Brunger, A. T. (1994) *Proteins 19*, 277−290.
35. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
37. Frishman, D., and Argos, P. (1995) *Proteins 23*, 566−579.
38. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J. 7*, 2593−2599.
39. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jaenicke, R., and Oschkinat, H. (1993) *J. Mol. Biol. 234*, 1048−1059.
40. Engh, R. A., Dieckmann, T., Bode, W., Auerswald, E. A., Turk, V., Huber, R., and Oschkinat, H. (1993) *J. Mol. Biol. 234*, 1060−1069.
41. Fisher, G. (2000) *Chem. Soc. Rev. 29*, 119−127.
42. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
43. Wishart, D. S., and Case, D. A. (2001) *Methods Enzymol. 338*, 3−34.
44. Ritonja, A., Kopitar, M., Jerala, R., and Turk, V. (1989) *FEBS Lett. 255*, 211−214.
45. Nagata, K., Kudo, N., Abe, K., Arai, S., and Tanokura, M. (2000) *Biochemistry 39*, 14753−14760.
46. Martin, J. R., Craven, C. J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V., and Waltho, J. P. (1995) *J. Mol. Biol. 246*, 331−343.
47. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) *EMBO J. 9*, 1939−1947.
48. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001) *Nat. Struct. Biol. 8*, 316−320.
49. Ekiel, I., and Abrahamson, M. (1996) *J. Biol. Chem. 271*, 1314−1321.
50. Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R., Craven, C. J., and Waltho, J. P. (2001) *EMBO J. 20*, 4774−4781.
51. Liu, Y., and Eisenberg, D. (2002) *Protein Sci. 11*, 1285−1299.
52. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI027133C