Structure of the functional domain of φ29 replication organizer: insights into oligomerization and DNA binding

Juan Luis Asensio1§*, Armando Albert2§, Daniel Muñoz-Espín3§, Carlos Gonzalez4, Juan Hermoso2, Laurentino Villar3, Jesús Jiménez-Barbero5, Margarita Salas3 and Wilfried J.J. Meijer3*

Running title: Structure of the functional dimeric domain of φ29 p16.7 protein

1Departamento de Química Orgánica Biológica, Instituto de Química Orgánica General, CSIC, 28006, Madrid, Spain
2Grupo de Cristalografía Macromolecular y Biología Estructural, Instituto de Química-Física “Rocasolano”, CSIC, 28006, Madrid, Spain
3Instituto de Biología Molecular “Eladio Viñuela” (CSIC), Centro de Biología Molecular “Severo Ochoa” (CSIC-Universidad Autónoma Madrid, 28049 Canto Blanco, Madrid, Spain
4Grupo de RMN, Instituto de Química-Física “Rocasolano”, CSIC, 28006, Madrid, Spain
5Grupo de Biología Estructural de Proteínas, Centro de Investigaciones Biológicas (CSIC), 28040, Madrid, Spain
§ These authors contributed equally to this work.
*: Corresponding authors.
Juan Luis Asensio, Departamento de Química Orgánica Biológica, Instituto de Química Orgánica General, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain. Phone: (+34) 91 561 8806 ext: 329, Fax: (+34) 91 564 4853, email: iqoa110@iqog.csic.es
Wilfried J.J. Meijer, Centro de Biología Molecular “Severo Ochoa”, CSIC-UAM, 28049 Madrid, Spain. Phone (+34) 91 497 8434, Fax (+34) 91 497 4799, email: wmeijer@cbm.uam.es

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 eukaryotic 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 multimerisation 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 multimerisation and DNA binding are coupled.

Despite extensive studies on DNA replication, relatively little is known about its in vivo organisation, which, in prokaryotes, occurs at the cell membrane (1-4). The well-studied B. 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 phage 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 in 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.** *E. coli* strain JM109 (F’ traD36 lacIq (lacZ)M15 proA+B+/c14-(McrA-) (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 for overexpression of proteins. Kanamycin was added to *E. coli* cultures and plates at final concentration of 20 µg/ml.

**DNA techniques.** All DNA manipulations were carried out according to Sambrook *et al.* (15). Restriction enzymes were used as indicated by the suppliers. [γ-32P]ATP (3000 Ci/mmol) was obtained from Amersham International plc. Plasmid DNA was isolated using Wizard® Plus DNA purification kit (Promega, Madison, USA). DNA fragments were isolated from gels using the Qiaquick Gel Extraction Kit (Qiagen Inc., Chatsworth, USA). 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® 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 Ni²⁺-NTA resin column as described before (9,12). Protein p16.7Cb was obtained by digestion of protein p16.7C with the thrombin endoprotease using the Thrombin Cleavage Capture Kit (Novagen, Merck Biosciences, Darmstadt, Germany) as described before (12). To obtain 15N and 13C/15N-labelled p16.7C proteins, cells were grown in U-15N and U-13C/15N and Bio-express medium, respectively (Cambridge Isotope Laboratories, Andover, MA, USA).

**Protein concentration.** The protein (monomer) concentration was determined by UV spectrophotometry, measuring the absorbance at 280 nm. The extinction coefficients ε₂₈₀ of p16.7C, p16.7Cb and pE72Q are 11360 M⁻¹.cm⁻¹ and that of pR98W is 16800 M⁻¹.cm⁻¹.

**Cross-linking assays.** Cross-linking reactions using 10 µM of protein and disuccinimidyl suberate (DSS) 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, analysed by PAGE in the presence of SDS. The proteins were visualised by Coomassie blue staining.

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

**Nuclease digestion assays.** Nuclease digestion assays were performed using end-labelled 297 bp DNA fragments corresponding to the φ29 right end genome. Labelled fragments were either used directly (dsDNA) or after heat-denaturation (ssDNA). Micrococcal nuclease digestions of the ssDNA nucleoprotein complexes were performed as described before (11). After cross-linking, proteins were precipitated upon the addition of 1 volume of ice-cold 20% (w/v) trichloroacetic acid and, after resuspension, analysed by PAGE in the presence of SDS. The proteins were visualised by Coomassie blue staining.
with ethanol in the presence of 15 µg linear polyacrylamide as carrier. Next, the resuspended DNA was analyzed in denaturing 6% polyacrylamide gels. Finally, gels were dried and subjected to autoradiography.

Analytical Ultracentrifugation experiments – Sedimentation Equilibrium. The experiments were performed using a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics, using an An50Ti rotor. Whole-cell buoyant molecular masses (M_w,a) were determined by fitting a sedimentation equilibrium model for a single sedimenting solute to individual datasets with the program EQASSOC (17) (supplied by Beckman).

NMR experiments. 1H-NMR spectra were recorded in 85:15 1H2O:2H2O on Bruker Avance 800, Bruker AMX-600 and Varian Unity 500 MHz spectrometers. Protein concentrations for the structural analysis of the protein dimer were in the 400-800 µM range in 200 mM NaCl, 10 mM sodium phosphate buffer and 4 mM DTT at pH 5.0. Protein assignments were obtained using a set of 2D- and 3D- homo- and heteronuclear NMR experiments performed on the unlabelled, 15N-labelled and 13C/15N double-labelled molecules. Thus, HNCA, HNCO and HN(CO)CA spectra were employed for backbone assignment. The side-chain assignments were completed with 3D HCCH-TOCSY experiments. NOE distance restraints were obtained from 15N- or 13C- edited 3D-NOESY spectra. In addition, 2D-NOESY, 13C-selected-12C-filtered experiments were performed on a heterolabelled dimer to analyze the interprotein contacts. This protein sample was generated by mixing equimolar amounts of unlabelled and 13C/15N double-labelled protein to a global protein concentration of 100 µM in 200 mM NaCl, 1mM DTT, 10 mM phosphate buffer and pH 5.0. The sample was incubated at 40 °C for 24h and then concentrated for the NMR analysis.

Upper limits for proton-proton distances were obtained from NOESY cross-peak intensities at three mixing times, 50, 75, and 150 ms. Cross peaks were classified as strong, medium, and weak corresponding to upper limits of 2.5, 3.5, and 5.0 angstroms. The lower limit for proton/proton distances was set as the sum of the van der Waals radii of the protons. Structure calculations were performed using the program DYANA (18). A set of 2180 constraints (204 interprotein) were used in the final round of calculations. The thirty best DYANA structures in terms of target function were submitted to a simulated annealing protocol with the AMBER 5.0 package with the Cornell et al. force field (19).

Protein oligomerization was detected employing DOSY (20) experiments. The average diffusion coefficients of the protein were determined at two different protein concentrations (350 µM and 3500 µM) and two different pH values (pH 5.0 and pH 7.0) in 200 mM NaCl, 10 mM sodium phosphate buffer. In order to get structural information about the protein oligomers 2D-NOESY experiments were also carried out with these samples. In addition, 13C-HSQC and 15N-HSQC spectra were recorded at high and low protein concentration. The high protein concentration sample was prepared by mixing unlabelled with 13C/15N double-labelled protein (4:1 ratio) to a final protein concentration of 3500 µM. Finally 2D-NOESY, 13C-selected-12C-filtered experiments were performed on this sample to analyze the interprotein contacts responsible for oligomerization.

Protein crystallization and data collection. Crystallography experiments were carried out with a p16.7C solution at 10 mg/ml containing 50 mM phosphate buffer pH 7.5, 200 mM NaCl. Protein crystallization was achieved by dialysis of the protein solution against a solution containing 5 mM phosphate buffer pH 7.5, 75 mM NaCl. Prismatic crystals (0.05x0.05x0.2 mm) appear after two or three days of incubation at 40 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 CCD detector using the ESRF Grenoble synchrotron radiation source at wavelength 0.92 at the BM16 beam-line. Diffraction data were processed using MOSFLM (21) (Table I).

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 a density average protocol performed with DM (23). The averaged density map was good enough
to manually refine the NMR model from residues 66a to 86b. This model was then refined 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 (TLS) refinement (28). The stereochemistry of the model was 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 LIGPLOT package (31).

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 in 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'-labelled 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 Figs. 1b and c). In the absence of p16.7Cb nearly all the end-labelled ssDNA fragment was degraded into small oligonucleotides (Fig. 1b, lane 6). In the presence of p16.7Cb, in the range of 5 till 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 4-5) 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 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 ss 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 ºC), 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 (supplementary Table I) was obtained using standard 2D and 3D NMR techniques on 15N and 13C/15N labelled samples. In order to distinguish intra- from intermolecular NOEs, half-filtered NOESY experiments (Fig. 2a) were carried out on a heterolabelled dimer obtained by mixing equivalent amounts of unlabelled and 13C/15N labelled protein. A summary of the experimental constraints employed and the characterisation 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 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 Materials and Methods). The RMSD values for superimposition show that the NMR and crystal structures of p16.7C dimer are very similar (0.87 and 1.04 Å for the monomer and dimer backbone superimposition respectively in ordered regions; i.e. residues 68-125). The combined NMR and X-ray 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 Figures 2b and 2c, 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 (82-87; loop I), are oriented in an antiparallel fashion with a crossing angle of ~160°. Helix H3 is connected to H2 by a seven-residue loop (96-102; loop II) and packs against both helix H1 and H2 with crossing angles of 54° and 139°, respectively. The C-terminal region (residues 122-128) adopts an extended structure. Finally, residues 63-66 and 129-130 are disordered (supplementary Fig. 1b).

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 Brookhaven Protein Data Bank exhibits high structural homology with p16.7C according to DALI (32) and SCOP (33). The two monomers, related by a non-crystallographic two-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 L124 of each monomer, which is totally buried in a hydrophobic patch formed by Y79, L78, and L106 of the other protein monomer. In addition, there are polar intermolecular contacts between R113/K122, S88a/S88b and R85/Y115. 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 W116 indol ring of both polypeptides (W116a and W116b, supplementary Fig. 2). 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. Figure 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 (K122a/K122b, K123a/K123b, R126a/R126b and K130a/K130b). 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 p16.7 contains a region(s) responsible for p16.7C dimer-dimer interaction. p16.7C dimers form a fibre around a crystallographic three-fold screw axis in the crystals obtained (Fig. 4a). This structural organisation 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 neighbouring p16.7C dimers in the crystal (see Fig. 4b) involve two salt bridges (E72a/R98b and the symmetry related pair) and six hydrogen bonds (R98a/V95b, N67a/N96b, C71a/Q100b and the symmetry related pairs) and cover 881 Å$^2$ of the dimer surface (11% of the total solvent accessible area). Interestingly, the E72 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 E72 confers extra stability to its interaction with R98. 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 are involved in oligomerization 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 dimerisation, these mutants were subjected in parallel with p16.7C to in vitro cross-linking analysis using the bifunctional cross-linking agent disuccinimidyld suberate (DSS). The results, presented in Figure 5a, show that similar amounts of cross-linked dimers were obtained for each protein demonstrating that the mutations do not have major effects on dimerisation. 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 DSS 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 E72 and R98 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 concentration 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 ($\Delta \log D = 0.17$ 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 N$^{15}$/C$^{13}$ 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 N83, Y108, Y115 and Q112 were extremely affected by line broadening and almost disappeared (some examples are shown in supplementary Figs. 3 and 4b). Moreover, chemical shift changes (> 0.025 ppm) between the high and low concentration samples were measured for some aliphatic proton signals in residues N83, R85, Q97, Y108, Q112, Y115, E119, K123 and Y125 (some examples are shown in supplementary Fig. 4a and b). Interestingly, they are located at the same extended surface of the protein dimer (see Fig. 6a and supplementary Fig. 5). 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 supplementary Fig. 6). All these contacts (shown in Fig. 6b and listed in supplementary Table II) were in agreement with the pattern of concentration dependent line broadening and chemical shift changes described above. For example, the Y125 aromatic side-chain shows clear NOEs with the methyl group of T111. 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 N83/Y115, L99/Y125, and Y125/Y108. 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 which involves larger surface areas.

Most interestingly, the addition of dsDNA (A14T14) to a low p16.7C concentration sample (350 µM) at pH 7.0 produced remarkably similar changes on the 15N.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. Figure 7 shows a comparison between two regions of a p16.7C 15N-HSQC spectrum under three different conditions; low (middle frames) and high (lower frames) protein concentration without dsDNA and low protein concentration in the presence of stochiometric amounts of dsDNA (upper frames). It can be observed that the NH-signals that are most affected by increasing the p16.7C concentration (as Y115, Y108, Q112 or N83 side-chain) are also the ones most altered at low protein concentration in the presence of DNA. In contrast, several residues involved in the interdimer contacts observed in the crystallographic fibre, as the side chain of Q100 or N67, 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 supplementary Fig. 7). 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 Y125 presents hydrophobic contacts with Y108 and the methyl group of T111, (ii) helix H3 and loop 1 would participate in a number of polar interprotein interactions involving residues N83, R85, E119 and Y115, 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 to 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 (around 20 and 40% of identity and similarity, 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 structures of p16.7C and homeodomains are very similar. Moreover, as in homeodomains, 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 organisation of the 3 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 with 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 towards the center of the cell, and thus is compatible with a role in organizing φ29 DNA replication at the membrane of the infected cell.

Dimers of p16.7C are arranged forming a protein fibre 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 fibre formation, a possible biological function for these contacts can not 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 organisation 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 (supplementary Fig. 8), 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. And 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 also 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.

**Accession Numbers**

The coordinates corresponding to the NMR and X-ray structures of p16.7C dimer have been deposited in the PDB (XXXX and XXXX codes, respectively).

**Acknowledgements**

This investigation was supported by research grants of the Spanish "Plan Nacional" (MCYT) CTQ2004-04994/BQU, BQU2001-3693-C02-02 and BMC2002-04011-C05-03 to J.L.A, C.G. and A.A. respectively, 2R01 GM27242-24 from the National Institutes of Health and BMC2002-03818 from the Spanish Ministry of Science and Technology to M.S. and an Institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular “Severo Ochoa”. D.M. is holder of a predoctoral fellowship from “Fondo de Investigación Sanitaria”. The “Ramón y Cajal” program of the Spanish Ministry of Science and Technology supported W.J.M. A. A. thanks the ESRF (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. USA* **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) (M.L.DePamphilis, ed), pp. 131-176, Cold Spring Harbor Laboratory press, New York
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, 6733-6742
11. Serna-Rico, A., Muñoz-Espín, D., Villar, L., Salas, M., and Meijer, W. J. J. (2003) *EMBO J.* 22, 2297-2306
12. Muñoz-Espín, D., Mateu, M. G., Villar, L., Marina, A., Salas, M., and Meijer, W. J. J. (2004) *J. Biol. Chem.* 1, 1-10
13. Yanish-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* 33, 103-119
14. Studier, F. W., and Moffat, B. A. (1986) *J. Mol. Biol.* 189, 113-130
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
16. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467
17. Minton, A. P. (1994) *Modern analytical centrifugation*, 81-93
18. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J Mol Biol.* 273, 283-298
19. Cornell, W. D., Cieplack, P. C., Bayly, I., Gould, I. R., Merz, K., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc.* 117, 5179
20. Johnson, C. S. (1999) *Prog. NMR spectrosc.* 34, 203-256
21. Leslie, A. G. W. (1987) in *Proceedings of the CCP4 Study Weekend* (Machin, J. R., and Papiz, M. Z., eds), pp. 39-50, SERC Daresbury Laboratory, Warrington, UK
22. Navaza, J. (1994) *Acta Cryst. A* 50, 157-163
23. Cowtan, K., and Main, P. (1998) *Acta Cryst. D* 54, 487-493
24. Brunger, A. T., and Adams, P. D. (1998) *Acta Cryst. D* 54, 905-921
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Cryst. D* 53, 240-255
26. Jones, T. A., Zon, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Cryst. A* 47, 110-119
27. Bailey, S. (1994) *Acta Cryst. D* 50, 760-763
28. Merritt, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr. D* 50, 869-873
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283-291
30. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946-950
31. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* 8, 127-134
32. Holm, L., and Sander, C. (1998) *Nucl. Acids Res.* 26, 316-319
33. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol.* 247, 536-540
34. Neidigh, J. W., Fesinmeyer, R. M., and Chothia, C. (1995) *J. Mol. Biol.* 107, 145-158
35. Hofrichter, J., and Richmond, T. J. (1995) *Nat. Struct. Biol.* 9, 425-430
36. Record, M. T., Lohman, M. L., and De Haseth, P. (1976) *J. Mol. Biol.* 107, 145-158
37. Kalodimos, C. G., Biris, N., Bonvin, A. M. J. J., Levadoski, M. M., Guennuegues, M., Boelens, R., and Kaptain, R. (2004) *Science* 305, 386-389
38. Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem.* 63, 487-526
39. Pellegrini, L., Tan, S., and Richmond, T. J. (1995) *Nature* 376, 490
40. Santelli, E., and Richmond, T. J. (2000) *J. Mol. Biol.* 297, 437
Figure legends

Figure 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-c) Nuclease treatment of nucleoprotein complexes formed by p16.7Cb. The end-labelled 297 bp right end φ29 DNA fragment was used directly (c) or after heat denaturation (b) in nuclease digestion assays. The labelled probes were preincubated either in the absence or presence of increasing amounts of p16.7Cb. After complex formation, samples were treated with micrococcal nuclease (b) or DNase I (c) as described in Materials and Methods, and the DNA products fractionated through polyacrylamide gels under denaturing conditions. The digestion pattern observed in Fig. 1b is the consequence of preferential cleavage by micrococal nuclease at AT-rich regions. 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 DSS. After cross-linking, the sample was subjected to SDS-PAGE and Western blot analyses using polyclonal antibodies against p16.7.

Figure 2.- Solution and crystal structure of p16.7C.  a) Double-filtered (left) and half-filtered (right) NOESY experiments corresponding to p16.7C at 35 °C, pH 5.0, 200 mM NaCl and 10 mM sodium phosphate. Interprotein NOEs involving the methyl group of L124 are clearly observable in the half-filtered spectrum. b) Backbone atoms (N,CA,C) of 25 superimposed NMR-derived structures of p16.7C dimer. Two different views corresponding to a 90° rotation around y axis are shown. c) Stereo ribbon representation of p16.7C crystal structure.

Figure 3.- Location of the putative p16.7C DNA-binding site. Molecular surface representations of p16.7C showing the electrostatic potential (blue is positive and red is negative). Both views (a and b) are related by a 90° rotation around x axis. The probable orientation of the coiled-coil and transmembrane domains in full length p16.7 is schematically represented (not at scale).

Figure 4.- p16.7C forms a protein fibre in the crystals obtained.  a). A representation of the p16.7C fibre structure observed in the crystallographic study. The three fold screw axis relating p16.7C dimers is displayed as a black line. b). A detail of the interdimer interface within the fibre. Interatomic distances below 3.2 Å are joined with a dashed line.

Figure 5.- Protein p16.7C mutants E72Q and R98W are not affected in their dimerization or multimerization abilities. In vitro DSS cross-linking was carried out in the absence (a) or presence (b) of dsDNA as described in Materials and Methods. After crosslinking, the samples were either subjected to SDS-PAGE and Coomassie blue staining (a) or to Western blot analyses using polyclonal antibodies against p16.7 (b). The calculated molecular mass of p16.7C and the mutant pE72Q is 10.3 kDa and that of mutant pR98W is 10.4 kDa. The positions of the denatured monomers and cross-linked dimeric and multimeric species are indicated.

Figure 6.- NMR analysis of the multimerization interface in solution.  a) The location of those residues significantly affected by line-broadening and/or chemical shift changes (>0.025 ppm),
on the p16.7C surface, is shown (in green). b) Schematic representation of the main interprotein contacts between p16.7C dimers deduced from the analysis of 2D-NMR spectra at high protein concentration.

**Figure 7.** – DNA binding induces the same oligomerization process detected at high p16.7C concentration in the absence of DNA. Comparison between two regions of a p16.7C $^{15}$N-HSQC spectrum under three different conditions; 350 μM protein concentration in the presence of stochiometric amounts of dsDNA (upper panels) and 350 and 3500 μM protein concentration without dsDNA (middle and lower panels, respectively). The most affected residues are shown in red. Residues involved in interdimeric contacts observed in the X-ray protein fibre are shown in blue.

**Figure 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 x 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 x axis.
Table I.- Characterisation of the 25 NMR structures of the p16.7C dimer retained for structural analysis.

| Distance constraints                      | Number  |
|-------------------------------------------|---------|
| NOE-derived upper distance limits (Å)     | 2180    |
| Intraresidue                              | 530     |
| Short-Range                               | 466     |
| Medium-Range                              | 670     |
| Long-Range                                | 514     |
| Intramonomer                              | 1976    |
| Intermonomer                              | 204     |

R.m.s deviation from ideal geometry

- Bonds: 0.0086
- Angle: 2.16

Ramachandran Analysis (%)

- Residues in most favourable regions: 81.9
- Residues in additionally allowed regions: 18.1

R.m.s deviation between pairs of structures (Å)

- Backbone (68-125): 0.77/0.96
- Heavy Atoms (68-125): 1.46/1.56

The average maximum violation for the upper distance constraints was 0.27 Å (average number 4.1). For the R.m.s deviations between pair of structures the two values correspond to superimposition of the monomeric (residues 68 to 125) and dimeric (residues 68a to 125a and 68b to 125b) structures.
Table II.- X-Ray data collection and refinement statistics.

Data collection

|                      |                  |
|----------------------|------------------|
| Space group          | P6₁              |
| Unit cell (Å)        | a=95.40 c=47.57  |
| Resolution limits (Å)| 20-2.9           |
| I/σ(I)               | 12.3 (3.2)       |
| R<sub>merge</sub> (%)| 4.9 (16.7)       |
| Completeness (%)     | 99.1 (99.8)      |
| Multiplicity         | 7.2 (5.2)        |
| Temp (K)             | 120              |

Refinement

|                      |                  |
|----------------------|------------------|
| Resolution range (Å) | 20-2.9           |
| R factor (%)         | 23.9 (32.0)      |
| R free (%)           | 26.5 (39.0)      |
| Number of atoms      | 1094             |
| Number of reflections| 5257             |

Model

|                               |                  |
|-------------------------------|------------------|
| Molecules per a.u / Solvent content (%) | 2 / 65          |
| Protein/water atoms          | 1064/19          |
| RMS deviation from ideal bond lengths/angles (Å/deg)) | 0.014/1.64 |
| Observations                  | No electron density for residues from 61 to 66 |

*<sub>R</sub><sub>merge</sub> = ΣₕₖₗΣᵢ|Iₕₖᵢ - Iₕₖ|/ΣₕₖᵢIₕₖᵢ . Values in parentheses are for the highest resolution shell: 2.97 – 2.9 Å
Figure 1: Asensio et al.
Figure 2: Asensio et al.
Figure-3 Asensio et al.
Figure 5: Asensio et al.
Figure-7 Asensio et al.
Supplemental Material

**Figure 1.** a) Ribbon representation of the crystal structure of p16.7C. Residues are coloured according to their respective average atomic temperature factor (red and blue colours mean high and low temperature factors, respectively). b) Backbone $^{1}H$-^{15}N heteronuclear NOEs of p16.7C (lower values correspond to highly mobile regions).

**Figure 2.** Ensemble of 25 NMR structures showing the stacking interaction between P87a/P87b and the W116a/W116b indol rings at the dimerization interface. Both views are related by a 90° rotation around x axis.

**Figure 3.** Regions from $^{13}$C-HSQC (a) and $^{15}$N-HSQC (b) spectra at 350 µM (left) and 3500 µM (right) protein concentration and pH 5.0. Due to the extreme line broadening associated to the protein oligomerization some signals (highlighted in red) disappear.

**Figure 4.** Two different regions (a and b) of 2D-NOESY experiments corresponding to p16.7C at 3500 µM (upper panels) and 350 µM (lower panels). The line broadening and/or chemical shift changes for some proton signals in residues 83, 108 and 112 are indicated.

**Figure 5.** a) The location of those residues significantly affected by line-broadening and/or chemical shift changes (>0.025 ppm), on the p16.7C surface, is highlighted in green. b) Ribbon representation of p16.7C with the same orientation as that shown in a.

**Figure 6.** Two different regions (a and b) of 2D-NOESY experiments at 350 µM (left panels) and 3500 µM (middle panels) p16.7C concentration. The right panels show the same regions (a and b) in a half-filtered NOESY experiment performed at high protein concentration. Interprotein contacts non-compatible with the structure of the dimer are shown in red. Interprotein contacts compatible with the structure of the dimer are shown in blue.

**Figure 7.** a) Schematic representation of the main interprotein contacts between p16.7C dimers deduced from the analysis of 2D-NMR spectra at high protein concentration. b) NMR-derived model for the p16.7C tetramer. An ensemble of 15 structures is shown. It is clear from these data that 30 experimental constraints is enough to define the relative orientation of the two p16.7C dimers within the supramolecular complex.

**Figure 8.** Proposed model for p16.7C interaction with dsDNA.
Table I. $^1$H, $^{15}$N and $^{13}$C assignments of p16.7C at 308 K and pH 5.0. TSP was used as internal reference.

| Residue | NH | N | Hα | Cα | Hβ | Hγ | Others |
|---------|----|---|-----|-----|-----|-----|--------|
| His 61 | 8.60 | 119.99 | 4.69 | 55.55 | 3.30 | 3.18 | - |
| Met 62 | 8.31 | 120.78 | 4.44 | 55.53 | 2.05 | 1.96 | 2.54 | 2.47 |
| Asp 63 | 8.28 | 121.29 | 4.58 | 54.31 | 2.74 | 2.66 | - |
| Lys 64 | 8.23 | 120.97 | 4.36 | 56.39 | 1.89 | 1.78 | 1.44 | 1.43 | Hδ 1.66/1.65, Hε 3.00/2.99 |
| Thr 65 | 8.20 | 113.94 | 4.29 | 62.35 | 4.19 | - | 1.17 | - |
| Val 66 | 8.73 | 120.81 | 4.14 | 61.87 | 2.04 | - | 0.89 | 0.87 |
| Asn 67 | 8.36 | 121.96 | 4.81 | 52.77 | 4.19 | 3.18 | - |
| Leu 68 | 7.93 | 122.21 | 4.59 | 54.08 | 1.67 | 1.58 | 1.57 | - |
| Ser 69 | 8.46 | 118.71 | 4.55 | 56.79 | 4.32 | 3.95 | - |
| Ala 70 | 8.88 | 123.45 | 4.11 | 55.70 | 1.51 | - | - |
| Cys 71 | 8.46 | 115.12 | 4.42 | 60.57 | 2.87 | 2.75 | - |
| Glu 72 | 7.38 | 119.85 | 4.05 | 59.30 | 2.59 | 2.55 | 1.85 | 1.80 |
| Val 73 | 8.46 | 118.91 | 3.49 | 67.05 | 2.06 | - | 1.01 | 0.93 |
| Ala 74 | 7.87 | 119.88 | 4.16 | 54.95 | 1.50 | - | - |
| Val 75 | 7.45 | 117.34 | 3.77 | 66.38 | 2.16 | - | 1.28 | 1.16 |
| Leu 76 | 8.52 | 121.19 | 4.30 | 59.05 | 1.99 | 1.84 | 1.59 | - |
| Asp 77 | 8.60 | 120.06 | 4.58 | 57.19 | 2.98 | 2.75 | - |
| Leu 78 | 7.69 | 120.79 | 4.07 | 58.06 | 1.98 | 1.56 | 1.85 | - |
| Tyr 79 | 7.66 | 118.89 | 3.90 | 64.06 | 3.22 | 3.03 | - |
| Glu 80 | 8.05 | 120.75 | 4.17 | 59.96 | 2.32 | 2.32 | 2.50 | 2.41 |
| Gln 81 | 8.53 | 119.38 | 4.13 | 58.31 | 2.25 | 2.18 | 2.59 | 2.52 |
| Ser 82 | 7.38 | 113.38 | 4.57 | 58.53 | 4.17 | 3.89 | - |
| Asn 83 | 8.16 | 116.69 | 4.35 | 54.08 | 3.19 | 2.75 | - |
| Ile 84 | - | 118.96 | 4.21 | 61.08 | 1.81 | - | 0.89 | 1.61/1.60 |
| Arg 85 | 9.63 | 129.08 | 4.33 | 55.53 | 1.33 | 1.77 | 1.59 | 1.56 |
| Ile 86 | 8.40 | 125.74 | 4.66 | 54.55 | 1.92 | - | 0.78 | 1.24/1.02 |
| Pro 87 | - | - | 3.90 | 62.53 | 1.29 | 1.14 | 1.68 | 1.65 |
| Ser 88 | 10.02 | 122.81 | 3.93 | 62.43 | 3.77 | 3.70 | - |
| Asp 89 | 8.11 | 116.84 | 4.22 | 56.82 | 2.83 | 1.91 | - |
| Ile 90 | 7.06 | 114.23 | 4.04 | 64.94 | 1.90 | - | 0.87 | 1.66/1.31 |
| Ile 91 | 6.85 | 119.14 | 3.53 | 64.57 | 2.02 | - | 0.89 | 1.56/1.12 |
| Glu 92 | 8.08 | 116.33 | 3.94 | 59.07 | 2.06 | 2.01 | 2.40 | 2.32 |
| Asp 93 | 7.51 | 117.83 | 4.67 | 56.55 | 3.14 | 2.94 | - |
| Leu 94 | 8.56 | 121.61 | 3.74 | 58.24 | 1.91 | 1.89 | 1.81 | - |
| Val 95 | 7.58 | 114.79 | 3.92 | 64.83 | 2.24 | - | 1.03 | 0.99 |
| Asn 96 | 7.37 | 116.26 | 4.77 | 53.79 | 2.95 | 2.78 | - |
| Gln 97 | 7.60 | 117.34 | 4.17 | 54.66 | 1.86 | 1.67 | 2.34 | 1.27 |
| Arg 98 | 8.17 | 119.44 | 4.33 | 55.67 | 1.81 | 1.74 | 1.58 | 1.52 |
| Leu 99 | 8.18 | 123.45 | 4.53 | 54.27 | 1.74 | 1.24 | 1.38 | - |
| Gln 100 | 8.77 | 117.56 | 4.45 | 56.45 | 2.31 | 2.03 | 2.41 | 2.33 |
| Ser 101 | 7.52 | 112.27 | 4.92 | 56.46 | 4.29 | 3.95 | - |
| Glu 102 | 9.49 | 122.71 | 3.75 | 60.50 | 2.08 | 2.00 | 2.54 | 2.46 |
| Gln 103 | 8.72 | 117.53 | 3.95 | 59.32 | 2.18 | 2.01 | 2.46 | 2.38 |
| Glu 104 | 7.68 | 117.52 | 4.16 | 59.28 | 2.26 | 2.18 | 2.46 | 2.39 |
| Val 105 | 8.10 | 120.91 | 3.74 | 66.93 | 2.38 | - | 1.16 | 1.06 |
| Leu 106 | 8.72 | 121.08 | 4.14 | 59.13 | 1.81 | 1.80 | 1.71 | - |
| Asn 107 | 8.30 | 115.71 | 4.42 | 56.25 | 2.94 | 2.93 | - |
| Tyr 108 | 8.21 | 121.72 | 4.19 | 62.60 | 3.24 | 3.12 | - | - | Hδ 7.01, Hγ 6.62 |
| Residue | NH | N | Hα | Cα | Hβ | Hy | Others |
|---------|----|---|----|----|----|----|--------|
| Ile 109 | 9.02 | 120.14 | 3.32 | 67.18 | 2.27 | - | 0.72 | 2.29/2.29 | H° 0.81 |
| Glu 110 | 9.09 | 118.38 | 4.14 | 59.58 | 2.36 | 2.16 | 2.66 | 2.65 | - |
| Thr 111 | 8.06 | 116.41 | 4.08 | 66.95 | 4.30 | - | 1.25 | - | - |
| Gln 112 | 7.93 | 121.58 | 4.16 | 57.41 | 2.01 | 1.82 | 2.40 | 2.28 | H° 6.58/6.48 N° 112.64 |
| Arg 113 | 9.37 | 123.45 | 4.09 | 61.13 | 2.41 | 1.93 | 2.13 | 1.40 | H° 3.35/2.87 |
| Thr 114 | 7.93 | 117.05 | 3.93 | 67.44 | 4.36 | - | 1.27 | - | - |
| Tyr 115 | 7.60 | 121.30 | 3.98 | 61.89 | 3.07 | 2.97 | - | - | H° 6.18 H° 5.65 |
| Trp 116 | 8.66 | 118.91 | 4.27 | 59.87 | 3.21 | 3.13 | - | - | H° 7.43 H° 6.78 H° 6.90 N° 130.23 |
| Lys 117 | 9.08 | 125.98 | 4.02 | 61.28 | 2.28 | 1.92 | 1.70 | 1.36 | H° 1.85/1.56 H° 3.02/2.90 |
| Leu 118 | 7.70 | 116.90 | 3.97 | 57.45 | 1.89 | 1.35 | 1.85 | - | H° 0.87/0.84 |
| Glu 119 | 8.38 | 122.17 | 3.74 | 59.51 | 1.89 | 1.44 | 2.11 | 1.91 | - |
| Asn 120 | 7.92 | 112.33 | 4.67 | 55.54 | 3.52 | 2.61 | - | - | H° 6.70/6.30 N° 111.12 |
| Gln 121 | 7.51 | 116.16 | 4.34 | 55.41 | 2.25 | 2.13 | 2.59 | 2.48 | H° 7.37/6.89 N° 110.72 |
| Lys 122 | 7.05 | 120.81 | 4.05 | 58.26 | 1.91 | 1.72 | 1.83 | 1.37 | H° 1.74/1.73 H° 2.95/2.94 |
| Lys 123 | 8.56 | 122.40 | 4.42 | 55.26 | 1.70 | 1.70 | 1.43 | 1.29 | H° 1.69/1.68 H° 2.99/2.99 |
| Leu 124 | 8.61 | 126.44 | 4.05 | 54.80 | 1.45 | 1.20 | 1.02 | - | H° 0.33/0.31 |
| Tyr 125 | 8.30 | 125.43 | 4.58 | 57.81 | 2.99 | 2.96 | - | - | H° 7.17 H° 6.84 |
| Arg 126 | 8.16 | 123.82 | 4.19 | 55.76 | 1.80 | 1.61 | 1.56 | 1.44 | H° 3.11/3.10 H° 7.17 |
| Gly 127 | 7.41 | 108.11 | 3.84/3.83 | 45.32 | - | - | - | - | - |
| Ser 128 | 8.16 | 115.80 | 4.44 | 58.28 | 3.89 | 3.86 | - | - | - |
| Leu 129 | 8.12 | 124.25 | 4.35 | 55.19 | 1.65 | 1.64 | 1.63 | - | H° 0.88/0.83 |
| Lys 130 | 7.88 | 126.41 | 4.14 | 57.39 | 1.81 | 1.68 | 1.39 | 1.38 | H° 1.68/1.67 H° 3.00/3.00 |
Table II.- Intermolecular NOEs observed at 3500 µM protein concentration representative of p16.7C oligomeric species.

| Intermolecular NOE          |
|-----------------------------|
| N83 Hα / Y115 Hζ           |
| N83 Hβ2 / Y115 Hζ          |
| N83 Hβ3 / Y115 Hζ          |
| N83 Hδ21 / Y115 Hζ         |
| N83 Hδ22 / Y115 Hζ         |
| Y125 Hδ1 / Y108 Hε1        |
| Y125 Hε1 / Y108 Hδ1        |
| Y125 Hδ1 / T111 Meγ2       |
| Y125 Hε1 / T111 Meγ2       |
| Y125 Hα / T111 Meγ2        |
| Y125 Hβ2 / T111 Meγ2       |
| Y125 Hβ3 / T111 Meγ2       |
| Y125 HN / T111 Meγ2        |
| Y125 Hδ1 / L99 Meδ2        |
| Y125 Hε1 / L99 Meδ2        |
**Supplemental Material**

**Figure 1.** a) Ribbon representation of the crystal structure of p16.7C. Residues are coloured according to their respective average atomic temperature factor (red and blue colours mean high and low temperature factors, respectively). b) Backbone $[^1H]-[^{15}N]$ heteronuclear NOEs of p16.7C (lower values correspond to highly mobile regions).

**Figure 2.** Ensemble of 25 NMR structures showing the stacking interaction between P87a/P87b and the W116a/W116b indol rings at the dimerization interface. Both views are related by a 90° rotation around x axis.

**Figure 3.** Regions from $^{13}C$-HSQC (a) and $^{15}N$-HSQC (b) spectra at 350 μM (left) and 3500 μM (right) protein concentration and pH 5.0. Due to the extreme line broadening associated to the protein oligomerization some signals (highlighted in red) disappear.

**Figure 4.** Two different regions (a and b) of 2D-NOESY experiments corresponding to p16.7C at 3500 μM (upper panels) and 350 μM (lower panels). The line broadening and/or chemical shift changes for some proton signals in residues 83, 108 and 112 are indicated.

**Figure 5.** a) The location of those residues significantly affected by line-broadening and/or chemical shift changes (>0.025 ppm), on the p16.7C surface, is highlighted in green. b) Ribbon representation of p16.7C with the same orientation as that shown in a.

**Figure 6.** Two different regions (a and b) of 2D-NOESY experiments at 350 μM (left panels) and 3500 μM (middle panels) p16.7C concentration. The right panels show the same regions (a and b) in a half-filtered NOESY experiment performed at high protein concentration. Interprotein contacts non-compatible with the structure of the dimer are shown in red. Interprotein contactcs compatible with the structure of the dimer are shown in blue.

**Figure 7.** a) Schematic representation of the main interprotein contacts between p16.7C dimers deduced from the analysis of 2D-NMR spectra at high protein concentration. b) NMR-derived model for the p16.7C tetramer. An ensemble of 15 structures is shown. It is clear from these data that 30 experimental constraints is enough to define the relative orientation of the two p16.7C dimers within the supramolecular complex.

**Figure 8.** Proposed model for p16.7C interaction with dsDNA.
Table I. $^1$H, $^{15}$N and $^{13}$C assignments of p16.7C at 308 K and pH 5.0. TSP was used as internal reference.

| Residue | NH | N | \(\alpha\) | C \(\alpha\) | H\(\beta\) | H\(\beta\)3 | H\(\gamma\) | H\(\gamma\)3 | Others |
|---------|----|---|---------|-------------|---------|---------|---------|---------|---------|
| His 61  | 8.40 | 119.99 | 4.69 | 55.53 | 3.30 | 3.18 | - | - | - |
| Met 62  | 8.31 | 120.78 | 4.44 | 55.53 | 2.05 | 1.96 | 2.54 | 2.47 | - |
| Asp 63  | 8.38 | 121.09 | 4.58 | 56.39 | 1.89 | 1.78 | 1.44 | 1.43 | H\(\delta\) 1.66,1.65, H\(\epsilon\) 3.00,2.99 |
| Lys 64  | 8.23 | 120.97 | 4.36 | 62.35 | 4.19 | - | 1.17 | - | - |
| Thr 65  | 8.20 | 113.94 | 4.29 | 61.87 | 2.04 | - | 0.89 | 0.87 | - |
| Val 66  | 7.83 | 120.81 | 4.14 | 52.77 | 2.84 | 2.66 | - | - | - |
| Asn 67  | 8.36 | 121.96 | 4.81 | 54.08 | 1.67 | 1.58 | 1.57 | - | H\(\delta\) 0.80,0.76 |
| Leu 68  | 7.93 | 122.21 | 4.59 | 59.30 | 2.59 | 2.55 | 1.85 | 1.80 | - |
| Ser 69  | 8.46 | 118.71 | 4.55 | 67.05 | 2.06 | - | 1.01 | 0.93 | - |
| Ala 70  | 8.88 | 123.45 | 4.11 | 56.79 | 4.32 | 3.95 | - | - | - |
| Cys 71  | 8.46 | 115.12 | 4.42 | 70.57 | 2.87 | 2.75 | - | - | - |
| Glu 72  | 8.38 | 119.85 | 4.05 | 54.95 | 1.50 | - | - | - | - |
| Leu 73  | 7.45 | 117.34 | 3.77 | 66.38 | 2.16 | - | 1.28 | 1.16 | - |
| Leu 76  | 8.52 | 121.19 | 4.30 | 59.05 | 1.99 | 1.84 | 1.59 | - | H\(\delta\) 0.81,0.80 |
| Asp 77  | 8.60 | 120.06 | 4.58 | 57.19 | 2.98 | 2.75 | - | - | - |
| Tyr 79  | 7.66 | 118.89 | 3.90 | 64.06 | 3.22 | 3.03 | - | - | - |
| Glu 80  | 8.05 | 120.75 | 4.17 | 59.96 | 2.32 | 2.32 | 2.50 | 2.41 | - |
| Ser 82  | 7.38 | 113.38 | 4.57 | 58.53 | 4.17 | 3.89 | - | - | - |
| Asn 83  | 8.16 | 116.69 | 4.35 | 54.08 | 3.19 | 2.75 | - | - | - |
| Arg 85  | 9.63 | 129.08 | 4.33 | 55.53 | 1.83 | 1.77 | 1.59 | 1.56 | H\(\delta\) 3.13,3.11 |
| Ile 86  | 8.40 | 125.74 | 4.66 | 54.95 | 1.92 | - | 0.78 | 1.24,1.02 | H\(\delta\) 0.02 |
| Pro 87  | - | - | 3.90 | 62.53 | 1.29 | 1.14 | 1.68 | 1.65 | H\(\delta\) 3.99,3.67 |
| Asp 89  | 8.11 | 116.84 | 4.22 | 56.82 | 2.83 | 1.91 | - | - | - |
| Asp 93  | 8.38 | 119.38 | 4.13 | 58.31 | 2.25 | 2.18 | 2.59 | 2.52 | H\(\delta\) 7.49,6.79,N\(\delta\) 112.1 |
| Leu 94  | 8.56 | 121.61 | 3.74 | 58.24 | 1.91 | 1.89 | 1.81 | - | H\(\delta\) 0.94,0.87 |
| Ser 101 | 7.52 | 114.79 | 3.92 | 64.83 | 2.24 | - | 1.03 | 0.99 | - |
| Asn 107 | 9.63 | 129.08 | 4.33 | 55.53 | 1.83 | 1.77 | 1.59 | 1.56 | H\(\delta\) 3.13,3.11 |
| Gln 108 | 8.72 | 121.72 | 4.19 | 60.50 | 2.08 | 2.00 | 2.54 | 2.46 | - |
| Gly 110 | 8.72 | 121.53 | 3.95 | 59.32 | 2.18 | 2.01 | 2.46 | 2.38 | H\(\delta\) 7.47,7.01,N\(\delta\) 113.31 |
| Arg 111 | 8.06 | 116.41 | 4.08 | 59.28 | 2.26 | 2.18 | 2.46 | 2.39 | - |
| Glu 112 | 7.93 | 121.58 | 4.16 | 57.41 | 2.01 | 1.82 | 2.40 | 2.28 | H\(\delta\) 6.58,6.48,N\(\delta\) 112.64 |
| Residue | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg     | 9.37| 123.45| 4.09| 61.13| 2.41| 1.93| 2.13| 1.40| H^δ 3.35/2.87 |
| Thr     | 7.93| 117.05| 3.93| 67.44| 4.36| 1.27| -   |   -   | H^δ 6.18 H^ε 5.65 |
| Tyr     | 7.60| 121.30| 3.98| 61.89| 3.07| 2.97| -   |   -   | 7.42, H^c 7.43 H^ε 6.78 H^δ 6.90, N^c 130.23 |
| Trp     | 8.66| 118.91| 4.27| 59.87| 3.21| 3.13| -   |   -   | H^δ 7.25 H^δ 10.96 H^ε |
| Lys     | 9.08| 125.98| 4.02| 61.28| 2.28| 1.92| 1.70| 1.36| H^δ 1.85/1.56 H^ε 3.02/2.90 |
| Leu     | 7.70| 116.90| 3.97| 57.45| 1.89| 1.35| 1.85| -   | H^δ 0.87/0.84 |
| Glu     | 8.38| 122.17| 3.74| 59.51| 1.89| 1.44| 2.11| 1.91| -   |
| Asn     | 7.92| 112.33| 4.67| 55.54| 3.52| 2.61| -   |   -   | H^δ 6.70/6.30 N^δ 111.12 |
| Gln     | 7.51| 116.16| 4.34| 55.41| 2.25| 2.13| 2.59| 2.48| H^δ 7.37/6.89 N^c 110.72 |
| Lys     | 7.05| 120.81| 4.05| 58.26| 1.91| 1.72| 1.83| 1.37| H^δ 1.74/1.73 H^ε 2.95/2.94 |
| Lys     | 8.56| 122.40| 4.42| 55.26| 1.70| 1.70| 1.43| 1.29| H^δ 1.69/1.68 H^ε 2.99/2.99 |
| Leu     | 8.61| 126.44| 4.05| 54.80| 1.45| 1.20| 1.02| -   | H^δ 0.33/0.31 |
| Tyr     | 8.30| 125.43| 4.58| 57.81| 2.99| 2.96| -   |   -   | H^δ 7.17 H^ε 6.84 |
| Arg     | 8.16| 123.82| 4.19| 55.76| 1.80| 1.61| 1.56| 1.44| H^δ 3.11/3.10 H^ε 7.17 |
| Gly     | 7.41| 108.11| 3.84/3.83| 45.32| -   |   -   |   -   |   -   |
| Ser     | 8.16| 115.80| 4.44| 58.28| 3.89| 3.86| -   |   -   | -   |
| Leu     | 8.12| 124.25| 4.35| 55.19| 1.65| 1.64| 1.63| -   | H^δ 0.88/0.83 |
| Lys     | 7.88| 126.41| 4.14| 57.39| 1.81| 1.68| 1.39| 1.38| H^δ 1.68/1.67 H^ε 3.00/3.00 |
Table II.- Intermolecular NOEs observed at 3500 μM protein concentration representative of p16.7C oligomeric species.

**Intermolecular NOE**

- N83 Hα / Y115 Hζ
- N83 Hβ2 / Y115 Hζ
- N83 Hβ3 / Y115 Hζ
- N83 Hδ2 / Y115 Hζ
- N83 Hδ2 / Y115 Hζ
- Y125 Hδ1 / Y108 He1
- Y125 He1 / Y108 Hδ1
- Y125 Hδ1 / T111 Meγ2
- Y125 He1 / T111 Meγ2
- Y125 Hα / T111 Meγ2
- Y125 Hβ2 / T111 Meγ2
- Y125 Hβ3 / T111 Meγ2
- Y125 HN / T111 Meγ2
- Y125 Hδ1 / L99 Meδ2
- Y125 He1 / L99 Meδ2
Supplementary figure-1 Asensio et al.
Supplementary figure-2 Asensio et al.
Supplementary figure-3 Asensio et al.
Supplementary figure 4 Asensio et al.
Supplementary figure-5 Asensio et al.
Supplementary figure-6 Asensio et al.
Supplementary figure-7 Asensio et al.
Supplementary figure-8 Asensio et al.
Structure of the functional domain of \( \phi \) replication organizer: insights into oligomerization and DNA binding
Juan Luis Asensio, Armando Albert, Daniel Muñoz-Espín, Carlos Gonzalez, Juan Hermoso, Laurentino Villar, Jesús Jimenez-Barbero, Margarita Salas and Wilfried Meijer

*J. Biol. Chem.* published online March 16, 2005

Access the most updated version of this article at doi: [10.1074/jbc.M501687200](https://doi.org/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