Phage φ29 DNA Replication Organizer Membrane Protein p16.7 Contains a Coiled Coil and a Dimeric, Homeodomain-related, Functional Domain*

Received for publication, March 24, 2004, and in revised form, September 7, 2004
Published, JBC Papers in Press, September 14, 2004, DOI 10.1074/jbc.M403297200

Daniel Muñoz-Espín†, Mauricio G. Mateu, Laurentino Villar, Anabel Marina, Margarita Salas§, and Wilfried J. J. Meijer‡

From the Instituto de Biología Molecular “Eladio Vinueza” (CSIC) Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM) Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

The Bacillus subtilis phage φ29-encoded membrane protein p16.7 is one of the few proteins known to be involved in prokaryotic membrane-associated DNA replication. Protein p16.7 contains an N-terminal transmembrane domain responsible for membrane localization. A soluble variant lacking the N-terminal membrane anchor, p16.7A, forms dimers in solution, binds to DNA, and has affinity for the φ29 terminal protein. Here we show that the soluble N-terminal half of p16.7A can form a dimeric coiled coil. However, a second domain, located in the C-terminal half of the protein, has been characterized as being the main domain responsible for p16.7 dimerization. This 70-residue C-terminal domain, named p16.7C, also constitutes the functional part of the protein as it binds to DNA and terminal protein. Sequence alignments, secondary structure predictions, and spectroscopic analyses suggest that p16.7C is evolutionarily related to DNA binding homeodomains, present in many eukaryotic transcriptional regulator proteins. Based on the results, a structural model of p16.7 is presented.

In the case of bacteria, replication of the chromosome, plasmids, and infecting phages occurs at the cell membrane (for review see Ref. 1). Besides acting as a scaffold and inherently compartmentalizing DNA replication, membranes are also likely to play an important role in the organization and function of the DNA replication complex (for recent reviews see Refs. 2 and 3). Nevertheless, very little is known about proteins participating in these processes. We started to investigate this fundamental process by using the Bacillus subtilis phage φ29, one of the best studied phages (for review see Ref. 4), as a model system. The genome of φ29 encodes a membrane protein linked at each 5′ end. Initiation of φ29 DNA replication occurs via a so-called protein-primed mechanism (4–6). First, the TP-containing DNA ends are recognized by a φ29 DNA polymerase/TP heterodimer. Then, after a transition step, these two proteins dissociate, and the DNA polymerase continues processive elongation, which is coupled to strand displacement, until replication of the nascent DNA strand is completed. Consequently, the DNA replication intermediates can contain extremely long stretches of ssDNA. Ivarie and Pène (7) provided evidence that early expressed φ29 protein(s) are required for membrane-associated φ29 DNA replication. Gene 16.7, present in an early expressed operon, is conserved in all φ29-related phages studied so far (4). Previously, we have shown (4, 8, 9) that p16.7 (130 amino acids) is involved in the organization of membrane-associated φ29 DNA replication. Protein p16.7 is a membrane protein, and its N-terminal transmembrane domain is responsible for membrane localization (8). Studies using a soluble variant in which the membrane anchor is replaced by a histidine tag, p16.7A, revealed (i) that it is a dimer in solution, (ii) that it has both single-stranded (ss) and dsDNA binding activity, and (iii) that it has affinity for the φ29 TP (8–10).

Here we determine the regions responsible for p16.7A dimerization and identify the functional DNA and TP binding domain. We provide evidence that this functional domain may be evolutionarily related to eukaryotic homeodomains.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—Escherichia coli strains JM109 (11) and BL21(DE3)pLysS (12) and B. subtilis strain 110NA (13) were used for cloning and overexpression of proteins. Chloramphenicol, kanamycin, and ampicillin were added to cultures and plates at final concentrations of 10, 20, and 100 μg/ml, respectively.

DNA Techniques—All DNA manipulations were carried out according to Sambrook et al. (14). [γ-32P]ATP (3000 Ci/mmol) was obtained from Amersham Biosciences. PCRs to amplify the N- and C-terminal 16.7 regions were done using proofreading-proficient Vent DNA polymerase (New England Biolabs, Beverly, MA) as described previously (9). The purified PCR products were digested with NdeI and NotI and cloned into the pET-28b(+) vector (Novagen) digested with the same enzymes resulting in vectors pET-16.7N and pET-16.7C. Site-directed mutants of 16.7A were obtained using the QuikChange™ site-directed mutagenesis kit (Stratagene). Mutant gene 16.7A4 was constructed by stepwise introduction of mutations changing a Leu into an Arg codon.

† This work was supported in part by Research Grant 2R01 GM27242-24 from the National Institutes of Health (to M. S.) and by Grants BIO2003-04445 (to M. G. M.) and BMC2002-03518 (to M. S.) from the Spanish Ministry of Science and Technology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
‡ Recipient of a predoctoral fellowship from “Fondo de Investigaciones Sanitarias.”
§ To whom correspondence should be addressed. Tel.: 34-91-497-8435; Fax: 34-91-497-8490; E-mail: msalas@cbm.uam.es.
¶ Supported by the “Ramón y Cajal” program of the Spanish Ministry of Science and Technology.

The abbreviations used are: dsDNA, double-stranded DNA; TP, terminal protein; DSS, disuccinimidyl suberate; GdmHCl, guanidinium hydrochloride; ssDNA, single-stranded DNA; NTA, nitroltitroacetic acid; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; RP-LC/MS, reverse phase-liquid chromatography coupled to mass spectrometry; IT, ion trap.
pHusing appropriate complementary oligonucleotides and plasmid
PUSH16.7a (8) as starting template DNA.

Overexpression and Purification of p16.7a and Its Derivatives and of TP—Purification of p16.7a and all its derivatives, except p16.7cb, containing N-terminal His tag (see Fig. 1A). Therefore, these proteins were purified from the corresponding E. coli strains in which they were overexpressed using Ni2+-NTA resin columns as described before (8). Thus, overnight cultures of E. coli harboring the appropriate plasmid were diluted 100-fold in fresh prewarmed LB medium and grown to an OD600 of 0.6–0.7. Expression, induced upon isopropyl 1-thio-β-β-nitrogalactopy-
ranoside addition to a final concentration of 1 mM, was allowed for 2 h. Cells were harvested by centrifugation and stored at −70 °C until further use. Frozen cells were thawed and kept at 4 °C with the their weight of aluminia powder (Merck) for 20 min. The slurry was resuspended in buffer A (50 mM Tris-HCl, pH 7.5, 0.5 mM NaCl, 1 mM EDTA, 5% glycerol, and 7 mM β-mercaptoethanol) using 4 volumes per g of cells. To remove the aluminia and intact cells, the mixture was centrifuged at 2500 × g. The pellet was resuspended in 2 volumes of buffer A and centrifuged again as before. The pooled supernatants were allowed to reach chemical and thermal equilibrium. The data were fitted to a folded

Percent helicity of the protein was calculated by using the following equation (17): % helicity = θ222/θ222(1 − kθ222), where θ222 is the molar ellipticity per residue at 222 nm; θ222(i) is the molar ellipticity per residue at 222 nm–helical proteins (i.e., −34,500 deg cm2/dmmol); k is a constant factor (k = 2.57 for λ = 222 nm); and n is the number of peptide bonds in the protein. Chemical denaturation equilibri um experiments were carried out by measuring the ellipticity at 222 nm of p16.7N solutions at 548 μM in 23 mM phosphate buffer, pH 7.5, and 250 mM NaCl, at the indicated concentrations. The temperature was kept constant at 25, 15, or 4 °C, and each sample was allowed to reach chemical and thermal equilibrium. The data were fitted to a folded
dimer-to-unfolded monomer, two-state transition. The equations de-
scribed in Mateu (18) and the program Kaleidagraph (Abelbeck Software) were used.

Size-exclusion Chromatography—A calibrated Superdex 75 HR 10/30 fast protein liquid chromatography column (Amersham Biosciences) was used. The chromatographic process was carried out at constant room temperature (23 °C) in 50 mM Tris-HCl buffer, pH 7.5, 300 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol at 1 ml/min. For an estimation of dissociation constants, analytical gel filtration experi-
ments were carried out as described (19, 20).

Partial Protoplectic Digestion of p16.7a—Protoplectic digestion was performed in a final volume of 25 μl, containing 50 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 10% (v/v) glycerol, 9 μg of purified p16.7a protein, and the indicated amount (ng) of proteinase K. After incubation for 1 min at 30 °C, proteolysis was stopped by addition of 4X loading buffer after which the samples were subjected to SDS-PAGE and Coo-

mase Blue staining.

RESULTS

The p16.7 Region Spanning Residues 21–68 Is Able to Dimerize as a Low Affinity Coiled Coil—Protein p16.7, in which the N-terminal transmembrane domain has been replaced by a histidine tag (Fig. 1A, p16.7a), is a dimer in solution (8). Analy-

sis of the p16.7a sequence revealed that the region spanning amino acid residues −25–60 (residue numbering is referred to the complete p16.7 sequence) has a high probability to form a coiled coil (8). A coiled coil consists of at least two amphipathic α-helices that are wound into a superhelix having a hydropho-

bic interface (26). The protein sequences of coiled coils are characterized by a heptad repeat of amino acids, denoted a–g. Residues at positions “a” and “d,” which are located at the same face of the helix, are predominantly hydrophobic and form the helix-helix interphase. A helical wheel showing the amphipathic character of this p16.7 region is presented in Fig. 1A.

The following reasons made this region of p16.7 the prime candidate to be responsible for p16.7 dimerization. First, coiled-coils are ubiquitous dimerization domains found in a wide range of structural and regulatory proteins. Second, no other region was detected by analyses of the p16.7 protein
sequence that a priori would hint to a possible involvement in dimerization. To test the possibility that the putative coiled-coil region is responsible for p16.7 dimerization, plasmids pET-16.7N and pET-16.7N4 were constructed and used to purify proteins p16.7N and p16.7N4 (Fig. 1A). Protein p16.7N encompasses the wild type sequence of the predicted p16.7 coiled-coil region. Protein p16.7N4 contains the same region but includes four Leu (Leu-36, Leu-39, Leu-50, and Leu-53) to Arg substitutions that would completely disrupt the hydrophobic face of the putative $\alpha$-helix (Fig. 1B) and therefore would be unable to form a coiled-coil-mediated dimer.

To determine their oligomeric state in solution, proteins p16.7N and p16.7N4 were subjected to in vitro cross-linking (Fig. 2A). Whereas a band with a molecular weight corresponding to a dimer was observed in the case of the wild type p16.7N protein after DSS treatment, no cross-linked species were observed for the mutant p16.7N4 protein. These results indicate that, contrary to mutant protein p16.7N4, the wild type protein p16.7N is able to dimerize in solution. The secondary structure of the wild type and mutant proteins was analyzed by far-UV CD spectroscopy. The CD spectrum of p16.7N and p16.7N4 at low and intermediate concentrations (15 and 75 $\mu$M) at 25 °C was characteristic of a random coil (not shown). However, the spectrum of protein p16.7N obtained at high concentration (590 $\mu$M) revealed a substantial helical content at 25 °C, which increased progressively at lower temperatures (Fig. 3A). In sharp contrast, the CD spectrum of the mutant p16.7N4 remained characteristic of a random coil at high concentration (590 $\mu$M), even at 4 °C (Fig. 3A). The estimated helical content of the proteins at the various conditions tested are given in Table I. A maximum helical content of 44% was obtained for p16.7N at 590 $\mu$M and 4 °C. Because p16.7N is 72 amino acids long, about 31 amino acids must be contained in a helical structure under these conditions, which is consistent with the sequence-based prediction of a coiled coil spanning residues 29–57 of p16.7 (see above and Fig. 1).
Circular dichroism analyses of p16.7N, p16.7N4, p16.7A, p16.7A4, and p16.7C. A, far-UV CD spectra of protein p16.7N (solid, dashed, and dotted lines) and p16.7N4 (dot-dashed lines) at high protein concentration (590 μM) and different temperatures: 4 (solid and single dot-dashed lines), 15 (dashed line), and 25 °C (dotted and double dot-dashed lines). Far-UV spectra of p16.7N and p16.7N4 at low (15 μM) or intermediate (75 μM) concentrations at 25 °C were highly similar to those of the presented p16.7N4 spectra (not shown). The spectra, as well as those shown in C, are truncated at a low wavelength where the measurement became unreliable. The background has been subtracted, and the spectra have been normalized. B, GdmHCl-induced dissociation/denaturation of p16.7N at 4 °C and 548 M protein (monomer) concentration followed by ellipticity measurements at 222 nm. The curve was fitted to a native dimer-to-unfolded monomer transition (see "Experimental Procedures"). C, far-UV CD spectra of protein p16.7A (solid line), p16.7A4 (dashed line), and p16.7C (dotted line). The total protein concentration was 15 μM (25 °C).

The oligomerization state of p16.7A (14.5 kDa) and p16.7A4 (14.65 kDa) was confirmed by analytical gel filtration. Both proteins behaved essentially as dimers, even at low concentrations. Their respective apparent molecular masses were 35.5 and 36.3 kDa, close to the calculated values (29.0 and 29.3 kDa, respectively). Analytical gel filtration was also used to estimate the order of magnitude of the dissociation constant. The elution volumes at the extremely low protein concentrations where the monomeric form would predominate could not be determined because of the very low signal-to-noise ratio. Thus, only the final part of the association curve could be traced. Nevertheless, a very good fitting to a dimer-monomer equilibrium was obtained (Fig. 4). Also, the fitting for p16.7A yielded an apparent molecular mass of 13.5 kDa for the monomeric form, which is very close to the calculated value (14.5 kDa). The dissociation constant thus obtained for p16.7A was about 20 nM. The actual value could be even somewhat lower, because of the possible dilution of the sample in small zone elution experiments (20). Most interestingly, the dissociation constant of p16.7A4 was also in the nanomolar range. These results can be summarized as follows. First, the dissociation constant obtained for p16.7A was less than that of p16.7N. Second, similar amounts of dimers were observed for p16.7A and p16.7A4 after DSS cross-linking, indicating that the mutant protein p16.7A4 does form dimers in solution, despite having a disrupted coiled-coil region (see also below).

The equilibrium between an essentially nonstructured state and a partly helical conformation of p16.7N and its dependence on protein concentration indicate that the transition corresponds to a coupled folding-dimerization process. Dimerization of p16.7N is a very low affinity process, however, because the dimeric form is favored only at high protein concentrations and low temperatures. Moreover, the presence of an isodichroic point for the CD spectra obtained at different temperatures is consistent with a two-state transition between unfolded monomer and folded dimer, with no stable intermediates (27, 28).

The dissociation/unfolding equilibrium of p16.7N was quantitated in chemical dissociation/denaturation experiments. Samples of p16.7N (548 μM) were treated with increasing concentrations of GdmHCl, and the ellipticity of the samples at 222 nm was obtained from the corresponding CD spectra (Fig. 3B). Because of the very low dimerization affinity, the equilibrium was studied at 4 °C. The data fitted well a two-state monomer-dimer equilibrium with a free energy difference in the absence of denaturant ΔG°,H2O = 5.7 ± 0.6 kcal/mol in the standard (1 M) state at 4 °C, and an m value (the increase in ΔG° upon increasing the denaturant concentration by 1 M), m = (−2.7) ± 0.7 kcal/mol-M. The value of ΔG°,H2O corresponds to an equilibrium dissociation/unfolding constant Kd = 33 μM at 4 °C.

In summary, the results clearly indicate that an isolated protein containing amino acid residues 21–68 of p16.7 is able to form, albeit with low affinity, a dimeric coiled coil of about 30 residues in length, in a coupled folding and association process. Fig. 3C is described below.

**Table I. α-Helical content of p16.7A and its derivatives**

| Protein  | Concentration | Temperature °C | θ222a (deg cm2 dmol−1) | Helix % |
|----------|---------------|---------------|------------------------|--------|
| p16.7A  | 15            | 25            | −20133                 | 59     |
| p16.7A4 | 15            | 25            | −9174                  | 27     |
| p16.7N  | 75            | 25            | −3688                  | 11     |
| p16.7N4 | 590           | 25            | −4550                  | 13     |
| p16.7C  | 590           | 25            | −11460                 | 34     |
| p16.7C  | 590           | 15            | −14800                 | 44     |
| p16.7C  | 590           | 4             | −3565                  | 10     |
| p16.7C  | 75            | 25            | −4398                  | 13     |
| p16.7C  | 590           | 25            | −3294                  | 10     |
| p16.7C  | 590           | 4             | −3389                  | 10     |
| p16.7C  | 15            | 25            | −13693                 | 40     |

*a* θ222 is the molar ellipticity per residue (in degrees cm² dmol⁻¹).

**Protein p16.7C Forms High Affinity Homodimers**—To study whether the coiled coil of p16.7A is the main dimerization determinant, a mutant 16.7A gene was constructed encoding a protein, p16.7A4, that contains the same four Leu to Arg substitutions present in p16.7N4 (Fig. 1A). In vitro DSS cross-linking analysis of p16.7A and p16.7A4 (Fig. 2B) revealed two interesting features. First, the amount of cross-linked dimeric species was much higher for these proteins as compared with that obtained with p16.7N at the same protein concentration, which suggests that p16.7A and p16.7A4 dimerize much more readily than p16.7N. Second, similar amounts of dimers were obtained for p16.7A and p16.7A4 after DSS cross-linking, indicating that the mutant protein p16.7A4 does form dimers in solution, despite having a disrupted coiled-coil region (see also below).
region other than the coiled coil is responsible for the high p16.7A dimerization affinity.  

To study whether the isolated C-terminal half of p16.7 is able to dimerize, plasmid pET-16.7C was constructed and used to purify protein p16.7C (Fig. 1A). In vitro cross-linking analyses (Fig. 2C) and analytical gel filtration studies (Table II) demonstrated that p16.7C forms dimers in solution. The values obtained for the weight average partition coefficient corresponded very well with that calculated for the dimeric form (Table II) and remained constant over the protein concentration range tested (from 50 down to 0.1 μM). As no significant amount of the monomeric form could be detected at concentrations as low as 100 nM, the dissociation constant cannot be higher than about 10 nM (at 25 °C). This value is in the same order of magnitude as that of p16.7A and p16.7A4 at the same temperature and is more than 3 orders of magnitude lower than the dissociation constant of p16.7N, even though the latter value was determined at a lower temperature (4 °C).

The Coiled Coil Is Formed in the Context of p16.7A—The results described above show that the coiled-coil-mediated dimerization of p16.7N is a very low affinity process. High affinity dimerization of p16.7A through the C-terminal halves of the protein will restrict the mobility and orientation of the two N-terminal regions, which would increase the frequency of productive collisions and thereby shift the equilibrium of the N-terminal regions toward association in a parallel coiled-coil structure. As a consequence, the coiled coil might therefore be formed at lower concentrations and higher temperature in the complete p16.7A protein as compared with that observed in its isolated form present in p16.7N. To test this possibility, the helical content of p16.7A and p16.7A4 was estimated by CD. The calculated helical content of p16.7A and p16.7A4 (at 15 μM and 25 °C) were 59 and 27%, respectively (Fig. 3C and Table 1). The reduction in helical content in p16.7A4 (~39 residues) corresponds roughly to the size of the coiled-coil region that would be disrupted by the mutations introduced. Thus, the N-terminal coiled coil is formed in the context of p16.7A, even at a low concentration and at 25 °C.

The Coiled Coil and the C-terminal Region Are Separated by a Protease-sensitive Linker—The results described above show that the isolated p16.7 regions spanning residues 21–68 (present in p16.7N), and 63–130 (present in p16.7C) contain a low and high affinity dimerization domain, respectively, and indicate that each region constitutes a separate domain. To verify this latter assumption and to study whether a protease-sensitive linker connects these two domains, p16.7A was subjected to partial proteolysis using the nonspecific cleavage site protease, proteinase K (see “Experimental Procedures”). Partial digestion of p16.7A gave two major proteolytic fragments with apparent molecular masses of ~7.5 and ~7 kDa (Fig. 5A), which, for simplicity, will be referred to as proteinase K fragment A and B, respectively. To gain insight in their nature, the tryptic peptide maps of these two proteolytic fragments were compared with the tryptic map of proteinase K-untreated p16.7A. Therefore, the two proteolytic p16.7A fragments as well as the complete p16.7A protein were digested in situ with trypsin, and the resulting peptides were subsequently analyzed by MALDI-TOF-type mass spectrometry. The results of these analyses are schematically presented in Fig. 5B. Comparison of the experimentally obtained peptide masses for p16.7A with the theoretically predicted ones after trypsin digestion of p16.7A allowed us to assign unambiguously eight tryptic peptides, which together covered almost the entire p16.7A protein. Only three regions, ranging between four and eight residues, were not detected. The expected peptide sequence of the tryptic fragments was verified by RP-LC/ESI-IT mass spectrometry (not shown). Most interestingly, high levels of the tryptic fragments 5–7, which allocate to the C-terminal region (down to p16.7 residue 65), were detected in proteinase K fragment B. Reciprocally, high levels of the N-terminally located tryptic fragments 2 and 3 (up to p16.7 residue 60) were detected in proteinase K fragment A. Thus, partial proteolysis of protein p16.7A with proteinase K generated two main products corresponding to the N-terminal region encompassing the coiled-coil region, and the C-terminal region containing the high affinity dimerization domain, demonstrating that these are two separate structural modules that are connected by a short proteinase K-sensitive linker region. The inference that proteinase K cleaved preferentially in the p16.7 region spanning residues 61–65 (corresponding to the protein sequence “SIDK”) is further supported by the observation that fragment 4 was only detected in the proteinase K-untreated p16.7A sample. Finally, these results demonstrate that the proteins p16.7N and p16.7C encompass the complete coiled-coil and C-terminal domain, respectively.

The C-terminal Region Constitutes the Functional Domain of Protein p16.7—Protein p16.7A can bind both ssDNA and dsDNA and has affinity for the φ29-encoded TP (9, 10). We have tested whether these functional properties are specifically associated with the C-terminal 65 residues (Fig. 1A) of p16.7C. First, to exclude the possibility that the positively charged histidine tag of p16.7C (Fig. 1A) might be (partly) responsible for DNA binding or for the interaction with TP, protein p16.7C was cleaved with thrombin, and the resulting purified protein, p16.7Cb (Figs. 1A and 6A), was used. Mass spectrometry analyses of p16.7Cb
confirmed that cleavage of the histidine tag had occurred at the correct place (see “Experimental Procedures” for details).

In vitro DSS cross-linking experiments were performed to determine whether p16.7Cb (8.5 kDa) has affinity for TP (31 kDa). The cross-linked samples were subjected to SDS-PAGE followed by Western blot analysis using polyclonal antibodies against TP (Figs. 1A and 6B). Similar to results obtained with p16.7A (9), besides signals corresponding to TP monomers and dimers, an additional signal was detected when the sample contained both TP and p16.7Cb. This band, which has an apparent molecular weight corresponding to a 16.7CbTP heterodimer (~40 kDa), was also detected when the blot was stripped and re-used with antibodies against p16.7, confirming that it contains both TP and p16.7Cb (not shown). Thus, the C-terminal half of p16.7 is sufficient to interact with TP.

Gel retardation studies were performed to determine whether p16.7Cb has ssDNA and dsDNA binding capacity. Thus, the 297-bp right end fragment of the ϕ29 genome was end-labeled and used directly or after heat denaturation in gel mobility shift assays. The results (Fig. 6, C and D) show that p16.7Cb and p16.7A have similar DNA binding capacities. Together, these results demonstrate that the basic functional properties of p16.7, binding to DNA and to ϕ29 TP, reside in the C-terminal region of the protein.

The Functional Domain of Protein p16.7 Has a Helical Structure and May Be Evolutionarily Related to Eukaryotic Homeodomains—A sequence-based homology search was carried out for the functional domain of p16.7. Most interestingly, the sequences with the highest similarity (around 20 and 40% of identity and similarity, respectively) corresponded to DNA binding homeodomains, which are present in a large family of eukaryotic transcription factors (29, 30). Homeodomains contain about 60 amino acid residues and are composed of three α-helices that are folded into a tight globular structure (31). An alignment between the C-terminal region of protein p16.7 and a consensus sequence based on 346 homeodomain sequences (Fig. 7) shows that various residues highly conserved in most homeodomains, and critical for either structure or function, are also conserved in the functional domain of p16.7. For instance, homeodomains contain four highly conserved residues (Trp-48, Phe-49, Asn-51, and Arg-53) in their DNA recognition helix III. Of these, Arg-53 is conserved in p16.7 (corresponding to Arg-113). In addition, the aromatic homeodomain residue Trp-48 corresponds to the aromatic tyrosine residue 108 in p16.7, and the apolar homeodomain residue Phe-49 corresponds to p16.7 residue isoleucine 109. Moreover, the invariant homeodomain residue Leu-16 is conserved in p16.7 (corresponding to Leu-76). In addition to sequence similarity, p16.7C and homeodomains

Fig. 5. Protein p16.7A contains two structurally separate domains as assessed by partial proteolysis. A, Coomassie-stained SDS-PAGE of proteolytic fragments resulting from digestion of p16.7A (9 µg) for 1 min at 30 °C with the indicated amounts of proteinase K (see “Experimental Procedures”). C, protein standards; 0, undigested protein p16.7A (14.5 kDa). The two major proteolytic fragments of ~7.5 and ~7 kDa are indicated as fragment A and B, respectively. B, schematic overview of MALDI-TOF-type mass spectrometry analysis of the “in gel” trypsin digestion-generated products of intact p16.7A and those of the proteinase K-generated proteolytic fragments A and B. The positions of the theoretical trypsin cleavage sites and those that were actually cleaved by trypsin are indicated with black and red arrows, respectively, below p16.7A, which is depicted as a thick black bar. The eight tryptic fragments (numbered 1–8), detected by analysis of intact trypsin-digested p16.7A and their position with respect to the protein p16.7A sequence, are indicated with thin bars. The experimentally obtained molecular mass of all fragments deviated less than 0.5 Da from the calculated one, and their expected sequence was verified by RP-LC/ESI-IT mass spectrometry (not shown). The following small tryptic fragments were not detected: KKQEAR (located between fragments 1 and 2), VVQR (located between fragments deviated less than 0.5 Da from the calculated one, and their expected sequence was verified by RP-LC/ESI-IT mass spectrometry and their position with respect to the protein p16.7A sequence, are indicated with thin bars. The experimentally obtained molecular mass of all fragments deviated less than 0.5 Da from the calculated one, and their expected sequence was verified by RP-LC/ESI-IT mass spectrometry (not shown). The following small tryptic fragments were not detected: KKQEAR (located between fragments 1 and 2), VVQR (located between fragments 2 and 3), and KLYRGSILK (extreme C-terminal fragment). The tryptic fragments corresponding to the N- and C-terminal part of p16.7A are shown in red and blue, respectively. Although trace amounts of the C-terminal tryptic fragments were detected in the sample containing fragment A and, vice versa, N-terminal tryptic fragments in the fragment B sample, there was at least a 10-fold difference in the relative level of the respective peptides between both samples. Because of their similar molecular weights, minor amounts of one fragment might have contaminated the other one during excision of the gel. Fragment 4, shown in black, was only detected in the intact p16.7A processed sample. The SIKD sequence located between the N-terminal fragment 3 and C-terminal fragment 5 is shown enlarged at the top. The p16.7 coiled-coil (red) and the C-terminal containing region (blue), present in proteins p16.7N and p16.7C, respectively, are indicated above the bar representing p16.7A.
are also similar in their secondary structure, as deduced from the following observations. First, various secondary structure prediction programs invariably predicted three α-helical segments within the C-terminal domain of p16.7, the position and length of which corresponded neatly to those present in homeodomains (Fig. 7). Second, experimental evidence that is consistent with the above predictions was obtained by CD and NMR spectroscopy. The CD spectrum of p16.7C indicates a helical content of about 40% (Fig. 3C and Table I), very similar to that derived from the secondary structure predictions and also similar to that found in homeodomains. In addition, preliminary analyses by NMR spectroscopy of p16.7C confirm the helical content estimated by CD and the predicted positions of the helical segments in the sequence. The homeodomain of the human Pbx1 protein, for which the three-dimensional structure is known, is similar to that of p16.7C in terms of secondary structure and overall conformation. However, certain differences in the amino acid sequence and the specific interactions with target DNA may lead to variations in the functional properties of these proteins.

2 J. L. Aensio, D. Muñoz-Espín, C. Gonzalez, L. Villar, J. Jiménez-Barbero, M. Salas, and W. J. J. Meijer, unpublished results.
The results clearly show that the C-terminal domain is responsible for the high affinity dimerization of p16.7. At present, the residues of the C-terminal domain involved in the dimerization interface are unknown. However, a likely candidate is a solvent-exposed hydrophobic patch formed mainly by residues belonging to helix I in the predicted model of this domain (Fig. 8). Proteolytic analyses demonstrated that the coiled-coil region and the high affinity C-terminal part of p16.7 are two separate structural domains, which are separated by a proteinase K-sensitive linker.

Protein p16.7C is able to interact with ø29 TP, ssDNA, and dsDNA, and thus constitutes the essential functional domain of p16.7. This conclusion is also supported by the observation that, within the ø29 family of phages, the C-terminal half of the p16.7 homologues has a higher level of conservation when compared with the N-terminal half (4).

A combination of experimental results and theoretical predictions allows us to propose that the C-terminal functional domain of p16.7 is evolutionarily related to a typical eukaryotic DNA binding domain, the homeodomain, present in hox proteins (34). This evidence includes the following: (i) a sequence similarity of about 40% with homeodomains (Fig. 7) (no other protein in the data bases scored a higher identity); (ii) the conservation of many residues that are also conserved in homeodomains and that are critical for their structure and/or function; (iii) the predicted and experimentally confirmed helical content and position of the three helices found in all homeodomains; and (iv) the homology-based generation of a structurally sound model for the approximate tertiary structure of the C-terminal domain of p16.7 based on the Pbx1 homeodomain x-ray structure (Fig. 8).

Despite structural similarities there are also differences between p16.7C and homeodomains, which probably are related to their distinct functions. Whereas homeodomain proteins are transcriptional regulators playing critical roles in eukaryotic development and body plan formation, protein p16.7 is involved in membrane-associated organization of phage ø29 DNA replication. Thus, although nearly all homeodomains bind exclusively dsDNA in a sequence-specific way, p16.7C is able to bind both dsDNA and ssDNA without apparent sequence specificity. These differences may be due to the oligomeric state of these proteins. Protein p16.7 functions as a homodimer. On the other hand, homeodomain proteins usually bind dsDNA with high affinity and specificity as a heterodimer consisting of two homeodomain-containing proteins (35). DNA binding of homeodomains involves residues that make nonspecific interactions with the sugar phosphate backbone and other residues that make specific interactions with the bases (34). The fact that homeodomain residues involved in nonspecific DNA binding are preferentially conserved in p16.7C may also explain, at least in part, the lack of sequence specificity of p16.7. Finally, the functional differences between homeodomains and p16.7C could also be related to some difference in the tertiary structure.

Similar to ø29, the genome of the eukaryotic adenoviruses consists of a linear dsDNA having a TP at each DNA 5′ end and replicating via a protein-primed mechanism (for reviews see Refs. 6 and 36). The cellular transcription factor Oct1 plays an important role in initiation of adenovirus DNA replication by recruiting the DNA polymerase-TP heterodimer complex to the origin of replication. This recruitment occurs through specific contacts between the Oct1 homeodomain and the adenovirus TP (37). The fact that the homeodomain-like C-terminal region of p16.7 interacts with the ø29 TP reflects a remarkable structure-function parallelism between Oct1 and p16.7.

In conclusion, the results presented here, together with those obtained previously (8–10), show that dimeric p16.7 is com-

**DISCUSSION**

Here we show that an isolated protein containing the predicted p16.7 coiled-coil region can form a dimeric coiled coil but with a very low affinity. Most interestingly, p16.7C, lacking the coiled-coil region, forms high affinity homodimers. Thus, the main dimerization interface in p16.7A is not located in the N-terminal domain but in the C-terminal domain. However, despite the low dimerization affinity of p16.7N, the coiled coil is formed in p16.7A. Dimerization of p16.7 through the C-terminal region will restrict the mobility and orientation of the two N-terminal segments, thus increasing the frequency of productive collisions and shifting the equilibrium toward association in a parallel coiled-coil structure. In the native protein, p16.7, the membrane anchor domain may restrict even more the relative mobility and orientation of the polypeptide chain and, hence, further facilitate formation of a coiled-coil.

The coiled-coil region, although not the primary dimerization domain, may be structurally relevant in several ways. It may be important to position the functional C-terminal domain at a certain distance of the cell membrane. Alternatively or additionally, the solvent-exposed residues of the coiled coil may be involved in interactions with other proteins.

The low dimerization affinity of the p16.7 coiled-coil domain is probably a consequence of the presence of both Arg-46 residues that in the dimeric parallel coiled coil are predicted to be located on the hydrophobic face of each helix, near each other (Fig. 1B), and which might cause a mutual electrostatic repulsion. Protein p16.7 homologues are found in all ø29-related phages studied so far (4, 8). All of them are predicted to have a modular structure similar to that of p16.7, including a coiled-coil domain that invariably contains a charged residue at the position corresponding to Arg-46 of ø29 p16.7. Thus, it is tempting to propose that the low dimerization affinity of the p16.7 coiled-coil domain may constitute an evolutionarily selected trait.
posed of three clearly distinguishable domains as follows: (i) a 20-amino acid-long N-terminal transmembrane domain that serves as an anchor for the viral replication components to be located at specific sites within the cell; (ii) an intermediate, ∼30-amino acid-long coiled-coil domain, which could serve as a rigid spacer between the membrane and the functional domain, and/or to interact with other protein(s); and (iii) a functional C-terminal domain that may be evolutionarily related to homeodomains, which binds TP and DNA and acts, in addition, as the main determinant for p16.7 dimerization. This modular organization would effectively allow p16.7 to contribute to the organization of the φ29 DNA-replication complex at the membrane of the infected cell.

Acknowledgments—We thank Dr. F. Pazos for initial modeling of p16.7C and Professor C. Alonso and M. A. Fuertes for the use of and help with the spectropolarimeter and the Proteomics Department, Centro de Biología Molecular “Severo Ochoa,” for help with the mass spectrometry analyses. The Centro de Biología Molecular “Severo Ochoa” was the recipient of an institutional grant from Fundación Ramón Areces.

REFERENCES

1. Firshein, W. (1989) Annu. Rev. Microbiol. 43, 89–120
2. Sueoka, N. (1998) Prog. Nucleic Acids Res. Mol. Biol. 59, 35–53
3. Cronan, J. E. (2003) Annu. Rev. Microbiol. 57, 203–234
4. Meier, W. J. J., Horcajadas, J. A., and Salas, M. (2001) Microbiol. Mol. Biol. Rev. 65, 261–287
5. Salas, M. (1993) Annu. Rev. Biochem. 62, 39–71
6. 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
7. Ivarie, R. D., and Pené, J. J. (1973) Virology 52, 351–362
8. Meier, W. J. J., Serna-Rico, A., and Salas, M. (2001) Mol. Microbiol. 39, 731–746
9. Serna-Rico, A., Muñoz-Espin, D., Villar, L., Salas, M., and Meier, W. J. J. (2003) EMBO J. 22, 2297–2306
10. Serna-Rico, A., Salas, M., and Meier, W. J. J. (2002) J. Biol. Chem. 277, 6733–6742
11. Yanish-Perron, C., Vieira, J., and Messing, J. (1985) Gene (Amst.) 33, 103–119
12. Studier, F. W., and Moffat, B. A. (1986) J. Mol. Biol. 189, 113–130
13. Moreno, F., Camacho, A., Vituela, E., and Salas, M. (1974) Virology 62, 1–16
14. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
15. Bravo, A., Hermoso, J. M., and Salas, M. (1994) Gene (Amst.) 148, 107–112
16. Zaballón, A., Lázaro, J. M., Méndez, E., and Salas, M. (1989) Gene (Amst.) 83, 187–195
17. Chen, Y., Yang, J. T., and Chau, K. H. (1974) Biochemistry 13, 3350–3359
18. Mateu, M. G. (2002) J. Mol. Biol. 318, 519–531
19. Darling, P. J., Holt, J. M., and Ackers, G. K. (2000) Biochemistry 39, 11500–11507
20. Valdes, R., Jr., and Ackers, G. K. (1979) Methods Enzymol. 57, 59–71
21. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Anal. Chem. 68, 850–858
22. Roepstorff, P., and Fohlman, J. (1984) Biomed. Mass Spectrom. 22, 87–91
23. Engel, M., Williams, R. W., and Erickson, B. W. (1991) Biochemistry 30, 3161–3169
24. Guex, N., and Peitsch, M. C. (1997) Electrophoresis 18, 2714–2723
25. Peitsch, M. C. (1995) Bio/Technology 13, 658–660
26. LaRonde-LeBlanc, N. A., and Wolberger, C. (2003) Adv. Genet. 48, 1–12
27. Engel, M., Williams, R. W., and Erickson, B. W. (1991) Biochemistry 30, 3161–3169
28. Krylov, D., Mikhailenko, I., and Vinzenz, C. (1994) EMBO J. 13, 2849–2861
29. Gehring, W. J. (1987) Science 236, 1245–1252
30. Kaufman, T. C., Seeger, M. A., and Olsen, G. (1990) Adv. Genet. 27, 309–362
31. Gehring, W. J., Affolter, M., and Burglin, T. (1994) Annu. Rev. Biochem. 63, 3161–3169
32. Piper, D. E., Batchelor, A. H., Chang, C., Cleary, M. L., and Wolberger, C. (1999) Cell 96, 587–597
33. LaBonde-LeBlanc, N. A., and Wolberger, C. (2003) Genes Dev. 17, 2060–2072
34. Burglin, T. (1994) in Guidebook to the Homeobox Genes (Duboule, D., ed) pp. 25–71, Oxford University Press, Oxford
35. Mann, R. S., and Chan, S. K. (1996) Trends Genet. 12, 258–262
36. de Jong, R. N., and van der Vliet, P. C. (1999) Gene (Amst.) 236, 1–12
37. de Jong, R. N., Mysiak, E. M., Meier, L. A., van der Linden, M., and van der Vliet, P. C. (2002) EMBO J. 21, 725–735