Sequence Recognition, Cooperative Interaction, and Dimerization of the Initiator Protein DnaA of *Streptomyces*

Received for publication, August 29, 2000, and in revised form, October 18, 2000
Published, JBC Papers in Press, November 9, 2000, DOI 10.1074/jbc.M007876200

Jerzy Majka‡‡, Jolanta Zakrzewska-Czerwińska‡, and Walter Messer§

From the ‡Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Weigla 12, 53-114 Wroclaw, Poland and §Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, Berlin-Dahlem D-14195, Germany

Using a combined PCR-gel retardation assay, the preferred recognition sequence of the *Streptomyces* initiator protein DnaA was determined. The protein showed a preference toward DNA containing two *Escherichia coli*-like DnaA boxes in a head-to-head arrangement (consensus sequence TTATCCACA, whereas the consensus sequence of the DnaA boxes found in the *Streptomyces* oriC region is TTGTCCACA). In quantitative band shift experiments, the kinetics of the *Streptomyces* DnaA-DnaA box interaction was characterized. The DnaA protein can form dimers while binding to a single DnaA box; dimer formation is mediated by the domain III of the protein, and the dissociation constant of this process was between 35 and 115 nM. *Streptomyces* initiator protein DnaA interacts in a cooperative manner with DNA containing multiple binding sites. For the cooperativity effect, which seems to be independent of the distance separating the DnaA boxes, domain I (or I and II) is responsible. The cooperativity constant is moderate and is in the range of 20–110.

The initiator protein DnaA plays an essential role in the initiation of bacterial chromosome replication. The interaction of DnaA protein with its chromosomal origin (oriC) is best understood in *Escherichia coli*. The *E. coli* DnaA protein (52 kDa) binds to five nonpalindromic, nonamer sequences, the DnaA boxes. Binding of 10–20 DnaA monomers promotes a local unwinding of an adjacent AT-rich region. The unwound region provides an entry site for the DnaB/DnaC helicase complex followed by other proteins required to form a replication fork (1–4).

Apart from its primary function as a replisome organizer, the DnaA protein acts as a transcription factor that represses or activates several genes or terminates transcription, depending on the location and arrangement of DnaA boxes (5).

Both functions of the DnaA protein, replisome organizer and transcription factor, are mediated by its interaction with target DNA. The DNA binding domain of *E. coli* DnaA has been localized in the 94 C-terminal amino acids. A potential helix-loop-helix motif has been reported within this part of the protein (6). However, because x-ray high resolution structure analysis is not yet available for DnaA proteins, the detailed interaction with the DNA is still poorly understood. The consensus sequence of the *E. coli* DnaA box differs depending on the method used for its evaluation. The most stringent definition for the DnaA box sequence comes from a determination of binding constants: 5′-TTAT(T/C)TNCACA-3′ (7). Tp, Tp′, Tp′′, and Tp″ were found to be directly involved in DNA-protein interaction (8). *E. coli* DnaA does not dimerize in solution and interacts with a single DnaA box as a monomer, as measured by gel retardation (7) and by surface plasmon resonance (9). However, binding to DnaA boxes that differ from the stringent consensus sequence by one or two base pairs requires two such boxes and an interaction of DnaA proteins bound to them (9). DnaA binds ATP and ADP with high affinity. Both forms of DnaA protein, ATP-DnaA and ADP-DnaA, recognize DNA in a similar fashion; however, only ATP-DnaA is active in the DNA replication process. Recently, it has been shown that *E. coli* ATP-DnaA protein recognizes also a hexamer sequence, the ATP-DnaA box 5′-AGATCT-3′ or close match of it (9).

*Streptomyces* (Gram-positive soil bacteria) differ from other prokaryotic organisms in their mycelial life cycle and in possessing a large (8-megabase pair), linear and GC-rich (about 72%) chromosome (10, 11). Recent discoveries suggest that replication of the linear chromosome of *Streptomyces coelicolor* A3/2 proceeds bidirectionally from the centrally located oriC region toward the ends of the chromosome (12).

The key elements of initiation of the *Streptomyces* chromosome replication, oriC region and DnaA protein, show higher complexity than those of *E. coli*. The *Streptomyces* oriC region contains numerous DnaA boxes, which are grouped into two clusters (13, 14). The *Streptomyces* DnaA protein consists, like all other DnaA proteins, of four domains. In contrast to the other bacteria, the *Streptomyces* DnaA protein is larger (70–73 kDa), since it comprises an additional stretch (~230 amino acids) of predominantly acidic or hydrophobic amino acids within domain II. The residues lower the isoelectric point of the entire *Streptomyces* DnaA protein (pI = 5.7) (15, 16). As it was shown for the *E. coli* and *Bacillus subtilis* DnaA proteins, domain III and the C-terminal part (domain IV) of the *Streptomyces* DnaA protein are responsible for binding of ATP and DNA, respectively (15, 17). The consensus sequence of the *Streptomyces* DnaA box in oriC is 5′-TTGTCCACA-3′, which differs at the third position (A/G) from the *E. coli* DnaA box (13). In contrast to *E. coli*, the *Streptomyces* DnaA protein can form a dimer when binding to a single DnaA box. Recently, it has been shown that the domains I and III are independently involved in dimerization of the *Streptomyces lividans* DnaA protein molecules. The interaction of *Streptomyces* DnaA protein with two DnaA boxes is cooperative and accompanied by strong DNA bending (16, 18). However, we do not know the contribution of the different domains of the *Streptomyces* DnaA protein to cooperative binding.
In this work, we apply a combined PCR-EMSA\(^1\) technique (19) to elucidate DNA sequence requirements for *Streptomyces* DnaA protein binding. Using EMSA, we define details of the binding of the *Streptomyces* wild type DnaA and its truncated forms to DNA. We reveal that the domain I participates in cooperative DnaA protein-DNA interactions. Quantitative analysis of gel mobility shifts allowed us to determine binding constants for dimerization and cooperative DNA protein interactions. In addition, we have varied the spacing between two DnaA boxes and examined the consequences on dimerization and cooperative binding of DnaA protein to these boxes.

**EXPERIMENTAL PROCEDURES**

Protocols and DNA—The His-tagged wild type DnaA proteins of *S. lividans* and its truncated mutants DnaA(III-IV) comprising the domain III and the DNA binding domain were purified on a N\(^{-}\)-nitrilotriacetic acid-agarose column (Qiagen) as described earlier (15). The DNA binding domain of the DnaA protein DnaA(BD) was expressed as a C-terminal glutathione S-transferase fusion and purified using glutathione-Sepharose 4B beads (Amersham Pharmacia Biotech) followed by factor X\(_c\) cleavage as described earlier (16).

Oligonucleotides were chemically synthesized and purified by high performance liquid chromatography. For the kinetic analysis of DnaA protein binding complementary nucleotides were annealed by heating at 95 °C for 10 min and gradually cooling to room temperature. The oligonucleotides were end-labeled with [\(^{32}\)P]ATP and T4 polynucleotide kinase and purified from 5% nondenaturing polyacrylamide gels.

In **Vitro DNA Binding Site Selection**—The recognition motif of the DnaA protein was determined using several cycles of amplification and selection, essentially as described previously (19). The library of 88-bp oligonucleotides for the first selection cycle was prepared by PCR using template oligonucleotides: 5'-GGCCGATCTCGACTAGGCGCGGTTAGTTCTGGCGGC-3' and the primer pair 5'-GCGCGATCCTCGGACTAGGCGC-3' and 5'-CGCGGATCCTAGGCGGC-3' (restriction sites BamHI and SacI are in italic type). The amplification reaction was carried out in 100 µl using 1 pmol of template oligonucleotide (1 pmol corresponds to the number of all possible combinations of 20 degenerated bases) and 100 pmol of each primer for 20 cycles, with each cycle consisting of 15 s at 96 °C, 30 s at 60 °C, and 30 s at 72 °C. Double-stranded oligonucleotides were separated on 3% agarose gels, electrophoresed for 30 min at room temperature. The complexes were selected and amplified by repeating each experiment three or four times. Different kinetic analysis experiments showed variations up to 15%. Equilibrium constants were determined using the modified statistical mechanical model (20) supplemented with the protein dimerization module. For the DNA substrate containing a single binding site A, the concentration of each species in the gel was expressed by the following equations (model 1),

\[
S = 1/Z_1 \quad (\text{Eq. 1})
\]

\[
SL = k_S L/Z_1 \quad (\text{Eq. 2})
\]

\[
SL_2 = k_SL_1 Z_2 \quad (\text{Eq. 3})
\]

\[
Z_1 = 1 + k_SL_1 L + k_{SL_1} L^2 \quad (\text{Eq. 4})
\]

where \(S\), \(SL\), and \(SL_2\) represent concentrations of free DNA, monomer, and dimer respectively; \(L\) is protein (ligand) concentration; \(k_S\) is the microscopically equilibrium affinity constant for binding site A, and \(k_{SL_1}\) is the affinity constant describing the dimerization process.

For the DNA containing binding sites A and B, the concentrations of the species presented in the gel are described by the following equations (model 2),

\[
S = 1/Z_2 \quad (\text{Eq. 5})
\]

\[
SL = (k_A + k_B L)/Z_2 \quad (\text{Eq. 6})
\]

\[
SL_2 = (k_A + k_B L)/Z_1 \quad (\text{Eq. 7})
\]

\[
Z_1 = 1 + (k_A + k_B L + (k_A L + k_B L + k_{SL_1} L_1^2) \quad (\text{Eq. 8})
\]

where \(k_A\) is the intrinsic affinity constant for the binding site B, and \(k_{SL_1}\) is the cooperativity constant describing interaction of protein molecules occupying both binding sites.

Experimental data were simultaneously fitted to the equations using Scientist\textsuperscript{®} for Windows\textsuperscript{TM} software (MicroMath).

\[\text{Table I}\]

**Oligonucleotides used in the gel retardation experiments**

The black arrow shows the strong *Streptomyces* DnaA box; the white arrow shows the *E. coli* DnaA box R1/R4; the grey arrow shows the box2 found by the *in vitro* DNA binding sites selection assay; and the crossed out arrow shows the scrambled DnaA box.

| Name   | Sequence and character |
|--------|------------------------|
| 1s-0   | AGACCTCTTCACGAGAGACTTGTTAAGA |
| 2s     | AGACCTCTTCACGAGAGACTTGTTAAGA |
| 2s+15  | AGACCTCTTCACGAGAGACTTGTTAAGA |
| 2s+10  | AGACCTCTTCACGAGAGACTTGTTAAGA |
| Sel 1s-0 | GGGCACTTACGACGTTAGTTCTGGCGGC-3' |
| Sel 2s  | GGGCACTTACGACGTTAGTTCTGGCGGC-3' |
| Sel 1s-1/2 | GGGCACTTACGACGTTAGTTCTGGCGGC-3' |

\[\text{The abbreviations used are: bp, base pair(s); EMSA, electrophoretic mobility shift assay; PCR, polymerase chain reaction; SPR, surface plasmon resonance.}\]
RESULTS

Determination of the Optimal DnaA Protein Binding Site—
The consensus sequence of the DnaA box identified within the Streptomyces oriC region is 5'-TTGTCCACA-3'. The Streptomyces oriC region contains a higher number of the DnaA boxes (19 boxes) than the E. coli oriC region (five boxes). A DNase I footprint of the oriC region with DnaA showed protection of all boxes. However, when analyzed individually by surface plasmon resonance (SPR), only some DnaA boxes were specifically recognized by the DnaA protein (14, 15). To characterize in detail the recognition properties of the Streptomyces DnaA protein, we applied a binding site selection technique based on the combined EMSA and polymerase chain reactions (Fig. 1A; Ref. 19).

The purified DnaA protein was incubated with a substrate pool of double-stranded oligonucleotides, in which a random 20-bp region was flanked by defined 19-bp sequences containing restriction sites to facilitate subsequent cloning. The protein-DNA complexes were separated from the unbound DNA on a 5% polyacrylamide nondenaturing gel. Only one protein-bound fraction was visible. The DNA recovered from the complex was amplified by PCR with the pair of primers complementary to the defined sequences of the oligonucleotides (Fig. 1B) and used as a substrate for the renewed binding assay. In the third selection cycle, we observed a second retarded band with a higher electrophoretic mobility than the previous one. Here, the bands with lower and higher electrophoretic mobility are called a “dimer” (Fig. 1C, ** complex) and a “monomer” (Fig. 1C, * complex), respectively. The DNA from both bands was recovered and used independently in the next selection cycles. The alignments of the selected oligonucleotides after the fourth, fifth, and sixth cycle are shown in Tables II and III.

The sequences selected from the “dimer” band after four rounds reveal remarkable features: 63 out of the 67 oligonucleotides analyzed carry two DnaA boxes, one complete 9-bp DnaA box, the “box 1,” and one incomplete DnaA box, the “box 2” (“box” as defined by sequence). The incomplete box 2 consists of 4–7 bp (3’-part of the DnaA box). Both boxes face each other...
### TABLE II

**Selected DNA binding sites for DnaA ("dimer" band)**

The sequence of cloned binding sites for DnaA, derived after the fourth, fifth, and sixth selection cycles, were aligned for maximum match to the canonical DnaA box definition. The region of 20 originally random nucleotides is shown in uppercase letters, whereas nucleotides from adjacent primer regions are indicated by lowercase letters (only five bases from each side of the random region are shown). Letters in boldface type match the DnaA box consensus sequence. Letters in italic type indicate the lower strand of the DnaA box. For the consensus sequence, letters in boldface represent nucleotides present more than 85%. Uppercase letters indicate nucleotides appearing in the frequency range 50–85%. Nucleotides that are represented at less than 15% were considered as not significant at the indicated position.

| Cycle 4 | Box 1 | Box 2 |
|---------|-------|-------|
| 20× | tagcgGCAG TTATCCACA | TGTGgata gctac |
| 10× | tagcgcGg TTATCCACA | TGTGgata GAgctac |
| 3× | tagcggGAAgGg TTATCCACA | TGTG gctac |
| 2× | tagcgcCTAgGg TTATCCACA | TGTGgata cgctac |
| 2× | tagcgcTGCAaGg TTATCCACA | TGTGgata ct |
| 2× | tagcgcGGgATGg TTATCCACA | TGTG gctac |
| 2× | tagcgGgTTGg GCCTACA | TGTG gctac |
| 2× | tagcgTGGgCA TTATCCACA | TGTG gctac |
| 2× | tagcggCGgA TTATCCACA | TGTGgata GAGctac |
| 2× | tagcggGgTTGg GCCTACA | TGTG gctac |
| 2× | tagcgGgGCgA TTATCCACA | TGTG gctac |
| 2× | tagcggGgTTGg GCCTACA | TGTG gctac |
| 2× | tagcgGgGCgA TTATCCACA | TGTG gctac |
| 2× | tagcgGgTTGg GCCTACA | TGTG gctac |
| 2× | tagcgGgGCgA TTATCCACA | TGTG gctac |
| 2× | tagcgGgTTGg GCCTACA | TGTG gctac |
| 2× | tagcgGgGCgA TTATCCACA | TGTG gctac |

| Cycle 5 | Consensus sequence | A | T | CCCACA |
|---------|--------------------|---|---|--------|
| 37× | tagcgGCgA TTATCCACA | TGTGgata | gctac |
| 12× | tagcgcGg TTATCCACA | TGTGgata | GAgctac |
| 4× | tagcgcCTAgGg TTATCCACA | TGTGgata | cgctac |
| 4× | tagcgcTGGA TTATCCACA | TGTGgata | gctac |
| 4× | tagcgcGGgGgTGA TTATCCACA | TGTGgata | GAgctac |
| 4× | tagcgGgTGcGg TTATCCACA | TGTGgata | cgctac |
| 4× | tagcgcGGgGgTGA TTATCCACA | TGTGgata | GAgctac |
| 4× | tagcgcGGgGgTGA TTATCCACA | TGTGgata | cgctac |
| 4× | tagcgcGGgGgTGA TTATCCACA | TGTGgata | GAgctac |

| Cycle 6 | Consensus sequence | C | G | CCCACA |
|---------|--------------------|---|---|--------|
| 29× | tagcgcGgGgA TTATCCACA | TGTGgata | gctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | gctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | gctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |
| 3× | tagcgcGgGgA TTATCCACA | TGTGgata | GAgctac |

| Consensus sequence | CAG | TTATCCACA | TGTGgata | GCTAC |
and are adjacent (except for a few oligonucleotides). 80% of the boxes 1 exhibit the *E. coli* type of DnaA box (T or A at the third position), and only 20% of them exhibit the *Streptomyces* type (G at the third position). During the next two rounds, the selection progressively narrowed the spectrum of observed oligonucleotide sequences; after six rounds the box 1 is exclusively represented by the *E. coli* DnaA box R1/R4 (5'-TTATCCACA-3'), and the partial box 2 contains in nearly all cases the

**TABLE III**

*Selected DNA binding sites for DnaA ("monomer" band)*

The sequence of cloned binding sites for DnaA, derived after the fourth, fifth, and sixth selection cycles, were aligned for maximum match to the canonical DnaA box definition. The region of 20 originally random nucleotides is shown in uppercase letters, whereas nucleotides from adjacent primer regions are indicated by lowercase letters (only five bases from each side of the random region are shown). Letters in boldface type match the DnaA box consensus sequence. Letters in italic type indicate the lower strand of the DnaA box. For the consensus sequence, letters in boldface represent nucleotides present more than 85%. Uppercase letters indicate nucleotides appearing in the frequency range 50–85%. Nucleotides that are represented at less than 15% were considered as not significant at the indicated position.

| Cycle 4 | Box 1 | Box 2 |
|---------|-------|-------|
| 20×     | tagcgGCAG | TTATCCACA | TGGGAGT | gctac |
| 2×      | tagcgG   | TTATCCACA | TGGGAGT | GCAgctac |
| 8×      | gtagccCAAC | TTATCCACA | GCGAAGGcgtac |
| 2×      | ttagcCAAG | TTATCCACA | GCGAAGGcgtac |
| 2×      | gtagccCAGACA | TTATCCACA | GGGCGAAGGcgtac |
| tagcgCATGG | TTATCCACA | GTGGGAGT | gctac |
| tagcgGTAG | TTATCCACA | AGTGGAGT | gctac |
| 8×      | gtagccCAAC | TTATCCACA | GCGAAGGcgtac |
| 2×      | ttagcCAAG | TTATCCACA | GCGAAGGcgtac |
| 2×      | gtagccCAGACA | TTATCCACA | GGGCGAAGGcgtac |
| tagcgCATGG | TTATCCACA | GTGGGAGT | gctac |
| tagcgGTAG | TTATCCACA | AGTGGAGT | gctac |

**Consensus sequence**

| Cycle 5 | Box 1 | Box 2 |
|---------|-------|-------|
| 8×      | tagcgGCAG | TTATCCACA | TGGGAGT | gctac |
| 2×      | tagcgG   | TTATCCACA | TGGGAGT | GCAgctac |
| 19×     | gtgcgcCAAC | TTATCCACA | GGGCGAAGGcgtac |
| 2×      | ttagcGTAG | TTATCCACA | GGGCGAAGGcgtac |
| 2×      | tagcgCATGG | TTATCCACA | GGGCGAAGGcgtac |
| tagcgTAG | TTATCCACA | GGGCGAAGGcgtac |
| gtgcgcCTAG | TTATCCACA | GGGCGAAGGcgtac |
| taggcCAAG | TTATCCACA | GAACCGAAGGcgtac |
| taggcCAAG | TTATCCACA | GAACCGAAGGcgtac |
| gtgcgcCCACAG | TTATCCACA | GGGCGAAGGcgtac |
| taggcCCACAG | TTATCCACA | GGGCGAAGGcgtac |

**Consensus sequence**

| Cycle 6 | Box 1 | Box 2 |
|---------|-------|-------|
| 25×     | tagcgGCAG | TTATCCACA | GGGCGAAGGcgtac |
| 4×      | ttagccCAAC | TTATCCACA | GGGCGAAGGcgtac |
| 3×      | ttagccCAGACA | TTATCCACA | GGGCGAAGGcgtac |
| 2×      | taggcCAAC | TTATCCACA | GGGCGAAGGcgtac |
| taggcCAAC | TTATCCACA | GGGCGAAGGcgtac |
| taggcCATAG | TTATCCACA | GGGCGAAGGcgtac |
| taggcCAAC | TTATCCACA | GGGCGAAGGcgtac |

**Consensus sequence**

*and are adjacent (except for a few oligonucleotides). 80% of the boxes 1 exhibit the *E. coli* type of DnaA box (T or A at the third position), and only 20% of them exhibit the *Streptomyces* type (G at the third position). During the next two rounds, the*
essential T nucleotides at the 7'– and 9'-positions.

More than half of the oligonucleotides selected from the monomer band after four rounds contained also two boxes. However, the monomer band had only appeared after the third round of the selection (Fig. 1C), and therefore it still contained traces of the oligonucleotides from the dimer band that were subsequently amplified. After the next two selection rounds, only one out of the 44 oligonucleotides contained two boxes. All of the boxes 1 from the monomer band exhibit the E. coli type DnaA box (R1/R4).

The data show that under the conditions of the assay, the Streptomyces DnaA protein interacts only with DnaA boxes; no other consensus sequences (e.g. the newly identified 6-bp ATP-DnaA box (9)) have been found. The results, presented in Tables II and III, show that DnaA protein from Streptomyces possesses a higher affinity toward DnaA boxes from E. coli than toward those from its own oriC region. Binding of the DnaA protein as a dimer apparently requires “head-to-head” orientation of the boxes.

Kinetic Studies with Streptomyces DnaA Protein—Upstream of the promoter region of the Streptomyces dnaA gene are two closely spaced DnaA boxes: a strong one (with the preferred Streptomyces sequence: 5'-TTGTCCACA-3') and a weak one (5'-TTGTCCCCCA-3') in head-to-head arrangement with 3 bp in between (21). Interaction of the DnaA protein with these boxes creates an autoregulatory circuit similar to that known for the E. coli dnaA gene (21). Recently, we have shown that binding of the DnaA protein to both DnaA boxes exhibits a cooperative character (16). Domains I and III independently participate in the dimerization of the DnaA protein molecules (18). To evaluate in detail the kinetics of binding of the DnaA protein to DnaA boxes from the dnaA gene promoter region, we applied gel mobility shift assays to determine the binding constants for cooperativity as well as for dimerization (20). This assay permits the quantitative analysis of the individual protein-DNA complexes. In our gel retardation experiments, we used the wild-type DnaA protein and its truncated forms, the DnaA(III-IV) lacking the two N-terminal domains (I and II) and the DnaA(BD) containing only the DNA binding domain IV. As a prerequisite for the kinetic studies, two DNA substrates containing either a single strong DnaA box, Is-0 (the weak DnaA box was scrambled), or two DnaA boxes, 2s, derived from the promoter region of the dnaA gene were designed in such a way that the DnaA box(es) is flanked on both sides by sequences corresponding to those within the dnaA promoter region (Table 1). To simplify kinetic calculations, the 2s substrate contains two identical boxes; the weak DnaA box was replaced by the strong one (Table 1). For the analysis, we used a fixed input DNA concentration and various protein concentrations, spanning 4 orders of magnitude (see legend to Fig. 2). The concentration of DNA in the reaction mixture was chosen to be at least 5 times lower than the lowest protein concentration used.

According to the statistical mechanical approach (20), the interaction between the DnaA protein and the DNA substrate composed of two binding sites could be described by the following equations,

\[
S = 1/Z 
\]

\[
SL = k_A + k_B L / Z 
\]

\[
SL_{2s} = k_A k_B L^2 / Z 
\]

\[
Z = 1 + (k_A + k_B) L + k_A k_B L^2 
\]

where \( S \), \( SL \), and \( SL_{2s} \) are concentrations of free DNA, monomer, and dimer bands, respectively; \( L \) is protein (ligand) concentration; \( k_A \) and \( k_B \) are the microscopic equilibrium affinity constants for binding sites A and B, respectively; and \( k_{AB} \) is the cooperativity constant describing the interaction of protein molecules occupying both binding sites.

These equations contain parameters \( k_A \), \( k_B \), and \( k_{AB} \) that occur only in combination and therefore cannot be unambiguously determined. However, if one of the pair \( k_A \) or \( k_B \) is known, it is possible to determine the remaining two. One way to achieve that goal is carrying out the gel shift experiment using a substrate containing only one binding site. In our analysis, it is the Is-0 oligonucleotide. Concentration of free DNA or DNA-protein complexes in this system is given by the Langmuir isotherm.

\[
SL = 1/S = k_A L / (1 + k_A L) 
\]

(Eq. 13)

The interaction of the Is-0 DNA fragment with increasing amounts of the wild type DnaA and DnaA(III-IV) led to two complexes (Fig. 2A, a and c), whereas incubation with the DnaA(BD) resulted in a single complex even at the highest protein concentration (Fig. 2A, e). Formation of a dimer band by the first two DnaA derivatives could be explained by an unspecific interaction with the scrambled DnaA box or another sequence motif of the Is-0 oligonucleotide or a protein dimerization process occurring on the DNA with only one-half of the dimer bound to DNA. The first possibility can be excluded, because domain IV alone does not form slower migrating complexes on that substrate, and all three proteins do not interact with DNA in which both DnaA boxes are scrambled (data not shown). Taking into account the dimerization process, we modified the original equations by adding a constant describing the oligomerization reaction. As a result, we obtained two sets of equations (model 1 and model 2) for the interaction with a single (Is-0) and with two DnaA boxes (2s), respectively (see “Experimental Procedures”).

The affinity constant for the single DnaA box, \( k_A \), and the affinity constant describing the dimerization process, \( k_{AB} \), were calculated by analyzing the interaction of the wild-type DnaA (or its truncated forms) with the Is-0 substrate according to model 1 or its simplified version, the Langmuir isotherm for the DnaA(BD). The values obtained from those experiments allowed us to determine \( k_A \) and \( k_{AB} \), the intrinsic affinity constant for the second DnaA box in substrate 2s, and the cooperativity constant, respectively (Fig. 2B).

The wild type DnaA protein and its truncated forms exhibit similar affinity to the single DnaA box, the Is-0 substrate (Table IV; \( K_j = 4–10 \) nM). The affinity constants calculated for the second DnaA box from the 2s substrate (\( k_B \)) were found to be nearly identical to the \( k_A \) values. Thus, the affinities seem to be independent of flanking sequences of the DnaA box(es).

The calculated cooperativity constants, \( k_{AB} \), showed that only the wild type DnaA is able to interact cooperatively with the two DnaA boxes; the cooperativity effect increases the affinity of the DnaA protein toward the DnaA boxes over 40 times. The values obtained for DnaA(III-IV) and DnaA(BD) are very close to 1; therefore, we consider these interactions as noncooperative. The results demonstrate that domain I or II or both of them are responsible for the cooperativity effect.

The dimerization dissociation constant for the wild-type protein (35 nM) is about 3 times lower than for a protein lacking domains I and II, DnaA(III-IV) (114 nM).

The Distance Separating DnaA Boxes Does Not Influence the Cooperativity of the DnaA-DnaA Box Interaction—The arrangement of the 19 DnaA boxes occurring within the Streptomyces oriC regions is highly conserved. The spacing between DnaA boxes varies from 3 to 20 bp or even more (except for two DnaA boxes that are adjacent). To evaluate the influence of the distance that separates binding sites on the cooperativity, we have...
designed three double-stranded oligonucleotides with two strong DnaA boxes separated by various spacings. The first one is the previously analyzed 2s substrate that contains 3 bp between the two DnaA boxes. The 2s mimics the spacing between the fifth and sixth DnaA boxes within the Streptomyces oriC region. The other two, 2s +5 and 2s +10 (Table I), contain 5 and 10 additional base pairs between the DnaA boxes, respectively. In the 2s +10 substrate, the increase of the distance separating both boxes by 10 bp does not affect their positioning on the DNA helix, while in the 2s +5 variant, the two DnaA boxes lie on the opposite face of the helix. The calculated cooperativity constants for the three substrates are presented in Table V.

Adding a turn of the helix (2s +10) increased the cooperativity about 2.5 times. Surprisingly, the addition of 5 bp (a half helix turn) caused negligible change of the cooperativity. As it
TABLE IV

Characterization of DNA binding properties of three DnaA protein forms to DnaA boxes derived from the promoter region of the dnaA gene

For every constant determination, S.D. is given (expressed as a percentage). * corresponding dissociation constants $K_d$ (reciprocal values of the affinity constants).

| Protein       | Intrinsic DnaA box affinity $K_d$ (M$^{-1}$) | Dimerization constant | Cooperativity constant |
|---------------|---------------------------------------------|-----------------------|------------------------|
| DnaA:         |                                             |                       |                        |
|               | $9.61 \times 7.5 \pm 5.7$                   | $2.83 \times 7.2 \pm 22$| $44 \pm 28\%$          |
| DnaA(III-IV): | $1.75 \times 8 \times 7.5 \pm 5.71$         | $8.77 \times 6 \pm 43$ | $2.5 \pm 14\%$         |
| DnaA(BD):     | $2.40 \times 8.6 \pm 6.6$                   | $1.47 \times 10$      | $0.53 \pm 18\%$        |

TABLE V

Comparison of the cooperativity constants for three DNA substrates that differ in the length of the spacer separating DnaA boxes

For every constant determination, S.D. is given (expressed as a percentage).

| DNA substrate | 3 bp | 6 bp | 13 bp |
|---------------|------|------|-------|
| Protein       |      |      |       |
| DnaA:         | 4482 \%| 4148\%| 10757\%|
| DnaA(III-IV): | 2.5 \pm 14\%| 0.7 \pm 20\%| 2.4 \pm 21\%|
| DnaA(BD):     | 0.53 \pm 18\%| 1.0 \pm 13\%| 0.7 \pm 25\%|

was shown for the “wild-type” substrate, 2s, the interaction of the truncated forms of the DnaA protein, DnaA(III-IV) and DnaA(BD), with the 2s +5 and 2s +10 substrates exhibited a noncooperative character (Table V).

Streptomyces DnaA Protein Prefer DnaA Boxes of the E. coli Type—The binding selection assay suggested that the Streptomyces DnaA protein prefers the DnaA box of E. coli type (5'-TTATCCACA-3') over its own DnaA box (5'-TTGTCCACA-3'). To evaluate in detail the interaction of the DnaA protein with the E. coli type DnaA box, quantitative gel retardation assays were performed. As before, we used a DnaA box system that enabled us to evaluate cooperativity in addition to intrinsic affinity constants. Thus, two substrates were used for the calculations: Sel 1s-0 and Sel 2s, containing the single box 1 (E. coli DnaA box) or two adjacent boxes 1 in head-to-head orientation, respectively (Table I). The Sel 1s-0 substrate was used to determine the affinity constant for the single binding box and the dimerization affinity constant. These values were then used for the interaction analysis of the DnaA protein with two boxes (substrate Sel 2s). Since the wild-type DnaA protein formed dimers during the interaction with a single DnaA box, the same mathematical models (models 1 and 2; see “Experimental Procedures”) were applied. In addition to the substrate with two identical boxes, the Sel 1s-1/2 oligonucleotide selected after the sixth cycle of the binding assay was also analyzed. The results of this analysis and their comparison with the constants obtained for the wild-type DnaA boxes from the promoter region are summarized in Table VI.

The affinity of the wild-type DnaA to the single DnaA box of E. coli is approximately 4 times higher than its affinity to the strong Streptomyces DnaA box. It explains why such DnaA recognition sequences had been found in the selection binding assay. As expected, the dimerization constant does not depend on the DNA recognition sequence; its value is nearly identical for both types of DnaA boxes (Table VI). The cooperativity constant for the Sel 2s substrate is approximately 2 times lower than the cooperativity for the wild type Streptomyces DnaA boxes (2s substrate). This may be due to the fact that the selected sequences did not have the 3-bp spacer.

DISCUSSION

Despite extensive work on the mechanism of initiation of DNA replication in prokaryotic and eukaryotic systems, several critical aspects of this mechanism and its control still remain obscure. One significant gap is a lack of understanding of the biological and biochemical roles of multiple initiator protein binding sites required by a large group of plasmid (interons) and chromosomal replicons (DnaA boxes). The oriC region of the Streptomyces linear chromosome consists of 19 DnaA boxes that serve as binding sites. Therefore, Streptomyces provides a good model for studying the interaction of initiator protein with multiple cognate binding sites. The initiator protein DnaA of Streptomyces contains two dimerization domains (I and III) that are separated by a long flexible domain II. According to our preliminary results (16, 22), interaction of the DnaA protein with two DnaA boxes exhibits cooperativity (e.g., the DnaA protein specifically binds to DNA fragments with two “weak” DnaA boxes, which are not bound as individ-
TABLE VII
Dissociation constants for the DnaA protein interactions with different types of DnaA box(es)

| DnaA boxes                        | $K_d$       |
|-----------------------------------|-------------|
| Single DnaA (colony)              | 10–78 nM    |
| R1/R4 E. coli DnaA box            | 3.4 nM      |
| Two DnaA: DnaA boxes (different)  | 1.7–3 nM    |
| Three DnaA boxes                  | 1.3 nM      |
| R1/R4 DnaA box; DnaA protein from E. coli | 0.9–1.2 nM |

- Intrinsic dissociation constant measured in gel retardation assays (10.4 nM for TGTTCACA) or apparent dissociation constant from SPR assay (12 nM for TGTTCACA, up to 78 nM for weak DnaA boxes).
- Apparent dissociation constant measured in gel retardation assay.
- Intrinsinc dissociation constant measured in gel retardation assay.
- Apparent dissociation constant measured in SPR assay.

Influence of the binding affinity. The naturally occurring DnaA box exhibit low diversity. Therefore, the possibility cannot be excluded that these sequences do not affect the ability of the DnaA protein to bind in a cooperative manner. The spacing between DnaA boxes within the oriC region varies from 3 to 20 bp.

The wild-type DnaA protein binds to two DnaA boxes separated by various spacings with a cooperativity parameter ranging from 23 to over 100. Surprisingly, cooperativity does not depend severely on the spacer length separating both binding sites (Table V). Adding a turn of the helix (2s or 10) increased the cooperativity about 2.5 times, whereas the addition of 5 bp (a half of a helix turn) had a modest effect on the cooperativity. Removing of 3 bp (Sel 2s) resulted in only 50% reduction of the cooperativity parameter. However, the Sel 2s oligonucleotide consists of the 4,5 type DnaA boxes and therefore cannot be directly compared with other DNA fragments. The moderate influence of the spacer length on the cooperativity may suggest that the protein domains involved in the intramolecular reactions are very flexible. Therefore, changing the spacing between adjacent binding sites does not affect the ability of the DnaA protein to bind in a cooperative manner. The spacing between DnaA boxes within the oriC region varies from 3 to 20 bp. Thus, in theory, the DnaA protein could be able to interact cooperatively with each pair of the adjacent DnaA boxes. However, in the binding site selection assay, only DnaA boxes that face each other have been selected. Therefore, we speculate that cooperativity occurs only when the binding sites are oriented head-to-head (such a box arrangement has been found in the promoter region of the dnaA gene as well as in the oriC regions, e.g. the fifth and sixth DnaA boxes). Our Nase I footprinting experiments (17) and electron microscopy studies (18) corroborate this hypothesis. Nase I footprinting experiments showed that in the oriC region, at the low protein concentration, DnaA binds first to the fifth and sixth DnaA boxes. According to the electron microscopy studies, the highest incidence of protein binding occurred at the middle of the first cluster of DnaA boxes, which corresponds to the location of DnaA boxes 5 and 6. Therefore, we assume that cooperativity at close distance determines at least the start of DnaA-oriC complex formation. Additional long range interactions may be formed subsequently and may be responsible for loop formation.

We established the kinetic constants for dimerization of the DnaA protein. The Streptomyces DnaA is the first chromosomal initiator for which the kinetics of two dimerization domains, I and III, have been determined. These domains dimerize independently (18). The dimerization of the DnaA protein does not occur in the absence of DNA. The wild-type DnaA protein reveals 3 times higher dimerization capability than the truncated DnaA protein containing only one dimerization domain.

The intermolecular interactions of the DnaA protein are...
3–20 times weaker than the interactions between DnaA protein and its DNA target (Table IV). Probably, it facilitates effective interactions of the DnaA molecules with DNA containing multiple recognition sequences (e.g., within the oriC region) and further formation of the nucleoprotein complex. For *E. coli* DnaA, the N-terminal domain 1 has been shown to promote oligomerization (25). However, as for *Streptomyces* DnaA, a second interaction face has been postulated in domain 3 or 4 (26).

**Why Does the Streptomyces oriC Region Contain so Many DnaA Boxes?**—The GC content of 57 *Streptomyces* DnaA boxes derived from three *Streptomyces* oriC regions is about 10 and 20% lower than the GC content of the oriC region (63%) and the overall GC content of *S. lividans* DNA (72%), respectively. However, it is still significantly higher than the GC content of the average *E. coli* DnaA box (9–30%). This difference in GC content explains the difference in the DnaA box consensus sequence between *Streptomyces* (5'-TTGTCCACA-3') and *E. coli* (5'-TTATCCACA-3'). It also results in flanking sequences of *Streptomyces* DnaA boxes that are relatively rich in GC.

As shown in Table VII, the dissociation constant ($K_d$) for specific binding of individual DnaA boxes derived from the *S. lividans* oriC region varies between 10 and 200 nM; a few additional DnaA boxes are not recognized by DnaA if they are analyzed outside the context of oriC ($K_d$ exceeding 200 nM). Interestingly, the affinity of the *S. lividans* DnaA protein for the R1/R4 *E. coli* DnaA box ($K_d = 3.4$ nM) is ~3 times higher than its affinity for the strong *Streptomyces* DnaA box. The apparent dissociation constant for binding of the *E. coli* DnaA protein to the DnaA box R1/R4 was calculated to be $K_d = 1.1$ nM (7).

Despite the strong difference in GC content between *Streptomyces* and other microorganisms, including *E. coli*, the domains I and III and the binding domain of the *Streptomyces* DnaA are highly conserved. A consequence of the high GC content of *Streptomyces* structural genes (72–74%) is the very nonrandom codon usage, with an extreme paucity of codons with A or T in the third position (C or G is usually at the third codon position). Like *E. coli* DnaA protein, the binding domain of *Streptomyces* DnaA protein contains the same putative DNA binding motive: two amphipathic α-helices with the basic loop in between formed by a third long α-helix (6, 16).

The affinity of the *S. lividans* protein for DNA fragments containing two or three closely spaced DnaA boxes is 6–10 times higher than its affinity for the single strong DnaA box (Table VII) and is comparable with the affinity of the *E. coli* DnaA protein to the R1/R4 box (0.9–1.2). The data suggest that efficient binding of the *Streptomyces* DnaA protein to DNA requires the presence of more than one *Streptomyces* DnaA box. Therefore, due to the high GC pressure exerted during the course of *Streptomyces* evolution, the structure and the sequence of the oriC region has been changed.

The relatively low affinity of *Streptomyces* DnaA protein for a single *Streptomyces* DnaA box seems to be compensated by a high number of DnaA boxes that are bound in a cooperative manner. In comparison with the lac repressor, the cooperativity parameter of the wild-type DnaA protein (40–100) is moderate; lac repressor by binding to its three operator sites shows cooperativity in the range of 250–950 (20). Such high values in a system with 19 binding sites (*Streptomyces* oriC) might cause irreversible binding of DnaA protein to the oriC region and consequently would block subsequent replication steps.

**REFERENCES**

1. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, W. H. Freeman and Co., New York
2. Skarstad, K., and Boye, E. (1994) *Biochim. Biophys. Acta* **1217**, 111–130
3. Messer, W., and Weigel, C. (1996) in *Escherichia coli and Salmonella : Cellular and Molecular Biology* (Neidhardt, F. C., Curtiss, R., III, Ingraham, J., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) pp. 1779–1801, American Society for Microbiology, Washington, D. C.
4. Kaguni, J. M. (1997) *Mol. Cells* **7**, 145–157
5. Messer, W., and Weigel, C. (1997) *Mol. Microbiol.* **24**, 1–6
6. Roth, A., and Messer, W. (1995) *EMBO J.* **14**, 2106–2111
7. Schaper, S., and Messer, W. (1995) *J. Biol. Chem.* **270**, 17622–17626
8. Speck, C., Weigel, C., and Messer, W. (1997) *Nucleic Acids Res.* **25**, 3242–3247
9. Speck, C., Weigel, C., and Messer, W. (1999) *EMBO J.* **18**, 6169–6176
10. Kutzner, H. J. (1981) in *The Prokaryotes* (Starr, M. P., Stolp, H., Truper, H. G., Balows, A., and Schlegel, H. G., eds) pp. 2028–2090, Springer, Berlin
11. Lin, Y. S., Kieser, H. M., Hopwood, D. A., and Chen, C. W. (1995) *Mol. Microbiol.* **10**, 923–933
12. Musialowski, M. S., Flett, F., Scott, G. B., Hobbs, G., Smith, C. P., and Oliver, S. G. (1994) *J. Bacteriol.* **176**, 5123–5125
13. Zakrzewska-Czerwińska, J., and Schrepf, H. (1992) *J. Bacteriol.* **174**, 2688–2693
14. Jakimowicz, D., Majka, J., Messer, W., Speck, C., Fernandez, M., Martin, M. C., Sanchez, J., Schauwecker, F., Keller, U., Schrepf, H., and Zakrzewska-Czerwińska, J. (1998) *Microbiology* **144**, 1281–1290
15. Majka, J., Messer, W., Schrepf, H., and Zakrzewska-Czerwińska, J. (1997) *J. Bacteriol.* **179**, 4236–4242
16. Majka, J., Jakimowicz, D., Messer, W., Schrepf, H., Lisowski, M., and Zakrzewska-Czerwińska, J. (1999) *Eur. J. Biochem.* **260**, 325–335
17. Majka, J. (1997) *Characterization of the Streptomyces lividans initiator protein DnaK*, Ph.D. thesis, Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wroclaw, Poland
18. Jakimowicz, D., Majka, J., Konopa, G., Wgrzyn, G., Messer, W., Schrepf, H., and Zakrzewska-Czerwińska, J. (2000) *J. Biol. Chem.* **275**, 4407–4417
19. Dolník, A., and Baláž, J. (2002) in *Streptomyces*, 3rd edn., pp. 1–31, Blackwell Publishing
20. Bucan, M., and Bašta, M. (2003) *Microbiol. Mol. Biol. Rev.* **67**, 329–355
21. Baláž, J., and Balážová, M. (2004) *Streptomyces*, 3rd edn., pp. 1–31, Blackwell Publishing
22. Jakimowicz, D., Majka, J., Lis, B., Konopa, G., Wgrzyn, G., Messer, W., Schrepf, H., and Zakrzewska-Czerwińska, J. (2000) *J. Mol. Biol.* **301**, 889–902
23. Balážová, M., and Baláž, J. (2004) *Streptomyces*, 3rd edn., pp. 1–31, Blackwell Publishing
24. Bell, C. E., Frescura, P., Hochschild, A., and Lewis, M. (2000) *Cell* **101**, 801–811
25. Messer, W., Blasing, F., Majka, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tüngler, D., Wegrzyn, G., Weigel, C., Welzcek, M., and Zakrzewska-Czerwińska, J. (1999) *Biochimie* **81**, 819–825
Sequence Recognition, Cooperative Interaction, and Dimerization of the Initiator Protein DnaA of *Streptomyces*

Jerzy Majka, Jolanta Zakrzewska-Czerwińska and Walter Messer

*J. Biol. Chem.* 2001, 276:6243-6252.
doi: 10.1074/jbc.M007876200 originally published online November 9, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M007876200

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

This article cites 22 references, 7 of which can be accessed free at http://www.jbc.org/content/276/9/6243.full.html#ref-list-1