The prokaryotic Cys_2His_2 zinc-finger adopts a novel fold as revealed by the NMR structure of *Agrobacterium tumefaciens* Ros DNA-binding domain

Gaetano Malgieri*, Luigi Russo*, Sabrina Esposito*, Ilaria Baglivo*, Laura Zaccaro†, Emilia M. Pedone†, Benedetto Di Blasio*, Carla Isernia*, Paolo V. Pedone*, and Roberto Fattorusso*‡

*Dipartimento di Scienze Ambientali, Seconda Università degli Studi di Napoli, Via Vivaldi 43, 81100 Caserta, Italy; and †Istituto di Biostrutture e Bioimmagini, Consiglio Nazionale delle Ricerche, Via Mezzocannone 16, 80134 Napoli, Italy

Edited by Gary Felsenfeld, National Institutes of Health, Bethesda, MD, and approved September 18, 2007 (received for review July 16, 2007)

The first putative prokaryotic Cys_2His_2 zinc-finger domain has been identified in the transcriptional regulator Ros from *Agrobacterium tumefaciens*, indicating that the Cys_2His_2 zinc-finger domain, originally thought to be confined to the eukaryotic kingdom, could be widespread throughout the living kingdom from eukaryotic, both animal and plant, to prokaryotic. In this article we report the NMR solution structure of Ros DNA-binding domain (Ros87), providing 79 structural characterization of a prokaryotic Cys_2His_2 zinc-finger domain. The NMR structure of Ros87 shows that the putative prokaryotic Cys_2His_2 zinc-finger sequence is indeed part of a significantly larger zinc-binding globular domain that possesses a novel protein fold very different from the classical fold reported for the eukaryotic classical zinc-finger. The Ros87 globular domain consists of 58 aa (residues 9–66), is arranged in a $\beta\beta\beta\alpha\alpha$ topology, and is stabilized by an extensive 15-residue hydrophobic core. A backbone dynamics study of Ros87, based on ^{15}N R_1, ^{15}N R_2, and heteronuclear ^{15}N-{^1H}-NOE measurements, has further confirmed that the globular domain is uniformly rigid and flanked by two flexible tails. Mapping of the amino acids necessary for the DNA binding onto Ros87 structure reveals the protein surface involved in the DNA recognition mechanism of this new zinc-binding protein domain.

DNA binding proteins | NMR spectroscopy | Ros protein

Eukaryotic Cys_2His_2 (or classical) zinc-finger domain is one of the most important structural motifs involved in protein–DNA interactions and is also known to be involved in binding of RNA, lipids, and proteins (1–5). The classical zinc-finger is a small domain consisting of ≈30 aa in which a zinc ion, crucial for its stability, is tetrahedrally coordinated by two cysteines and two histidines. Its amino acid consensus sequence is $(F/Y)XCX_{2-5}CX_3(F/Y)X_5\Psi X_2HX_{3-5}H$, where X represents any amino acid and Ψ is any hydrophobic amino acid; Ψ forms with the other two hydrophobic residues (F/Y) a small hydrophobic core that together with the zinc ion stabilizes a compact 3D structure, consisting in an antiparallel β-sheet faced by an α-helix ($\beta\beta\alpha$ fold) (2). The α-helix is constituted of three turns including the two coordinating histidines on two successive turns at the C-terminal part of the finger, whereas the β-sheet occurs at the N-terminal part and contains the two coordinating cysteines. Structural studies accomplished on classical zinc-finger protein–DNA complexes have revealed that sequence-specific recognition is achieved by contacts between the α-helix of the zinc-finger and bases in the major groove of the DNA.

A single zinc-finger domain in itself is not sufficient for high-affinity binding to a specific DNA target sequence. In fact, proteins containing multiple zinc-finger domains usually require a minimum of two zinc-fingers for high-affinity DNA binding (1, 6). Nevertheless, the single zinc-finger domain present in the *Drosophila* GAGA transcription factor (7, 8), as well as the QALGGH single zinc-finger domain of the *Arabidopsis thaliana* SUPERMAN protein (9, 10), are capable of sequence-specific DNA binding when flanked by basic regions.

Recently, the first putative prokaryotic Cys_2His_2 zinc-finger domain has been identified in a transcriptional regulator, the Ros protein, from *Agrobacterium tumefaciens* (11), indicating that the classical zinc-finger domain, originally thought to be confined to the eukaryotic kingdom, could be widespread throughout the living kingdom from eukaryotic, both animal and plant, to prokaryotic. *A. tumefaciens* is a Gram-negative bacterium able to infect a large number of plants. The infection leads to crown gall tumors caused by a horizontal transfer of genes, similar to bacterial conjugation (12), from the bacterium to the plant. The transferred genes, 25 kb called T-DNA and contained in the 200-kb Ti plasmid, encode products that catalyze the formation of plant growth hormones (indoloacetic acid and cytokinin) in the transformed plant cells (11).

The protein Ros negatively regulates the *virC* and *virD* operons (13), present on the Ti plasmid, whose products are involved in the processing of the T-DNA. It binds a 40-bp sequence, named Ros box, present in the promoter of *virC* and *virD* and in the promoter of *ros* gene itself (14, 15). Ros also regulates the expression of the *ipt* oncogene located on the T-DNA region (11). Mutation in the *ros* gene causes increased expression of *virC* and *virD*, cold temperature sensitivity, and derepression of the *ipt* oncogene (11).

Ros is a 15.5-kDa protein with an isoelectric point of 7.13. The N-terminal part of the protein is negatively charged and contains many hydrophobic amino acid residues whereas the C-terminal part is positively charged and hydrophilic. Analysis of Ros primary structure revealed the presence of the sequence $IXCX_2CX_3FX_2LX_2HX_3HH$ (Fig. 1), which significantly resembles the consensus sequence of an eukaryotic Cys_2His_2 zinc-finger domain. Interestingly, this zinc-finger-like domain contains three histidine residues, and the 9-aa region between the second cysteine and the first histidine is shorter than the canonical 12-aa spacer invariantly observed in eukaryotic zinc-finger. We have recently demonstrated (16) that the putative zinc-finger domain is essential for Ros DNA binding and is part of a larger DNA-binding domain (region 56–142, Ros87) that includes four basic regions located on either side of the finger, one at the N terminus and three at the C terminus. We have also shown that Cys-79 (Cys-24 in Ros87),

Author contributions: B.D.B., C.I., P.V.P., and R.F. designed research; G.M., L.R., S.E., I.B., L.Z., E.M.P., C.I., P.V.P., and R.F. performed research; G.M., L.R., S.E., I.B., L.Z., E.M.P., B.D.B., C.I., P.V.P., and R.F. analyzed data; and G.M., L.R., C.I., P.V.P., and R.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviation: HSQC, heteronuclear single quantum correlation.

Data deposition: The NMR chemical shifts have been deposited in the BioMagResBank, www.bmrb.wisc.edu (accession no. 15373), and the atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2JSP).

‡To whom correspondence should be addressed. E-mail: roberto.fattorusso@unina2.it.

This article contains supporting information online at www.pnas.org/cgi/content/full/0706659104/DC1.

BIOCHEMISTRY

Fig. 1. Sequence alignments of Ros87 with the six best homologue proteins found in the databases. Amino acid identities are indicated by asterisks. Conservative and nonconservative homologies are indicated by double and single dots, respectively. The putative Cys_2His_2 zinc-finger region is in red. Basic regions (BR) necessary for Ros87 DNA-binding activity and secondary structure elements as derived from the NMR structure ensemble are also indicated. Interestingly, Ros87 homologues here reported (3) are all prokaryotic transcriptional regulator proteins belonging to bacteria strongly related to plants.

Cys-82 (Cys-27), His-92 (His-37), and His-97 (His-42) are involved in the zinc coordination and that His-96 (His-41) can replace His-97 in the coordination sphere, when His-97 is mutated to alanine. In this article we report the NMR solution structure of the Ros DNA-binding domain, providing a structural characterization of a prokaryotic Cys_2His_2 zinc-finger domain. The obtained high-resolution structure shows that the putative zinc-finger sequence (Fig. 1) is part of a larger domain that assumes a fold very different from the classical fold reported for the eukaryotic classical zinc-finger. Ros DNA-binding domain, in fact, consists of a globular domain comprising 58 aa and stabilized by an extensive 15-residue hydrophobic core. A backbone dynamics study of Ros87, based on ^{15}N R_1, ^{15}N R_2, and heteronuclear ^{15}N-{^1H}-NOE measurements, has further confirmed that the globular domain is uniformly rigid, whereas the two tails are flexible. Mapping of the amino acids necessary for the DNA binding onto Ros87 structure reveals the protein surface involved in the DNA recognition mechanism of this new zinc-binding protein domain that by sequence alignment is shown to be highly conserved in a number of prokaryotic proteins identified so far.

Results

Structure Determination. Absolute estimates of molecular mass and translational diffusion coefficient of Ros87 were determined by using a combination of size exclusion chromatography, multiple-angle light scattering, and quasi-elastic light scattering and compared with the translational diffusion coefficient obtained through the DOSY experiments, indicating that Ros87 is monomeric up to the NMR concentration. A nearly complete assignment of Ros87 ^1H, ^{13}C, and ^{15}N resonances has been obtained by using standard triple resonance experiments (see *Materials and Methods*). Secondary structural elements of Ros87 (Fig. 1) were initially identified by the analysis of the chemical shift index, and successively the NOE pattern and the hydrogen exchange confirmed the results [see supporting information (SI) Fig. 7 and *SI Materials*]. Protein structures were calculated based on 1,643 experimental constraints, derived by the NOESY experiments and the coupling constants measurements, and on 40 residual dipolar couplings (Table 1).

Solution Structure of Ros87. The NMR structure of Ros87 is of high quality and consists of a very well defined globular domain (rmsd = 0.417 Å) of 58 aa, ranging from Pro-9 to Tyr-66, and two disordered tails at the N and C termini (Fig. 2). Ros87 globular fold is stabilized by an extended hydrophobic core and has a $\beta\beta\beta\alpha\alpha$ topology, which is flanked by a series of well defined β-turns. The N-terminal region of the domain is constituted by a short loop (loop1, residues 9–13), followed by a distorted type I β-turn (residues Val-13, Arg-14, Lys-15, and Ser-16) preceding the first β-strand. β_1 strand (β_1, formed by Val-17 and Gln-18), β_2 strand (β_2, formed by His-21, Ile-22, and Val-23), and β_3 strand (β_3, formed by Ser-30 and Phe-31) constitute an antiparallel β-sheet that partially faces α-helix 1 (α1). The exposed surface of the β-sheet is constituted by side chains of Gln-18, His-21, whose side chain is a $N\delta_1$-H tautomer (16), stabilized by $N\delta_1$-H hydrogen bond with Asp-19 backbone carbonyl group (Fig. 3 *Right*), Val-23, and Ser-30. β_1 and β_2 are connected by a type II β-turn, which contains two acidic residues. The loop connecting β_2 and β_3 (loop 2) is in part constituted by a well defined type II β-turn (formed by Cys-24, Leu-25, Glu-26, and Cys-27) and contains the two cysteines coordinating the zinc ion. A short two-residue loop (loop 3) links β_3 and α1, which is constituted by slightly more than two turns, ranging from Leu-34 to His-42. The zinc ion resides on a tip of the globular fold and is tetrahedrally coordinated by Cys-24 and Cys-27 thiolates and by His-37 and His-42 side chain Nε nitrogens (Fig. 2 *Right*). A three-residue loop (loop 4) connects α1 to α-helix 2 (α2), which includes residues from Pro-46 to Trp-53 and whose axis is nearly orthogonal with the α1 axis. α2 is followed by two tight turns, a type II β-turn, formed by residues Leu-55, Pro-56, Val-57, and Asp-58, and a type I β-turn, formed by residues Ala-63, Pro-64, Ala-65, and Tyr-66; those two turns are linked by a four-residue loop (loop 5), and this protein

Table 1. NMR structural statistics

NMR constraints	
Distance	1,215
Intraresidue	203
Sequential ($\|i-j\|=1$)	321
Medium-range ($\|i-j\|<5$)	346
Long-range ($\|i-j\|\geq 5$)	345
Dihedral angle restraints	428
Hydrogen bonds	19
Residual dipolar couplings (^1H-^{15}N)	40
Structure statistics	
rmsd from idealized covalent geometry	
Bond length, Å	0.0038 ± 0.0006
Bond angle, °	0.526 ± 0.051
rmsd from distance restraints, Å	0.0069 ± 0.0004
rmsd from dihedral restraints, °	0.162 ± 0.017
rmsd from RDCs, Hz	1.20 ± 0.01
CYANA target function, Å^2	2.1 ± 0.2
AMBER, kcal/mol	
Total	−747 ± 11
Van der Waals	−503 ± 3
Electrostatic	−546 ± 63
Coordinate precision	
rmsd from mean structure (residues 9–66), Å	
All backbone atoms	0.417
All heavy atoms	0.820
Ramachandran analysis (residues 9–66), % residues	
Most favored regions	64
Additional allowed regions	28
Generously allowed regions	6
Disallowed regions	2

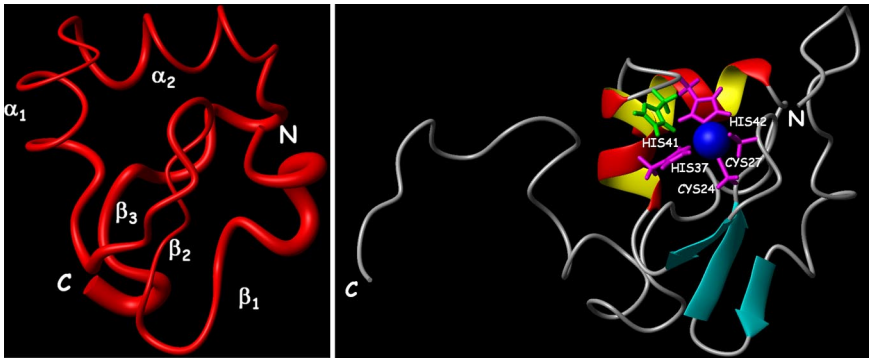

Fig. 2. The solution structure of Ros87. (*Left*) Sausage representation of the globular fold (residues 9–66) of Ros87 NMR ensemble of structures. The secondary structure elements are indicated. (*Right*) Ribbon drawing of one representative conformer of the Ros87 NMR structure. The zinc ion, the four coordinating side chains (magenta), and the His-41 side chain (green) are shown.

region is strongly anchored through a backbone hydrogen bond (Ala-63 HN → Lys-32 CO) to loop 3. The hydrophobic core is well resolved in the solution structure and is constituted by 15 side chains of residues positioned quite uniformly along the entire domain backbone chain, particularly by Pro-9, Val-17, Ile-22, Leu-25, Phe-31, Leu-34, Leu-38, Met-44, Tyr-49, Trp-53, Leu-55, Pro-56, Tyr-59, Met-61, and Pro-64 (Fig. 3). The 67–72 region, although being predicted as an helix on the basis of the CSI (see SI Fig. 7), does not fold in any predominant secondary structure element and is, on the contrary, structurally disordered.

The Globular Domain of Ros87 Shows a Previously Uncharacterized Fold. It is important to point out that no other structures having a fold homologous to that of the Ros87 globular domain have been at this point in time reported in the Protein Data Bank. An extensive search in the structural database, using the Dali method and the CATH database (17–19), did not produce any significant match.

Backbone Dynamics of Ros87. The three relaxation parameters ^{15}N R_1, ^{15}N R_2, and heteronuclear ^{15}N-{^1H}-NOE of Ros87 have been measured. The graphs of the relaxation parameters vs. residue numbers are reported in *SI Materials* (see SI Fig. 8). Relaxation parameters are generally constant along the whole globular domain as expected for a rigid structure, whereas they are well below the mean values in the N- and C-terminal regions. Interestingly, R_2 values higher than the mean have been found for backbone amides at the N-terminal region of the globular domain (residues 11–16). The measured relaxation data were used in the ModelFree software to determine the parameters characterizing the internal mobility. Five models were used to appropriately fit the dynamical parameters to the experimental relaxation data. The model selection strategy of Mandel *et al.* (20) was used to select the correct model for each residue (see SI Table 2 and *SI Materials*), and the axially symmetric diffusion tensor of the molecule has been chosen as the best fitting the collected relaxation data. The initial estimations of the overall molecular correlation time τ_m (6.88 ± 0.1 ns) were calculated on the basis of R_2/R_1 ratio and later optimized with the ModelFree protocol; the calculated dynamics parameters, S^2 and τ_e, vs. the polypeptide sequence of the two proteins are reported in Fig. 4.

Discussion

Ros protein is a transcriptional regulator from *A. tumefaciens* (11), containing the first identified putative prokaryotic Cys_2His_2 zinc-finger domain. After Ros identification, a number of Ros homologue proteins have been found in other prokaryotic organisms, which share a very high sequence identity (Fig. 1). We recently reported (16) the complete functional characterization of the Ros DNA-binding domain, demonstrating that in the single zinc-binding motif present in the Ros protein the metal ion is coordinated by two cysteine and two histidine residues (Fig. 2 *Right*). Moreover, we have shown that the putative Cys_2His_2 zinc-finger domain is essential for Ros DNA binding and is part of a larger DNA-binding

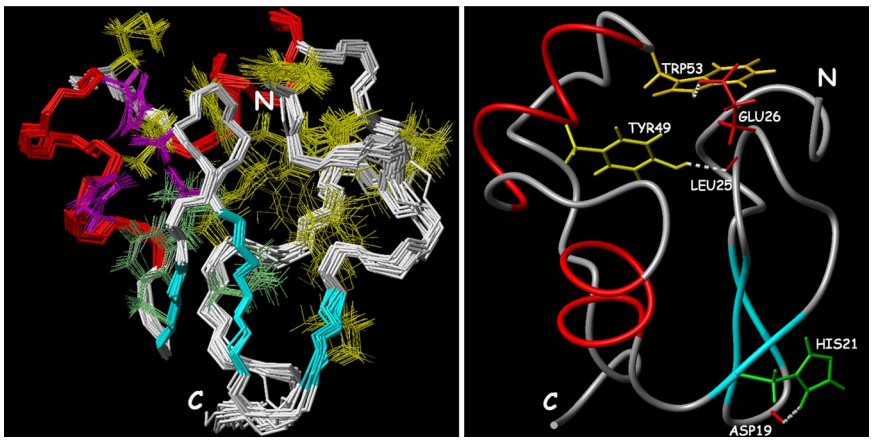

Fig. 3. The hydrophobic core of Ros87. (*Left*) Superposition of the best NMR structures (residues 9–66) to show the polypeptide backbone, the four zinc-coordinating residues, and the 15 hydrophobic core side chains. The four zinc-coordinating residues are depicted in magenta, the three corresponding to the eukaryotic hydrophobic core are in cyan, and the others are in yellow. (*Right*) Three relevant hydrogen bonds (white) forming in Ros87 NMR structure.

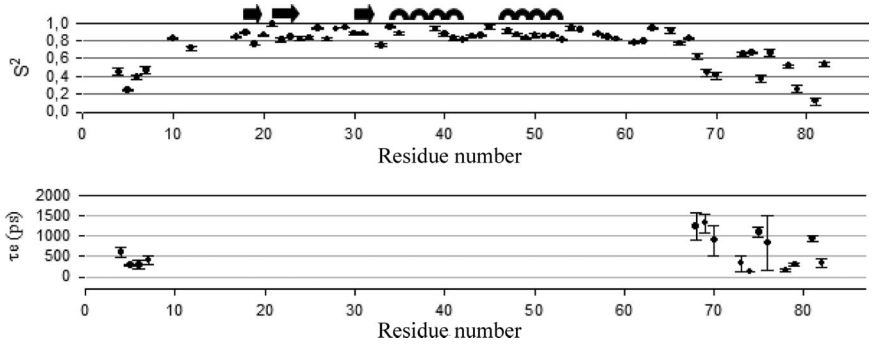

Fig. 4. The order parameters, S^2 and τ_e, defining the backbone dynamics of Ros87 are plotted as a function of the residue numbers.

domain that includes four basic regions located on either side of the finger, one at the N terminus and three at the C terminus. Here we present the solution structure of Ros DNA-binding domain (Ros deletion mutant 56–142, named Ros87), which represents a high-resolution structural characterization of the prokaryotic Cys_2His_2 zinc-finger domain. The NMR structure of Ros87 clearly shows that the putative prokaryotic Cys_2His_2 zinc-finger sequence is indeed part of a significantly larger zinc-binding globular domain, which possesses a novel protein fold. Ros87 globular domain consists of 58 aa and is arranged in a $\beta\beta\beta\alpha\alpha$ topology (Fig. 2). To better appreciate the differences between prokaryotic and the eukaryotic Cys_2His_2 zinc-finger domains, we superimposed Ros87 globular domain with the first zinc-binding domain of Tramtrack (21), which possesses a triple β-sheet similarly to Ros87, aligning their zinc-coordinating residues (Fig. 5). The two zinc coordination spheres are extremely similar; in particular, in Ros87 Cys-24 and Cys-27, located on the β-hairpin, together with His-37 and His-42, positioned at the middle and at the C terminus of α1, tetrahedrally coordinate the zinc ion through their thiolate sulfurs and indole Nε nitrogens, respectively. His-41, able to coordinate the zinc ion when His-42 is mutated to Ala (16), is also included in α1 and is close to the coordination sphere, having the possible role to further protect the zinc ion from the water bulk (Fig. 2 *Right*). The relative orientation of Ros87 triple β-sheet and α1 is also very similar to that observed in Tramtrack zinc-finger 1 (Fig. 5); on the contrary, α1 in the Ros structure is indeed one turn shorter than the α-helix in Tramtrack and in all of the other eukaryotic classical zinc-finger

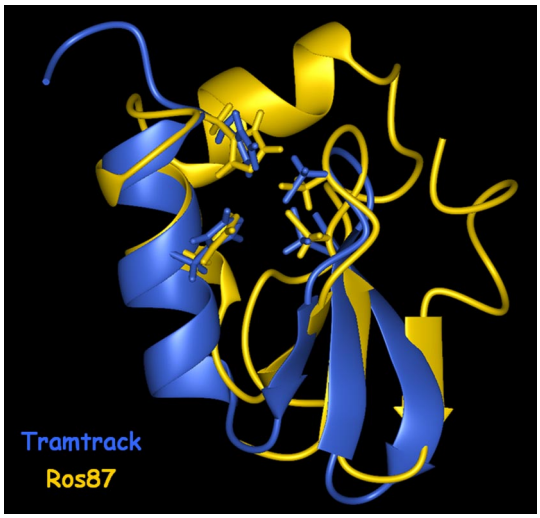

Fig. 5. Superposition of Ros87 globular domain with the first zinc-finger domain of Tramtrack protein (Protein Data Bank ID code 2DRP), obtained by aligning the four zinc-coordinating residues.

domains. This missing turn is clearly due to the linker between the second cysteine and first histidine, which in Ros87 is three residues shorter but is still able to orient the four zinc-coordinating residues in the same relative orientation as in the eukaryotic zinc-finger domain. Moreover, in Ros87, α2 bends over the $\beta\beta\beta\alpha$ region with an axis nearly orthogonal to the α1 axis and contributes to form the enlarged compact hydrophobic core. In the eukaryotic Cys_2His_2 zinc-finger domain the zinc coordination and the small three-residue hydrophobic core contribute similarly to the fold stabilization, whereas Ros87 contains an extensive and highly conserved (Fig. 1) 15-residue hydrophobic core, which appears to play a major role in stabilizing the globular fold (Fig. 3 *Right*). Particularly, residues included in each of the secondary structure elements of the $\beta\beta\beta\alpha\alpha$ motif are involved in the hydrophobic core, and two hydrogen bonds anchor the α2 to the β-hairpin, further stabilizing the globular domain (Fig. 3 *Right*). On the contrary, amino acids of the N-terminal (residues 1–8) and of the C-terminal (residues 67–87) tails do not make any relevant interaction with the globular domain and are almost completely disordered.

The ModelFree analysis based on the measured NMR relaxation parameters and on the heteronuclear ^{15}N-{^1H}-NOE values provided an optimized Ros87 τ_m value of 6.6 $\pm$ 0.2 ns, corresponding, through the Debye equation, to a hydrodynamic radius (r_h) value of 1.9 $\pm$ 0.1 nm, which is in a good agreement with the r_h values derived from the DOSY translational diffusion coefficient (2.1 $\pm$ 0.1 nm) and from the Ros87 NMR structure analyzed with HYDRO software (2.1 $\pm$ 0.1 nm). The obtained S^2 values (Fig. 4) in the 10–66 region (residue 9 is a proline) are uniformly rather high and significantly drop in the two terminal regions, confirming that Ros87 consists of a compact globular domain and two flexible tails. In particular, the global average S^2 value is 0.86 $\pm$ 0.01 in the region 10–66 and 0.39 $\pm$ 0.05 and 0.50 $\pm$ 0.06 in the N and C termini, respectively. Moreover, exchange terms (R_{ex}) are required for only two residues of the globular domain (0.351 $\pm$ 0.172 s^{-1} for Val-23 and 0.303 $\pm$ 0.169 s^{-1} for Gly-28), and effective internal correlation times (τ_e) are needed for 4 and 11 residues of the N and C termini, respectively (Fig. 4). Interestingly, Arg-14 and Lys-15, which are necessary for DNA binding of Ros87 (Fig. 6, BR1), are included in a region that shows R_2 values higher than the mean; therefore, they should be affected by chemical exchange processes occurring on slow microsecond-to-millisecond time scales, which have been already reported to characterize residues involved in nonspecific and specific protein–DNA interactions (22).

Surface mapping of the amino acids that have been demonstrated to be essential for Ros87 high-affinity DNA binding (16) is shown in Fig. 6. Lys-14 and Arg-15, located in the basic region at the N terminus of the zinc-binding motif (BR1), Lys-35 and Arg-36 in α1, and Arg-50 and Lys-52 in α2 (BR2) are included in the globular domain, whereas Arg-70 and Arg-72 (BR3) are just at the beginning of C-terminal tail and Arg-82, Arg-83, and Lys-84 (BR4) are at its end. BR2 side chains are involved in ionic interactions with Asp-58

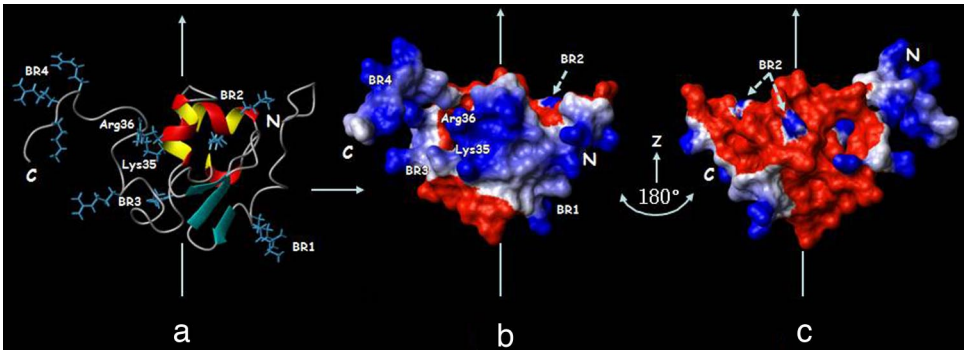

Fig. 6. DNA binding surface of Ros87. (*a*) Mapping of the residues necessary for DNA binding, as previously determined, onto Ros87 structure, shown as a ribbon drawing. (*b* and *c*) Solvent-accessible surface of Ros87 in the same orientation as in *a* (*b*) and its rotation of 180° around the *z* axis (*c*). Surface properties of Ros87 are blue for positively charged residues and red for negatively charged residues.

and Glu-26 carboxylate oxygens, respectively, playing therefore a clear structural role in the stabilization of Ros87 globular domain. On the contrary, BR1, BR3, BR4, Lys-35, and Arg-36 side chains are solvent-exposed and form a basic face, as is shown in Fig. 6; as a result, their relevance in Ros87 DNA-binding activity could be well explained by a direct involvement in Ros87–DNA interaction. Interestingly, the BR3 region is included in the 67–72 fragment that has been shown by the CSI prediction (SI Fig. 7) to have some tendency to assume an helical conformation, which is not clearly present in Ros87 solution structure but could be further stabilized by the interaction with the DNA. We therefore propose that Ros87 interacts with its DNA specific target through a surface including the N-terminal region of the globular domain and the α1 and through its C-terminal tail that could wrap around the double helix. In this way, Ros87 could likely contact and recognize more than three DNA bases, not necessarily contiguous. Moreover, the sequence alignment of the Ros homologues (Fig. 1) indicates that they should preserve Ros87 globular domain, and they probably recognize very similar or even identical DNA target sequences, because the amino acids involved in the DNA recognition are highly conserved.

The eukaryotic classical zinc-finger domains recognize their specific target sequence mostly by contacts between the α-helix and bases in the major groove of the DNA, with each finger being able to fold independent of the rest of the protein and contacting a triplet of the DNA target site; also in the Ros DNA-binding domain amino acids of α1 are important for high-affinity DNA binding, but the presence of amino acids involved in DNA binding also in other regions of the 58-aa globular domain suggests a different DNA-binding modality.

Conclusions

When a putative Cys_2His_2 zinc-finger domain was discovered in the Ros protein, possible structural differences with their eukaryotic counterparts were predicted on the basis of the shorter distance between the second cysteine and the first histidine residues (11). In this article we show by a structural and dynamics NMR study that Ros DNA-binding domain adopts a novel protein fold, which comprises ≈60 aa and is structurally very different from the eukaryotic Cys_2His_2 zinc-finger domains. In particular, Ros87 shows a globular domain characterized by a conserved extensive 15-residue hydrophobic core, which should play in the fold stabilization a much more relevant role than the zinc coordination. The $\beta\beta\alpha$ topology of the region that folds around the zinc ion resembles the structure of the eukaryotic Cys_2His_2 zinc-finger domain (Fig. 5), but, differently from the eukaryotic counterpart, which clearly folds independent of the rest of the protein, in the Ros DNA-binding domain it is part of a significantly larger globular domain. Nonetheless, the similarity of Ros87 zinc-binding region with the eukaryotic Cys_2His_2 zinc-finger domain suggests that the two domains could be evolutionarily related. *A. tumefaciens* is well known for its unique ability to transfer and incorporate foreign DNA into plants; through this mechanism some plant could have acquired from *A. tumefaciens* or from some other plant-infecting bacterium the region encoding the ros (or a ros homologue) zinc-binding motif, which, in the eukaryotic organisms, could have been modified and mainly used in multiple contiguous copies to recognize DNA sequences. On the contrary, such an event might have taken place in reverse during the course of evolution, and the bacterial genomes may have acquired the region encoding the zinc-finger motif from an eukaryotic source and then used it in a different fashion as part of a larger protein domain.

Materials and Methods

NMR Sample Preparation. Single-labeled (^{15}N Ros87) and double-labeled (^{15}N-^{13}C Ros87) proteins were overexpressed and purified as previously published (16).

NMR samples typically contained 1 mM ^{15}N Ros87 or ^{15}N-^{13}C Ros87, 20 mM phosphate buffer (pH 6.8), 0.2 M NaCl, and 90% H_2O/10% ^2H_2O or 100% ^2H_2O. Gel electrophoresis and mass spectrometry were used to verify the identity, purity, and isotopic labeling of the protein.

NMR Spectroscopy. NMR experiments were acquired at 298 K on four different spectrometers: Bruker Avance 500 MHz with cryoprobe and 800 MHz at the European Magnetic Resonance Center of the University of Florence (Florence, Italy), Varian Unity INOVA 600 MHz at the Institute of Biostructures and Bioimages of Consiglio Nazionale delle Ricerche (Naples, Italy), and Varian Unity INOVA 500 MHz at the Environmental Science Department of the Second University of Naples (Naples, Italy). Triple-resonance NMR experiments including 3D HNCA (23, 24), 3D CBCANH (25), and 3D CBCA(CO)NH (25) were collected to enable sequence-specific backbone and C_β resonances assignment. The side-chain ^{1}H and ^{13}C NMR signals were assigned from (H)CCH-TOCSY experiments (26). NOE were evaluated from 3D ^{15}N- and ^{13}C-edited NOESY spectra and 2D [^{1}H,^{1}H]-NOESY. All of the NOESY spectra have been acquired with a mixing time of 100 ms. Slowly exchanging amide protons were identified in an ^{15}N-heteronuclear single quantum correlation (HSQC) spectrum recorded immediately after exchanging the protein into a buffer prepared with ^2H_2O. Vicinal (three-bond) H_N-Hα coupling constants ($^3J_{HNH\alpha}$) were evaluated from cross-peak intensities in quantitative J-correlation (HNHA) spectra (27). Residual dipolar couplings (H_N-N) were measured by using an in-phase/antiphase HSQC experiment (28) on ^{15}N-^{13}C Ros87 in a liquid crystalline medium of 7% polyacrilamide, 0.1% ammonium persulfate, and 0.5% TEMED. The translation diffusion coefficient (D_t) was

measured by using the pulsed-field gradient spin-echo DOSY experiment (29). A correction factor was introduced to keep in count of the major viscosity of the solution 90% H_2O and 10% ^2H_2O. The Stokes–Einstein equation was used to calculate the hydrodynamic radius. The hydrodynamic properties were also evaluated by using HYDRO software (30).

NMR experiments were processed by using Varian (VNMR 6.1B) or Bruker (XWIN NMR) software. ^1H, ^{13}C, and ^{15}N chemical shifts were calibrated indirectly by using external references. The program XEASY (31) was used to analyze and assign the spectra.

Structure Calculations. NOE-derived distance constraints, coupling constants, and residual dipolar couplings were used to calculate Ros87 structures with the program CYANA (32, 33). The input data for the final structure calculation are reported in Table 1. The zinc ion was not included in the calculations. A total of 100 structures was calculated, and the 20 conformers with the lowest CYANA target function were further refined by means of unrestrained energy minimizations with the program SPDB (34).

The small number of residual constraint violations (Table 1) indicates that the input data represent a self-consistent set and that the constraints are well satisfied in the calculated conformers. The global rmsd value calculated for the backbone atoms of the region 9–66 (Table 1) shows that an overall high precision of the structure determination has been achieved. The structures were visualized and evaluated by using the programs MOLMOL (35) and PROCHECK-NMR (36). The chemical shift assignments are available from the BioMagResBank (accession no. 15373), and the final atomic coordinates are available from the Protein Data Bank (ID code 2JSP).

Relaxation Data Processing and Analysis. The relaxation parameters were evaluated by recording and analyzing the following set of experiments: inversion recovery ^1H-^{15}N HSQC for the evaluation of R_1; spin echo ^1H-^{15}N HSQC for the evaluation of R_2; and two ^1H-^{15}N HSQCs for the evaluation of the ^{15}N-$\{^1H\}$ steady-state heteronuclear NOE (in one the protons were unsaturated, and in the other the protons were saturated for 3 s). R_1 and R_2 rates were determined by fitting the peak heights at multiple relaxation delays (37). Uncertainties in R_1 and R_2 were obtained from the error fit. ^{15}N-$\{^1H\}$ steady-state NOEs were calculated as the ratio of ^1H-^{15}N correlation peak heights in the spectra acquired with and without proton saturation, and their uncertainties were set to 5%. S^2 values were derived from a model free analysis of the R_1, R_2, and heteronuclear NOE data using the ModelFree software package (20, 38). An initial estimate of the magnitude and orientation of the diffusion tensor was obtained from the ratios of ^{15}N R_2 and R_1 values by using the programs QUADRIC_DIFFUSION (39, 40) and R_2R_1_1.1 (41). Residues with large-amplitude fast internal motions were excluded from the calculation. Among the remaining residues, those with significant conformational exchange on the microsecond to millisecond time scale were also excluded.

Hydrodynamic Properties. Ros87 (100 μl, 1.0 mM) in 20 mM phosphate buffer (pH 6.8) and 0.2 M NaCl solution was loaded onto an S-75 16/60 column (GE Health Biosciences), preequilibrated with the same buffer, and eluted at room temperature at a flow rate of 1 ml/min. The column was connected downstream to a multiangle laser light (690.0 nm) scattering DAWN EOS photometer (Wyatt Technology). Quasi-elastic (dynamic) light scattering data were collected at a 90° angle by using a Wyatt quasi-elastic light scattering device. Data were analyzed by using Astra 4.90.07 software (Wyatt Technology).

We thank Mr. Fabio Calogiuri, Mr. Massimo Lucci, and Dr. Leonardo Gonnelli of the European Magnetic Resonance Center for their assistance in the acquisition and processing of the NMR and multiple-angle light scattering/quasi-elastic light scattering experiments. We also thank Mr. Leopoldo Zona, Dr. Vincenzo Piscopo, Mr. Maurizio Muselli, and Mr. Marco Mammucari for the excellent technical assistance. Access to the 800- and 500-MHz spectrometers at the European Magnetic Resonance Center of the University of Florence was provided by the Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine Paramagnetiche (Contract MIUR-RBLA032ZM7). This work was partially funded by Ministero dell'Istruzione, dell'Università e della Ricerca Grants PRIN 2005 (to C.I.), PRIN 2006 (to R.F. and P.V.P.), and FIRB 2003 (to P.V.P.) and by L.R. 5 2003 from Regione Campania (to P.V.P.).

1. Klug A, Schwabe JW (1995) *FASEB J* 9:597–604.
2. Wolfe SA, Nekludova L, Pabo CO (2000) *Annu Rev Biophys Biomol Struct* 29:183–212.
3. Isernia C, Di Blasio B, Baglivo I, Pedone PV, Fattorusso R (2006) in *Recent Development in Bioinorganic Chemistry: Metal Complexes of Bioactive Molecules* (Transworld Research Network, Kerala, India).
4. Brown RS (2005) *Curr Opin Struct Biol* 15:94–98.
5. Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP, Crossley M (2007) *Trends Biochem Sci* 32:63–70.
6. Berg JM, Godwin HA (1997) *Annu Rev Biophys Biomol Struct* 26:357–371.
7. Pedone PV, Ghirlando R, Clore GM, Gronenborn AM, Felsenfeld G, Omichinski JG (1996) *Proc Natl Acad Sci USA* 93:2822–2826.
8. Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM, Clore GM (1997) *Nat Struct Biol* 4:122–130.
9. Dathan N, Zaccaro L, Esposito S, Isernia C, Omichinski JG, Riccio A, Pedone C, Di Blasio B, Fattorusso R, Pedone PV (2002) *Nucleic Acids Res* 30:4945–4951.
10. Isernia C, Bucci E, Leone M, Zaccaro L, Di Lello P, Digilio G, Esposito S, Saviano M, Di Blasio B, Pedone C, *et al.* (2003) *ChemBioChem* 4:171–180.
11. Chou AY, Archdeacon J, Kado CI (1998) *Proc Natl Acad Sci USA* 95:5293–5298.
12. Kado CI (1992) *Molecular Signals in Plant–Microbe Communications*, ed Verma DPS (CRC Press, Boca Raton, FL), pp 201–208.
13. Kado CI (2002) *Plasmid* 48:179–185.
14. D'Souza-Ault MR, Cooley MB, Kado CI (1993) *J Bacteriol* 175:3486–3490.
15. Cooley MB, D'Souza MR, Kado CI (1991) *J Bacteriol* 173:2608–2616.
16. Esposito S, Baglivo I, Malgieri G, Russo L, Zaccaro L, D'Andrea L, Mammucari M, Di Blasio B, Isernia C, Fattorusso R, *et al.* (2006) *Biochemistry* 45:10394–10405.
17. Holm L, Sander C (1993) *J Mol Biol* 233:123–138.
18. Dietmann S, Park J, Notredame C, Heger A, Lappe M, Holm L (2001) *Nucleic Acids Res* 29:55–57.
19. Orengo CA, Pearl FM, Bray JE, Todd AE, Martin AC, Lo Conte L, Thornton JM (1999) *Nucleic Acids Res* 27:275–279.
20. Mandel AM, Akke M, Palmer AG (1995) *J Mol Biol* 246:144–163.
21. Fairall L, Schwabe JWR, Chapman L, Finch JT, Rhodes D (1993) *Nature* 366:483–487.
22. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R (2004) *Science* 305:386–389.
23. Ikura M, Kay LE, Bax A (1990) *Biochemistry* 29:4659–4667.
24. Grzesiek S, Bax A (1992) *J Magn Reson* 96:432–440.
25. Grzesiek S, Bax A (1992) *J Am Chem Soc* 114:6291–6293.
26. Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD (1993) *J Magn Reson B* 101:333–337.
27. Vuister GW, Bax A (1993) *J Am Chem Soc* 115:7772–7777.
28. Ottiger M, Delaglio F, Bax A (1998) *J Magn Reson* 131:373–378.
29. Stejskal EO, Tanner JE (1965) *J Chem Phys* 42:288–292.
30. de la Torre JG, Huertas ML, Carrasco B (2000) *Biophys J* 78:719–730.
31. Bartels C, Xia T, Billeter M, Wüthrich K (1995) *J Biomol NMR* 5:1–10.
32. Herrmann T, Güntert P, Wüthrich K (2002) *J Mol Biol* 319:209–227.
33. Guntert P (2004) *Methods Mol Biol* 278:353–378.
34. Guex N, Peitsch MC (1997) *Electrophoresis* 18:2714–2723.
35. Koradi R, Billeter M, Wüthrich K (1996) *J Mol Graphics* 14:29–32.
36. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) *J Biomol NMR* 8:477–486.
37. Viles JH, Duggan BM, Zaborowski E, Schwarzinger S, Huntley JJA, Kroon GJA, Dyson HJ (2001) *J Biomol NMR* 21:1–9.
38. Palmer AG, Rance M, Wright PE (1991) *J Am Chem Soc* 113:4371–4380.
39. Lee LK, Rance M, Chazin WJ, Palmer AG (1997) *J Biomol NMR* 9:287–298.
40. Brüschweiler R, Liao X, Wright PE (1995) *Science* 268:886–889.
41. Tjandra N, Feller SE, Pastor RW, Bax A (1995) *J Am Chem Soc* 117:12562–12566.