Structure of the Functional Domain of φ29 Replication Organizer

INSIGHTS INTO OLIGOMERIZATION AND DNA BINDING*S

Received for publication, February 14, 2005
Published, JBC Papers in Press, March 16, 2005, DOI 10.1074/jbc.M501687200

Juan Luis Asensio,a,b,c Armando Albert,b,d Daniel Muñoz-Espin,b,e,f Manuel Grande,a,b,c Juan Hermoso,a Lauretta Villar,a Jesús Jiménez-Barbero,b Margarita Salas,e and Wilfried J. J. Meijer,e,f

From the aDepartamento de Química Orgánica Biológica, Instituto de Química Orgánica General, the bGrupo de Cristalografía Macromolecular y Biología Estructural, Instituto de Química-Física Rocasolano, and the cGrupo de RMN, Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas (CSIC), 28006, Madrid, the dInstituto de Biología Molecular Eladio Vizuela, CSIC, Centro de Biología Molecular Severo Ochoa (CSIC-Universidad Autónoma Madrid), 28049 Canto Blanco, Madrid, and the eGrupo de Biología Estructural de Proteínas, Centro de Investigaciones Biológicas, CSIC, 28040, Madrid, Spain

The Bacillus subtilis phage φ29-encoded membrane protein p16.7 is one of the few proteins involved in prokaryotic membrane-associated DNA replication that has been characterized at a functional and biochemical level. In this work we have determined both the solution and crystal structures of its dimeric functional domain, p16.7C. Although the secondary structure of p16.7C is remarkably similar to that of the DNA binding homeodomain, present in proteins belonging to a large family of euukaryotic transcription factors, the tertiary structures of p16.7C and homeodomains are fundamentally different. In fact, p16.7C defines a novel dimeric six-helical fold. We also show that p16.7C can form multimers in solution and that this feature is a key factor for efficient DNA binding. Moreover, a combination of NMR and x-ray approaches, combined with functional analyses of mutants, revealed that multimerization of p16.7C dimers is mediated by a large protein surface that is characterized by a striking self-complementarity. Finally, the structural analyses of the p16.7C dimer and oligomers provide important clues about how protein multimerization and DNA binding are coupled.

Despite extensive studies on DNA replication, relatively little is known about its in vivo organization, which, in prokaryotes, occurs at the cell membrane (1–4). The well studied Bacillus subtilis phage φ29 (5) is one of the few systems for which this fundamental process has been investigated. The genome of φ29 consists of a linear double-stranded DNA (dsDNA) that contains a terminal protein covalently linked at each 5’-end. Initiation of φ29 DNA replication, as well as that of most other linear genomes containing a terminal protein attached to its DNA ends, occurs via a so-called protein-primed mechanism (5–7). The φ29 genome encodes most, if not all, proteins required for phase DNA replication making φ29 an attractive system to study membrane-associated DNA replication.

The early-expressed φ29 gene 16.7, which is conserved in all φ29-related phages studied so far, encodes a well characterized membrane protein, p16.7 (130 amino acids), that plays an important role in membrane-associated φ29 DNA replication (5, 8, 9). It contains an N-terminal transmembrane domain which is responsible for membrane localization (9) (a schematic organization of protein p16.7 is shown in Fig. 1A). Analyses of a soluble variant lacking the N-terminal membrane anchor, p16.7A, revealed that it has affinity for both single-stranded (ss) and dsDNA, as well as for the φ29 terminal protein. Moreover, p16.7A, which is a dimer in solution, can form multimers, especially upon DNA binding, and multimerization is important for the mode by which it binds DNA (9–11). Recently, it was shown that the dimerization and DNA-binding activities of p16.7 are confined to the C-terminal half of the protein, p16.7C (12).

Here, we show that p16.7C is able to form multimers in solution and that this process is enhanced in the presence of DNA. To gain insight into the multiple features and functions of p16.7C we determined its solution and crystal structures. In addition, the protein oligomers have been analyzed both by x-ray and NMR methods. The structural features of the monomer, dimer, and oligomers and their implications for DNA binding are discussed.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—Escherichia coli strain JM109 (F’ traD36 lacIq lacZΔM15 proAB +/− relA1 supE44 recA1 Δ(lac-proAB) thi gyrA96 (Nalr) endA1 hsdR17 (rK− mK+) relA1 supE44 recA1 (13)) was used for cloning, and E. coli strain BL21(DE3)pLysS (14) was used

*S The on-line version of this article (available at http://www.jbc.org) contains Figs. S1–S8 and Tables SI and SII.

**Supported by the Ramón y Cajal program of the Spanish Ministry of Science and Technology. To whom correspondence may be addressed. Tel.: 34-91-561-8806 (ext. 329); Fax: 34-91-564-4853; E-mail: iqla@iqog.csic.es.

†Holder of a predoctoral fellowship from Fondo de Investigación Sanitaria.

§The abbreviations used are: dsDNA, double strand DNA; ssDNA, single strand DNA; 2D, two-dimensional; 3D, three-dimensional; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single quantum coherence; r.m.s.d., root mean square deviation.

□ This paper is available on line at http://www.jbc.org

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in U.S.A.
for overexpression of proteins. Kanamycin was added to E. coli cultures and plates at a final concentration of 20 \( \mu \text{g/mL} \).

**DNA Techniques**—All DNA manipulations were carried out according to the method of Sambrook et al. (15). Restriction enzymes were used as indicated by the supplier. \( \gamma^{32} \text{P} \)-labeled ATP (3000 Ci/mol) was obtained from Amersham Biosciences. Plasmid DNA was isolated using a Wizard\textsuperscript{TM} Plus DNA purification kit (Promega, Madison, WI). DNA fragments were isolated from gels using the QiAquick gel extraction kit (Qiagen). The dideoxynucleotide chain-termination method (16) with Sequenase (United States Biochemicals sequencing kit) was used for DNA sequencing. Site-directed Mutagenesis of p16.7C—Site-directed mutations in gene 16.7C were obtained by polymerase chain reaction using the QuikChange\textsuperscript{TM} Site-D mutagenesis kit (Stratagene). Plasmid pET-16.7C (12) was used as template DNA.

**Overexpression and Purification of p16.7C and Its Derivatives**—Protein p16.7C and its derivatives were overexpressed and purified using a Ni\textsuperscript{2+}-NTA resin column as described before (9, 12). Protein p16.7C was obtained by mixing unlabeled with \( ^{13} \text{C}/^{15} \text{N} \)-labeled p16.7C proteins, cells were grown in U-\( ^{15} \text{N} \) and U-\( ^{13} \text{C} \)/\( ^{15} \text{N} \)-Bio-express medium, respectively (Cambridge Isotope Laboratories, Andover, MA).

**Protein Concentration**—The protein (monomer) concentration was determined by absorbance spectrophotometry, measuring the absorbance at 280 nm. The extinction coefficients \( \varepsilon_{280} \) of p16.7C, p16.7Cb, and pE72Q were 11360 \( \text{M}^{-1} \cdot \text{cm}^{-1} \) and that of pR98W was 16800 \( \text{M}^{-1} \cdot \text{cm}^{-1} \).

**Cross-linking Experiments**—Cross-linking reactions using 10 \( \mu \text{M} \) protein and disuccinimidyl suberate as cross-linking agent were performed as described before (9). After cross-linking, proteins were precipitated upon the addition of 1 volume of ice-cold 20\% (w/v) trichloroacetic acid and, after resuspension, analyzed by PAGE in the presence of SDS. The proteins were visualized by Coomassie Blue staining.

**Gel Mobility Shift Assays**—Gel retardation assays were performed as described before (11).

**Nuclease Digestion Assays**—Nuclease digestion assays were performed using end-labeled 297-bp DNA fragments corresponding to the p29 right end genome. Labeled fragments were either used directly (dsDNA) or after heat-denaturation (ssDNA). Micrococcal nuclease digestion of the ssDNA nucleoprotein complexes was performed as described before (11). DNase I reactions contained, in 20 \( \mu \text{l} \), besides the end-labeled dsDNA fragment and the indicated amount of protein, 25 \( \mu \text{M} \) Tris-HCl (pH 7.5) and 10 \( \mu \text{M} \) MgCl\textsubscript{2}. Binding reactions were incubated for 10 min at 37 \( ^\circ \text{C} \) before 0.05 unit of DNase I (Promega) was added. Digestion was allowed to proceed for 2 min at 37 \( ^\circ \text{C} \) after which the reaction was stopped upon EDTA addition to a final concentration of 20 \( \mu \text{M} \). Micrococcal nuclease and DNase I digested samples were analyzed on a 1% agarose gel in TAE buffer.

**Protein Crystalization**—The protein (monomer) concentration was determined by UV spectroscopy and sedimentation equilibrium analysis. The sedimentation coefficient was measured by analytical ultracentrifugation at 86000 \( \text{g} \) for 12 h. The molecular mass and sedimentation coefficient were determined independently for each protein. These data were used to calculate the molecular weight of the protein using an equation based on the assumptions that the protein is monodisperse and spherical. The protein concentration was determined by absorbance at 280 nm. The extinction coefficients of the protein were determined at two different protein concentrations (350 and 3500 \( \mu \text{M} \)) and two different pH values (pH 5.0 and 7.0) in 200 \( \text{mM} \) NaCl, 10 \( \text{mM} \) sodium phosphate buffer. To get structural information about the protein oligomers 2D-NOESY experiments were also carried out with these samples. In addition, \( ^{13} \text{C} \)-HSQC and \( ^{13} \text{N} \)-HSQC spectra were recorded at high and low protein concentration. The high protein concentration sample was prepared by mixing unlabeled with \( ^{13} \text{C} \)-doubled-labeled protein (4:1 ratio) to a final protein concentration of 3000 \( \mu \text{M} \). Finally 2D-NOESY, 3D-NOE and 2D-NOESY-filtered experiments were performed on this sample to analyze the interprotein contacts responsible for oligomerization.

**Protein Crystalization and Data Collection**—Crystallization experiments were carried out with a p16.7C solution at 10 mg/ml containing 200 \( \text{mM} \) sodium phosphate buffer, pH 7.5, 200 \( \mu \text{M} \) NaCl. Protein crystallization was achieved by dialysis of the protein solution against a solution containing 5 mM phosphate buffer, pH 7.5, 75 \( \mu \text{M} \) NaCl. Prismatic crystals (0.05 \( \times \) 0.2 mm) appear after 2 or 3 days of incubation at 4 \( ^\circ \text{C} \). Crystals were successively transferred to a series of solutions with increased percentage of glycerol (from 5\% to 35\%) and mounted in a fiber loop and frozen at 100 K in a nitrogen stream. X-ray diffraction data were collected in a charge-coupled device detector using the European Synchrotron Radiation Facility (Grenoble, France) synchrotron radiation source at wavelengths of 0.92 and 1.61\( \AA \) at the BM16 beam-line. Diffraction data were processed using MOSFLM (21) (Table 1). Structure Determination and Refinement—As mentioned above, the x-ray structure of p16.7C was solved by molecular replacement using the coordinates of an ensemble of 10 preliminary NMR models (AMoRe (22)). The preliminary electron density map was improved by density modification and solvent flattening. The final electron density map was calculated by using the simulated annealing routine of CNS (24). Several cycles of restrained refinement with REFMAC5 (25) and iterative model building with O (26) were carried out. Water structure was also modeled. Calculations were performed using CCP4 programs (27). The final model was refined by iterative maximum likelihood positional and translation, libration, and screw rotation displacement refinement (28). The stereochemistry of the model was

### Table 1

**Characterization of the 25 NMR structures of the p16.7C dimer for structural analysis**

| Distance constraints | Number |
|----------------------|--------|
| NOE-derived upper distance limits (\( \AA \)) | 2180 |
| Intrasidue | 530 |
| Short range | 466 |
| Medium range | 670 |
| Long range | 514 |
| Intramonomer | 1976 |
| Intermonomer | 204 |
| r.m.s.d. from ideal geometry Bonds | 0.0086 |
| Angle | 2.16 |
| Ramachandran analysis (%) Residues in most favored regions | 81.9 |
| Residues in additionally allowed regions | 18.1 |
| r.m.s.d. between pairs of structures (\( \AA \)) Backbone (68–125) | 0.770/0.96 |
| Heavy atoms (68–125) | 1.46/1.56 |

---

M. Finally 2D-NOESY, 13C-selected/12C-filtered experiments were performed on a 500-MHz spectrometers. Protein concentrations for the structural analysis were determined by fitting a sedimentation equilibrium model for a single sedimenting solute to individual datasets with the program EQASSOC (17) (supplied by Beckman).
Structure of the Functional Dimeric Domain of φ29 p16.7 Protein

FIG. 1. The functional domain of p16.7 can form multimers. A, schematic representation of the organization of protein p16.7 (130 residues). TM, transmembrane region. B and C, nuclease treatment of nucleoprotein complexes formed by p16.7Cb. The end-labeled 297-bp right end φ29 DNA fragment was used directly (C) or after heat denaturation (B) in nuclease digestion assays. The labeled probes were preincubated either in the absence or presence of increasing amounts of p16.7Cb. After complex formation, samples were treated with micrococcus nuclease (B) or DNase I (C) as described under “Experimental Procedures,” and the DNA products fractionated through polyacrylamide gels under denaturing conditions. The digestion pattern observed in B is the consequence of preferential cleavage by micrococcal nuclease at AT-rich regions (11, 39). Digestion of the ssDNA fragment without protein during a shorter period of time or using lower micrococcus nuclease concentrations gave patterns highly similar to those obtained in the presence of limited amounts of p16.7Cb. Protein p16.7Cb concentrations used in (B) were 5, 10, 20, 40, and 80 μM; those in C were 5, 10, and 20 μM. D, protein p16.7Cb was cross-linked in vitro in the presence of dsDNA using the cross-linking agent disuccinimidyl adipate. After cross-linking, the sample was subjected to SDS-PAGE and Western blot analyses using polyclonal antibodies against p16.7.

RESULTS

The Functional Domain of Protein p16.7 Can Form Multimers—Gel retardation studies demonstrated that the C-terminal half of protein p16.7 (p16.7C, containing p16.7 residues 63–130) has ss- and dsDNA-binding activities (12). These experiments were performed with protein p16.7Cb, the derivative of p16.7C lacking its N-terminal (His6) tag, to exclude possible effects of this positively charged region on DNA binding. Although these experiments unequivocally demonstrated DNA-binding activity, they did not provide insight as to the ability of p16.7Cb to form multimers or the mode of DNA binding, which can be deduced by nuclease digestion analyses of the nucleoprotein complexes. Therefore, increasing amounts of p16.7Cb were incubated with 5'-labeled ss- or dsDNA probes. Next, the nucleoprotein complexes formed with ss- and dsDNA were challenged to micrococcal nuclease and DNase I digestion, respectively, after which the fragments were fractionated through denaturing polyacrylamide gels (see Fig. 1, b and c). In the absence of p16.7Cb nearly all the end-labeled ssDNA fragment was degraded into small oligonucleotides (Fig. 1b, lane 6). In the presence of p16.7Cb, in the range of 5–20 μM, however, the amount of small oligonucleotides decreased giving rise to a variety of larger ssDNA digestion products (Fig. 1b, lanes 1–3). At higher p16.7Cb concentrations, hardly any degradation products were observed (lanes 4 and 5) indicating that the ssDNA was (almost) completely protected from micrococcal nuclease attack. Together, these results confirm that p16.7Cb binds to ssDNA and, importantly, provide strong evidence that, at elevated concentrations, p16.7Cb forms multimers upon ssDNA binding leading to the generation of a continuous array of protein protecting the entire DNA fragment from micrococcal nuclease attack. Similar results were obtained when dsDNA was used (Fig. 1c). Thus, whereas a typical DNase I digestion pattern was observed in the absence or in the presence of a low amount (5 μM) of p16.7Cb (lanes 1 and 2, respectively), full protection of the dsDNA fragment was seen at higher p16.7Cb concentrations (lanes 3 and 4). Contrary to the analyses with ssDNA, though, no intermediate levels of nuclease protection were observed with dsDNA. This might indicate that multimer formation of p16.7Cb is enhanced upon binding to dsDNA, although this difference may also be due to different digestion characteristics of DNase I versus micrococcal nuclease. In vitro cross-linking carried out in the presence of dsDNA followed by Western blot analyses confirmed the ability of p16.7Cb to form multimers (Fig. 1d). Similar results were obtained using ssDNA instead of dsDNA or using p16.7C (not shown).

Structure Determination of p16.7C Dimer—Analytical ultracentrifugation experiments showed that p16.7C was mainly in its dimeric form at the temperature (30–40 °C), pH (5–7), ionic-strength (200 mM NaCl), and concentration (0.4–0.8 mM) conditions employed for the NMR studies described below. Full backbone and side-chain assignment for p16.7C (Supplemental

verified with PROCHECK (29). Ribbon figures were produced using MOLSCRIPT (30) and RASTER (28). The accessible surface area of p16.7C dimer and protomer was calculated with the program “naccess” from the LIGPLOT package (31).
Table SI) was obtained using standard 2D and 3D NMR techniques on \(^{15}\)N- and \(^{13}\)C/\(^{15}\)N-labeled samples. To distinguish intra- from intermolecular NOEs, half-filtered NOESY experiments (Fig. 2A) were carried out on a heterolabeled dimer obtained by mixing equivalent amounts of unlabeled and \(^{13}\)C/\(^{15}\)N-labeled protein. A summary of the experimental constraints employed and the characterization of the final NMR ensemble are shown in Table I.

In the course of the NMR studies, diffraction quality crystals of the protein were obtained in a dialysis membrane. These conditions were then used as a starting point for crystallographic studies. A preliminary NMR-derived ensemble of ten models was initially employed in the x-ray structural determination of p16.7C by molecular replacement at 2.9-Å resolution (see Table II and “Experimental Procedures”). The r.m.s.d. values for superimposition show that the NMR and crystal structures not only provide validation of one another but also give a more complete picture of the structure and dynamics of p16.7C than either structure alone.

Architecture of the Dimer—The solution and crystallographic 3D structures of p16.7C dimer are shown in Fig. 2 (B and C), respectively. Each polypeptide chain (green and red) contains three α helices (corresponding to p16.7 residues 72–81 (H1), 88–95 (H2), and 103–121 (H3)). Helices H1 and H2, connected by a six-residue loop (62–87; loop 1), are oriented in an antiparallel fashion with a crossing angle of \(\sim 160^\circ\). Helix H3 is connected to H2 by a seven-residue loop (96–102; loop 2) and packs against both helix H1 and H2 with crossing angles of \(54^\circ\) and \(139^\circ\), respectively. The C-terminal region (residues 122–128) adopts an extended structure. Finally, residues 63–66 and 129–130 are disordered (Supplemental Fig. 1B). The secondary and tertiary structure of each monomer is stabilized by formation of a hydrophobic core resulting from the packing of the three helices.

According to both NMR and x-ray data, p16.7C forms a symmetric dimer that corresponds to a novel six-helical fold. Indeed, none of the proteins present in the Rutgers Protein Data Bank exhibit high structural homology with p16.7C according to DALI (32) and SCOP (33). The two monomers, related by a non-crystallographic 2-fold symmetry, are held together by a combination of extensive hydrogen bonding and hydrophobic and electrostatic interactions. In the following description of the main intermolecular contacts, letters \(a\) and \(b\) will be employed to specify the polypeptide chain only when essential to avoid ambiguities. The primary dimer interface is formed by helices H3a/H3b and the extended C-terminal region of both monomers, which are oriented in an antiparallel fashion. Thus, the extended C-terminal region of each protein packs against helices H1 and H3 of the other, being involved in both hydrophobic and polar intermolecular contacts. Especially relevant for the dimer stability seems the role of Leu-124 of each monomer, which is totally buried in a hydrophobic patch.
formed by Tyr-79, Leu-78, and Leu-106 of the other protein monomer. In addition, there are polar intermolecular contacts between Arg-113/Lys-122, Ser-88a/Ser-88b, and Arg-85/Tyr-115. Recent studies have highlighted the influence of CH–π interactions involving proline and aromatic side chains on protein stability (34). In p16.7C the P87a and P87b side chains, 4 Å apart at the dimer interface, pack against the Trp-116 indol ring of both polypeptides (Trp-116a and Trp-116b, Supplementary Fig. S2). This particular arrangement is likely to make a significant contribution to the stability of the complex.

It has been shown that nonspecific binding of proteins to polyelectrolytes such as DNA is purely electrostatic and can be quantitatively explained in terms of competitive condensed counterion displacement from DNA by polycationic regions of the protein (35). In this sense, the distribution of basic residues on the surface of a protein can provide important clues regarding nucleic acid binding sites. Thus, the surface electrostatic distribution of the p16.7C dimer was calculated to determine the most probable DNA binding site location. Fig. 3 shows that the negatively charged residues are clustered at the surface defined by helices H1/H2. In contrast, the opposite side of the dimer exhibits a moderate positive potential and therefore constitutes the most likely DNA binding site. Indeed, this protein region, defined by helices H3a/H3b plus the C-terminal regions (residues 122–130), contains several basic residues (Lys-122a/Lys-122b, Lys-123a/Lys-123b, Arg-126a/Arg-126b, and Lys-130a/Lys-130b). The eight positive charges available for protein-DNA interactions seem low in comparison with those observed in other nonspecific DNA-binding proteins (36). Nevertheless, it is important to mention that protein oligomerization may multiply the DNA recognition surface and therefore enhance DNA binding. This view agrees with the apparent cooperative binding observed for p16.7C (see Fig. 1).

**Structural Basis for p16.7C Oligomerization**—Recent studies have shown that p16.7 multimerization is a prominent feature in its DNA binding mode (11). Here we have demonstrated that p16.7C retains the capacity to form multimers, implying that the C-terminal half of p16.7C contains one or more regions responsible for p16.7C dimer–dimer interaction. p16.7C dimers form a fiber around a crystallographic 3-fold screw axis in the crystals obtained (Fig. 4A). This structural organization may reflect the multimerization mode required for efficient DNA binding and therefore we initiated our studies on the protein surface involved in DNA-binding induced multimerization by examining the interdimeric p16.7C contacts observed in the crystal.

**The Interdimeric Contacts Observed in the p16.7C Crystal Are Not Essential for DNA Binding-induced Multimerization**—The interactions between two neighboring p16.7C dimers in the crystal (see Fig. 4B) involve two salt bridges (Glu-72a/Arg-98b and the symmetry related pair) and six hydrogen bonds (Arg-98a/Val-95b, Asn-67a/Asn-96b, Cys-71a/Gln-100b, and the symmetry related pairs) and cover 881 Å² of the dimer surface (11% of the total solvent-accessible area). Interestingly, the Glu-72 side chain is buried into the hydrophobic interface formed by helices H1, H2, and H3, being a rather unusual orientation for a polar residue. Probably, the poor solvation of Glu-72 confers extra stability to its interaction with Arg-98. Consequently, the two salt bridges seem to be especially relevant for the stability of the complex.

To study whether the interdimeric p16.7C contacts observed in the crystal reflect the multimerization mode involved in efficient DNA binding we have constructed and purified two p16.7C mutants, pE72Q and pR98W, and analyzed their features. In E72Q, the salt bridges observed in the wild-type protein would be replaced by simple polar interactions between a charged and a neutral side chain. In R98W, the interaction between residues 72 and 98 would be totally absent. In addition, the large steric volume of the tryptophan side chain is expected to completely disrupt most of the polar interactions detected at this multimerization interface. To assure that the mutations introduced do not affect dimerization, these mutants were subjected in parallel with p16.7C to *in vitro* cross-linking analysis using the bifunctional cross-linking agent disuccinimidyl suberate. The results, presented in Fig. 5A, show that similar amounts of cross-linked dimers were obtained for each protein demonstrating that the mutations do not have major effects on dimerization. Gel retardation and footprinting assays showed that the DNA binding characteristics of both mutants was similar to that of p16.7C (not shown). Finally, Western blot analysis of *in vitro* disuccinimidyl suberate cross-linked samples in the presence of DNA conclusively showed that both mutants retained their ability to form multimers (Fig. 5B). Together, these results indicate that residues Glu-72 and Arg-98 are not crucial for the DNA binding-induced multimerization of p16.7C.

**Determination of the p16.7C Surface Involved in Oligomerization by NMR**—Solution studies were then undertaken to get information about the p16.7C multimerization mode. First, DOSY (20) experiments were performed at low (350 μM) and high (3500 μM) protein concentrations and two pH values (pH 5.0 and 7.0). Significant differences in the diffusion coefficient
of the protein were detected between the high and low protein concentration NMR samples (Δlog $D = 0.17$ at pH 7 and 0.1 at pH 5.0), indicating a clear difference in the oligomeric state of p16.7C. Second, 2D NOESY and N15/C13 HSQC NMR spectra were carried out with the 350 and 3500 μM protein NMR samples. Interestingly, despite the expected general broadening of the resonances at high protein concentration due to the limited oligomerization occurring under these conditions, clear sequence-dependent differences were detected (these were more evident at pH 5.0, when the extent of oligomerization is lower, according to the diffusion coefficient). Thus, several proton signals in residues Asn-83, Arg-85, Gln-97, Tyr-108, Gln-112, Tyr-115, Glu-119, Lys-123, and Tyr-125 (some examples are shown in Supp. Fig. S4, A and B). Interestingly, they are located at the same extended surface of the protein dimer (see Fig. 6A and Supplemental Fig. S5). Finally, close inspection of conventional NOESY and half-filtered NOESY spectra performed at high protein concentration (3500 μM) resulted in the unambiguous identification of 15 intermolecular NOEs that are non-compatible with the 3D structure of the dimer (some examples are shown in Supplemental Fig. S6). All these contacts (shown in Fig. 6B and listed in Supplemental Table SII) were in agreement with the pattern of concentration-dependent line broadening and chemical shift changes described above. For example, the Tyr-125 aromatic side chain shows clear NOEs with the methyl group of Thr-111. Both residues are more than 21 Å apart in the dimer structure.
and were not detected in the diluted protein sample. Similarly, clear intermolecular contacts were observed between Asn-83/Tyr-115, Leu-99/Tyr-125, and Tyr-125/Tyr-108. Interestingly, all the identified residues involved in interdimeric contacts fall at either side of the p16.7C dimer. Altogether, these data indicate the existence of a multimerization interface that is different from the one detected by x-ray crystallography and that involves larger surface areas.

Most interestingly, the addition of dsDNA (A_{14}T_{14}) to a low p16.7C concentration sample (350 μM) at pH 7.0 produced remarkably similar changes on the 1^{15}N-HSQC spectrum as those observed at high p16.7C concentration (3500 μM) in the absence of DNA. According to cross-link and band-shift experiments, DNA binding promotes protein oligomerization. Thus, the observed line broadening of p16.7C signals in the presence of dsDNA is likely to result from a combination of two different association processes, DNA binding and protein oligomerization. Fig. 7 shows a comparison between two regions of a p16.7C 1^{15}N-HSQC spectrum under three different conditions; low (middle panels) and high (lower panels) protein concentration without dsDNA and low protein concentration in the presence of stoichiometric amounts of dsDNA (upper panels). It can be observed that the NH signals that are most affected by increasing the p16.7C concentration (as the Tyr-115, Tyr-108, Glu-119, or Asn-83 side chain) are also the ones most altered at low protein concentration in the absence of DNA. In contrast, several residues involved in the interdimer contacts observed in the crystallographic fiber, as the side chain of Glu-100 or Asn-67, remain totally unaffected in both cases. These results strongly suggest that DNA binding promotes the same multimerization process detected in solution without DNA at much higher p16.7C concentration.

An NMR-based model of a p16.7C tetramer was calculated by employing the experimental sets previously measured for the dimeric species and 30 additional interprotein NOE-derived constraints to define the relative orientation of both p16.7C dimers within the complex (see Fig. 8A and Supplemental Fig. S7). A particularly striking feature of this model is the self-complementary shape of the p16.7C dimer. It can be observed that the two most internal helices H3 plus the extended 122–127 regions belonging to separate p16.7C dimers are aligned in an antiparallel fashion (crossing angle of 0°). The overall orientation of the two protein dimers within the tetramer differs by 35°. Various experimentally detected interdimeric interactions are apparent from the model. Thus, (i) the aromatic side chain of Tyr-125 presents hydrophobic contacts with Tyr-108 and the methyl group of Thr-111, (ii) helix H3 and loop 1 would participate in a number of polar interprotein interactions involving residues Asn-83, Arg-85, Glu-119, and Tyr-115, and (iii) the helix H2 solvent-exposed surface might present several polar contacts with helix H1.

A model for a larger multimer was obtained by simply extending this oligomerization mode in both directions (Fig. 8B). Due to the 35° twist present at each step, the protein oligomer would define a large helical structure with 10 dimers per turn. The inner part of the helix is characterized by a positive electrostatic potential and its shape and charge complementarity with dsDNA surface strongly suggests that this region of the supramolecular complex participates in DNA recognition. This organization implies that the coiled-coil and transmembrane domains, present in the full-length p16.7 protein, would be exposed to the outer face of the helix. It seems clear that such a configuration would be compatible with attachment to the membrane only for protein oligomers of limited size (up to four p16.7 dimers).

**DISCUSSION**

The ϕ29 dimeric membrane protein p16.7 has been shown to be involved in the organization of membrane-associated ϕ29 DNA replication (8, 9). This small protein (130 residues) presents a remarkable diversity of activities. First, it binds nonspecifically to both ss- and dsDNA (10–12). Second, p16.7 can form multimers, both in vitro and in vivo, and it has been shown that multimerization is crucial for its DNA binding mode (11). Third, it has affinity for the ϕ29 terminal protein. Interestingly, the C-terminal half of protein p16.7 (p16.7C) is responsible for protein dimerization and contains the ϕ29 terminal protein and DNA binding capacities (12). Here, we have shown that protein p16.7C is also capable of forming multimers. Altogether, these results indicate that p16.7C constitutes the functional domain of p16.7. This conclusion is in line with the observation that the C-terminal half of all p16.7 homologues known has a higher level of conservation when compared with the N-terminal half (5).

To gain insight into its various features and functions we have determined the solution and crystal structures of p16.7C. The primary sequence of p16.7C has homology with the DNA binding homeodomains (~20 and 40% of identity and similar-
ity, respectively), which are present in a large family of eukaryotic transcription factors. Especially, several residues highly conserved in most homeodomains and critical for either structure or function are also conserved in p16.7C. In addition, computer-assisted analyses suggested that the secondary structure p16.7C may be similar to that of homeodomains (12). Determination of the structure indeed reveals that the secondary and tertiary structures are stabilized by the formation of a well defined hydrophobic core, resulting from the packing of the three helices. Nevertheless, the spatial organization of the three α-helices in p16.7C is fundamentally different from that of homeodomains (37), and, in contrast to them, p16.7C forms stable dimers. Thus, despite its similarities to homeodomains regarding the primary and secondary structures, the functional domain of p16.7 defines a novel dimeric six-helical fold.

The positively charged surface formed by the H3 helices and its following extended regions probably constitutes the DNA binding site. Considering that native p16.7 is anchored at the membrane by its N terminus, this putative binding surface of p16.7 will be directed toward the center of the cell, and thus is compatible with a role in organizing DNA replication at the membrane of the infected cell.

Dimers of p16.7C are arranged forming a protein fiber in the crystals obtained. Mutagenesis studies, however, show that this organization is not crucial for DNA recognition. Nevertheless, given the large variety of different activities displayed by p16.7 and the apparent strength of the interactions that promote fiber formation, a possible biological function for these contacts cannot be ruled out.

The overall structural features of p16.7C oligomers in solution have been determined by NMR methods. Interestingly, the oligomers display a striking charge and shape complementarity with dsDNA, which is in agreement with a role in promoting binding of p16.7C to dsDNA. The fact that the same multimerization mode was also detected at low p16.7C concentration in the presence of dsDNA strongly supports this view.

The observed protein oligomers present a rather low stability as shown by analytical ultracentrifugation experiments. However, under natural conditions, compartmentalization of native p16.7 by attachment of its N-terminal membrane anchor to the membrane (9) restricts its translation diffusion to two dimensions and increases its local concentration and hence is expected to stimulate multimerization. Local protein concentration will also increase upon DNA binding. Moreover, DNA binding may further enhance multimerization by restricting p16.7C flexibility and/or by introducing subtle changes in the p16.7C dimer.

It is worth mentioning that, although feasible in vitro, the helical arrangement resulting from extensive protein oligomerization is unlikely to be compatible with the anchorage to the bacterial membrane for the full-length p16.7 protein. Fluorescence microscopy showed that, especially at early infection times, strong foci of p16.7 signals were obtained throughout the membrane (8). No indication, however, was obtained that p16.7
would form ring or helical structures at the membrane. Rather, the dispersed patches suggest clustering of limiting amounts of p16.7 molecules at the membrane.

Based on all this information we propose the following model for the protein-DNA complex. The structural organization of p16.7C dimers precludes deep penetration of the H3-helices, located at the probable DNA-binding surface, into the major groove of dsDNA in a similar way to that observed for homeodomains or prokaryotic helix-turn-helix proteins, implying that it binds dsDNA in a fundamentally different way. A protein-DNA complex compatible with the structural features of the p16.7C dimer and oligomers can be built, however, by spanning the binding surface of the protein dimer above the major or minor groove of a canonical B-form of DNA with helices H3a/H3b plus the extended C-terminal regions sitting on the polyphosphate backbone. This proposed DNA binding mode (Supplemental Fig. S8), which is highly similar to that of the so-called MADS-BOXES (38, 39), is attractive because it allows interaction of the arginine and lysine side chains, located at the putative binding surface, with the antiparallel phosphate backbone that forms the groove edges of the dsDNA, which may explain the lack of sequence specificity of p16.7C. Moreover, this model agrees with several other features determined for p16.7C. First, the moderate number of basic residues located at the probable p16.7C DNA binding surface explains why p16.7C oligomerization is required for efficient dsDNA recognition. Second, this binding mode is compatible with the multimerization mode detected in solution permitting consecutive dimers of p16.7C within the oligomer to bind to adjacent DNA regions. Third, due to extension of the DNA binding surface in the protein oligomer the protein-DNA interactions will be multiplied, which explains the observed non-linear increase of in vitro DNA binding with respect to protein concentration and the observed DNA-enhanced multimer formation of p16.7C. In fact, the protein concentration-dependent effects on DNA binding, observed in gel retardation and nuclease protection assays, strongly indicate that p16.7C binds DNA in a cooperative manner. A cooperative way of DNA binding is observed for most proteins that act transiently during DNA replication, and p16.7 is believed to fall into this class of proteins.

Acknowledgment—A.A. thanks the European Synchrotron Radiation Facility (Spanish beam line BM16) for access to the synchrotron radiation source.

REFERENCES

1. Lemon, K. P., and Grossman, A. D. (1998) Science 282, 1516–1519
2. Noirot-Gros, M. F., Dervyn, E., Wu, L. J., Mervelet, P., Errington, J., Ehrlich, S. D., and Noirot, P. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 8342–8347
3. Cook, P. R. (1999) Science 284, 1790–1795
4. Firshein, W. (1989) Annu. Rev. Microbiol. 43, 89–120
5. Meijer, W. J. J., Horcajadas, J. A., and Salas, M. (2001) Microbiol. Mol. Biol. Rev. 65, 261–287
6. Salas, M. (1991) Annu. Rev. Biochem. 60, 39–71
7. Salas, M., Miller, J. T., Leis, J., and DePamphilis, M. L. (1996) in DNA Replication in Eukaryotic Cells (DePamphilis, M. L., ed) pp. 131–176, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
8. Meijer, W. J. J., Lewis, P. J., Errington, J., and Salas, M. (2000) EMBO J. 19, 4182–4190
9. Meijer, W. J. J., Serna-Rico, A., and Salas, M. (2001) Mol. Microbiol. 39, 731–746
10. Serna-Rico, A., Salas, M., and Meijer, W. J. J. (2002) J. Biol. Chem. 277, 6735–6742

FIG. 8. Structural features of p16.7C oligomers. A, ribbon representation of an NMR-derived model for the p16.7C tetramer. Two different views related by a 90° rotation around y-axis are shown. B, model for a larger protein multimer based on the NMR data. The two views are related by a 90° rotation around the y-axis.
Structure of the Functional Domain of φ29 Replication Organizer: INSIGHTS INTO
OLIGOMERIZATION AND DNA BINDING
Juan Luis Asensio, Armando Albert, Daniel Muñoz-Espín, Carlos Gonzalez, Juan
Hermoso, Laurentino Villar, Jesús Jiménez-Barbero, Margarita Salas and Wilfried J. J.
Meijer

J. Biol. Chem. 2005, 280:20730-20739.
doi: 10.1074/jbc.M501687200 originally published online March 16, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M501687200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/03/31/M501687200.DC1

This article cites 35 references, 8 of which can be accessed free at
http://www.jbc.org/content/280/21/20730.full.html#ref-list-1