Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 8455–8460, July 1998
Biochemistry

The solution structure of the DNA-binding domain of Skn-1

MEI-CHU LO, SHA HA, ISTVÁN PELCZER, SANTONA PAL, AND SUZANNE WALKER†

Department of Chemistry, Princeton University, Princeton, NJ 08544

Communicated by JoAnne Stubbe, Massachusetts Institute of Technology, Cambridge, MA, May 26, 1998 (received for review March 26, 1998)

ABSTRACT **Skn-1 is a maternally expressed transcription factor that specifies the fate of certain blastomeres early in the development of *Caenorhabditis elegans*. It has been reported that the DNA-binding domain is a molten globule and that the structure cannot be defined because there are no long-range nuclear Overhauser effects (NOEs). Working with short Skn domain fragments and using ^{13}C-labeled proteins, we have been able to identify 28 long-range NOEs that establish a tertiary fold for the Skn domain. The internal region of the Skn domain consists of three stable helices and one conformationally labile helix organized into a nascent helix–turn–helix–turn–helix–turn–helix motif. The N and C termini of the Skn domain are unstructured and emerge from the same end of the folded domain. This structure is consistent with biochemical data on binding of the Skn domain to DNA, which shows that the N and C termini bind in the adjacent minor and major grooves from the same face of the DNA helix. The NMR solution structure of the Skn domain should be useful for developing a complete understanding of the DNA recognition event, including any conformational changes that take place upon binding.**

Because of the central importance of DNA in biology, interest in proteins with new DNA-binding motifs remains high. Each new structure reveals additional details about how sequence-specific DNA recognition can be achieved and provides a blueprint for the *de novo* design of DNA binders with predictable binding specificities and behavior (1). In 1992, the amino acid sequence of a transcription factor found in *Caenorhabditis elegans* was reported (2). This transcription factor, Skn-1, is a maternal gene product that acts at a very early stage in the development of *C. elegans* embryos. The amino acid sequence revealed the presence of a basic region at the C terminus of the protein, suggesting that it might belong to the bZIP class of transcription factors. However, there was no leucine zipper or other dimerization domain, and biochemical studies showed that the protein binds to DNA as a monomer (3). Further analysis and additional experiments on the protein showed that the binding motif is contained within an 85-aa C-terminal fragment of the protein and that this fragment includes a "homeodomain arm" at the N terminus in addition to the C-terminal basic region. Hence, it was suggested that Skn-1 represents a distinct class of transcription factors in which recognition elements found in known transcription factor families—bZIP proteins and homeodomains—are combined to form a hybrid protein that recognizes a hybrid DNA sequence. Consistent with this, studies have shown that the preferred Skn-1 recognition site is ATTGTCAT (3), which is a combination of a homeodomain half-site and a basic region half-site. The amino-terminal arm of the Skn domain contacts the minor groove at the ATT half-site, and the basic region contacts the major groove at the TCAT half-site, undergoing a coil-to-helix transition upon binding (3). Thus, the binding behavior of the Skn domain is precisely as expected for a hybrid protein composed of familiar DNA recognition elements.

Nevertheless, there remain some important questions regarding the structure and function of the Skn domain. One of these questions involves the role of the ≈50-aa polypeptide that links the basic region and the N-terminal arm of the Skn domain. This region of the protein has limited homology to structurally characterized DNA-binding motifs. This linking domain could function as a scaffold to orient the N-terminal arm and the basic region of Skn for DNA binding, but it may play other roles in binding as well.

CD and preliminary NMR studies from our lab (4) as well as Blackwell's and Wagner's (5) have shown that the internal region of the Skn domain is largely helical. However, no information about the tertiary fold of the Skn domain has been reported. In fact, Carroll *et al.* (5) reported that no long-range nuclear Overhauser effects (NOEs) could be identified in the Skn domain, making it impossible to establish the orientation of the helices with respect to one another.

We have continued our NMR studies on the solution structure of the Skn domain and have been able to identify 28 long-range NOEs ($|j-i| \geq 5$) between different parts of the Skn domain, which have allowed us to determine the tertiary fold. The NMR solution structure is consistent with the biochemical data on binding (3) and should be useful for developing a complete understanding of the DNA recognition event.

MATERIALS AND METHODS

Sample Preparation. Skn domain fragments Skn-73 and Skn-60 were expressed from pET3a plasmids in *Escherichia coli* strain BL21(DE3)pLysS (Novagen) and purified over carboxymethyl-sephadex (CM-Sephadex) as described previously (4). Uniformly ^{15}N-labeled Skn proteins were obtained by growing the cells in M9 medium containing $^{15}NH_4Cl$ (1 g/liter, Cambridge Isotope Labs) as the sole nitrogen source and supplemented with 0.25 g $MgSO_4 \cdot 7\ H_2O$/0.0014 g $CaCl_2 \cdot 2\ H_2O$/0.005 g $FeCl_3 \cdot 6\ H_2O$/0.001 g thiamine/10 g glucose per liter (6). Uniformly ^{13}C-labeled Skn proteins were made in the same way except using 2 g/liter of D-glucose-$^{13}C_6$ (Isotec) in place of the 10 g/liter unlabeled D-glucose.

NMR Experiments. NMR spectra were acquired on a Varian Unity/INOVA spectrometer equipped with a $^1H/^{13}C/^{15}N$ probehead and three-axis gradients at 600 MHz at 25°C unless otherwise noted. Protein samples were prepared at 0.8–1.0 mM concentration in 250 mM NaCl/10 mM phosphate buffer, pH 7.0. Homonuclear two-dimensional (2D) correlation experiments included gradient-selected homonuclear multiple quantum coherence spectroscopy (7) with several excitation delays between 10 and 200 ms both in H_2O and ^2H_2O; total correlation spectroscopy (TOCSY) (8) in H_2O (30-, 60-, and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; 2D and 3D, two- and three-dimensional; HSQC, heteronuclear single quantum coherence.

†To whom reprint requests should be addressed. e-mail: swalker@pucc.princeton.edu.

100-ms spin-lock times) and ^2H_2O (30- and 60-ms spin-lock times); and NOE spectroscopy (NOESY) spectra (with 60-, 100-, and 200-ms mixing times in H_2O, and with 100- and 200-ms mixing times in ^2H_2O) (9). For samples in H_2O, the water signal was suppressed by using the WATERGATE scheme (10), whereas radiation damping during t_1 was avoided using a weak gradient echo (11), reconstructing 2*–4* missing points by backwards linear prediction (LP) during data processing. Complex acquisition was used in both dimensions by using States–Redfield phase incrementation (12, 13) for sign discrimination in the remote dimension.

Heteronuclear experiments were conducted on ^{15}N-labeled, or $^{15}N/^{13}C$-doubly labeled samples. Gradient-selected, sensitivity-enhanced ^{15}N,^1H-heteronuclear single quantum coherence (HSQC) experiments (14) with a flip-back pulse for the water (15). ^{13}C,^1H-HSQC experiments also were acquired with a 4,000-Hz spectral window for the ^{13}C dimension (t_1^{max} = 200 ms). Resolution in F_1 was enhanced by LP extension of the time domain data in both cases. ^{15}N relaxation (T_1 and T_2) and heteronuclear NOE experiments were conducted by using similar conditions following published methodology (16). ^{15}N T_1 values were measured by using 10 relaxation delays of 10.8, 43.4, 86.7, 119.2, 195.1, 271.0, 379.4, 460.7, 650.4, and 1084.0 ms. ^{15}N T_2 values were determined from spectra recorded with delays T = 14.4, 28.8, 43.2, 57.6, 72.0, 86.4, 100.8, 129.6, and 158.4 ms. The relaxation rates were determined by fitting the measured peak intensities to a single exponential decay curve in NMRVIEW (17). ^1H-^{15}N steady-state NOE values were obtained by recording spectra with and without NOE enhancement by presaturation of amide protons and calculating the ratio of peak intensities. Amide proton exchange rates were measured at 10°C by dissolving the ^{15}N-labeled protein, lyophilized after dialysis against pH 5.0 buffer, in ^2H_2O and monitoring the disappearance of the amide correlation peaks in ^{15}N,^1H-HSQC spectra. The pH* of the NMR sample, measured immediately after the experiments, was 5.8.

The following three-dimensional experiments also were collected and analyzed: a ^{15}N-edited TOCSY (18, 19) with 61-ms spin-lock time, a ^{15}N-edited NOESY (18) with 150-ms NOE mixing time, a simultaneous $^{13}C/^{15}N$-edited-^1H,^1H-NOESY (20) with 150-ms NOE mixing time, and a HNHA experiment (21).

All spectra were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) indirectly (22) by using standard parameters from the BioMagResBank (http://www.bmrb.wisc.edu). Multidimensional data processing and data analysis were done in NMRPIPE (23) and NMRVIEW (17).

Structure Determination. The structure of the Skn domain was calculated with the program DYANA (24), which uses a simulated annealing procedure combined with molecular dynamics in torsion angle space. Distance constraints were generated from 2D ^1H-^1H NOESY spectra recorded in H_2O with mixing times of 60 and 100 ms, and in ^2H_2O with a mixing time of 100 ms. To transfer NOESY cross-peak volumes into upper bounds on proton–proton distances, the program CALIBA (25) was used, which divides NOE peaks into three classes: one for backbone protons, one for side-chain protons, and one for methyl groups. The calibration functions used for these three classes are $V = A/r^6$, $V = B/r^4$, $V = C/r^4$, respectively, where V is the peak volume and r is the corresponding distance. Parameters A, B, C were calculated automatically by the program (26). The simple automatic calibration was used without further tightening the calibration curve. Six hundred and fifteen upper-limiting distance constraints were generated after calibration. Twenty-five additional distance constraints from the 3D $^{13}C/^{15}N$-edited-^1H,^1H-NOESY spectrum were added as looser constraints according to their peak intensities, which include four sequential NOEs, 10 medium-range NOEs ($2 \leq |j\text{-}i| < 5$), and 11 long-range NOEs ($|j\text{-}i| \geq 5$). Then, these 640 upper-limiting distance constraints were modified to account for the absence of stereospecific assignments (25). The $^3J_{HN\alpha}$ coupling constants were measured from the 3D HNHA spectra and converted to dihedral angle constraints by the program HABAS (27). Fifty-one ϕ angle constraints, 51 ψ angle constraints, and 21 χ angle constraints were used in the calculations. Forty structures were calculated, and the 20 conformers with lowest values of pseudopotential energies were selected for analysis.

RESULTS AND DISCUSSION

Initial biochemical studies on Skn-1 by Blackwell *et al.* (3) localized the DNA-binding domain to the C-terminal 85 aa. We subsequently established that the first four and the last eight residues of the 85-residue polypeptide can be removed to produce a 73-aa fragment that displays the same binding characteristics as the longer protein (4). Hence, for most of the NMR experiments reported below we used a 73-aa fragment (Skn-73) that represents the shortest fully functional DNA-binding motif. This 73-aa protein contains one amino acid change from the native sequence, a Cys-66 → Ser mutation, which obviates the presence of reducing agents and facilitates handling. The mutation does not alter the DNA-binding specificity of the protein.

For some NMR experiments, we used an even shorter 60-aa polypeptide, Skn-60, that is lacking an additional 13 residues from the C terminus. Previous CD and NMR studies have shown that the 13 C-terminal residues in Skn-73 are unstructured and that removing them does not change the folding pattern of the internal region (4). Thus, although Skn-60 does not bind to DNA, it contains the same tertiary fold as the longer protein. Because it is significantly shorter and contains fewer unstructured resonances, Skn-60 was used as necessary to simplify or otherwise facilitate structural analysis of the Skn domain.

Resonance Assignments. Sequence-specific ^1H assignments were made for residues 9–57 from homonuclear ^1H-^1H NOESY, double-quantum coherence spectroscopy, and TOCSY spectra by using standard assignment strategies (28). Several of the remaining residues also were assigned from the homonuclear spectra; however, sequence-specific assignments

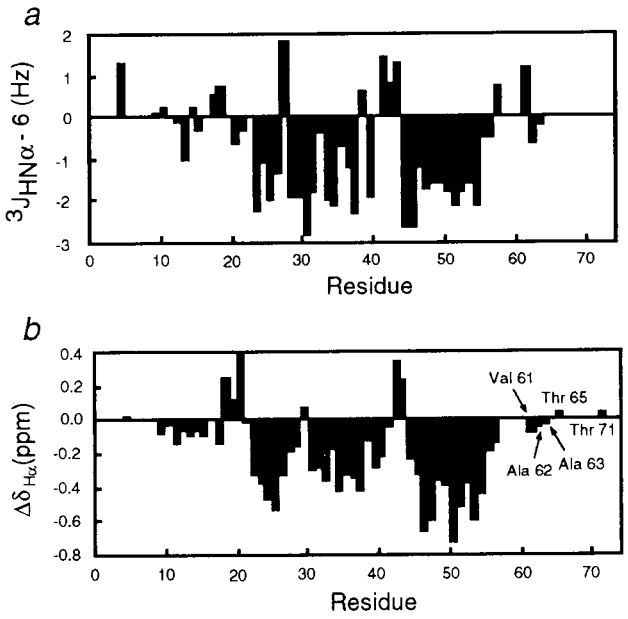

FIG. 1. Data for Skn-73. (*a*) $^3J_{HN\alpha}$ coupling constant deviations from 6 Hz. (*b*) Hα chemical-shift perturbations ($\Delta\delta = \delta^{Skn73} - \delta^{unstructured\ peptides}$) (32). Residues that have not been sequence-specifically assigned are not shown.

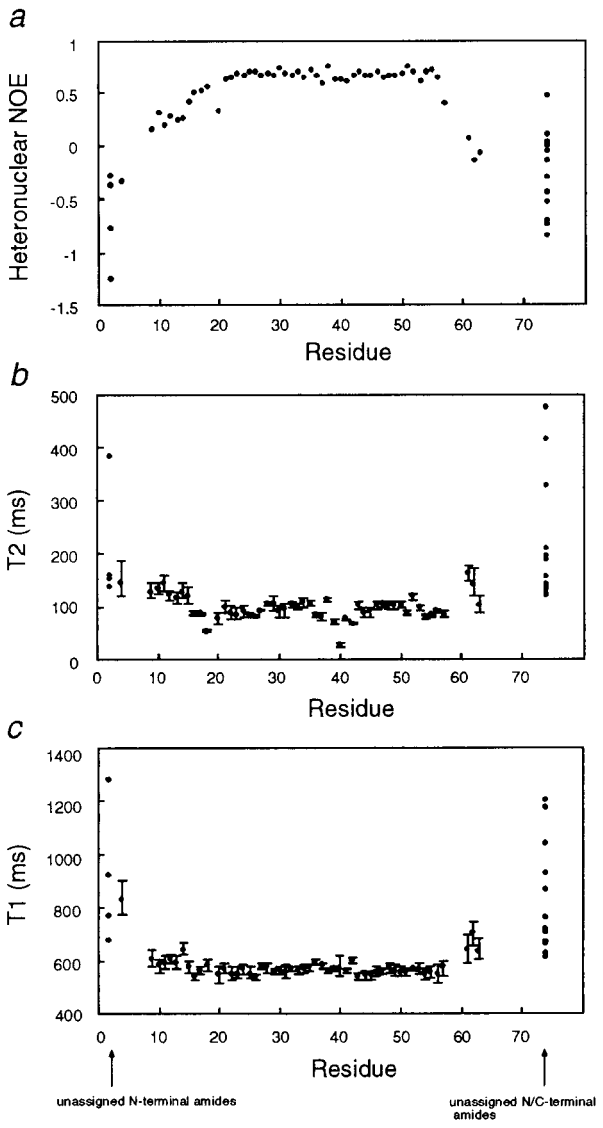

FIG. 2. Relaxation data for Skn-73 as a function of residue number. Amide resonances at the N and C termini that have been identified but not sequence-specifically assigned are shown at the same residue position. (*a*) Heteronuclear NOEs. (*b*) ^{15}N T_2 values. (*c*) ^{15}N T_1 values.

could not be completed for the N and C termini because they are unstructured (see below). Nevertheless, the remaining backbone amide resonances were identified from ^{15}N,^{1}H-HSQC spectra, and several of them could be assigned to the N terminus by comparing the spectra of Skn-60 and Skn-73.

Structural Analysis. The secondary structure of the Skn domain was established from a combination of ^{1}H-^{1}H NOEs, $^{3}J_{HN\alpha}$ coupling constants, Hα chemical shifts, ^{1}H-^{15}N NOEs, and ^{15}N T_1 and T_2 relaxation data (28–31). The data show that the protein is unstructured from residues 1–8 and 57–73, as indicated by near-random coil chemical shift values (32) for the Cα protons of spin systems that can be assigned, including the two threonines, two alanines, and one valine in the basic region (Fig. 1*b*). In addition, the amides assigned to the N and C termini of the protein have negative ^{1}H-^{15}N NOEs (Fig. 2*a*) and significantly longer ^{15}N T_1 and T_2 relaxation times (Fig. 2 *b* and *c*) than amides in the internal region of the protein. The lack of structure leads to extensive resonance overlap and weak sequential NOEs, so it was not possible to complete sequence-specific assignments for the N- and C-terminal residues from homonuclear experiments. However, for generating a solution structure and relating this structure to the structure of the protein in the DNA complex, it is sufficient to know simply that these residues are disordered.

The internal region of the protein is largely helical, as indicated by the NMR data summarized in Fig. 3. The data show that the Skn domain contains three well defined helical segments spanning residues approximately at 22–27, 30–38, and 44–56. Most of the $^{3}J_{HN\alpha}$ coupling constants for residues in these helices are below 6 Hz (Fig. 1*a*); furthermore, extensive NOE networks that characterize helices are present. For example, all three of these peptide segments contain strong sequential amide–amide NOEs as well as αN(i, i + 3) and $\alpha\beta$(i, i + 3) NOEs. For the latter two segments there are also several αN(i, i + 4) NOEs. The peptide segment spanning residues 9–15 also shows some helical characteristics. There are several sequential amide–amide NOEs, αN(i, i + 3) and $\alpha\beta$(i, i + 3) NOEs. However, the $^{3}J_{HN\alpha}$ coupling constants for most of these residues are between 6 and 7 Hz (Fig. 1*a*), values that typically are found in flexible peptides and in loops that are not constrained in α or β conformations or are subject to conformational averaging (30). In addition, the chemical shifts of the amide and Cα protons in this peptide segment do not deviate as much from the random coil values as those in the other helices (Fig. 1*b*). The NMR data suggest that this peptide segment fluctuates between a helix and an unfolded structure. Consistent with the proposed conformational lability, the ^{15}N T_2 relaxation times are longer and the heteronuclear NOEs are weaker for residues 9–15 than for residues involved in the stable helices (Fig. 2). Thus, the secondary structure of the Skn domain consists of three stably folded helices and a more dynamic "nascent" helix separated by relatively long turns, and

FIG. 3. Summary of the resolved sequential and medium-range NOE connectivities for the first 60 residues of the Skn domain. The width of the bar for sequential connectivities reflects the relative NOE intensities. Medium-range connectivities, d_{NN}(i, i + 2), $d_{\alpha N}$(i, i + 2), $d_{\alpha N}$(i, i + 3), $d_{\alpha\beta}$(i, i + 3), $d_{\alpha N}$(i, i + 4), are represented by lines spanning the interacting residues. The approximate boundaries of the four helices are shown. Solid lines indicate helical NOE patterns and coupling constants <6 Hz; dashed lines indicate helical NOE patterns alone.

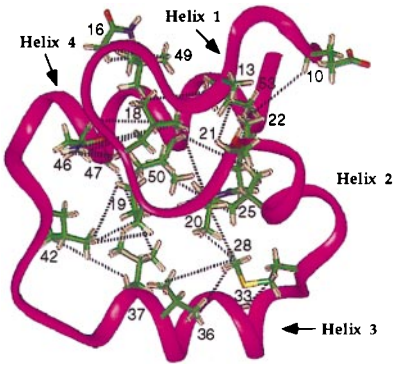

FIG. 4. Ribbon drawing of residues 9–56 of the Skn domain in the lowest pseudopotential energy structure. Dashed lines indicate long-range NOEs.

flanked by completely unstructured N- and C-terminal extensions.

The Tertiary Fold. The tertiary fold of the Skn domain was determined based on long-range NOEs identified between elements of secondary structure, i.e., between the helices as well as between various helices and the intervening turns (Fig. 4). Seventeen long-range NOEs were identified from 2D homonuclear NOESY spectra and verified in 3D $^{13}C/^{15}N$-edited ^{1}H-^{1}H NOESY spectra; 11 additional long-range NOEs were identified from 3D $^{13}C/^{15}N$-edited ^{1}H-^{1}H NOESY spectra. Forty NMR structures for the Skn domain were calculated by using the 28 long-range NOEs along with several hundred shorter-range NOEs as well as the measured coupling constants (Table 1). The 20 conformers with the lowest pseudopotential energies were superimposed as shown in Fig. 5. The average of the pairwise rms deviation (rmsd) of these 20 conformers relative to the mean coordinates is 1.4 Å for the backbone heavy atoms of residues 9–56, and 1.0 Å for the more ordered region of residues 22–56. The average rmsd for all heavy atoms of residues 9–56 is 2.0 Å, and 1.7 Å for the more ordered region of residues 22–56. It should be noted that the remaining 20 conformers, although they have slightly higher pseudopotential energies, have the same overall fold as the lowest-energy conformers.

As shown in Fig. 4, the tertiary fold consists of a nascent helix–turn–helix–turn–helix–turn–helix. There are two hydrophobic clusters that may help stabilize the individual elements of secondary structure. One hydrophobic cluster is formed by

Table 1. NMR-derived constraints and statistics of NMR structures

Summary of constraints	
Total NOE constraints	640
Intraresidue NOEs	360
Sequential NOEs	156
Medium-range NOEs	96
Long-range NOEs	28
Dihedral angle ϕ	51
Dihedral angle φ	51
Dihedral angle χ 1	21
Average rms deviations, Å	
Backbone heavy atoms	
Residue 9–56	1.4
Residue 22–56	1.0
All heavy atoms	
Residue 9–56	2.0
Residue 22–56	1.7
Average sums of constraint violations	
Upper limiting distances	1.1 Å
van der Waals distances	0.2 Å
Dihedral angle constraints	0.5°

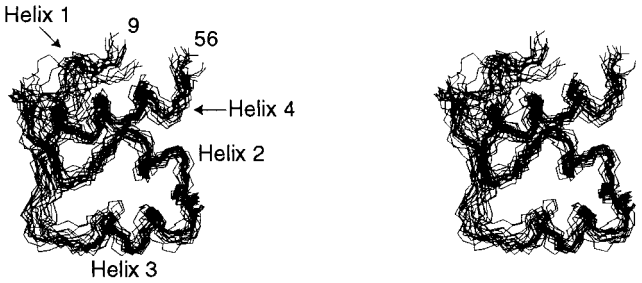

FIG. 5. Stereoviews showing the peptide backbone for residues 9–56 of the Skn domain in the 20 energy-minimized DYANA conformers after superposition of the backbone heavy atoms.

residues Leu-18, Ala-22, Ile-25, Ile-50, and Ile-53. Thus, helix 4, the helix immediately preceding the unstructured basic region, appears to be stabilized by hydrophobic contacts to amino acids in helix 2 as well as to amino acids in the preceding turn. Helix 3, which is strongly amphipathic, containing leucines and valine along one face, shows contacts to the side chains of residues in the adjacent turns. For example, Leu-33 and Val-36 contact Met-28, and Leu-37 contacts Leu-42. These contacts form the second hydrophobic cluster in the Skn domain. As the structure shows, helices 3 and 4 are roughly antiparallel, and helix 2 crosses helix 4 at an oblique angle. Although helix 1 is not stably folded and its orientation cannot be clearly defined, as indicated by the poor overlap of this peptide segment in the superimposed low-energy structures (Fig. 5), there is a reversal in the direction of the peptide backbone between helices 1 and 2 in every calculated structure. Because of the four helix fold, the N and C termini emerge from the same end of the folded domain (Fig. 6). As discussed below, this tertiary fold is consistent with biochemical data on DNA binding, which shows that the N and C termini bind in the adjacent minor and major grooves from the same face of the DNA helix.

Hydrogen Exchange. Amide hydrogens that are involved in hydrogen-bonded structure can exchange with solvent only when some kind of unfolding reaction takes place. Therefore, hydrogen exchange rates can provide useful information about protein structure and dynamics. To evaluate the hydrogen exchange behavior of the Skn domain, a series of ^{15}N,^{1}H-HSQC spectra for Skn-73 was acquired at 10°C after dissolution of the lyophilized protein in $^{2}H_2O$. After 18 min, 17

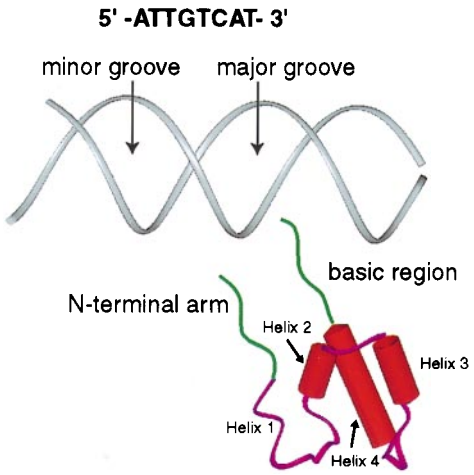

FIG. 6. A schematic showing that the folded Skn internal region orients the N-terminal arm and the basic region for binding in the adjacent minor and major grooves from the same face of the DNA helix. The parts of the structure shown in magenta and red are the same as in the Fig. 4, but are displayed in a different orientation. The unstructured N and C termini are shown in green.

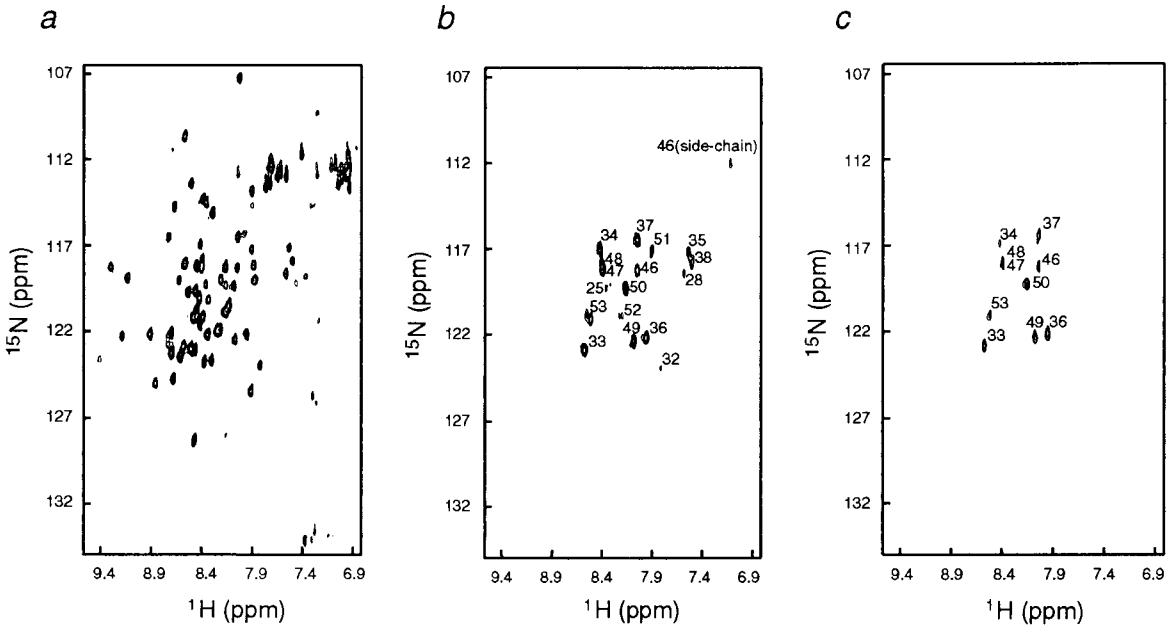

FIG. 7. The time course of amide proton exchange for Skn-73 at 10°C. (*a*) ^{15}N,^{1}H-HSQC spectrum in H_2O. (*b*) ^{15}N,^{1}H-HSQC spectrum after 18 min in $^{2}H_2O$. (*c*) ^{15}N,^{1}H-HSQC spectrum after 66 min in $^{2}H_2O$.

backbone amide protons within helices 2, 3, and 4 were still present (Fig. 7*b*). No amide protons from nascent helix 1 remained, a result consistent with the greater mobility of this peptide segment. After 1 hr, 10 amide protons still remained (Fig. 7*c*), all of them located in helices 3 and 4. Seven of these amide protons were still present after almost 2 hr when the experiments were terminated. All the amides that remain protected from exchange for more than 1 hr belong to residues that show long-range NOEs: Leu-33, Val-36, Leu-37, Gln-46, Arg-47, Leu-49, Ile-50, and Ile-53. Thus, the hydrogen exchange data are consistent with the secondary structure we have determined and also support the tertiary fold.

It is worth noting that the inability of certain alanine mutants of the Skn domain to bind to DNA can also be explained by the solution structure we have presented above. For example, Carroll *et al.* (5) have reported that mutating the hydrophobic residues Ile-50 and Ile-53 to alanine weakens or even abolishes DNA binding. These residues are not believed to contact the DNA directly. Our solution structure shows that these residues are in the core of the protein and show long-range contacts that presumably help stabilize the tertiary fold. Replacing them with alanine likely disrupts the folding pattern of the protein, and this would be expected to have a deleterious effect on DNA binding.

Evaluation of the NMR Structure. Although we were able to identify long-range NOEs that establish the relative orientation of the helices in the Skn domain, the resolution of our solution structure is low. The structure is underdetermined because the number of long-range NOEs observed is relatively small. Other helical proteins of comparable size, e.g., the Antp homeodomain (33), show many more long-range NOEs. We considered two possible explanations for the small number of long-range NOEs. The first explanation relies on the fact that the Skn domain does not have any aromatic residues in the core; both Phe-23 and Tyr-45 are surface-exposed. One consequence of this is that the long-range contacts only involve aliphatic side chains, with a large number of methyl–methyl contacts. Because there are no aromatic residues in the core, the chemical shift dispersion of the methyl groups in the Skn domain is limited. Hence, many NOEs that may be present cannot be unambiguously identified because of resonance overlap.

The second explanation invokes flexibility as a cause for an anomalously small number of NOEs. Weiss and coworkers (34, 35) have carried out extensive investigations of the des-pentapeptide insulin analog (DPI) and have noted that only 25% of the predicted number of resolved long-range NOEs were actually observed in NOESY spectra of the protein. They have proposed that DPI actually can be represented by an ensemble of structures, each having the same secondary structure and a similar tertiary fold, but differing in the details of the side-chain packing. Although a large number of long-range NOEs are expected for each structure, it was proposed that rapid fluctuations in interproton vectors quench many of them. The only NOEs that are observed in the experimental NOESY spectra are those that are common among all of the different structures.

To evaluate these explanations, we simulated NOESY spectra for several of our structures by using the program CORMA (complete relaxation matrix analysis) (36, 37). The results show that a large number of resolved long-range NOEs from helix 1 and the adjacent turn to the rest of the protein in the simulated spectra are missing in the experimental spectra. Helix 1 has been shown to be conformationally labile in solution, so this finding, which supports the explanation that parts of the structure are underdetermined because of flexibility, is not surprising. For residues 22–56, which comprise the better-defined region of the protein, an insignificant number of long-range NOEs that are predicted to be resolved are missing. However, a significant number of predicted NOEs turned out to be between overlapping resonances, and it would not be possible to establish whether they are present without resorting to four-dimensional experiments (38), or using residue-specific labeled protein.

Although we cannot attribute the relatively small number of NOEs observed for the better-defined region of the protein to flexibility, there is some evidence that this protein may be more flexible than is typical for a native protein. For example, the thermal unfolding transition for the Skn domain is not highly cooperative (5), suggesting that the side chains may not be tightly packed (39). Furthermore, the amide proton exchange rates are relatively rapid, with most amides exchanging within 2 hr. Conformational flexibility is common in proteins that are involved in interactions with other macromolecules

(40). Information on the structures of different Skn domain–DNA complexes and on the energetics of binding to different sites is required to evaluate the role of flexibility in the internal region in DNA binding.

Functional Relevance of the NMR Structure. A key question is whether the proposed structure will shed any light on the DNA recognition event and the function of this internal region in DNA binding. The preferred recognition sequence for Skn-1 has been identified as ATTGTCAT (3). The "footprint phenotype" that has been generated for Skn-1 binding shows that the N-terminal arm contacts the minor groove at the ATT half-site and the C-terminal basic region contacts the major groove at the TCAT half-site (3). Thus, the protein binds DNA from the same face of the helix. The tertiary fold we have determined shows that the N-terminal arm and the C-terminal basic region emerge from the same side of the folded domain (Fig. 6). Furthermore, the fold is such that when the basic region is docked in the TCAT half-site with the appropriate helical surface buried in the major groove, the N-terminal arm is positioned to contact the ATT half-site at the 5′ side of the recognition sequence. Therefore, the NMR structure implies that one function of the internal region of the Skn domain is to act as a scaffold that organizes the N-terminal arm and the C-terminal basic region so that they are presented appropriately to the DNA for binding (Fig. 6). This organized presentation of DNA-binding elements presumably facilitates scanning of the DNA by the Skn domain as it searches for appropriate binding sites.

In conclusion, we have shown that the Skn domain has a specific tertiary fold in solution as defined by long-range NOEs. The NMR structure we have presented above should be valuable for developing a complete understanding of the DNA recognition event, which includes an assessment of the conformational changes that take place upon DNA binding.

1. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995) *Science* **267,** 93–96.
2. Bowerman, B., Eaton, B. A. & Priess, J. R. (1992) *Cell* **68,** 1061–1075.
3. Blackwell, T. K., Bowerman, B., Priess, J. R. & Weintraub, H. (1994) *Science* **266,** 621–628.
4. Pal, S., Lo, M.-C., Schmidt, D., Pelczer, I., Thurber, S. & Walker, S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5556–5561.
5. Carroll, A. S., Gilbert, D. E., Liu, X., Cheung, J. W., Michnowicz, J. E., Wagner, G., Ellenberger, T. E. & Blackwell, T. K. (1997) *Genes Dev.* **11,** 2227–2238.
6. McIntosh, L. P. & Dahlquist, F. W. (1990) *Q. Rev. Biophys.* **23,** 1–38.
7. Pelczer, I. & Bishop, K. D. (1997) in *Methods for Structure Elucidation by High-Resolution NMR*, eds. Batta, G., Kövér, K. E. & Szántay, C., Jr. (Elsevier, Amsterdam), pp. 187–207.
8. Braunschweiler, L. & Ernst, R. R. (1983) *J. Magn. Reson.* **53,** 521–528.
9. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys.* **71,** 4546–4553.
10. Piotto, M., Saudek, V. & Sklenár, V. (1992) *J. Biomol. NMR* **2,** 661–666.
11. Sklenár, V. (1995) *J. Magn. Reson. Ser. A* **114,** 132–135.
12. Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989) *J. Magn. Reson.* **85,** 393–399.
13. Pelczer, I. & Carter, B. G. (1997) in *Methods in Molecular Biology: Protein NMR Techniques*, ed. Reid, D. G. (Humana, Totowa, NJ), Vol. 60, pp. 71–155.
14. Kay, L. E., Keifer, P. & Saarinen, T. (1992) *J. Am. Chem. Soc.* **114,** 10663–10665.
15. Grzesiek, S. & Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 12593–12594.
16. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994) *Biochemistry* **33,** 5984–6003.
17. Johnson, B. A. & Blevins, R. A. (1994) *J. Biomol. NMR* **4,** 603–614.
18. Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994) *J. Biomol. NMR* **4,** 845–858.
19. Clore, G. M., Bax, A. & Gronenborn, A. M. (1991) *J. Biomol. NMR* **1,** 13–22.
20. Boelens, R., Burgering, M., Fogh, R. H. & Kaptein, R. (1994) *J. Biomol. NMR* **4,** 201–213.
21. Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. (1994) *J. Biomol. NMR* **4,** 871–878.
22. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995) *J. Biomol. NMR* **6,** 135–140.
23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293.
24. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997) *J. Mol. Biol.* **273,** 283–298.
25. Güntert, P., Braun, W. & Wüthrich, K. (1991) *J. Mol. Biol.* **217,** 517–530.
26. Mumenthaler, C., Güntert, P., Braun, W. & Wüthrich, K. (1997) *J. Biomol. NMR* **10,** 351–362.
27. Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989) *J. Am. Chem. Soc.* **111,** 3997–4004.
28. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
29. Wishart, D. S. & Sykes, B. D. (1994) *Methods Enzymol.* **239,** 363–392.
30. Case, D. A., Dyson, H. J. & Wright, P. E. (1994) *Methods Enzymol.* **239,** 392–416.
31. Moreau, M., de Cock, E., Fortier, P.-L., Garcia, C., Albaret, Z., Blanquet, S., Lallemand, J.-Y. & Dardel, F. (1997) *J. Mol. Biol.* **266,** 15–22.
32. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *J. Biomol. NMR* **5,** 67–81.
33. Billeter, M., Qian, Y.-Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990) *J. Mol. Biol.* **214,** 183–197.
34. Hua, Q. X., Kochoyan, M. & Weiss, M. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2379–2383.
35. Hua, Q. X., Ladbury, J. E. & Weiss, M. A. (1993) *Biochemistry* **32,** 1433–1442.
36. Keepers, J. W. & James, T. L. (1984) *J. Magn. Reson.* **57,** 404–426.
37. Borgias, B. A., Thomas, P. D. & James, T. L. (1989) in *Complete Relaxation Matrix Analysis (CORMA)* (Univ. of California, San Francisco).
38. Clore, G. M. & Gronenborn, A. M. (1991) *Prog. NMR Spectrosc.* **23,** 43–92.
39. Ptitsyn, O. B. (1995) *Adv. Protein Chem.* **47,** 83–229.
40. Szwajkajzer, D. & Carey, J. (1997) *Biopolymers* **44,** 181–198.