Recognition of DNA by $\omega$ protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: analysis of binding to operator DNA with one to four heptad repeats

Ana B. de la Hoz, Florencia Pratto, Rolf Misselwitz, Christian Speck, Wilhelm Weihofen, Karin W elfle, Wolfram Saenger, Heinz W elfle and Juan C. Alonso*

Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología, CSIC, 28049 Madrid, Spain, Max-Delbrück-Centrum für Molekulare Medizin, 13092 Berlin, Germany, Max-Planck-Institut für Molekulare Genetik, 14195 Berlin, Germany and Institut für Kristallographie, Freie Universität Berlin, 14195 Berlin, Germany

Received February 19, 2004; Revised April 28, 2004; Accepted May 13, 2004

ABSTRACT

pSM19035-encoded $\omega$ protein forms a dimer ($\omega_2$) that binds to a set of 7-bp repeats with sequence 5'-NATCACN-3'. Upon binding to its cognate sites, $\omega_2$ regulates transcription of genes required for copy number control and stable inheritance of plasmids, and promotes accurate plasmid segregation. Protein $\omega_2$ binds poorly to one heptad but the affinity to DNA increases with two and more unspaced heptads in direct or inverted orientation. DNA titration of increasing numbers of heptads with $\omega_2$, monitored by circular dichroism measurements, indicates the binding of one $\omega_2$ to one heptad ($\omega_2$:heptad stoichiometry of 1:1). Spacing of two directly or inversely oriented heptads by 1 to 7 bp reduces the affinity of the protein for its cognate target site. The binding affinity of $\omega_2$ for two directly repeated heptads was severely reduced if one of the base pairs of the core 5'-ATCAC-3' sequence of one of the heptads was individually substituted by any other base pair. Hydroxyl radical footprinting shows a protection pattern at the 5'-ATCAC-3' core. These data suggest that each heptad defines an operator half-site and that tight binding of the symmetric $\omega_2$ to the central 5'-TCA-3' core of symmetric or asymmetric targets (differently oriented heptads) is probably achieved by structural changes of DNA and/or protein or both.

INTRODUCTION

Gene regulation in prokaryotes is achieved largely by proteins, having predominantly helix–turn–helix (HTH) DNA binding motifs, and interacting specifically with palindromic DNA sequences to repress or activate expression of associated genes or groups of genes (1). Recognition of arrays of short direct DNA repeats by specific transcriptional regulators is rare in bacteria (2–5). The ribbon–helix–helix (RHH) or $(\beta\alpha_2\beta)$ proteins are a growing family of DNA binding proteins, of which many have been studied by X-ray crystallography and NMR spectroscopy (6–12). Among the characteristics associated with the DNA-binding domain of $(\beta\alpha_2\beta)$ proteins are two-stranded antiparallel $\beta$-sheets that make sequence-specific contacts with the major groove of DNA, and several residues of the second $\alpha$-helix ($\alpha_B$) interact with phosphates of the DNA backbone. Most RHH proteins, which are symmetric dimers, bind as tetramers, or higher order oligomers, to operator sequences containing two or more binding sites that are usually arranged as inverted (Arc, CopG) or direct (MetJ) repeats for each dimeric DNA-binding protein (6,7,9). The MetJ interacts symmetrically with its directly repeated cognate site, whereas the Arc and CopG proteins interact asymmetrically with their palindromic targets (6,7,9).

The *Streptococcus pyogenes* pSM19035-encoded $\omega_2$ protein, which belongs to a family of RHH DNA-binding proteins, binds to upstream promoter regions of genes involved in copy number control ($P_{cop}$), partition ($P_{par}$/parS) and partition and post-segregational killing ($P_\omega$) (2,13,14). The $\omega_2$ protein target site consists of two ($P_\omega$) or three ($P_{cop}$/parS) copies of a block composed of two direct and one inverted 7-bp repeats (5'-$\Lambda_T$ATCAC$^\Lambda_T$-3') symbolized by $\longrightarrow\longleftrightarrow$, plus one additional 7-bp direct repeat $\longrightarrow$ downstream of the last block in $P_{cop}$, or in the inverted $\longleftrightarrow$ orientation downstream of the last block in $P_\omega$ (see Fig. 1). Upstream of $P_{par}$, which overlaps with the parS site, there are seven head-to-tail direct 7-bp repeats and two inverted 7-bp repeats (Fig. 1). Such organization of binding sites has been well documented for eukaryotic transcription factors that bind cooperatively to tandemly repeated units [e.g., TFIIIA (15), STAT family of proteins (16), HSF protein (17), IRF family of proteins (18), MR-GR proteins (19), NR superfamily of proteins (TRE, VDR, RAR, RXR) (20–22)]. Here the relative
orientation and spacing of the core recognition motifs play an essential role in the specificity of DNA-binding and transcriptional activation.

Using site-directed mutagenesis, electrophoretic mobility shift assay (EMSA), DNase I and hydroxyl radical footprinting, surface plasmon resonance (SPR), and spectroscopic studies we have examined the binding of ω2 to various arrays of the 7-bp repeat. These experiments show that ω2 binds with very low affinity and no specificity to a single 7-bp repeat. The binding of ω2 to DNA, albeit with varying individual affinity, is achieved by units composed of two differently oriented repeats [(→), (←−) and (←→)]. The inverted repeats (←→) and (←−) are palindromic as the sequences are related by a 2-fold rotation axis, whereas the directed repeats (→) are not. Clusters of three, four or more (as well as the full-length cognate recognition sites composed of up to 10 heptads) provide the highest stability to the DNA complexes with dissociation constants in the 5 nM range (2). Circular dichroism (CD) studies indicate that ω2 induces local conformational changes in DNA.

MATERIALS AND METHODS

Bacterial strains, plasmids and oligonucleotides

Escherichia coli BL21(DE3) and XL1-Blue strains, the Bacillus subtilis PB: lacZ, recA4 strain and plasmids pUC18, pH13-ω and pT712ω were previously described (2,23,24). pHP14-borne ω (PHP14-ω) and ωΔ19 gene (PHP14-ωΔ19) were used for β-galactosidase measurements. Oligonucleotides containing the specified number of heptads were synthesized with three flanking adenines (aaa) at the 5′-end and three guanosines (ggg) at the 3′-end [denoted in lower case letters, e.g., 5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATCACAAATCACAggg-3′, 60-bp (5′-aaaAATCACAAATCACAAATC

Figure 1. Comparison of the nucleotide sequences of the ω2 targets. The conserved −35 and −10 regions of the PcopS, P6 and Pα promoters are indicated as full line boxes. Bent arrows and +1 denotes known transcription start sites. The ω2 protein binding 7 bp repeats and their relative orientations are indicated by arrows below the nucleotide sequences. Putative palindromic sequences are indicated by dashed convergent arrows below the nucleotide sequences.
pCB443 (C6→A), pCB442 (C6→G); pCB446 (A7→C), pCB445 (A7→G).

Chemicals, DNA, enzymes, proteins and reagents

All chemicals were p.a. grade and purchased from Merck, Darmstadt, Germany. DNA restriction and modification enzymes, dNTPs, and poly[d(I-C)] were from Boehringer Mannheim, Germany. Ultrapure acrylamide was from Serva, Heidelberg, Germany. The broad protein molecular weight marker was obtained from Gibco-BRL, Barcelona, Spain.

The buffers used were: buffer A, 50 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 50 mM NaCl, 2 mM DTE, 0.2 mM phenylmethylsulfonylfluoride, 5% glycerol; buffer B, 50 mM Tris–HCl pH 7.5, 10 mM MgCl₂, 50 mM NaCl.

For EMSA and footprinting experiments, the HindIII–KpnI DNA segments (59 bp, two heptads, 66 bp, three heptads and 73 bp, four heptads) or EcoRI–SpIh DNA fragments (71 bp, three heptads; and 78 bp, four heptads) were end-labeled by the filling-in reaction using the Klenow enzyme at the HindIII or EcoRI sites, and gel purified.

For fluorography experiments single biotinylated (Bio) double stranded (ds) oligonucleotides containing the specified number of heptads, flanked by Bio-gtgaat at the 5′-end and tattca at the 3′-end [one (20 bp, e.g., 5′-Bio-gtgaatAATCACAT -tattca-3′, (→→) two (27 bp) in (→→) or (→→) orientation, three (34 bp) in the (→→) or (→→) and four heptads 41 bp (→→)] and their non-biotinylated complementary strands were annealed, purified and bound to the SPR chip.

Oligonucleotides for CD studies containing the specified number of heptads, flanked by gcg at the 5′-end and gg the 3′-end [one (12 bp, e.g., 5′-gcgAATCACAG -ggttaa-3′), two (19 bp) (→→), (→→) or (→→), three (26 bp) (→→), and four heptads (33 bp) (→→)] in the indicated orientation and their complementary strands were annealed and purified by ion exchange chromatography using a Mono Q column. The concentration of DNA is expressed as moles of nucleotides and was determined using a molar extinction coefficient of 6500 M⁻¹·cm⁻¹ at 260 nm.

The protein concentration is expressed in moles of Bio₂ dimers and was determined from the absorbance at 276 nm using an absorption coefficient of A₁%,₁cm = 3.63 (26).

Measurement of protein–DNA interactions

For EMSA, gel purified [³²P]HindIII–KpnI DNA (2 nM) was incubated with various amounts of Bio₂ and in the presence of 1 µg poly[d(I-C)] as non-specific competitor DNA in buffer B for 15 min at 37°C in 20 µl final volume as previously described (2). The mixture was then separated using an 8% non-denaturing polyacrylamide gel electrophoresis (ndPAGE). The ndPAGE were run with 0.5× Tris–borate-EDTA (24) at 45 V at 4°C and dried prior to autoradiography.

To obtain apparent dissociation constant (K_d,app) values from EMSA and DNase I footprint experiments, concentrations of free DNA and Bio₂–DNA complexes were densitometrically determined under non-saturating conditions from differently exposed autoradiographs of EMSA and DNase I footprinting gels. Bio₂ protein concentrations that transfer 50% of the ³²P-DNA into complexes or protect 50% from DNase I digestion are approximately equal to the K_d,app under conditions where the DNA concentration is much lower than K_d,app. Since DNA concentrations used in the EMSA and DNase I protection experiments are 2 nM, reliable results are expected for K_d,app values above 20 nM. In SPR experiments, the Bio₂ protein concentrations at 50% of the maximal resonance units (RU) values were taken as equivalent to K_d,app. The reaction conditions for DNase I footprint experiments were as for EMSA. DNase I treatment was as previously described (2).

SPR was measured in real time as previously described (27) using a Biacore 2000 instrument with biotinylated synthetic DNA immobilized on a Streptavidin sensor chip. By construction, all chips contained the same molar concentration of heptads, although the number of heptads per DNA molecule was different. After the DNA segments were coupled to the streptavidin-coated surface of the flow cells, kinetics of binding were measured with Bio₂ concentrations ranging from 0.8 to 100 nM. The SPR signal was measured in RU, 100 RU corresponding to 0.1 ng/mm² of Bio₂ bound to the chip. The rate and equilibrium binding constants and Hill coefficients were calculated using BIAevaluation 3.0 software.

Hydroxyl radical footprinting was performed as previously described (28). [α³²P]HindIII–KpnI or EcoRI–SpIh DNA (2 nM) and 1 µg of poly [d(I-C)] as non-specific competitor DNA were incubated with various amounts of Bio₂ protein in a total volume of 20 µl in buffer B. After 20 min incubation at 30°C, the footprint was started by addition of 3 µl of a freshly prepared solution containing 4 mM EDTA, 2 mM ammonium iron(II)-sulfate hexahydrate, 14 mM sodium ascorbate and 1.5% H₂O₂. After 4 min, the reaction was stopped by addition of 2 µl of 100 mM thiourea and 2 µl of 0.5 M EDTA. Samples were diluted 1:1 with water, and the DNA ethanol precipitated. The samples were resuspended in loading buffer separated in 15% dPAGE, and the gels were autoradiographed. For the analysis of the footprint the intensities of random bands (6 to 8), in each lane, outside the footprinting area were added and used to obtain a normalization factor. This factor served to correct any potential loading error in the different lanes. The normalized band intensities of protein-free control were then subtracted from the corresponding band intensities of the lanes of the protein-containing DNA.

CD spectra in the 320–250 nm region were measured at 20°C using a Jasco J720 spectropolarimeter. For titration of DNA with Bio₂ protein, aliquots of a concentrated Bio₂ solution in buffer B were added to the desired DNA solutions in a 5 mm path-length cuvette, mixed and incubated for 15 min at 20°C before measurements. For the calculation of the stoichiometry the ellipticity differences at 264 nm between Bio₂-DNA complexes and free DNA was plotted versus the molar ratio Bio₂:DNA.
RESULTS

$\omega_2$ protein binds to heptad repeats with different orientation

In pSM19035, upstream of the $\alpha$-P$\omega_2$ genes there are 10, nine and seven copies of the 7 bp repeat 5’-AATCAGAGT-3’, respectively (Fig. 1). These heptads are the suspected $\omega_2$ binding sites (2); however, alternative $\omega_2$ protein binding sites have not been formally ruled out (see dashed arrows in Fig. 1).

Using EMSA we studied the complex formation of $\omega_2$ with the 59 to 73 bp long [α-32P]HindIII–KpnI DNA fragments containing one to four unspaced heptads as a function of $\omega_2$ protein concentration. We failed to detect binding of $\omega_2$ (up to 500 nM) to DNA fragments containing only one heptad. The $K_d$ that were obtained from EMSA are summarized in Table 1.

The $\omega_2$ protein concentration required to bind 50% of the DNA containing two heptads in the head-to-head orientation is approximately lower than for binding DNA containing heptads in the $\omega_2$ or $\omega_2$ orientation. The $K_d$ values (equivalent to the $\omega_2$ concentration required to protect 50% of the DNA) are 20, 90 and 120 nM, respectively (Fig. 2B, A and C, Table 1). Only one shifted band was observed with heptads in the $\omega_2$, $\omega_2$ and $\omega_2$ orientation (Fig. 2A to C denoted by arrowheads).

The affinity of $\omega_2$ binding to DNA containing three and four heptads (Fig. 2D–H) is similar to full-length sites (4–12 nM) (2). With the exception of diheptads (Fig. 2A–C) and of DNA in the $\omega_2$, $\omega_2$ configuration where only one shifted band was observed (Fig. 2H), the formation of higher order complexes was observed for DNA with increased number of adjacent heptads. One to three shifted bands were observed when the DNA fragment encompassed three or four heptads, respectively (denoted by brackets in Fig. 2D–G). This indicates that the 7-bp heptads are the binding site for $\omega_2$ protein and that a minimum of two heptads is required for initial binding. This rules out the binding of $\omega_2$ protein to alternative sites, as the palindromic sequences denoted by dashed converging arrows in Figure 1. The observed heterogeneity of formed complexes suggests a varying occupancy of the heptad binding sites by $\omega_2$ and/or a distortion of the DNA segments.

$\omega_2$ protein protects large DNA segments against DNase I attack

To study the effect of $\omega_2$ on the protection of DNA fragments containing two to four heptads in different orientations, the $\omega_2$–DNA complexes were analyzed by DNase I footprinting. Protein $\omega_2$ failed to protect a DNA fragment containing only one heptad from DNase I attack (data not shown) whereas discrete regions of DNA containing two to four heptads were protected. In most of the studied DNA fragments the protected region is longer than the stretch of heptad sequences, namely ~18 bp in length for $\omega_2$ (Fig. 3A), ~22 bp for $\omega_2$ (Fig. 3B), ~28 bp for $\omega_2$ (Fig. 3D) and $\omega_2$ (Fig. 3E) and ~36 bp for $\omega_2$ (Fig. 3F) and $\omega_2$ (Fig. 3G). The protection is mainly extended at the upstream ends of the dsDNA. As shown in Figure 3C and H, the protected regions from DNase I attack were shorter than the heptad stretches with $\omega_2$ protein protecting only ~12 bp in a DNA segment containing two heptads in the $\omega_2$ or $\omega_2$ orientation and ~26 bp in a DNA fragment with four heptads in $\omega_2$ orientation. In the latter case, the upstream region and the first three heptads were protected from DNase I attack by $\omega_2$, but the fourth heptad (at the 3′-end) was not protected (Fig. 3H).

With the exception of heptads in the $\omega_2$, $\omega_2$ and $\omega_2$ orientations (Fig. 3C, G and H) where single sensitive sites interrupted by protected regions were observed, continuous protection from DNase I attack by $\omega_2$ were found with DNA containing two ($\omega_2$, $\omega_2$), three ($\omega_2$, $\omega_2$ and $\omega_2$) and four ($\omega_2$, $\omega_2$) heptads (Fig. 3A, B, D, E and F). Protection from DNase I attack was also observed in the upstream region, where heptads were not present, with DNA containing two ($\omega_2$ and $\omega_2$), three ($\omega_2$, $\omega_2$ and $\omega_2$) and four ($\omega_2$) heptads (Fig. 3A, B, D–F). As previously postulated (2), this is most likely caused by formation of a large nucleoprotein complex in which non-specific binding of $\omega_2$ to the flanking DNA regions is nucleated from the PcopS, Pα and Pβ or parS region (see Fig. 1).

The protection effect of $\omega_2$ was quantified by densitometric scanning of the autoradiographs (see Fig. 3). Protein $\omega_2$ binds with similar affinity to two heptads with the $\omega_2$ orientation ($K_d$ ~25 nM) and three and four heptads ($K_d$ ~5–12 nM) (Table 1).

There is a difference in the $K_d$ as determined by EMSA and DNase I for $\omega_2$–DNA complexes with the heptads in the $\omega_2$ orientation ($K_d$ 90 and 25 nM, respectively) when compared to the $\omega_2$ orientation ($K_d$ ~25 nM in both cases), which might suggest a faster off rate of $\omega_2$ for $\omega_2$ DNA under the conditions of EMSA. The $\omega_2$ concentrations required to protect 50% of the tail-to-tail $\omega_2$ or $\omega_2$ DNA segments are similar when determined with both EMSA and DNase I footprinting ($K_d$ ~130 nM) (Table 1).

Table 1. $K_d$ in nM of $\omega_2$ necessary to achieve half saturation of 2 nM DNA

<table>
<thead>
<tr>
<th>Orientation and number of heptads</th>
<th>EMSA</th>
<th>DNase I footprinting</th>
<th>SPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_2$</td>
<td>&gt;500</td>
<td>&gt;500</td>
<td>&gt;500</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>90</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>20</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>$\omega_2$ and $\omega_2$</td>
<td>120</td>
<td>140</td>
<td>ND</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>6</td>
<td>8</td>
<td>ND</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>14</td>
<td>14</td>
<td>ND</td>
</tr>
<tr>
<td>$\omega_2$ and $\omega_2$</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

ND, not determined.
The association ($k_a \sim 1 \times 10^7 \text{ ms}^{-1}$) and dissociation ($k_d \sim 0.1 \text{ s}^{-1}$) rate constants are similar for $\omega_2$ binding to DNA containing three heptads in $\rightarrow_2$ (Fig. 4C) and $\rightarrow_3$ orientation and to DNA containing four heptads in the $\rightarrow_4$ orientation (Fig. 4D). Although the different chips are coated with the same number of binding sites, the moles of bound DNA differ according to the different number of heptads per DNA molecule. We assume that differences in conformation of the DNA targets account for the small differences in $K_{d,\text{app}}$ between targets with two (20 nM) and three or four repeats ($K_{d,\text{app}} 6-10$ nM) and not the statistical degeneracy of the macroscopic binding constant (29). A similar binding affinity was observed when the full-length site (Fig. 1) containing seven to 10 repeats ($K_{d,\text{app}} 4-10$ nM) was studied (2).

### Binding of $\omega_2$ protein induces conformational changes in DNA

Titration of DNA containing two $\rightarrow_2$, three $\rightarrow_3$ and four $\rightarrow_4$ direct heptads (Fig. 5A–C) with $\omega_2$ was accompanied by spectral changes in the DNA region between 250 and 320 nm.
Isosbestic points were found at 283 nm for DNA with two heptads, \( \rightarrow 2 \), and at 282 nm for DNA with three \( \rightarrow 3 \) and at 281 nm for DNA with four heptads, \( \rightarrow 4 \). The positions of the isosbestic points indicate that the CD spectra are mainly determined by two components, namely free DNA and complexes with \( \omega_2 \). The CD spectra do not discriminate between complexes differing in the number of \( \omega_2 \) and position of \( \omega_2 \) occupying the two, three and four binding sites per DNA. At substoichiometric amounts of \( \omega_2 \) the CD spectra do not allow to deduce the possible formation of complexes with varying stoichiometry as suggested by the EMSA experiments shown in Figure 2.

Plots of \( \Delta [\theta]_{264 \text{nm}} \), the difference in the ellipticity at 264 nm between \( \omega_2 \)-DNA and free DNA, versus molar ratio of \( \omega_2 \) to DNA containing two \( \rightarrow 2 \) (Fig. 5D), three \( \rightarrow 3 \) (Fig. 5E) or four \( \rightarrow 4 \) heptads (Fig. 5F) showed stoichiometries of binding of 2.2, 3.1 and 4.5 \( \omega_2 \) per DNA with two, three and four heptads, respectively. These results, which are close to the binding of one \( \omega_2 \) to one DNA heptad, indicate that each dimer binds to each half of the minimal binding site and interacts with the neighboring dimer.

Sequence selective recognition by \( \omega_2 \) protein

To learn about the relevant nucleotides involved in \( \omega_2 \) binding, we have substituted every base pair of the upstream heptad (5'_A/TATCACA/T_3', positions 1 to 7) of the \( \rightarrow 2 \) binding site, whereas the downstream heptad (positions 1' to 7') was not modified (Table 2). The protein concentration required to protect 50% of the 59-bp HindIII–KpnI target DNA from DNase I attack was determined. The replacement of \( A \) at position 1 or 7 by \( T \) (A\( \rightarrow \)T, a natural one), A\( \rightarrow \)C or A\( \rightarrow \)G does not affect the \( \omega_2 \) footprinting (\( K_{d, \text{app}} \approx 20 \text{ nM} \)), and the presence of a \( C \) at positions 1 or 7 slightly increases (2-fold) \( \omega_2 \) protection to DNase I attack (see above). The A\( \rightarrow \)T or A\( \rightarrow \)G replacements at position 2 do not seem to affect the \( \omega_2 \) binding, but A\( \rightarrow \)C reduced by 8-fold \( \omega_2 \) protection from DNase I attack (Table 2). The T\( \rightarrow \)G replacement at position 3, C\( \rightarrow \)A or C\( \rightarrow \)G at position 4 and A\( \rightarrow \)G or A\( \rightarrow \)T replacements at position 5 or replacement of C at position 6 for any other base pair reduced \( \omega_2 \) protection from DNase I attack by >7-fold (Table 2). The other replacements revealed an intermediate affinity for \( \omega_2 \) (Table 2).

The base specific effects of changes in positions 2 and 3 and more general effects of base pair exchanges in positions 4 to 6, which drastically affect the binding of \( \omega_2 \) to a diheptad DNA, suggest that: (i) the \( \omega_2 \) operator site consists of two highly similar heptads, (ii) each heptad defines an operator half site.
and (iii) the pentameric central 5'-ATCAC-3' or its complementary sequence is relevant for \( \omega_2 \) binding. Identical results were obtained when the mutations were introduced in one of the repeats of two inversely oriented heptads (data not shown).

### \( \omega_2 \) protein interacts with the 5'-ATCAC-3' central core

Hydroxyl radical footprinting of \( \omega_2 \)-DNA complexes with DNA containing two (\( \rightarrow \)) and four (\( \rightarrow \)) heptads showed a different protection pattern at both DNA strands. The distribution of protected positions at the ‘top’ strand was different to that observed at the ‘bottom’ strand, suggesting that \( \omega_2 \) interacts differently with each strand. As revealed in Figure 6A and B, the bases at the ‘top’ strand that were protected by \( \omega_2 \) protein on a \( \rightarrow \) DNA segment from the attack of hydroxyl radicals cluster in the 5'-AT-3' or 5'-ATC-3' sequence (positions 2 and 3 or 2 to 4). The protected bases of the first heptad cannot be quantified due poor resolution of the gel. The protected regions on the three heptads are separated by four or five non-protected nucleotides (Fig. 6C).

Hydroxyl radical footprinting of a \( \omega_2 \)-DNA complex with DNA containing three heptads in the \( \rightarrow \) orientation was also assayed (Fig. 7A–C). On the ‘top’ strand the bases protected on the two directly repeated heptads cluster in the 5'-AT-3' or 5'-ATC-3' sequence and in the inversely repeated heptad they cluster in the 5'-GTG-3' sequence (Fig. 7C). On the ‘bottom’ strand the protected bases cluster in the 5'-GTG-3' sequence of the directly repeated heptads and in the 5'-AT-3' or 5'-ATC-3' sequence of the inversely repeated heptad (Fig. 7A and B).

### \( \omega_2 \) protein binds poorly to a spaced core binding motif

The \( \omega_2 \) protein probably binds to the DNA major groove with two anti-parallel \( \beta \)-strands, featuring two arginines (Arg31, Arg31') that point to the base pairs and possibly recognize both guanines (10). In a previous section it is shown that: (i) a mutation at the central core of the upstream heptad in \( \rightarrow \) configuration for any other nucleotide reduces \( \omega_2 \) binding >8-fold compared to the cognate site (Table 2), and (ii) \( \omega_2 \) protein specifically protects the 5'-ATCAC-3' segment of the heptads.

To address whether the protein binds to a spaced target site, the spacing of two direct \( \omega_2 \) core motifs was varied by 1 (\( \rightarrow \)1\( \rightarrow \)) to 7 bp (\( \rightarrow \)7\( \rightarrow \)), and the protection of DNA by bound \( \omega_2 \) protein was determined by DNase I footprinting. All these separated heptads failed to form measurable \( \omega_2 \)-DNA complexes in the presence of up to 40 nM \( \omega_2 \). Half maximal saturation was observed in the presence of 100–120 nM \( \omega_2 \) to diheptads separated either by 1 (\( \rightarrow \)1\( \rightarrow \)) to 7 bp (\( \rightarrow \)7\( \rightarrow \)) indicating a 5- to 6-fold reduced binding affinity when compared with the unspaced cognate \( \rightarrow \)2 site (\( K_{d,app} \approx 20 \) nM). Identical results were obtained when the spacing of two inversely oriented heptads...
Similarly the Arc repressor binds poorly to spaced half sites (30).

**DISCUSSION**

The DNA binding site of \( \omega_2 \) at the \( P_{copS}, P_6 \) and \( P_\omega \) promoter regions is composed of an array of adjacent unspaced heptads (see Fig. 1). Unlike the Arc dimers, which bind to each subsite with nanomolar affinities (31), the \( \omega_2 \) protein binds with low affinity \((K_{d,app} > 1 \mu M)\) to a subsite or single heptad. The \( \omega_2 \) protein binds with high specificity and affinity to two \((K_{d,app} \approx 20 \text{ nM})\) or more unspaced heptads \((K_{d,app} \approx 4-12 \text{ nM})\). The characterization of several heptads in different orientation has led to the identification of a consensus \( \omega_2 \) operator site composed of two conserved and unspaced heptads \((5'\-\Lambda^2_{\omega}ATCAC^4_{\omega}T-3')\) to which two \( \omega_2 \) bind with high affinity. This is consistent with stoichiometry experiments that show that 2.2, 3.1 and 4.5 \( \omega_2 \) molecules bind DNA segments containing two, three and four heptads, respectively. It is likely, therefore, that each heptad defines an operator half-site.

A reduced (3- to 6-fold) \( \omega_2 \) binding affinity to two heptads spaced by 1 to 7 bp was observed. Similar results have been reported for other members of the RHH protein family when the binding site was artificially spaced (30,32). Each \( \omega_2 \) protein contacts a 3 to 5 bp sequence in each DNA half site (heptad) (33; this work). Heptads in inverted and divergent orientations have palindromic symmetry, which allows the symmetry-related binding of two \( \omega_2 \) proteins to each of the two heptads, whereas \( \omega_2 \)-DNA complexes with...
Table 2. Binding of \( \omega_2 \) protein to DNA studied by DNase I protection experiments

<table>
<thead>
<tr>
<th>Modified position</th>
<th>Nucleotide position</th>
<th>([\omega_2] ) required to reach ( K_{d,\text{app}} ) (in nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 1' 2' 3' 4' 5' 6' 7'</td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>5'...A T C A C a a  A T C A C a a A T C A C a a 3'</td>
<td>–20</td>
</tr>
<tr>
<td>1</td>
<td>5'...C</td>
<td>–10</td>
</tr>
<tr>
<td>1</td>
<td>5'...G</td>
<td>–20</td>
</tr>
<tr>
<td>2</td>
<td>5'...T</td>
<td>–20</td>
</tr>
<tr>
<td>2</td>
<td>5'...C</td>
<td>&gt;160</td>
</tr>
<tr>
<td>2</td>
<td>5'...G</td>
<td>–20</td>
</tr>
<tr>
<td>3</td>
<td>5'...A</td>
<td>–80</td>
</tr>
<tr>
<td>3</td>
<td>5'...C</td>
<td>–80</td>
</tr>
<tr>
<td>3</td>
<td>5'...G</td>
<td>–80</td>
</tr>
<tr>
<td>4</td>
<td>5'...T</td>
<td>–80</td>
</tr>
<tr>
<td>4</td>
<td>5'...G</td>
<td>–20</td>
</tr>
<tr>
<td>4</td>
<td>5'...A</td>
<td>&gt;160</td>
</tr>
<tr>
<td>4</td>
<td>5'...G</td>
<td>&gt;160</td>
</tr>
<tr>
<td>5</td>
<td>5'...T</td>
<td>&gt;160</td>
</tr>
<tr>
<td>5</td>
<td>5'...C</td>
<td>–160</td>
</tr>
<tr>
<td>5</td>
<td>5'...G</td>
<td>–160</td>
</tr>
<tr>
<td>6</td>
<td>5'...G</td>
<td>&gt;160</td>
</tr>
<tr>
<td>6</td>
<td>5'...A</td>
<td>&gt;160</td>
</tr>
<tr>
<td>6</td>
<td>5'...G</td>
<td>&gt;160</td>
</tr>
<tr>
<td>7</td>
<td>5'...C</td>
<td>–10</td>
</tr>
<tr>
<td>7</td>
<td>5'...G</td>
<td>–20</td>
</tr>
</tbody>
</table>

The DNA contained two directly repeated heptads of which the upstream one was mutated by one base in consecutive positions as shown on each line. aThe heptad sequences are flanked by three adenines at the 5'-ends and three cytosines at the 3'-ends and surrounded by the multi-cloning site of the vector. The base pair difference with wt is highlighted.

adjacent heptads in tandem orientation \( \rightarrow_2 \) are not symmetry-related. The interfaces between \( \omega_2 \) in complexes with symmetry-related and non-symmetry-related heptads \( \rightarrow_2 \) might be different. Except EMSA, DNase I and chemical footprinting and SPR techniques showed that \( \omega_2 \) protein binds with comparable high affinity to diheptads in \( \rightarrow_2 \) or \( \leftarrow_2 \) orientation, \( K_{d,\text{app}} \sim 20 \) nM, but binds to a diheptad in the \( \leftarrow_2 \) orientation with ~6-fold lower affinity, \( K_{d,\text{app}} \sim 130 \) nM.

Three heptads in \( \rightarrow_3 \) or \( \leftarrow_2 \) orientation or four heptads in the \( \rightarrow_4 \), \( \leftarrow_2 \) or \( \leftarrow_2 \rightarrow_2 \) orientations form high affinity binding sites for \( \omega_2 \) protein, as judged by EMSA, DNase I footprinting or SPR. The affinity of these heptads is comparable to that observed with the full-length binding site (2). All these data are consistent with the observation that the affinity of \( \omega_2 \) for a specific promoter does not increase by increasing the number of heptads in the respective operator, provided that there are at least three or four heptad repeats (2; this work). Upon binding to its cognate site, \( \omega_2 \) protein at high concentration has the tendency to polymerize on the 5' region of the 'top' strand (2), an effect that has yet to be understood.

Genetic and biochemical experiments show that the bases of the 5'-ATCAC-3' core or the complementary 5'-GTGAT-3' sequence are essential for the interaction of \( \omega_2 \). Hydroxyl radical footprints show that \( \omega_2 \) mainly interacts with the central 5'TCA-3' sequence. This has been confirmed by Raman spectroscopy, showing that the central 5'TCA-3' motif of the heptads might be the main target site for \( \omega_2 \) binding to operator DNA (34).

The N-terminal regions (1–23 and 1–22 in subunits I and II, respectively) of \( \omega_2 \) are not defined in the electron density due to proteolysis during crystallization and partial disorder (10). In vivo experiments demonstrated that plasmid-based \( \omega_2 \) variants lacking the first 20 residues specifically repress utilization of a chromosomal-based P6 promoter (F. Pratto, unpublished results). It is likely therefore that the flexible N-terminus is not involved in binding of \( \omega_2 \) protein to operator sequences. According to a preliminary model \( \omega_2 \) protein binds to the DNA major groove with the two anti-parallel \( \beta \)-strands (formed by residues Lys28 to Val32 and Lys28' to Val32') (10).

The operator sites of other RHH proteins (namely the MetJ, Arc and CopG proteins) show an overall bend of ~50–60° in the protein–DNA complexes (7,9,33). This agrees with CD titration experiments that show conformational changes in DNA upon \( \omega_2 \) binding, and Raman spectra indicate an induced fit of both, \( \omega_2 \) and DNA, as shown by changes in vibrational modes of deoxyribose moieties and protein-induced DNA bending (34). By contrast, DNase I footprint experiments did not indicate the presence of hypersensitive sites upon \( \omega_2 \) binding to its cognate site (see Fig. 3).

Crystal structures of repressor–DNA complexes of MetJ (6), Arc (7,8) and CopG (9) have shown that these repressors bind as two dimers to palindromic operator sites. MetJ binds symmetrically to tandem binding sites (6), whereas the interaction of Arc and CopG with their palindromic cognate sites is asymmetric (7,9). The distances between the binding centers of the two dimers on the DNA are different: MetJ 8 bp apart, Arc 11 bp apart, Mnt and CopG each 9 bp apart. The \( \beta \)-ribbon of each of the repressors comprises seven to nine amino acid residues per monomer in MetJ, Arc, Mnt and CopG. By contrast, only five amino acid residues form the \( \beta \)-ribbon in \( \omega_2 \) protein (10) and the centers of the half-sites are only 7 bp apart in agreement with the heptad repeat sequence of operator DNA. For any heptad orientation (\( \rightarrow_2 \) or \( \leftarrow_2 \)) the 5'TCA-3' motifs are separated by a center-to-center spacing of 7 bp and rotated relative to each other by \( 7 \times 34° = 238° \). Therefore, like MetJ (32), \( \omega_2 \) bound to seven to 10 heptads should wrap around the DNA helix.

Figure 8 shows a model of the \( \omega_2 \)-DNA complex that is based on the MetJ–DNA complex as the RHH motifs are...
comparable (6,10,32). It illustrates the gross arrangement of three \( \omega_2 \) bound to a straight B-DNA segment composed of three heptads (\( \rightarrow_3 \)), each \( \omega_2 \) inserting its antiparallel \( \beta \)-ribbon into the major groove of the DNA. The \( \beta \)-ribbon of \( \omega_2 \) contains residues R31 on one \( \beta \)-strand and R31' on the other strand that are related by a 2-fold rotation axis relating the two monomers in the dimer (10). The side-chains of R31 and R31' are candidates for asymmetric contacts with guanines on the same DNA strand (motif 5'-CAC-3) in the \( \rightarrow_2 \) diheptad or for symmetric contacts with guanines on the opposite strand in the \( \rightarrow_3 \) diheptad (10). Since protein \( \omega_2 \) binds to seven to 10 heptads: \( \omega_2 \) possibly decorates the DNA helix, adjacent \( \omega_2 \) being rotated relative to each other by 252° around the DNA helix axis. The distance between \( \omega_2 \) \( \alpha \)-helices \( A \) can easily be reduced to become comparable to the distance in MetJ-DNA if DNA is bent (see above) or \( \omega_2 \) slightly rotated on the heptad binding site. Figure 8 also shows that due to the 2-fold symmetry in the \( \omega_2 \) dimer, two \( \omega_2 \) bound to two heptads in head-to-tail orientation, \( \rightarrow_2 \) is comparable to two \( \omega_2 \) bound to \( \rightarrow_4 \) DNA. However, depending on the heptad orientation, different helices would interact. In the \( \rightarrow_2 \) situation \( \alpha \)-helices \( A' \) and \( A \) were located in close neighborhood at the surfaces of the two dimers, whereas in the \( \rightarrow_4 \) case \( \alpha \)-helices \( A' \) of both \( \omega_2 \) should be involved in interdimer interaction. In any case possible interactions between neighboring \( \omega_2 \) bound to multiple heptads should sensitively depend on the spacing between the heptads, as actually observed (see above).

ACKNOWLEDGEMENTS

We thank Mrs B. Kannen for technical assistance during the spectroscopic measurements. This work was partially supported by grants BMC2003-00150 from MCT-DGI to J.C.A.,
REFERENCES


