Orientation of the LexA DNA-binding motif on operator DNA as inferred from cysteine-mediated phenyl azide crosslinking

(SOS response/gene regulation/LexA repressor/DNA recognition/4-azidophenacyl bromide)

PASCAL DUMOULIN, PASCALE OERTEL-BUCHHEIT, MICHELE GRANGER-SCHNARR, AND MANFRED SCHNARR

Institut de Biologie Moléculaire et Cellulaire, UPR 9005 du Centre National de la Recherche Scientifique, 15, rue René Descartes, F-67084 Strasbourg, France

Communicated by Jean-Marie Lehn, November 5, 1992 (received for review June 30, 1992)

ABSTRACT  To address the question how the recognition helix of the LexA repressor is positioned within the major groove of operator DNA we have applied a site-specific photocrosslinking approach using a LexA mutant repressor (LexA-C52) that harbors a single cysteine side chain in position 52, close to the COOH terminus of helix 3. The LexA-C52 mutant repressor has been purified and modified site-specifically with the photoreactive azido compound 4-azidophenacyl bromide, giving rise to LexA-C52*. Here we show that LexA-C52* may be selectively photocrosslinked with two adjacent bases within each operator half-site. The crosslinked bases are located, respectively, 10 and 11 base pairs from the dyad axis of the operator. The crosslinking data imply that the LexA recognition helix is oriented opposite to what is generally observed for helix–turn–helix proteins and that this helix should form a steeper angle with respect to the plane of the base pairs than is observed for standard helix–turn–helix proteins.

The LexA repressor from Escherichia coli regulates the transcription of about 20 different so-called SOS genes, which are mostly involved in DNA repair, mutagenesis, DNA replication, and cell division (for reviews see refs. 1–4). LexA is a two-domain protein of 202 amino acids that binds DNA via its NH2-terminal domain (5, 6). The protein dimerizes by means of its COOH-terminal domain with a fairly small dimerization constant of 2 × 10^4 M\(^{-1}\) (7, 8) and regulates transcription upon binding to its palindromic operator sequences such that two LexA monomers bind sequentially and cooperatively to the two operator half-sites (9). The NH2-terminal DNA-binding domain of LexA shows only weak sequence homology with known DNA-binding motifs, including helix–turn–helix (HTH) motif proteins (4, 10). A structural NMR investigation showed that the conformation of the NH2-terminal domain is not more closely related to the class of HTH proteins than to proteins without known DNA-binding activity (11). At present only low-resolution NMR data on the LexA-DNA complex are available (12). Sequence conservation among the SOS operators suggests that an optimal operator should have a twofold symmetric CTGT(A/T)\(_2\)ACAG sequence (13), and indeed this sequence binds LexA more tightly than any naturally occurring SOS operator (4). The major recognition elements should be comprised in the two outer elements (the CTGT motifs) as judged from the distribution of operator-down mutations (compiled in ref. 4), chemical protection and interference data (14), and NMR results (12).

A random mutagenesis study of the LexA protein suggested that helix 3 of LexA (aa 41–54) should be involved in DNA recognition (15). To test this hypothesis further, we have purified and chemically modified a LexA mutant protein containing a single cysteine residue instead of an arginine residue in position 52, close to the COOH terminus of helix 3. Site-specific incorporation of small bifunctional photocrosslinking or DNA-cleaving agents allows mapping of the position of an amino acid relative to the DNA bases within a protein-DNA complex, as shown earlier with EDTA-Fe (16–20) or phenanthroline adducts (21). Site-specific incorporation of photoactivatable crosslinking agents such as phenyl azide has been successfully used in the study of DNA triple-helix formation (22, 23) and more recently in the establishment of protein–DNA contacts in the case of the catabolite gene activator protein (24). Here we use the bifunctional photoactivatable reagent 4-azidophenacyl bromide to show that helix 3 of LexA may be crosslinked with operator DNA.

MATERIALS AND METHODS

Protein Purification. The mutant repressor LexA-C52 (i.e., Arg-52 → Cys) was purified essentially as described (7) except that the protein eluted from the phosphocellulose column at about 80 mM NaCl instead of 280 mM in the case of the LexA wild-type protein. Accordingly, the ionic strength of the loading buffer was reduced to 40 mM NaCl. The protein was stored at —80°C in 200 mM KCl/20 mM Tris-HCl, pH 7.3/5 mM 2-mercaptoethanol/5% (vol/vol) glycerol.

As judged from gel retardation experiments, LexA-C52 binds to a 50-bp SOS consensus operator fragment with about one-fourth to one-third the affinity of LexA wild-type repressor. The sequence of the synthetic operator duplex used for most of the experiments was

5′-gcaagtctggccctaccttaatctatataactagtaaccctgtaagggcgggg
3′-ctgctacgtagctggcaactttgatatatcgtcatcgctatggcattgcgcggcc,

where the two CTGT recognition motifs are shown in boldface letters. Some experiments were done with a slightly modified version of this duplex for which the two underlined base pairs were replaced by a GC-CG sequence to assess if all four nucleotides may be crosslinked in these positions.

Protein Modification with 4-Azidophenacyl Bromide. LexA-C52 (0.4 ml, 50 μM) was incubated for 3 hr with 3.5 mM 4-azidophenacyl bromide (Sigma) at 23°C in a buffer containing 100 mM KCl, 20 mM Mops at pH 8.0, 1.25% (vol/vol) glycerol, and 1.25 mM 2-mercaptoethanol, giving rise to LexA-C52*. Unreacted 4-azidophenacyl bromide was removed by chromatography on a Bio-Gel P-6DG (Bio-Rad) column (1 × 27 cm).

The degree of modification with phenyl azide was determined by measuring the amount of carboxymethylcysteine upon reaction with iodoacetic acid. The relative amount of carboxymethylcysteine with respect to the serine peak was 0 for LexA wild-type repressor containing no cysteine residue.

Abbreviations: HTH, helix–turn–helix; LexA-C52*, LexA mutant (Arg-52 → Cys) repressor modified with 4-azidophenacyl bromide. †To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
ues, 0.98 for LexA-C52 containing a single cysteine, and 0.9 again for LexA-C52 after modification with 4-azidophenacyl bromide, suggesting that essentially all the cysteine side chains have reacted with 4-azidophenacyl bromide and are no longer free to interact with iodoacetic acid.

**Protein-DNA Photocrosslinking.** LexA-C52* (3.2 μM) was incubated at 20°C for 30 min in a buffer containing 20 mM Mops at pH 7.3 and 100 mM KCl with an operator DNA fragment (50-mer), which was 32P-labeled at the 5′ end of either the top or the bottom strand. Aliquots (50 μl) were then withdrawn and irradiated with a 6-W Vilber Lourmat VL-6MC UV lamp (312 nm) at a distance of about 20 cm for 1, 3, or 10 min. The samples were further subjected to three extractions with phenol in which the aqueous phase was discarded, follow by a precipitation with ethanol and cleavage with piperidine (1 M) for 30 min at 90°C. No DNA cleavage products were observed without piperidine treatment, showing that phenyl azide crosslinking alone does not give rise to strand cleavage. Chemical G>A and G+A sequencing reactions were as described (25). High-resolution gel electrophoresis was carried out in 20% polyacrylamide denaturing gels.

**RESULTS**

In this study we use the photoactivatable crosslinking agent 4-azidophenacyl bromide to map LexA–DNA contacts. Since the wild-type LexA repressor does not contain any cysteine side chains, a mutant repressor had to be used for this study. In an earlier genetic study (15) we have identified a suitable candidate, LexA-C52, which harbors an Arg→Cys mutation in position 52. Gel retardation experiments with the purified LexA-C52 repressor showed that this LexA variant binds to the consensus SOS operator with about one-fourth to one-third the affinity of the LexA wild-type repressor (data not shown).

The purified LexA-C52 protein was modified with the cysteine-specific bifunctional photoactivatable crosslinking agent 4-azidophenacyl bromide under conditions in which essentially all the cysteine-52 side chains were modified, giving rise to a protein species that we will call in the following LexA-C52*. To assess protein-DNA crosslinking with this LexA derivative we have incubated LexA-C52* with a 32P-labeled duplex DNA harboring the consensus SOS operator sequence, irradiated the reaction mixture with 312-nm UV light, and analyzed the reaction products by SDS/polyacrylamide gel electrophoresis (Fig. 1). The amount of crosslinked DNA after 5 min of UV irradiation is about 2–3%, as judged from densitometry of the autoradiographs. Further irradiation does not increase the amount of crosslinked DNA, probably due to the photodegradation of the phenyl azide moiety.

To map the positions of the nucleotides which have been crosslinked to this LexA derivative, the crosslinked DNA has been subjected to cleavage with piperidine (25). Treatment of the crosslinked protein-DNA complex with piperidine at high temperature results in DNA cleavage at the crosslinked nucleotides. Fig. 2 shows that crosslinking on the top strand takes place preferentially at two adjacent nucleotides close to

---

**Fig. 1.** Autoradiograph of a denaturing (SDS) polyacrylamide gel, showing the formation, after UV irradiation for the indicated times, of a slowly migrating crosslinked complex between the purified LexA-C52 mutant repressor harboring a phenyl azide group in position 52 (LexA-C52*) and a 50-bp DNA duplex harboring the consensus operator sequence.

**Fig. 2.** Autoradiograph of a high-resolution denaturing polyacrylamide gel revealing the site-specific crosslinking of LexA-C52* to the top strand of a synthetic 50-bp DNA duplex containing the consensus SOS operator sequence. Two arrows indicate specific piperidine cleavage products for an adenine and a thymine base situated, respectively, 3 and 2 bases 5′ to the highly conserved CTGT motif.
the CTGT motif. The two symmetrically related nucleotides on the bottom strand are crosslinked with the same efficiency and specificity as those on the top strand (data not shown). Within both strands the crosslinked nucleotides are situated, respectively, 2 and 3 base pairs 5' to the first nucleotide of the highly conserved CTGT motifs. Crosslinking occurs with two adjacent A and T nucleotides on the top strand and two adjacent T nucleotides on the bottom strand.

For all four crosslinked nucleotides we observe two piperidine cleavage products with about equal intensity, one of which comigrates with the standard sequencing products whereas the other shows a slightly reduced electrophoretic mobility. It seems likely that the second species contains the same number of negatively charged phosphate groups but some additional mass, for example an additional deoxyribose moiety, leading to a slightly reduced electrophoretic mobility. Further work is necessary to elucidate the nature of the second species. The results obtained so far indicate that this species also when the crosslinked DNA is cleaved with NaOH (0.1 M, 30 min at 90°C) or with a piperidine concentration of 2 M instead of 1 M, or when photocrosslinking is induced with UV light of a different wavelength (254 nm or 360 nm instead of 312 nm).

To make sure that crosslinking is due to the formation of a specific LexA-DNA complex, two control experiments have been performed: First, no crosslinking was observed with a DNA fragment lacking the specific operator DNA site for LexA binding (data not shown). Second, photocrosslinking of LexA-C52Δ is suppressed upon addition of increasing amounts of unmodified LexA repressor (Fig. 3 Left). Several other control experiments (see Fig. 2) have been done, showing that crosslinking occurs only for LexA-C52Δ, not for LexA-C52 and LexA wild-type repressor.

A crosslinking agent might have some selectivity with respect to the target constituents. The experiments outlined above have shown that LexA-C52Δ may be crosslinked to both A and T. To assess if G and C may also be crosslinked in our system, we have replaced the reactive A and T on the top strand by G and C. Fig. 3 Right shows that G and C are indeed crosslinked in the same position and with comparable efficiency. We may conclude from these results that all four nucleotides may be efficiently photocrosslinked.

Fig. 4 summarizes the crosslinking contacts between the SOS consensus operator and the LexA derivative used in this study and suggests further how the putative LexA recogni-

Fig. 3. (Left) Autoradiograph of a high-resolution polyacrylamide gel showing that increasing amounts of unmodified LexA wild-type repressor suppress photocrosslinking of LexA-C52Δ to the top strand of the same DNA duplex as that used in Fig. 2. (Right) Autoradiograph of a high-resolution polyacrylamide gel showing that a 50-bp duplex harboring a GC sequence instead of AT in positions −11 and −10 results in the same crosslinking pattern as the AT-containing duplex.

**DISCUSSION**

We show that a derivatized amino acid side chain situated close to the COOH terminus of helix 3 of the LexA repressor

![Fig. 4. Diagram showing a juxtaposition of helix 3 of LexA and the SOS consensus operator sequence with a numbering system relative to the symmetry axis of the operator (•). The two CTGT recognition motifs are boxed and the crosslinking contacts are indicated by arrows. Amino acids are written in the one-letter code. The black dots indicate those phosphate groups for which ethylation suppresses complex formation (26).](image-url)
can be specifically crosslinked to nucleotides which are located close to the CTGT recognition motif of an SOS consensus operator. As inferred from methylation protection and interference experiments (6, 14, 27), LexA binds DNA predominantly within the major groove. Placing the COOH terminus of helix 3 within the major groove of standard B-DNA, such that the phenyl azide moiety in position 52 may contact the crosslinked nucleotides, very naturally brings the NH₂ terminus of helix 3 close to the CTGT recognition motif (Fig. 5). This finding is in good agreement with a recently discovered LexA mutant repressor with a different DNA recognition specificity (29), suggesting that the side chain of Glu-45 should recognize preferentially the C-G base pair in the first position of the CTGT recognition motif. Crosslinking and genetic data together allow us to realize a fairly precise docking of helix 3 into the major groove of the DNA and to compare its orientation with that of the recognition helix of a standard HTH protein. On the basis of the structure of the LexA DNA-binding domain we have built in a first step a physical model (using Maruzen molecular models) of the LexA-DNA complex, using a standard B-DNA duplex. Docking of the protein was achieved such that the exocyclic azide moiety of the phenyl ring attached to Cys-52 was positioned in the major groove of the DNA such that it may interact with the two crosslinked nucleobases. Glu-45 was postulated to form a hydrogen bond with the exocyclic NH₂ group of the cytosine base within the CTGT motif (base pairs ± 8 in the numbering system proposed in Fig. 4). Given the crosslinking data, docking of Glu-45 to the opposite guanine would lead to serious clashes between the NH₂ terminus of helix 3 and the DNA backbone. In a second step the physical model was reproduced on a Silicon Graphics workstation.

The result of this docking procedure is shown in Fig. 5. The left side shows our model for the complex between the LexA recognition helix and one operator half-site. The right side shows part of the cocrystal structure of the phage λ repressor-DNA complex (28). In both cases the protein

Fig. 5. (Left) An α-helix having the sequence of amino acids 39–54 of the LexA repressor (SPNAAEEHLKALARKG, residues 39–54) has been docked to a dodecamer of standard B-DNA (corresponding to the right half-operator shown in Fig. 4—a.e., base pairs 1–12) generated with INSIGHT II. According to recent NMR results helix 3 of LexA is essentially straight (R. Fogh, R. Kaptein, H. Rüterjans, M.G.-S., and M.S., unpublished work). The four base pairs of the CTGT motif are situated in the center of the dodecamer. The phenyl azide moiety (shown as the reactive nitrene intermediate formed after N₂ release upon UV irradiation) attached to the cysteine in position 52 (labeled as Cys) and the side chains of Glu-45, Asn-41, and Ser-39 are shown in yellow. To realize this docking the phenyl azide ring was placed into the major groove at about equal distance from the two crosslinked bases (labeled with red arrows) and Glu-45 was positioned close to the exocyclic NH₂ group of the cytosine base (labeled with a white arrow) of base pair 8. (Right) Part of the crystal structure of the complex of the phage λ repressor with its DNA-binding site (28). Only the recognition helix and two amino acids of the preceding and two amino acids of the following turn are represented (residues 42–54: MGQSGVGALFNGI). The first amino acid of the recognition helix (Gln) is labeled and, as in the case of the model of the LexA-DNA complex, the dyad axis of the operator is placed at the bottom of the figure.
segments (shown as purple ribbons) correspond to the recognition helix and the two NH$_2$-terminal amino acids of the preceding turn. Fig. 5 shows that the positioning of the LexA recognition helix should be fairly different from that observed for the phage $\lambda$ repressor recognition helix. The most obvious difference is that in the case of phage $\lambda$ repressor the axis of the recognition helix is almost parallel to the plane of the base pairs, whereas the LexA recognition helix is expected to form a pronounced angle with respect to the plane of the base pairs.

A second difference resides in the fact that the LexA recognition helix is oriented within the major groove of the DNA such that its NH$_2$ terminus is directed towards the palindromic center of the operator. All HTH proteins for which an x-ray structure of the complex is available are oriented in the opposite sense, with their NH$_2$ termini farther away from the center than their COOH termini (30). NMR, genetic, and affinity cleavage data suggest, however, that lac repressor (29, 31, 32) and possibly also tet repressor (33) might be oriented with their NH$_2$ termini directed towards the palindromic center.

Beyond the overall positioning of the LexA recognition helix within the major groove of the DNA, the model of the LexA-DNA complex shown in Fig. 5 suggests the following predictions concerning the recognition of the CTGT motif:

(i) Asn-41 should be located close to the adenine base of the second base pair of the CTGT motif (base pair 7), potentially allowing the formation of two hydrogen bonds between the amide group of Asn-41 and, respectively, the exocyclic NH$_2$ group and the N-7 atom of adenine 7. The methyl group of the opposite thymine base is also expected to play a role in LexA recognition, since its removal weakens DNA binding (14).

According to our model the formation of a hydrophobic contact between this methyl group and the methyl side chain of Ala-42 would be possible upon a minor distortion of the DNA structure.

(ii) The contacts with the third and fourth base pair of the CTGT motif (base pairs 6 and 5) are more hazardous to predict because these base pairs are farther from the crosslinking and genetic docking points. Ser-39 might contact the guanine base of the third base pair of the CTGT motif upon formation of a hydrogen bond between its hydroxyl group and the N-7 atom of guanine 6.

Independently of the precise details of recognition, the overall positioning of the LexA recognition helix is well established by the crosslinking and genetic data, and we may conclude that the binding mode of the LexA recognition helix is apparently different from the way the recognition helices of classical HTH proteins are inserted into the major groove.

We are grateful to Dr. R. H. Ebright for communicating many details concerning the cysteine-specific incorporation and the photocrosslinking of phenyl azide in the case of the catabolite gene activator protein prior to publication, to Dr. J. Reindolt for determining the carboxymethylcysteine content of the different LexA variants, and to Dr. R. Fogh for his introduction into the use of Insight II. This work was supported by grants from the European Community (ST2J-0254), the Institut National de la Santé et de la Recherche Médicale (871007), the Association de Recherche contre le Cancer, and the Ligue Nationale de la Lutte contre le Cancer.