NMR Solution Structure of the Antifungal Protein from *Aspergillus giganteus*: Evidence for Cysteine Pairing Isomerism[†,‡]

Ramón Campos-Olivas,[§] Marta Bruix,[§] Jorge Santoro,[§] Javier Lacadena,[||] Alvaro Martinez del Pozo,[||] José G. Gavilanes,[||] and Manuel Rico*[,§]

Instituto de Estructura de la Materia, Consejo Superior de Investigaciones Cientificas, Serrano 119, 28006 Madrid, Spain, and Departamento de Bioquimica y Biologia Molecular I, Facultad de Quimica, Universidad Complutense, 28040 Madrid, Spain

Received November 8, 1994; Revised Manuscript Received December 21, 1994[⊗]

ABSTRACT: The solution structure of the antifungal protein (AFP, 51 residues, 4 disulfide bridges) from *Aspergillus giganteus* has been determined by using experimentally derived interproton distance constraints from nuclear magnetic resonance (NMR) spectroscopy. Complete sequence-specific proton assignments were obtained at pH 5.0 and 35 °C. A set of 834 upper limit distance constraints from nuclear Overhauser effect measurements was used as input for the calculation of structures with the program DIANA. An initial family of 40 structures calculated with no disulfide constraints was used to obtain information about the disulfide connectivities, which could not be determined by standard biochemical methods. Three possible disulfide patterns were selected and the corresponding disulfide constraints applied to generate a family of 20 DIANA conformers for each pattern. Following energy minimization, the average pairwise RMSD of the 20 conformers of each family is 1.01, 0.89, and 1.01 Å for backbone atoms and 1.82, 1.74, and 1.81 Å for all heavy atoms. One of these three families contains the disulfide bridge arrangement actually present in the solution structure of AFP. Although the three families fulfill the NMR constraints, one of the disulfide patterns considered (cysteine pairs 7–33, 14–40, 26–49, 28–51) is favored among the others on the basis of previous chemical studies. It thus probably corresponds to the actual pattern of disulfide bridges present in the protein, and the corresponding family represents the solution structure of AFP. The folding of AFP consists of five antiparallel β strands connected in a −1, −1, +3, +1 topology and highly twisted, defining a small and compact β barrel stabilized by four internal disulfide bridges. A cationic site formed by up to three lysine side chains adjacent to a hydrophobic stretch, both at the protein surface, may constitute a potential binding site for phospholipids which would be the basis of its biological function. On the other hand, a second, minor form of AFP has been detected. NMR data, together with results from mass spectrometry, chemical analysis, and sedimentation equilibrium, suggest that this species differs from the major form in the pairs of cysteines involved in the four disulfide bridges.

Several small proteins with antifungal activity have been isolated from plants and are believed to be involved in a defense mechanism against phytopathogenic fungi by inhibiting fungal growth through diverse molecular modes, such as binding to chitin or permeabilizing fungal membranes or cell walls. They include the barley thionins (Bohlmann et al., 1988), stinging nettle lectin (Broekart et al., 1989), maize zeamatin (Roberts & Seletrennikoff, 1990), and rubber-tree hevein (Parijs et al., 1990). However, antifungal proteins are not exclusively produced by plants. A small, very basic protein with antifungal activity has been purified from the extracellular medium of the imperfect ascomycete *Aspergillus giganteus* (Olson & Goerner, 1965). The so-called antifungal protein (AFP)[1] displays an inhibitory activity against the growth of a variety of filamentous fungi, having no effect on the growth of mammalian cells, yeast, and eubacteria. It consists of 51 amino acids with a high content (12) of lysine residues, and eight cysteine residues which are all involved in disulfide bridges (Nakaya et al., 1990). The structural organization of the gene encoding AFP and its transfer into another filamentous fungi, *Aspergillus niger*, has also been described (Wnendt et al., 1994).

† This work was supported by the Direccion General de Investigacion Cientifica y Tecnica (Spain) (PB90/0120 and PB93/0090). R.C.O. was recipient of a predoctoral grant (FP92-02882770) from the Ministerio de Educacion y Ciencia (Spain).

‡ The atomic coordinates for the 40 "R" DIANA conformers have been deposited in the Brookhaven Protein Data Bank under the ident code 1AFP.

* Author to whom correspondence should be addressed (Fax: 34-1-5642431).

§ Instituto de Estructura de la Materia.

|| Universidad Complutense.

⊗ Abstract published in *Advance ACS Abstracts*, February 15, 1995.

1 Abbreviations: AFP, antifungal protein from *Aspergillus giganteus*; CITY, computer-improved total correlation spectroscopy; COSY, correlated spectroscopy; 2D, two dimensional; $d_{\alpha\alpha}(i,j)$, NOE connectivity between the C_αH proton on residue *i* and the C_αH proton on residue *j*; $d_{\alpha N}(i,j)$, NOE connectivity between the C_αH proton on residue i and the NH proton on residue *j*; $d_{\beta N}(i,j)$ NOE connectivity between the C_βH proton on residue *i* and the NH proton on residue *j*; DMPC, dimyristoylphosphatidylcholine; DMPS, dimyristoylphosphatidylserine; $d_{NN}(i,j)$, NOE connectivity between the NH proton on residue *i* and the NH proton on residue *j*; HPLC, high performance liquid chromatography; M_r, relative molecular mass; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; PAGE, polyacrylamide gel electrophoresis; ppm, parts per million; REM, restrained energy minimization; RMSD, root mean square deviation; ROESY, rotating frame Overhauser effect spectroscopy; SDS, sodium dodecyl sulfate; SNS, nuclease from *Staphylococcus aureus*; TFA, trifluoroacetic acid; TOCSY, total correlation spectroscopy; Tris, tris(hydroxymethyl)-aminomethane; TSP, sodium 3-trimethylsilyl[2,2,3,3-$^2H_4$]propionate; WATERGATE, water suppression by gradient-tailored excitation.

0006-2960/95/0434-3009$09.00/0

This small size antifungal protein can be isolated in a relatively high amount from *A. giganteus* culture medium. This fact and its high thermal stability make AFP a suitable structure for NMR studies. In addition, this protein displays interesting structural properties. Two out of the four disulfide bridges reported for AFP are of the $i-i+2$ type (Nakaya et al., 1990). To our knowledge, this is the first example in the literature where two $i-i+2$ disulfide bridges are reported to be present in the same polypeptide chain. Only in one other case, human chorionic gonadotropin α subunit, one -Cys-X-Cys- disulfide bridge has been identified (Mise & Bahl, 1980). These disulfide bridges are intrinsically unstable as shown by the experimentally determined disulfide exchange rate constants for the formation and opening of small disulfide loops in model peptides: the equilibrium constants for -Cys-X-Cys- loop closing present the smallest values, and, more significantly, the rate constants for loop opening show the highest values, compared to other peptides containing two cysteines in different position (Zhang & Snyder, 1989). Structural information about this uncommon motif has not been available since no three-dimensional structure was determined for the proteins above mentioned. Sequence comparison indicates that AFP may be related to the central part of phospholipase A_2 (Nakaya et al., 1990), although the similarity is not high, and thus it could be considered as the first member of a new family of proteins. A detailed picture of the conformation of AFP is obviously needed in order to study the suggested connection with phospholipase A_2. Very recently, the gen Y product from the filamentous fungus *Penicillium chrysogenum* has been shown to exhibit a 42% degree of sequence similarity with AFP, including six out of the eight cysteine residues of the molecule, and its antifungal activity was also reported (Vollebregt et al., 1994), therefore supporting the hypothesis of a new family of antifungal proteins.

Although the mechanism of action of AFP is still unknown, its antifungal activity immediately suggests several technological uses for this protein: food preservation, development of antifungal drugs, or the design of plants resistant against phytopathogenic fungi. Therefore, knowledge of the three-dimensional structure of AFP is again essential in order to gain insight on the function of this protein and its potential manipulation. These considerations, together with the contribution that a new protein structure can make to our knowledge about conformation, stability, and folding of proteins, provide the major impetus for the present study.

In this report we present the solution conformation of AFP, as determined from proton NMR spectroscopy data. Strong evidence in favor of the presence of a minor, isomeric form of AFP is also presented, and the nature of this form is discussed.

MATERIALS AND METHODS

Sample Preparation. AFP was produced at 30 °C as described (Olson & Goerner, 1965; Olson et al., 1965) in 1 L Erlenmeyer flasks each containing 250 mL of culture medium. After 70−80 h of fermentation the protein was purified from the extracellular medium of *A. giganteus*. The method was based on that described (Olson & Goerner, 1965; Gasset et al., 1991) for α-sarcine because both proteins are simultaneously produced and overlap during the purification procedure. To obtain homogeneous AFP one additional step was introduced, consisting of filtration on a Sephadex G-25 (Pharmacia) column (1.5 × 34 cm) equilibrated in 0.1 M acetic acid.

Biochemical Characterization. Protein hydrolyses were carried out at 108 °C in evacuated and sealed tubes for 24 h, with 5.7 N HCl, containing 0.1% (w/v) phenol. The amino acid analyses were performed on a Beckman 6300 automatic analyzer with an IBM-AT based System Gold enhancement. Automatic Edman degradation of the isolated protein was performed on an Applied Biosystems 477A sequencer. The resulting phenylthiohydantoin−amino acid derivatives were identified by using a 120A on-line analyzer and the standard Applied Biosystems software (Hewick et al., 1981). Reverse-phase chromatography was performed on a Beckman HPLC system. The protein was eluted in 0.1% trifluoroacetic acid with a linear gradient of 0−25% B over 30 min, where B is 0.1% trifluoroacetic acid in acetonitrile. The column employed was a Beckman Ultrasphere (5 μm, 4.6 × 150 mm). The peptides generated by proteolytic digestion were fractionated in the same system with a Pharmacia Super-Pack Pep-S column (5 μm, 4 × 250 mm). In this case, the gradient used was a 0−20% B linear gradient over 90 min, also in 0.1% TFA. Lysyl endopeptidase digestion was made by incubating the protein in 0.1 M Tris-HCl, pH 9.0, containing 8 M urea, at 37 °C for 30 min, and then diluting the sample 2-fold with distilled water. Lysyl endopeptidase was then added to this solution up to a final enzyme to substrate ratio of 1:100 and further incubated for 24 h at 37 °C. The reaction between native AFP and the Ellman's reagent was followed as reported elsewhere (Merola et al., 1989) to determine the number of reactive SH groups. Electrophoresis on polyacrylamide gels, containing 0.1% SDS, were carried out according to Laemmli (1970). Samples (1−5 μg) in denaturing buffer were electrophoresed toward the anode in 20% acrylamide separating gel/4% stacking gel by using a running buffer consisting of 25 mM Tris-HCl and 0.2 M Glycine, pH 8.0, containing 0.1% SDS.

Mass Spectrometry. Picomolar amounts of protein directly taken from the AFP sample used for the NMR experiments were analyzed by mass spectrometry at the European Molecular Biology Laboratory Peptide & Protein service (Heidelberg). Mass spectra were recorded in two different experimental conditions: native (NH_4HCO_3 buffer, pH 6.0) and denaturing conditions (49.5% methanol, 49.5% water, 1% formic acid).

Sedimentation Equilibrium. Sedimentation equilibrium experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge using a Ti60 rotor and double sector 12 mm epon-charcoal centerpieces. Samples of the protein (50 μL) in distilled water pH 3.5, with or without 0.1 M NaCl, were centrifugated at 40 000 rpm at 5 °C. Radial scans at wavelengths between 280 and 300 nm were taken after equilibrium was reached. The relative molecular masses (M_r) were obtained by analyzing the experimental data using the programs XLAEQ and EQASSOC [supplied by Beckman Instruments Inc.; see Minton (1994)]. The partial specific volume of the protein was 0.724 mL/g, calculated from the amino acid composition (Laue et al., 1992).

NMR Spectroscopy. NMR samples were prepared by dissolving the lyophilized protein in either H_2O/D_2O 9:1 or D_2O at a concentration of ca. 1 mM. Data were collected at three different experimental conditions: at 25 °C and pH 3.5, at 35 °C and pH 3.5, and at 35 °C and pH 5.0 (direct pH-meter reading without correction for isotope effects).

Most experiments were done in H_2O and a single set of spectra in D_2O enabled qualitative determination of the degree of protection against exchange of peptide protons. Sodium 3-trimethylsilyl[2,2,3,3-$^2H_4$]propionate (TSP) was used as an internal reference.

NMR spectra were recorded on a Bruker AMX-600 spectrometer operating at 600 MHz for the proton. All the two-dimensional spectra were acquired in the phase-sensitive mode using the time-proportional phase incrementation technique (Redfield & Kunz, 1975; Marion & Wüthrich, 1983). Water suppression was achieved either by selective presaturation of the water signal or by including a WATERGATE module (Piotto et al., 1992) in the original pulse sequences prior to adquisition. Conventional pulse sequences and phase-cycling procedures were used for COSY (Aue et al., 1976), NOESY (Kumar et al., 1980), and ROESY (Bothner-By et al., 1984) spectra. Mixing times of 50 and 150 ms were used in NOESY spectra. A unique ROESY spectrum with a mixing time of 20 ms was recorded at 35 °C and pH 5.0. TOCSY spectra were performed using the clean CITY (Briand & Ernst, 1991) spin-lock sequence and 60 ms mixing time.

The size of the acquisition data matrix was 2048 × 512 words in f_2 and f_1, respectively. Prior to Fourier transformation the 2D data matrix was zero-filled up to 4096 × 2048 words. Resolution enhancement methods used included phase-shifted sine-bell or square-sine-bell window functions multiplication in both dimensions. The corresponding shift was optimized for every spectrum. Qualitative measurements of amide hydrogen exchange were made by freshly dissolving the sample in D_2O and observing the intensity of peptide NH resonances after a 48 h period. Slowly exchanging protons were those still present after that period.

Distance Constraints. After the complete assignment of the backbone protons, as well as most side chain protons, was accomplished by using the standard sequence-specific procedure (Wüthrich, 1986), the NOESY spectrum recorded with a mixing time of 150 ms was analyzed and the assigned dipolar correlations were translated into upper limit distance constraints to be used in the structure calculation. Well resolved cross peaks were assigned and their intensities quantitatively translated into distances using the geminal protons cross peaks for calibration. Overlapping cross peaks were qualitatively translated into upper limits after visual inspection according to the following criterion: high, medium, and low intensity correspond with upper limits of 3.5, 4.5, and 5.0 Å, respectively. This conservative criterion was chosen to avoid being misled by too tight restraints arising from cross peaks with significant contributions from signals produced by the second form (see below). No stereospecific assignment of diastereotopic pairs was performed and the usual corrections for pseudoatoms (Wüthrich, 1986) were added in all cases.

Structure Calculation. Distance geometry calculations were performed with the program DIANA (Güntert et al., 1991). The REDAC strategy (Güntert & Wüthrich, 1991) was used in these calculations. Initially, only unambiguous NOEs involving backbone protons were used as distance constraints in the calculations. DIANA conformers complying with these restraints were subsequently calculated, and they were used to evaluate the proposed assignment, based on chemical shifts, for the nonassigned cross peaks. In favorable cases initial conformers ruled out some possible assignments for nonassigned cross peaks and allowed their unambiguous assignment. In this fashion, the number of assigned NOEs was increased until no further assignment was possible. Finally, DIANA conformers were subjected to restrained energy minimization (REM) with the GROMOS package (van Gunsteren & Berendsen, 1987). All calculations were performed on a Control Data Cyber 910 computer.

RESULTS

Assignment of the ^1H NMR Spectra. The first stage in the process of assigning the ^1H NMR spectra of a protein involves the identification of systems of scalarly coupled resonances belonging to a particular residue (Wüthrich, 1986), what was performed on the basis of the combined analysis of COSY and TOCSY spectra. AFP contains 51 residues, including five Gly and one Pro. Therefore, a maximum of 54 COSY cross peaks corresponding to 49 residues was expected in the fingerprint region involving the backbone NH$-C_\alpha$H resonances. The COSY spectrum in H_2O at 35 °C and pH 3.5 (fingerprint region) is shown in Figure 1, where at least 70 backbone cross peaks can be observed. A number of additional backbone signals could also be observed in the TOCSY spectrum (not shown). It was readily observed in both the COSY and the TOCSY spectra that intense cross peaks often have a weaker partner cross peak of similar chemical shifts. Moreover, the two members of each pair of backbone cross peaks have a very similar pattern of side chain resonances in the aliphatic region of the TOCSY spectrum, indicating that a double set of signals is present for most residues. The problem of assigning two related species simultaneously present at different concentrations was solved by the following strategy: the weaker signal of each unambiguous pair was initially ignored and only the most intense signals of each pair, as well as other unpaired but intense signals, were considered for assigning the major component; once the assignment of the major component was achieved, the minor component could be easily assigned by transferring the information previously obtained for the major component.

Several residues showed some peculiarities that complicated their assignment: Ser 44 presents a highly shielded C_αH proton (δ = 3.85 ppm) and a highly unshielded β proton (δ = 4.33 ppm), thus resonating at a lower field than the C_αH proton; Cys 51 resonances were only observed at pH 5.0 and did show themselves as very broad cross peaks; β resonances in Tyr 45 happened to be degenerate and only were identified at the sequential assignment stage. On the other hand, the C_αH protons of Tyr 8, Asp 11, Cys 33, and one of Gly 5, all of them resonating in the 4.65−4.75 ppm range, could only be unambiguously identified in the TOCSY spectrum, where excellent water suppression was achieved.

Most signals arising from the minor component were also classified at this stage. The weaker close neighbors to the signals from the main component were confirmed to be the corresponding signals of the minor component during the sequential assignment. Initially, a few backbone cross peaks from the minor form, which showed significant shift differences with respect to the corresponding major signals, appeared unpaired and were erroneously identified as belonging to the major form. Only in the sequential assignment procedure these unconsistencies were detected, leading to a right reassignment of those signals to specific residues of the minor component.

The second stage involves the assignment of the spin systems identified in the previous stage to specific residues

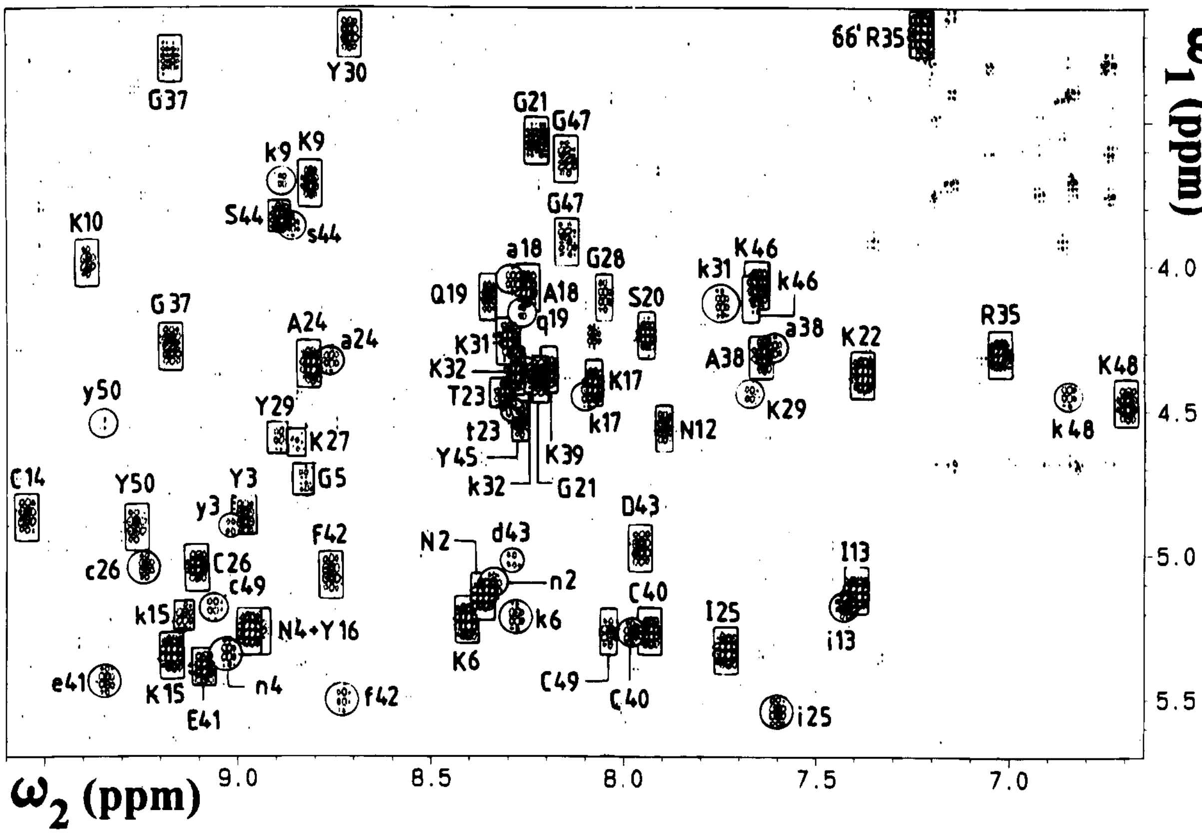

FIGURE 1: Fingerprint region of the 600 MHz COSY spectrum of the antifungal protein from *A. giganteus* (~1 mM, 9:1 H_2O/D_2O, pH 3.5, 35 °C). The boxes indicate cross peaks of the major form, while those from the minor form have been encircled.

in the protein sequence. The sequence-specific assignment is achieved by correlating one amino acid spin system with those of its neighbors, by means of sequential NH−NH, C_αH−NH, and/or C_βH−NH NOE connectivities observed in the NOESY spectrum (Wüthrich et al., 1984). The sequential assignment procedure applied to AFP is illustrated by Figure 2. A summary of the sequential connectivities identified and the number of such connectivities per residue have been represented in Figure 3. The two sequential backbone NOE connectivities were observed for many residue pairs, thus increasing the reliability of the assignments. Sequential C_αH−C_δH and C_αH−C_δH′ NOE correlations were observed between Cys 33 and Pro 34, indicating that the corresponding peptide bond is in a *trans* configuration. In addition, many assignments were corroborated by sequential side chain−NH NOE connectivities (Figure 3). Most assignments of the minor form, preliminary made by just transferring the assignment of the major form signals to their corresponding weak counterparts, were confirmed by the presence of the corresponding sequential and long-range NOEs in the minor form. Several examples of such sequential connectivities are indicated in Figure 2. The chemical shifts of all proton resonances of the major form of AFP and the chemical shifts of most backbone proton resonances, as well as some backbone and side chain resonances of the minor form of AFP, at pH 5.0 and 35 °C, are given in the supplementary material.

Secondary Structure Analysis. Secondary structure was readily delineated from a qualitative interpretation of C_αH conformational shifts (Jiménez et al., 1987; Wishart et al., 1991), together with an analysis of observed NOE patterns (Wüthrich et al., 1984) and NH-exchange data (Englander & Kallenbach, 1983). Figure 4a shows the Hα conformational shifts, as the differences, $\Delta\delta$, between observed chemical shifts, δ_{obs}, and those corresponding to the random coil state, δ_{RC} (Bundi & Wüthrich, 1979): $\Delta\delta = \delta_{obs} - \delta_{RC}$. The general trend for both forms of AFP is identical: five regions well distributed along the protein sequence clearly show positive conformational shifts, indicative of extended conformations. The observation of strong C_αH−NH sequential connectivities (Figure 3a) is another piece of information in favor of these segments being in an extended conformation. The relative orientation of the β-strands was determined via long-range interstrand C_αH−C_αH(i,j), C_αH−NH$(i,j+1)$, and NH−NH$(i+1,j-1)$ NOE connectivities, which provide additional support for the location and extension of the β strands, as can be seen in Figure 5a. As expected, most of the backbone amide protons within these strands exchange very slowly in D_2O while those lying outside individual β strands exchange rapidly. The 49 backbone amide protons can be classified according to their exchange properties as 25 slow (remaining after 48 h in D_2O), 4 medium (disappearing after 12 h in D_2O), and 20 fast exchanging (no observable in D_2O). From a joint consideration of Figures 4a and 5a it can be concluded that the five antiparallel β strands run between the following residues: Thr 2 to Lys 9 (β1), Asn 12 to Tyr 16 (β2), Ala 24 to Lys 27 (β3), Ala 38 to Asp 43 (β4), and Lys 48 to Cys 51 (β5). The β strands are connected in a $-1-1+3+1$ topology, therefore defining a Greek key motif followed by the hairpin-connected fifth strand, as represented in the schematic diagram of Figure 5b. The remainder of the protein consists of nonregular structure connecting the β strands (loops 1, 2, and 3, defined in Figure 5b). Negative conformational shifts and short distance contacts, indicative of helical or turn-like conformation, are also observed within these three loops (Figures 3a and 4a).

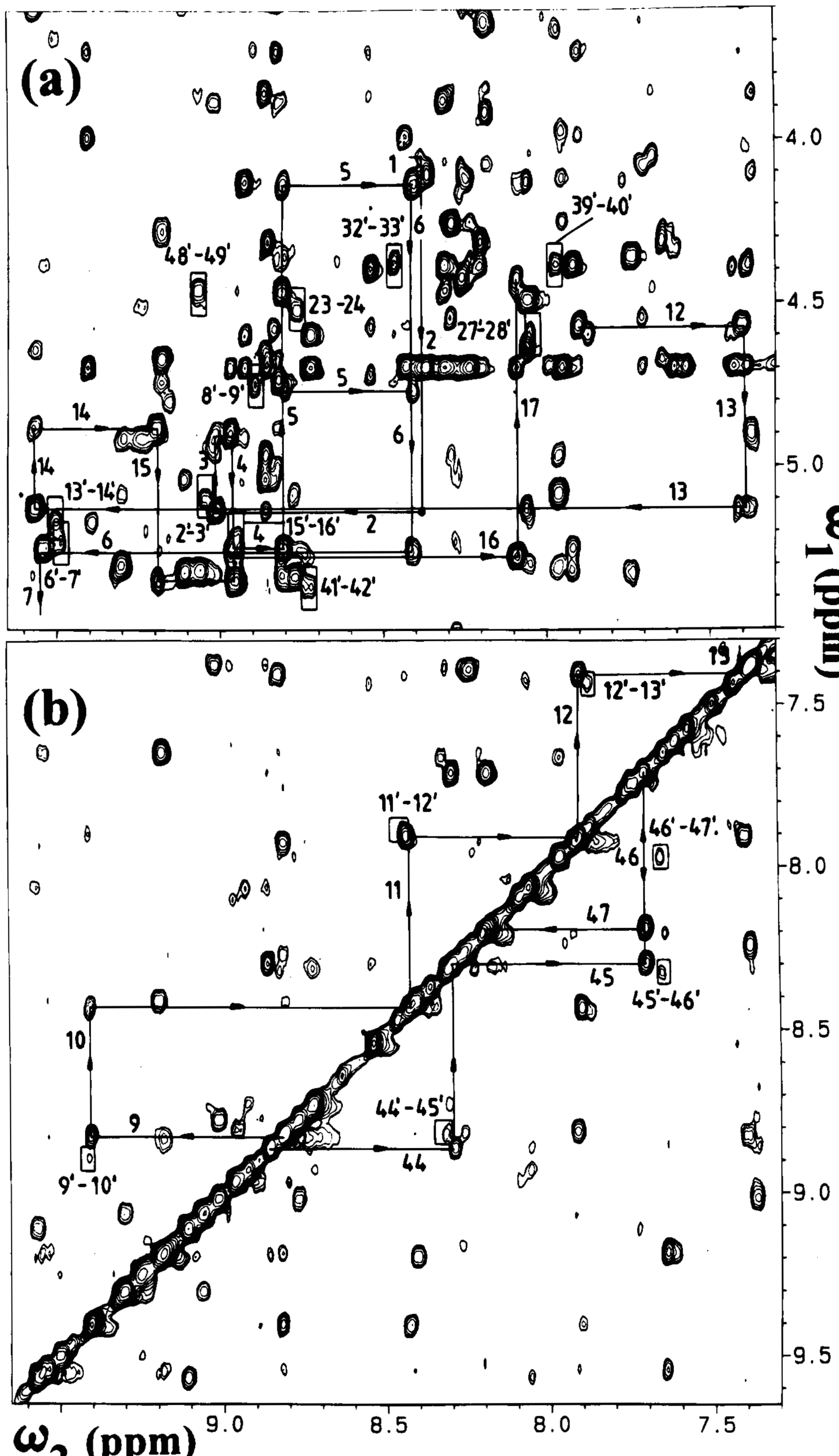

FIGURE 2: Portion of the NOESY spectrum of the antifungal protein from *A. giganteus* (~1 mM, H_2O/D_2O 9:1, pH 3.5, 35 °C, mixing time 150 ms). The assignment pathway throughout the sequential connectivities $d_{\alpha N}(i,i+1)$ (segments 1−7 and 12−17) and $d_{NN}(i,i+1)$ (segments 9−13 and 44−47) for the major form are shown in panels a and b, respectively. Several $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ connectivities from the minor form are indicated in boxes.

Tertiary Structure Determination. The assignments discussed above allowed the calculation of the solution structure of the major form of AFP. Using the procedure described in Materials and Methods, a total of 834 interproton upper distance constraints (195 intraresidue, 208 sequential, 97 short distance, and 334 long distance) conformationally significant were obtained. They are homogeneously distributed along the sequence, apart from the regions of β sheet,

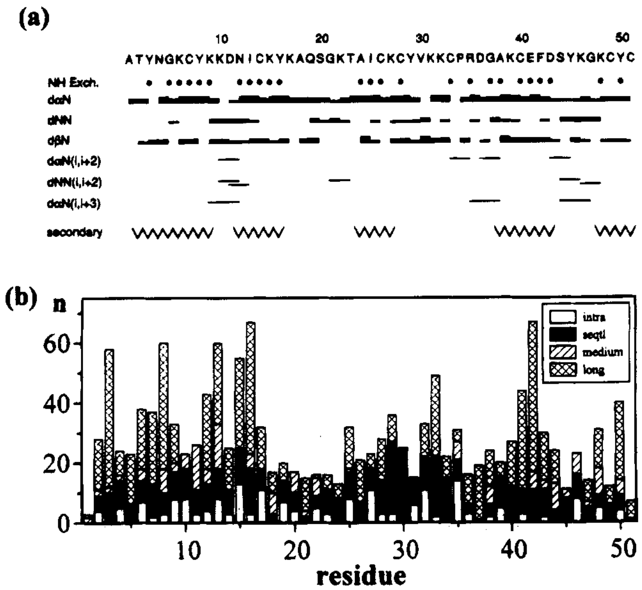

FIGURE 3: Statistical analysis of the conformationally significant NOEs and backbone amide protection. (a) Sequence of AFP together with a summary of observed short range interresidue NOEs involving NH, $C_\alpha H$, and $C_\beta H$ protons. Slowly exchanging NH protons with solvent deuterons are indicated by filled circles. The sequential NH−NH (d_{NN}) and $C_\alpha H$−NH ($d_{\alpha N}$) NOEs are classified into strong, medium, and weak by the thickness of the line. The β strands are identified at the bottom. (b) Plot of the number, *n*, and types of NOE constraints used in the calculation of the structure of AFP versus the amino acid sequence. The code used to define the different ranges of NOE constraints is the following: white, intraresidue; black, sequential; hatched, medium range; cross-hatched, long range.

where the number of long-range restraints increases (Figure 3b). Initially, the pattern of disulfide bridges reported (Nakaya et al., 1990) was assumed, and the corresponding restraints (lower/upper limits for $C\beta_i$−S_j and S_i−S_j distances of 3.0/3.1 and 2.0/2.1 Å, respectively) were applied for the DIANA calculations. Several sequential NOEs involving residues 27 and 28 (27 $C_\alpha H$−28 $C_\beta H'$, 27 $C_\beta H$−28 NH, 27 $C_\beta H'$−28 NH) were observed and unambiguously assigned. The corresponding interproton distances could not be accommodated by the first set of structures, generated with a short (and unambiguous) set of NOEs, and where the disulfide bridge 26−28 was imposed. Conversely, when the disulfide bridge restrain was relaxed, the 27−28 distance constraints were fully satisfied. Thus, these three NOEs are not compatible with the 26−28 disulfide bridge. Moreover, calculations on the oxidized Cys-Lys-Cys model peptide showed that the three constraints were consistently violated. Therefore, the above mentioned NOEs would not be observable, had the disulfide 26−28 been present. These cross peaks are not artifacts arising from spin diffusion because they are clearly observed in the 50 ms NOESY spectrum. On the other hand, the two dipolar connectivities involving the NH proton of Cys 28 are also observed in the ROESY spectra at 20 ms. Therefore, we can safely conclude that the disulfide bridge 26−28 previously proposed is not present in the major form of AFP, as seen in the present study. However, it could correspond to the minor form. This fact, together with chemical and spectroscopic evidence (see below) revealing the possible heterogeneity of disulfide bridges within the protein, led us to consider the proposed disulfide connectivities (Nakaya et al., 1990) as uncertain. Consequently, we decided to build up the structures imposing no disulfide-derived distance constraints in the DIANA calculations. A family of 40 "reduced" structures complying with all the experimental restraints were calculated by using the program DIANA and the REDAC strategy. All possible S−S pairs were excluded from steric overlap check in this calculation. This family is referred to as the R family of conformers. The statistical analysis of the R family of conformers is presented in Table 1.

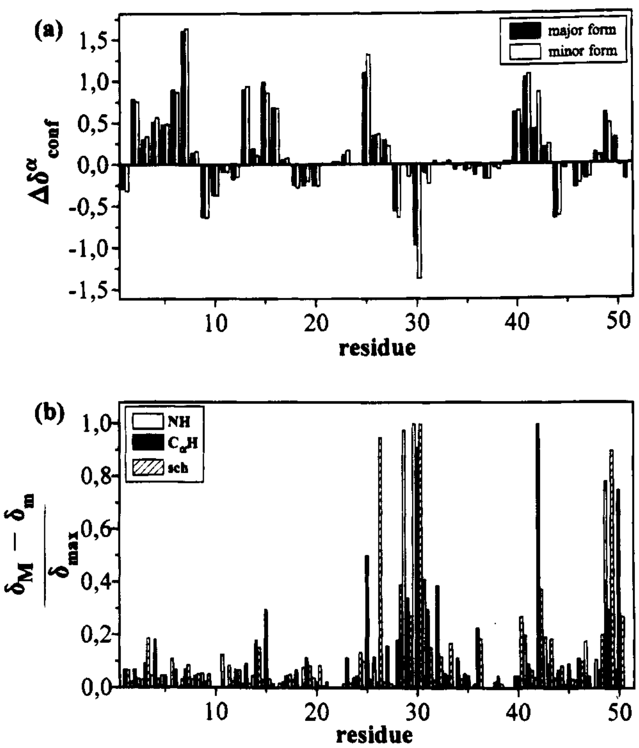

FIGURE 4: Analysis of proton chemical shift (δ) data of AFP. (a) α Proton conformational shifts for the major form (black) and the minor form (white) of AFP. This conformational shifts were evaluated as the differences between the observed δ (see supplementary material) and those accepted as random coil values for each amino acid residue (Bundi & Wüthrich, 1979). For Gly residues the average of the measured values was used as the observed δ. (b) Normalized chemical shift differences of backbone amide protons (white), α protons (black), and side chain aliphatic protons (hatched) between the major and the minor form of AFP. The vertical scales are different for each proton type, since the δ differences have been normalized to the corresponding highest value, which were 1.29, −0.44, and 0.59 ppm for NH, $C_\alpha H$, and side chain protons, respectively. For the side chain shifts the differences between δ values of corresponding protons in the third column of Table 1 in the supplementary material were calculated, and the resulting highest value was selected. The same criterion was used for $C_\alpha H$ shifts of Gly residues.

The most probable disulfide patterns were estimated according to the method proposed in Klaus et al. (1993). Briefly, pairing weights are assigned to all putative cysteine pairs in the protein according to a Gaussian-type distribution function from the $C\beta_i$−$C\beta_j$ interatomic distances in the preliminary structures; all conceivable disulfide patterns are then formed and weights assigned to all patterns from the weights of the participating cysteine pairs. The method did not give a unique answer, although there were three configurations clearly superior (Figure 6). These three configurations, from now on referred to as A, B, and C, present relative weights of 1:0.44:0.35 when the dispersion parameter was set to 0.5 Å (Klaus et al., 1993), whereas the next most likely pattern had a relative weight of only 0.20. The corresponding value for the pattern reported by Nakaya and co-workers (X in Figure 6) is only 7 × 10^{-12}. On that basis, only the disulfide bridge patterns A, B, and C were

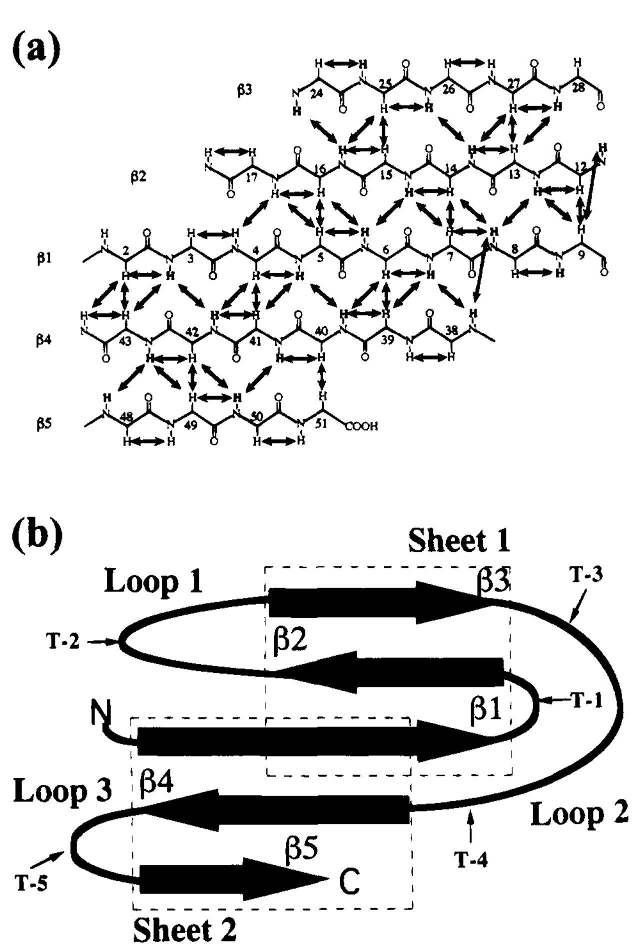

FIGURE 5: Definition and nomenclature of the secondary structure elements in AFP. (a) Schematic diagram showing the NOE connectivities between backbone protons which allow definition of the secondary structure of AFP. An arrow connecting two protons is drawn when a NOE between them is observed. Slow exchanging protons are represented in bold. (b) Representation of the arrangement of the five β sheets in a $-1,-1,+3,+1$ topology. The nomenclature proposed for the β strands is written at the head of the arrows used to represent them. The five turns are identified as "T", and the approximate position of the corresponding central residue is pointed by the thin arrows. The three loops and the two β sheets are also named.

considered to be compatible with the set of NOEs found for the major form of AFP. The solution structure was then modeled by independently taking into account these three possible configurations in an effort to obtain sufficiently different results which could discriminate between them and give a definite solution to the precise disulfide connectivity present in the major form of AFP. A family of 20 conformers for each one of the three most likely configurations were then calculated by DIANA imposing the corresponding disulfide-derived constraints. Finally, all 60 structures were subjected to restrained energy minimization using the GROMOS package (van Gunsteren & Berendsen, 1987), which includes a pseudoenergy term for distance constraints. Each family of 20 energy-minimized conformers, which will be called families A, B, and C in reference to the disulfide bridges present, were used to represent the solution conformation of alternative disulfide configurations of the major form of AFP. The statistics corresponding to the three converged families of structures are given in Table 2. It can be seen that the three disulfide bridge patterns are equally compatible with the large set of NOE constraints, all of them essentially conformed by the structures of the three families. In the same way, the total energy and its different contributions are very similar for families A, B, and C. With the data at hand, it is thus not possible to determine the actual pattern of disulfide bridges present in the major form of AFP.

Analysis of Structures. The four different families of conformers calculated for the "reduced" (R) and the three disulfide-cross-linked forms (A, B, C) of AFP have several common features, regardless of the pairing of cysteines considered. This is not surprising, since all structures comply with the large set of experimental interproton distance constraints obtained (Figure 3b). The overall folding of the four sets of structures can be compared by considering the mean structure of each family (see Table 3). The backbone conformation is almost identical (maximum pairwise RMSD of 0.75). A, B, and C families are very similar to the R conformers, with a maximum RMSD of 0.66 Å between the mean conformers of the B and the R families, while the differences in backbone conformation are only slightly larger when the three disulfide-cross-linked families are compared. In general, the four families are well defined, presenting average pairwise RMSD for backbone heavy atoms of around 1.00 Å (see Tables 1 and 2). Other common point between the four model families is the low definition in the region Ala 18–Ile 23, where large values of backbone local RMSD were obtained. In fact, two subfamilies of structures exclusively differing in the local conformation of the segment 18–23 and its relative position with respect to the other better defined regions could be defined in the R family of conformers (Table 1). When considered independently, the two subfamilies (R_1 and R_2 in Table 1) presented a better definition than when the whole R family was superimposed. Very probably, this is the result of the absence of long distance NOEs of enough quality in this region (Figure 3b) and the intrinsic flexibility of Gly 21, which make possible the existence of two alternative conformations equally acceptable.

On the basis of the high similarity observed between the four families of model structures calculated, and in view of our uncertainty about the actual pattern of disulfide bridges present in the major form of AFP, the R family of conformers will be used in the following for the description of the solution structure of the major form of AFP. A stereoview of the backbone heavy atoms of the best 20 DIANA structures of the R family is given in Figure 7a, and a schematic representation of one of them is provided in Figure 7b. The folding of AFP can be described as resulting from the orthogonal packing of two three-stranded β sheets connected by a long strand which is shared between both sheets. It can be readily visualized by mentally performing the following operation on the topological representation of AFP in Figure 5b: take the two small strands $\beta 4$ and $\beta 5$ as a block and pack them perpendicularly against the back of the other two small strands $\beta 2$ and $\beta 3$; to allow the $\beta 1$ strand to be shared by both sheets, introduce a 90° bend in the center of $\beta 1$ in such a way that the N-terminal half of $\beta 1$ ends up vertically oriented behind the $\beta 2-\beta 3$ loop and its C-terminal half remains in the original position. The resulting fold is essentially the one adopted by AFP, which has been represented in Figure 7. From that perspective, the sheet formed by the N-terminal part of $\beta 1$, $\beta 4$, and $\beta 5$ strands (sheet 2) is packed over the one formed by the C-terminal part of $\beta 1$, $\beta 2$, and $\beta 3$ (sheet 1). All three central strands ($\beta 2$, $\beta 1$, and $\beta 4$) are dramatically twisted so that $\beta 1$ can be shared by sheets 1 and 2. A number of nonrepetitive

Table 1: Structural Statistics of the R Family of Conformers of AFP[a]

	family R	family R	subfamily R_1	subfamily R_2
no. structures	40	40	24	16
superposition range	1−51	1−17 and 24−51	1−51	1−51
backbone RMSD	1.00 ± 0.32 (0.31, 1.63)	0.74 ± 0.14 (0.32, 1.19)	0.69 ± 0.13 (0.31, 1.03)	0.79 ± 0.16 (0.34, 1.16)
heavy atom RMSD	1.82 ± 0.34 (1.01, 2.73)	1.58 ± 0.19 (1.01, 2.35)	1.50 ± 0.17 (1.01, 1.94)	1.65 ± 0.18 (1.09, 2.34)
target function	1.05 ± 0.27 (0.48, 1.66)	1.05 ± 0.27 (0.48, 1.66)	1.17 ± 0.23 (0.70, 1.66)	0.87 ± 0.27 (0.48, 1.65)
NOE constraint violation				
number (>0.2 Å)	3 ± 1 (0, 6)	3 ± 1 (0, 6)	3 ± 1 (0, 6)	2 ± 1 (1, 4)
sum	5.10 ± 0.81 (3.6, 6.8)	5.10 ± 0.81 (3.6, 6.8)	5.46 ± 0.75 (4.0, 6.8)	4.57 ± 0.62 (3.6, 6.8)
maximum	0.29 ± 0.05 (0.20, 0.39)	0.29 ± 0.05 (0.20, 0.39)	0.30 ± 0.05 (0.20, 0.39)	0.28 ± 0.04 (0.20, 0.34)

[a] Average values and standard deviations of the pairwise root mean square deviations (RMSD) are listed for each group. The weighting factors for the NOE upper distance constraints were 1 and, for the van der Waals lower distance limits, 2 Å^2. RMSD values and NOE contraints violations are expressed in Å and target functions in Å^2. The maximum and minimum values of each magnitude are given in parentheses.

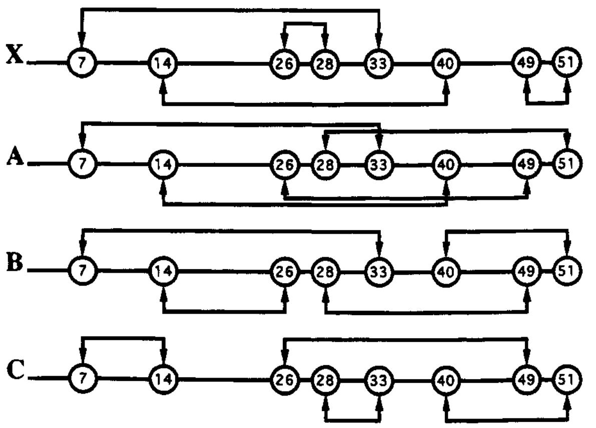

FIGURE 6: Representation of the four different disulfide bridge connectivities considered in this study: X is the previously reported pattern (Nakaya et al., 1990); A, B, and C are the three most likely patterns as determined by the C_β−C_β distance method (Klaus et al., 1993) applied to the preliminary 40 "reduced" conformers calculated in this study.

secondary structure motifs could also be identified (see Figure 5b for nomenclature): a β turn belonging to the I category (Richardson, 1981) links β1 and β2 (turn 1: Lys 9−Lys 10−Asp 11−Asn 12); one β-turn of the II type (Rose et al., 1985) is found in the second half of loop 2 (turn 4: Arg 35−Asp 36−Gly 37−Ala 38); β4 and β5 strands are connected by a short segment formed by residues 44−47 (turn 5: Ser 44−Tyr 45−Lys 46−Gly 47) which do not form the canonical hydrogen bond but can be classified as a turn on the basis of its short $C\alpha_i$−$C\alpha_{i+3}$ distance (lower than 6 Å for all structures); a five-membered turn could also be identified in the first half of loop 2 (turn 3: Cys 28−Tyr 29−Val 30−Lys 31−Lys 32), apparently arising from the disulfide bridges involving Cys 28 and Cys 33 (whatever they might be) pulling in the extremes of that segment to tightly pack in the core of the protein. The drastic change in direction experienced by the backbone from β2 to β3 also suggests the presence of a tight turn (turn 2), although the low resolution obtained in loop 1 precludes any deeper description.

The precision of the structures can be analyzed by considering the overall RMSD between the 40 structures (Figure 8, top) and the range of values of the ϕ,ψ and χ_1 angles adopted by them (Figure 8, center). Most regions of the protein are well defined, typically with backbone RMSD values lower than 1 Å together with short ranges of ϕ and ψ values: strands β1, β2, β3, β4, and β5 and turns 1, 4, and 5. In contrast, the loop joining β2 and β3 (loop 1) and turn 3 are poorly defined. All these conclusions can also be drawn from inspection of Figure 7a. When loop 2 was excluded from the comparison, the backbone RMSD dropped dramatically (Table 1).

The hydrophobic core and the surface topology of the protein can be analyzed by visual inspection of either the superimposed structures shown in Figure 7a or the mean structure of the R family of conformers, taking into account the experimental precision attained in the side chains (local RMSD and dihedral angles dispersion) and the degree of exposure of each residue (relative accessible surface), structural parameters plotted in Figure 8. All cysteines are deeply buried and constitute the major interface (center of the structures in Figure 7) between sheet 2 (Cys 7 in β1, Cys 14 in β2, and Cys 26 and 28 in β3) and sheet 1 (Cys 49 and 51 in β5). This hydrophobic core further extends (bottom right in Figure 7a) with a cluster of three aromatic residues surrounded by alanines, which results from the packing of sheet 1 (Ala 1 and Tyr 3 in β1 and Phe 42 in β4) against loop 1 (Tyr 16, Ala 18, and Ala 24). The low relative accessibilities observed for those residues correlate well with their higher definition, as expected, apart from the ones at the N and C termini. The remaining hydrophobic and aromatic residues are unexpectedly located on the surface of the protein and thus present high solvent accessibility values and, in general, lower definition. This is the case for the hydrophobic patch formed by turn 3 (Val 30 and Tyr 29), β5 (Tyr 50), and turn 5 (Tyr 45) (Figure 7c) and, to a lesser extent, for the shallow pocket defined by Tyr 8 (β1), Ile 13 (β2), Ala 24, and Ile 25 (β3). As it is commonly observed, Gly residues are located at strategic positions where advantage can be taken of their high conformational flexibility and small size: Gly 5 is in the center of β1 where the change in direction within this strand is more pronounced, whereas Gly 21, Gly 37, and Gly 47, all them presenting positive ϕ values, are located in turn 2, turn 4, and turn 5, respectively. The same reasoning applies for Asn 12. Pro 34 introduces a sharp bend in the central part of loop 2 and properly orients the rest of that loop. It should also be noticed that several nonglycine and non-asparagine residues (Ala 18, Thr 23, Ile 25, Val 30, Lys 31, and Cys 49) present positive ϕ values in a small fraction (typically <10%) of the structures, indicating that these uncommon conformations are also compatible with the set of NOE restraints. The polar residues are evenly distributed over the surface of the protein with their side chains pointing outward. More importantly, the surface of the protein is dominated by a high degree of positive charges, arising from the 12 lysines and the single arginine. Specially interesting is the cluster formed by Lys 9, Lys 10 (turn 1), and Lys 32 (turn 3): these side chains are not well defined in terms of χ^1, but their C_β atoms have defined positions as a result of the high definition of their

Table 2: Structural and Energetic Statistics of the A, B and C Family of Conformers of AFP[a]

	family A	family B	family C
no. structures	20	19	20
superposition range	1−51/1−17 and 24−51	1−51/1−17 and 24−51	1−51/1−17 and 24−51
backbone RMSD	1.01 ± 0.36/0.68 ± 0.13 (0.27, 1.55)/(0.28, 0.97)	0.89 ± 0.28/0.65 ± 0.13 (0.35, 1.36)/(0.34, 1.07)	1.01 ± 0.33/0.68 ± 0.15 (0.40, 1.62)/(0.37, 1.05)
heavy atom RMSD	1.82 ± 0.36/1.51 ± 0.16 (1.01, 2.52)/(1.05, 1.92)	1.74 ± 0.29/1.52 ± 0.19 (1.12, 2.36)/(1.09, 2.16)	1.81 ± 0.34/1.53 ± 0.18 (1.06, 2.49)/(1.04, 1.95)
GROMOS energy (kJ × 10^{-2})			
van der Waals	−10.4 ± 1.65 (−7.71, −13.0)	−10.3 ± 1.60 (−5.68, −12.0)	−10.4 ± 1.07 (−7.72, −11.9)
electrostatic	−6.07 ± 2.55 (−2.50, −12.1)	−7.03 ± 2.33 (−1.70, −9.89)	−9.03 ± 3.48 (−1.72, −17.0)
NOE restraints	0.38 ± 0.11 (0.62, 0.20)	0.45 ± 0.11 (0.75, 0.29)	0.47 ± 0.15 (0.75, 0.45)
total	−8.88 ± 3.18 (−3.87, −15.0)	−9.43 ± 2.78 (−4.26, −13.6)	−11.3 ± 3.76 (−3.92, −20.4)

[a] Average values and standard deviations of the pairwise root mean square deviations (RMSD) are listed for each group. The harmonic force constant used for the NOE pseudoenergy term in GROMOS was 6000 kJ mol^{-1} nm^{-2}. RMSD values are expressed in Å and energies in kJ × 10^{-2}. The maximum and minimum values of each magnitude are given in parentheses. The disulfide bridges of the A, B, and C families are given in Figure 6.

Table 3: Pairwise RMSD Comparisons of the Different Families of Structures Calculated for AFP[a]

	RMSD (Å) with				
family	⟨R_1⟩	⟨R_2⟩	⟨A⟩	⟨B⟩	⟨C⟩
⟨R⟩	0.43	0.64	0.28	0.66	0.35
⟨R_1⟩		1.07	0.65	1.04	0.57
⟨R_2⟩			0.50	0.40	0.72
⟨A⟩				0.57	0.38
⟨B⟩					0.75

[a] Mean structures for each family were calculated after superposition of the backbone heavy atoms of all the members of the family. The pairwise RMSD values were calculated for the backbone heavy atoms after complete superposition of the corresponding two mean structures. The disulfide bridges of the A, B, and C families are given in Figure 6.

backbone atoms and they are orientated toward a similar direction (Figure 7c), constituting a putative cationic site. The spatial proximity of this site to the stretch of hydrophobic residues spatially extending from turn 3 (Val 30 and Tyr 29) to turn 5 (Tyr 45), along β5 (Tyr 50), could be of functional relevance as these two groups of residues might constitute a potential phospholipid binding site (Figure 7c).

Once the solution structure of AFP was determined, the observed protection of backbone amide protons against exchange could be structurally explained either in terms of their low solvent accessibility (residues 9, 18, 20, 22, 25, and 35) or by their participation in hydrogen bonds in a significant fraction of the conformers (residues 3, 5, 6, 7, 8, 12, 13, 14, 15, 16, 24, 26, 28, 33, 37, 38, 40, 41, 42, 43, 48, and 50). In principle, the introduction of hydrogen bond-derived distance constraints in the structure calculation could solve the problem of determining the actual pattern of disulfides in the major form of AFP by excluding all but one of the possible configurations. However, this approach, although slightly improved the convergence of the structures (average pairwise backbone RMSD from 1.00 to 0.95 Å), did not prove fruitful for the disulfide bridge problem.

The Nature of the Minor Form. Analysis of AFP by other methods does not reveal the presence of two protein forms. Thus, the mass spectra of the protein sample at the two different conditions presented a set of peaks which correspond to a unique mass of 5795.5 (±0.5), satisfactorily matching the expected molecular mass for the native and neutral protein of 5794. Therefore, it is apparent from these results that the two forms observed by NMR are not distinguishable in terms of their mass and very probably present an identical amino acid composition. These results are further supported by the results of a variety of biochemical analyses (see below). Sedimentation equilibrium experiments were also performed in order to detect any potential protein association. At low ionic strength, the M_r of the protein did not change with concentration (from 0.1 to 0.5 mM) but is smaller (3964 ± 165 g/mol) than the expected monomeric M_r (5798 g/mol), which could be related to a nonideal behavior at low salt (Ralston, 1993). However, when the sedimentation equilibrium was performed at 0.1 M NaCl, the protein M_r was very close to the expected monomer value (5748 ± 245 g/mol) and those values did not vary with protein concentration, a good indication of no association behavior (McRorie & Voelker, 1993; Ralston, 1993).

As indicated above, the observed double set of signals for most cross peaks in the NMR spectra has been interpreted in terms of two protein forms in the AFP sample. The population ratio of the minor to the major form was estimated to be smaller than 1:10. The chemical shifts of most backbone protons and some side chain protons was the only information obtained for the minor form of AFP (see supplementary material). The secondary structure, as given by the α proton conformational shifts (Figure 4a), is largely coincident with that of the major form. In addition, the same slow backbone amide exchange in D_2O was observed for a significant number of residues in both forms (6, 9, 12, 13, 15, 24, 26, 35, 38, 40, 41, 42, 43, 48, 50). However, there are some differences in backbone chemical shifts between the major and the minor forms of AFP which may give insights into the nature of their isomerism. As can be seen in Figure 4b, the large variations are well localized in certain positions of the sequence (Ile 25, Tyr 29, Val 30, Lys 31, Phe 42, Cys 49 and Tyr 50), and, more importantly, the different positions are related: Ile 25, Tyr 29, and Tyr 50 have adjacent cysteine residues; Cys 49 is itself one of them; Phe 42 is facing Cys 49 as they belong to the interacting β4 and β5 strands, and there is an interstrand $C_\alpha H-C_\alpha H$ contact between them (Figure 5a); and Val 30 and Lys 31 backbone shifts could be easily perturbed by the aromatic ring of Tyr 29. The highest variations of side chain proton chemical shifts are also localized in these regions of the protein. Therefore, it appears that the largest chemical shift differences locally observed can be spatially correlated, and a picture is obtained in which the differences between the major and the minor forms of AFP are clearly related to the four cysteine residues which were reported to form two $i-i+2$ disulfide bridges (Cys 26−Cys 28 and Cys 49−Cys

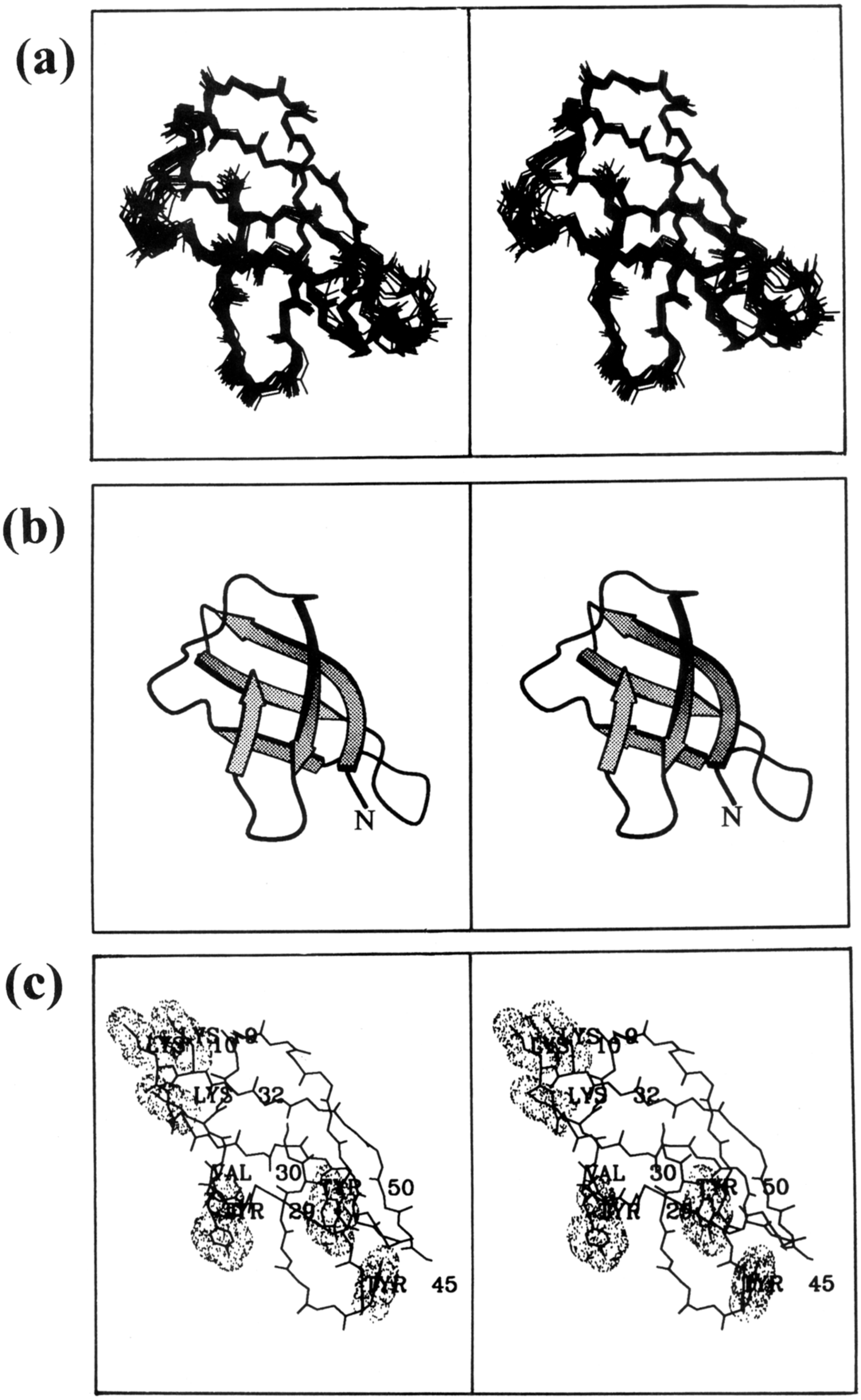

FIGURE 7: (a) Stereoview of the backbone superposition of 20 DIANA conformers of the R family, representing the solution structure of AFP. The 10 conformers that showed the smallest backbone RMSD deviation with respect to the mean structure of each subfamily (Table 1) were selected for the superposition. Only backbone heavy atoms are shown. (b) Ribbon representation of the three dimensional structure of AFP as represented by one of the 40 R conformers. The MOLSCRIPT program (Kraulis, 1991) was used for producing this drawing. The orientation is the same as that in panel a. (c) Stereoview of one of the structures of the R family of conformers highlighting the side chains of the residues which could constitute a phospholipid binding site. The orientation of the molecule has been slightly varied (~30° clockwise rotation around the vertical axis) with respect to the orientation in panels a and b to allow a better perspective of the mentioned side chains.

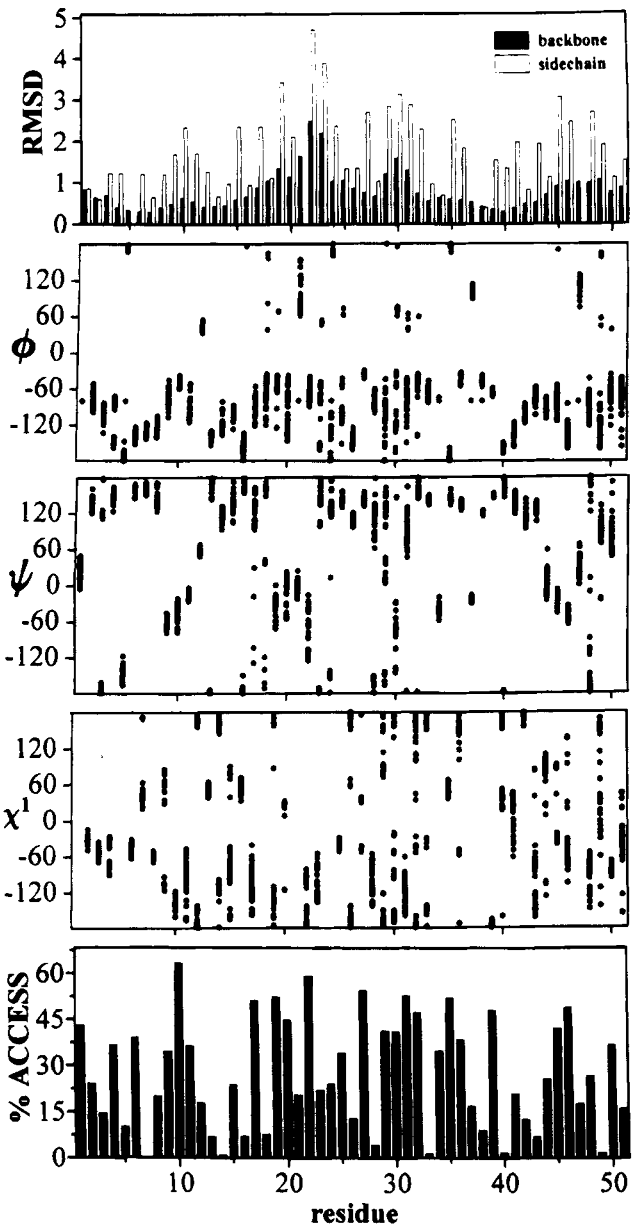

FIGURE 8: Statistical analysis of the 40 DIANA conformers (the R family) generated with the 834 NOE constraints summarized in Figure 4b and with no disulfide bridge-derived restraints. The averaged values for each residue have been calculated over the 40 conformers after optimum superposition of all backbone heavy atoms was performed. The following magnitudes are represented: top, global average pairwise RMSD for backbone heavy atoms (black) and side chain heavy atoms (white), expressed in Å; center, values of phi (ϕ), psi (ψ), and chi 1 (χ_1) angles adopted by the calculated conformers (one dot is plotted for each residue in each structure); bottom, average relative solvent accessible surface, as computed by considering the exposed surface of the isolated residue in the same conformation as that adopted in the protein and taking that value as 100% exposure.

51) in the native AFP (X configuration in Figure 6).

In order to test the hypothesis that the two forms may derive from two species in slow conformational exchange, a ROESY experiment with a mixing time of 20 ms was performed. A significant fraction of the signals previously assigned appeared as real dipolar peaks (negative sign with respect to the positive diagonal), and only two positive cross peaks were observed, those arising from exchange between the two side chain amide protons of Asn 4 and Gln 19 (the corresponding exchange cross peak for Asn 12 was not observed, presumably because the exchange process occurs in a longer time scale due to the hydrogen bond formed between Asn 12 Oδ and Cys 33 NH). On the other hand, the absence of exchange cross peaks in the NOESY and TOCSY spectra previously assigned was also confirmed. Therefore, if chemical exchange is the origin of the two forms of AFP, it would take place with a exchange lifetime longer than ~150 ms (minimum measurable difference between corresponding resonances of major and minor forms). No indication of such exchange was observed in any of the recorded spectra.

In addition, the two species observed from the NMR study cannot be evidenced by a set of biochemical analyses. No free SH groups were detected in native AFP by using the Ellman reagent. The protein behaves homogeneously by SDS–PAGE and HPLC under different conditions. Its amino acid composition and N-terminal sequence (first five residues of the protein), as well as the sequence of some of the peptides obtained by proteolysis, are identical to those previously described (Nakaya et al., 1990). These authors reported that native AFP is highly resistant to proteolytic digestion. In fact, SV-8 protease, trypsin, pepsin, proteinase K, carboxypeptidase Y, and thermolysin fail to degrade native AFP, even at protease to AFP ratios as high as 1:5, 24 h hydrolysis time, and boiling the AFP sample for 30 min prior to the protease treatment. The thermolysin digestion was performed at 65 °C for 4 h also with negative results. The protein was hydrolyzed by endoproteinase Lys C in 4 M urea, as reported (Nakaya et al., 1990), but a larger number of peptides than that reported by these authors was obtained. Analysis of the resulting peptides revealed that those lacking cysteine residues were present in the expected proportion. But cysteine-containing peptides were not quantitatively recovered. Characterization of these peptides showed that some cysteine residues are involved in more than one disulfide bridge, and even in up to three different ones. Therefore, the nature of the disulfide bridges identified suggests that some intramolecular rearrangement took place during the proteolytic protocol. Consequently, identification of the four disulfide bridges present in AFP could not be achieved by chemical means.

DISCUSSION

In this study, the solution structure of the antifungal protein from *A. giganteus* has been determined. The NMR spectra have revealed the presence of two forms of AFP, and the two species have been investigated. Along the research here presented, we faced two different, and in principle, independent problems: (i) determination of the disulfide bridges present in the major form of the protein; and (ii) determination of the nature of the minor form, as compared to the major form.

Independently of the actual pattern of disulfide bridges present in the major form of AFP, the three-dimensional structure can be precisely defined because all four disulfide bridges, whatever the coupling of cysteines could be, are adjacent in the hydrophobic core of the protein and the structure can accommodate some "rearrangements" of these disulfides with little distortion of the polypeptide chain. On the contrary, the detection and unambiguous assignment of a sufficiently large number of NOE cross correlations was not possible for the minor form. Therefore, its structural determination was precluded, and the only experimental data available about the nature of the minor form were the chemical shifts of most protons. In most cases the sequential cross peaks as well as some tertiary contacts observed for the major form were also found for the minor form. These facts, together with the qualitatively conserved frequency and shape of most equivalent signals of the major and the minor forms, suggest the identity of sequence and the similarity of tertiary fold for the two forms. Indeed, when the conformational shifts of both species are compared, the same secondary structure is inferred (Figure 4a), and protection against exchange was also conserved for a large fraction of the minor form amide backbone resonances. Significantly enough, as can be seen in Figure 4b, the few remarkable differences in backbone chemical shifts are located at certain positions in the sequence which are clearly connected with Cys 26, Cys 28, Cys 49, and Cys 51. These four cysteines were reported to form two $i-i+2$ disulfide bridges in the native AFP (Nakaya et al., 1990). *Cis−trans* isomerism of the single X−Pro peptide bond was excluded as the origin of the conformational heterogeneity observed because, at difference with a number of well documented cases (Evans et al., 1989; Akke et al., 1992; Scanlo & Norton, 1994; Ottiger et al., 1994), very small chemical shift variations between the two forms showed up in the region of Pro 34.

On the basis of the tertiary information gathered from the structure calculated with no disulfide constrains imposed (R family of conformers in Table 1), all the possible disulfide patterns were assessed, and only the three most likely are proposed as the actual disulfide bridges present in the major form of AFP. A plausible model emerges which considers the conformations generated (A, B, and C families of conformers in Table 2) assuming the three patterns (A, B, and C in Figure 6) as alternative and mutually exclusive solutions for the structure of the major form of AFP. On the basis of the chemical results previously reported (Nakaya et al., 1990) and in our own three-dimensional structure-based predictions, it is reasonable to suggest that A (Cys 7−Cys 33, Cys 14−Cys 40, Cys 26−Cys 49, and Cys 28−Cys 51) is the actual pairing of disulfide bridges present in the major form of AFP. However, it should be noticed that there is not direct evidence in favor of the disulfide connectivities assumed. The proteolytic digestion by Nakaya and co-workers, and our own, was performed at pH 9.0 due to the resistance of AFP to proteolysis, therefore under the risk of base-catalyzed disulfide rearrangements (Creighton, 1990). This may introduce artifacts leading to incorrect conclusions about the precise pairing of cysteines to form disulfide bridges.

In view of the evidence summarized above, the minor form of AFP presents the same sequence of amino acids as the major form, and the differences between them have their origin in either the precise pairing of cysteines to form disulfide bridges or the conformation (the rotational state around the S−S single bond) of the disulfide bridges present in each form. In fact, the changes must involve the four cysteines (or two disulfide bridges, if they are connected) for which the chemical differences between the two forms are largest. The occurrence of a disulfide bond isomerism in BPTI involving change in the chirality of the disulfide bridge has recently been described (Otting et al., 1993). The isomerism produced a double set of signals in the neighborhood of the two cysteines involved in that bridge. It is difficult to imagine, for AFP, that a flip in the chirality of one of the disulfide bridges has such a specific influence in the other bridge (if the other two cysteines are connected), unless there is a coupling between them and both change chirality simultaneously in going from one form to the other. The absence of chemical exchange for AFP in a time scale where disulfide bond isomerism clearly takes place for BPTI (Otting et al., 1993) puts serious objections to the hypothesis of conformational equilibria as the cause of the heterogeneity observed. Consequently, a more plausible hypothesis explains the correlation between the four cysteines and the absence of conformational exchange as the result of two nonexchangeable species which differ in the pairing of the cysteine residues to form the four disulfide bonds: the major form presents the pattern A, and the minor form is the result of a rearrangement of the bridges 26−49 and 28−51 in the major form to give 26−28 and 49−51 (or 26−51 and 28−49) in the minor form.

Apart from being the first member of a new and still not well described family of proteins with antifungal activity and significant degree of sequence similarity, AFP constitutes, to our knowledge, the first example in which its folding architecture constitutes a complete protein. On the contrary, the same folding motif was found as a domain in the larger protein *Staphylococcal aureus* nuclease (SNS; Arnone et al., 1971), which does not present disulfide bridges, and was described as a Greek key barrel with the less preferred handedness (viewed from the inside the Greek key pattern forms a counterclockwise swirl, as in Figure 5b) (Richardson, 1981). No significant sequence similarity is found for the two proteins after superposition of the matching secondary structure elements. The SNS β domain is relatively independent and accessible to the solvent, only interacting with the rest of the molecule through the face defined by the counterparts of loop 3, the second half of $\beta1$, and $\beta5$. Notwithstanding the above differences, the similarity in the arrangement of the β antiparallel core without a significant sequence homology is a fact to be noted and further analyzed in the context of possible autonomously folding motifs. In addition, the structural similarity between AFP and the β domain of SNS raises the question as to whether the disulfide bridges are responsible for the fold adopted by AFP: is the AFP fold the result of the disulfide bridges packing in the core of the molecule which force the rest of the molecule to adopt its native fold, by chance similar to that of SNS β domain? Or are the disulfide bridges formed after the polypeptide chain is guided to its native fold by specific, "hidden" interactions that are very similar (although their sequences are very different) for AFP and the SNS β domain? It seems that in this small, highly cross-linked protein the tertiary covalent connectivities fundamentally determine its high stability (together with the three aromatic residues constituting the second hydrophobic core), even when a substantial hydrophobic surface remains exposed, unfavorably interacting with the solvent. However, the high density of charged and polar groups on the surface of AFP compensate that unfavorable interaction and additionally results in a high solubility of the protein. The high symmetry of its constituent secondary structure elements and its high intrinsic stability allows us to propose the AFP basic scaffold as an attractive model system on which protein engineering experiments could be performed. Its small size and the simplicity of the structural elements on which its stability seems to be based also erect AFP as a suitable model for protein folding and stability studies.

Information about the biological function of AFP is scarce. This protein was simply described as antifungal agent, but the molecular basis of this action has not been elucidated (Olson & Goerner, 1965; Wnendt et al., 1994). In this regard, recent work on toxin γ from the venom of *Naja nigricollis* suggested that a cluster of hydrophobic side chains on one side of the three-stranded β sheet, together with a spatially close cationic site formed by three lysines, can form a conserved site by which cardiotoxins bind to membrane phospholipids (Gilquin et al., 1993). As described in the previous section, a similar site has been found for AFP: Lys 9, Lys 10 (turn 1), and Lys 32 (turn 3) constitute a plausible cationic site which is spatially adjacent to the hydrophobic stretch defined by Val 30 and Tyr 29 (turn 3), Tyr 50 (β5), and Tyr 45 (turn 5). The cationic site could accommodate the binding of a phosphate ion while the hydrophobic patch constitutes a possible binding site for the hydrophobic moiety of phospholipids. In fact, preliminary experiments performed in our laboratory indicate that AFP induces aggregation of DMPS vesicles, with no effect on liposomes composed of the zwiterionic phospholipid DMPC. Therefore, it is tempting to suggest that AFP could bind phospholipids through the previously described site and that the antifungal activity of AFP might reside on its ability to perturb biological membranes.

ACKNOWLEDGMENT

We thank Mr. A. Gómez, Mrs. C. López, and Mr. L. de la Vega for excellent technical assistance and G. Rivas and J. M. Andreu [CIB(CSIC), Madrid] for the sedimentation equilibrium experiments. We also thank Dr. C. Orengo for running her programs for the search of tertiary structure homology between our protein and the structures in the database.

SUPPLEMENTARY MATERIAL AVAILABLE

One table listing the ^{1}H chemical shifts of the major and the minor form of AFP at pH 5.0 and 35 °C (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Akke, M., Drakenberg, T., & Chazin, W. J. (1992) *Biochemistry 31,* 1011–1020.

Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Jr., Richardson, D. C., Richardson, J. S., & Yonath, A. (1971) *J. Biol. Chem. 246,* 2302–2316.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 5,* 2229–2246.

Bohlmann, H., Clausen, S., Behnke, S., Giese, H., Hiller, C., Reimann-Philipp, U., Schrader, G., Barkholt, V., & Apel, K. (1988) *EMBO J 7,* 1559–1565.

Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., & Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106,* 811–813.

Briand, J., & Ernst, R. R. (1991) *Chem. Phys. Lett. 185,* 276–285.

Broekart, W. F., Parijs, J. van, Leyns, F., Joos, H., & Peumans, W. J. (1989) *Science 245,* 1100–1102.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18,* 285–297.

Creighton, T. E. (1990) in *Protein Structure, a Practical Approach* (Creighton, T. E., Ed.) pp 155–167, Oxford University Press, Oxford, England.

Englander, S. W., & Kallenbach, N. R. (1983) *Q. Rev. Biophys. 16,* 521–655.

Evans, P. A., Kautz, R. A., Fox, R. O., & Dobson, C. M. (1989) *Biochemistry 28,* 362–370.

Gasset, M., Oñaderra, M., Martínez del Pozo, A., Schiavo, G. P., Laynez, J., Usobiaga, P., & Gavilanes, J. G. (1991) *Biochim. Biophys. Acta 1068,* 9–16.

Gilquin, B., Roumestand, C., Zinn-Justin, S., Ménez, A., & Toma, F. (1993) *Biopolymers 33,* 1659–1675.

Güntert, P., & Wüthrich, K. (1991) *J. Biomol. NMR 1,* 447–456.

Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217,* 517–530.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., & Dreyer, W. (1981) *J. Biol. Chem. 256,* 7990–7997.

Jiménez, M. A., Nieto, J. L., Herranz, J., Rico, M., & Santoro, J. (1987) *FEBS Lett. 221,* 320–324.

Klaus, W., Broger, C., Gerber, P., & Senn, H. (1993) *J. Mol. Biol. 232,* 897–906.

Kraulis, P. (1991) *J. Appl. Crystallogr. 24,* 946–950.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95,* 1–6.

Laemmli, U. K. (1970) *Nature 227,* 680–685.

Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., & Horton J. C., Eds.) pp 90–125, Royal Society of Chemistry, Cambridge.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113,* 967–974.

McRorie, D. K., & Voelker, P. J. (1993) in *Self-Associatiing Systems in the Analytical Ultracentrifuge,* Beckman Instruments Inc., Irvine, CA.

Merola, M., Martinez del Pozo, A., Ueno, H., Recsei, P., Didonato, A., Manning, J. M., Tanizawa, K., Masu, Y., Asano, S., Tanaka, H., Soda, K., Ringe, D., & Petsko, G. A. (1989) *Biochemistry 28,* 505–509.

Minton, A. P. (1994) in *Modern Analytical Ultracentrifugation* (Schuster, T. M., & Laue, T. M., Ed.) Birkhauser, Boston.

Mise, T., & Bahl, O. L. (1980) *J. Biol. Chem. 255,* 8516–8522.

Nakaya, N., Omata, K., Okahashi, I., Nakamura, Y., Kolkenbrock, H. J., & Ulbrich, N. (1990) *Eur. J. Biochem. 193,* 31–38.

Olson, B. H., & Goerner, G. L. (1965) *Appl. Microbiol. 13,* 314–321.

Olson, B. H., Jennings, J. C., Roga, V., Junek, A. J., & Schuurmans, D. M. (1965) *Appl. Microbiol. 13,* 322–326.

Ottiger, M., Szyperski, T., Luginbühl, P., Ortenzi, C., Luporini, P., Bradshaw, R. A., & Wüthrich, K. (1994) *Protein Sci. 3,* 1515–1526.

Otting, G., Liepinsh, E., & Wüthrich, K. (1993) *Biochemistry 32,* 3571–3582.

Parijs, J. van, Broekart, W. F., Goldstein, I. J., & Peumans, W. J. (1991) *Planta 183,* 258–264.

Piotto, M., Saudek, V., & Sklenar, V. (1992) *J. Biomol. NMR 2,* 661–665.

Ralston, G. (1993) in *Introduction to Analytical Ultracentrifugation,* Beckman Instruments Inc., Irvine, CA.

Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19,* 250–259.

Richardson, J. S. (1981) *Adv. Protein Chem. 34,* 167–363.

Roberts, W. K., & Seletrennikoff, C. P. (1990) *J. Gen. Microbiol. 136,* 1771–1778.

Rose, G. D., Gierasch, L. M., & Smith, J. A. (1985) *Adv. Protein Chem. 37,* 1–109.

Scanlon, M. J., & Norton, R. S. (1994) *Protein Sci. 3,* 1121–1124.

van Gunsteren, W. F., & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual,* Groningen, The Netherlands.

Vollebregt, A. W. H., Solingen, P. van, & Bovenberg, R. L. A. (1994) *2nd European Conference on Fungal Genetics,* Lunteren, The Netherlands.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222,* 311–333.

Wnendt, S., Ulbrich, N., & Stahl, U. (1994) *Curr. Genet. 25,* 519–523.

Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids,* John Wiley & Sons, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180,* 715–740.

Zhang, R., & Snyder, G. H. (1989) *J. Biol. Chem. 264,* 18472–18479.

BI942605B