

DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues

KENDALL L. KNIGHT AND ROBERT T. SAUER

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by Boris Magasanik, October 17, 1988

ABSTRACT The Arc and Mnt repressors of phage P22 are related proteins that bind to different operator DNA sites. By creating a hybrid Arc-Mnt protein, we show that the binding specificity of Mnt can be switched to that of Arc by replacing six residues at the N terminus of Mnt with the corresponding nine residues from Arc.

The Arc and Mnt repressors of bacteriophage P22 are small transcriptional regulatory proteins that show 40% sequence homology yet bind exclusively to different operator sequences (1-5). The three-dimensional structures of Arc and Mnt are not known, nor are the mechanisms by which these proteins bind to operator DNA. However, genetic and biochemical studies of mutant proteins have suggested that residues in the N-terminal regions of both proteins play important roles in operator DNA recognition and binding (6-8). The properties of an altered specificity mutant of Mnt indicate that residue 6 makes a specific contact with the *mnt* operator DNA (6, 7), whereas mutations at residues 2, 3, 4, 5, 8, and 10 of Arc cause large decreases in *arc* operator binding affinity without affecting the structure of the folded protein (8). To evaluate the extent to which these N-terminal residues serve as determinants of operator recognition, we have constructed a hybrid Arc-Mnt protein and have studied its operator binding properties. Our studies show that the binding specificity of Mnt can be switched to that of Arc by exchange of a small block of N-terminal residues.

MATERIALS AND METHODS

Strains, Plasmids, and Phages. All plasmids used in this study were propagated in *Escherichia coli* strain X90 (*ara*⁻ *Δlac-pro nala argEam rif^R thi*⁻; *F'*⁻ *lac*⁺ *lacI*^Q *pro*⁺) (9). Strain Y90 is a phage λ lysogen of X90. *E. coli* strain UA2F was used to assay for Arc activity *in vivo* (10). This strain contains a λimm21 prophage bearing a fusion of the Arc-repressible *P*_{ant} promoter to the gene for chloramphenicol acetyl transferase (*cat*). Plasmid pTM201 is a phage M13-origin plasmid that contains the wild-type *mnt* gene under transcriptional control of the *tac* promoter (3). Plasmid pTA200 is a related plasmid that contains the wild-type *arc* gene under transcriptional control of the *tac* promoter (3). λDA2 is a derivative of λKH54 (a *cI*⁻ phage), bearing the *immI* region of phage P22 in the *b2* region (11).

Construction of the Gene Encoding the Hybrid Protein. To construct the gene for the Arc-Mnt hybrid protein, oligonucleotide-directed mutagenesis was first used to introduce unique *Nco* I and *Sph* I restriction sites into the N-terminal coding region of the *mnt* gene of plasmid pTM201:

```

fMet Ala Arg Asp Asp Pro His Phe Asn Phe Arg Met Pro
GC ATG GCT AGA GAT GAT CCG CAC TTT AAC TTC CGT ATG CCT.
↓
CC ATG G GC ATG C
*****
Nco I Sph I
```

Directed mutagenesis procedures and detailed methods for oligonucleotide synthesis and purification are presented elsewhere (12). Two mutagenic oligonucleotides, each 17 bases in length and each carrying one of the single base changes indicated above, were phosphorylated and hybridized to pTM201 containing a single-stranded gap in the *P*_{*tac*}-*mnt* sequence, followed by extension with DNA polymerase I large fragment and ligation using T4 DNA ligase. The resulting plasmid, pTM201/NS3-3, was purified and digested with *Nco* I and *Sph* I, and the backbone was gel-purified. Oligonucleotides for a double-stranded cassette (see below) encoding the N-terminal 13 residues of Arc were synthesized by using an Applied Biosystems 380B DNA synthesizer.

```

Met Lys Gly Met Ser Lys Met Pro Gln Phe Asn Leu Arg Met
C ATG AAA GGA ATG AGC AAA ATG CCG CAG TTC AAT TTG CGC ATG
TTT CCT TAC TCG TTT TAC GGC GTC AAG TTA AAC GC
```

The oligonucleotides (0.2 nmol each) were mixed in a buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, and 100 mM NaCl and were annealed by heating to 70°C for 15 min and cooling successively at 37°C and room temperature for 15 min each. Equal molar amounts of this double-stranded cassette and the pTM201/NS3-3 backbone (0.2 pmol of each) were mixed in a buffer containing 50 mM Tris (pH 7.5), 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM spermidine, 1 mM ATP, and 100 μg of bovine serum albumin per ml and were incubated overnight at 16°C with 10 units of phage T4 DNA ligase. Transformation into strain X90 gave rise to ampicillin-resistant colonies containing the plasmid pTAM5-1, which carries the gene encoding the Arc-Mnt hybrid protein. Four of the 13 N-terminal residues of Arc (positions 8, 10, 11, and 13) are identical to the corresponding residues of Mnt when the sequences are aligned as shown in Fig. 1. In addition, hybrids bearing either Leu-12 (the Arc residue) or Phe-12 (the Mnt residue) have extremely similar properties (not shown), indicating that this substitution is functionally neutral. Thus, the results in this work are discussed in terms of a hybrid protein construct in which the N-terminal six residues of Mnt have been replaced by the N-terminal nine residues of Arc.

Protein Purification. The Arc-Mnt hybrid protein was purified from strain X90/pTAM5-1 after induction of *P*_{*tac*} with isopropyl β-D-thiogalactopyranoside. The hybrid protein was produced at a level of ≈2-5% of the total cell protein. Purification was performed essentially as described for the wild-type Mnt protein (13) with the following modifications. Crude lysate was dialyzed against a buffer containing 50 mM Tris (pH 7.5), 0.1 mM EDTA, 5% (vol/vol) glycerol, 1.4 mM

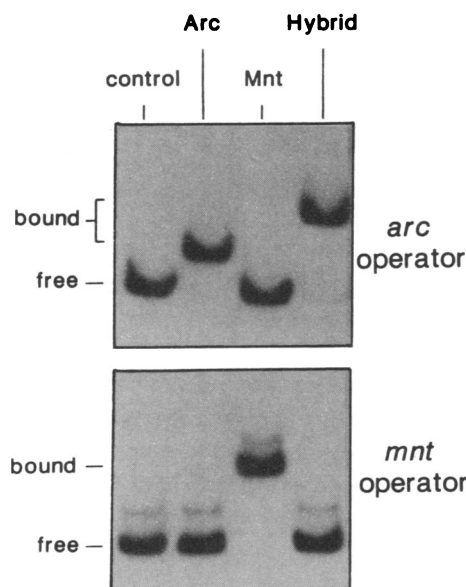


FIG. 2. Operator-binding specificity determined by using the gel mobility shift assay. End-labeled DNA containing either the *arc* operator (Upper) or *mnt* operator (Lower) was incubated in the absence of protein (control) or in the presence of Arc (0.2 μ M), Mnt (0.5 nM), or hybrid protein (0.5 nM). The products of each reaction were electrophoresed on a 5% acrylamide gel and visualized by autoradiography.

titrated against a constant amount of labeled DNA. The results of these studies are shown in Table 1. Both Arc and the hybrid protein bound strongly to the *arc* operator, whereas their binding to the *mnt* operator was no stronger than to nonoperator DNA (Table 1). Mnt, by contrast, bound strongly to the *mnt* operator and weakly to either the *arc* operator or nonoperator DNA. These results suggest that the information required for specific recognition of the *arc* operator is contained in the N-terminal nine residues of Arc. The hybrid protein actually showed stronger binding to the *arc* operator than did wild-type Arc (Table 1). This may be related in part to the different oligomeric states of the two proteins. Mnt and presumably the hybrid bind to the operator as preformed tetramers, whereas Arc must oligomerize before it binds the operator (4, 5). In addition, residues at the C terminus of Mnt contribute to high-affinity binding to the *mnt* operator (see Discussion) and may contribute similarly to the binding of the hybrid protein to the *arc* operator.

To compare the binding properties of the hybrid protein and Arc in greater detail, we performed hydroxyl radical protection experiments to probe contacts with the DNA backbone and methylation protection experiments to probe contacts in the major groove. Cleavage of DNA by hydroxyl radicals occurs specifically at deoxyribose moieties. Protection against this cleavage will result if a bound protein is in close proximity to the DNA backbone (16). In Fig. 3 we show

Table 1. DNA binding affinities of Mnt, Arc, and the hybrid proteins

DNA	Mnt	Arc	Hybrid
Operator			
<i>arc</i>	10,000	100	2.5
<i>mnt</i>	1	10,000	5,000
Nonoperator	10,000	10,000	5,000

These values are the relative protein concentrations at which half of the operator or nonoperator DNA is bound as determined by the gel mobility shift assay. Values for Mnt and the hybrid protein indicate concentrations of tetramer, whereas values for Arc indicate protein monomer. A value of 1 = 20 pM.

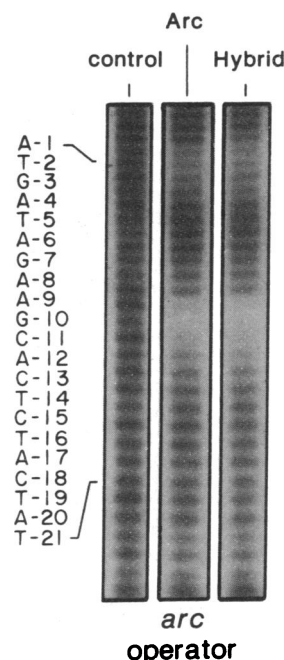


FIG. 3. Hydroxyl radical protection experiments. End-labeled DNA containing the *arc* operator was incubated in the absence of protein (control) or in the presence of Arc (0.1 μ M) or hybrid protein (1 nM). After reaction with hydroxyl radicals, DNA cleavage products were recovered, electrophoresed on an 8% acrylamide sequencing gel, and visualized by autoradiography. Identification of the bands was determined by alignment with Maxam and Gilbert sequencing controls (not shown).

hydroxyl radical cleavage of the *arc* operator in the presence of Arc or the hybrid protein. These experiments and similar ones in which the opposite strand was end-labeled showed that Arc and the hybrid protein protect the same set of deoxyribose positions in both the outer and central regions of the *arc* operator sequence. This indicates that the DNA binding regions of Arc and the Arc-Mnt hybrid protein interact with the backbone of the *arc* operator in the same manner. The backbone positions protected by both proteins lie along one face of the operator helix modeled as B-form DNA.

Methylation protection studies were performed by using a DNA fragment that contains both the *arc* and *mnt* operators, arranged tandemly as they are found in the phage P22 genome. Both Arc and the hybrid protein strongly protected guanine-18 and failed to protect guanine-13 and -11 in the *arc* operator (Fig. 4, lanes 3 and 4). As expected, neither protein protected bases in the *mnt* operator. Within the *arc* operator, the hybrid protein did not show the weak protection of guanine-15 seen with Arc, and differences were also observed at guanine-3 and -7 on the other strand (not shown). These differences in methylation protection indicate that the local major groove packing near certain guanine bases is somewhat different for Arc and the hybrid protein. However, contacts between Arc and the N-7 positions of these guanine bases cannot be critical, as mutations at position three or premethylation of guanine-15 or guanine-7 cause only slight decreases in Arc binding (3, 4). Because the hybrid protein bound to the *arc* operator with high specificity, we infer that it must make most if not all of the key recognition contacts with the operator.

Operator Specificity of the Hybrid Protein *in Vivo*. The results presented thus far show that Arc and the hybrid protein bind specifically to the *arc* operator *in vitro*. Does this specificity of operator recognition pertain *in vivo*? To answer

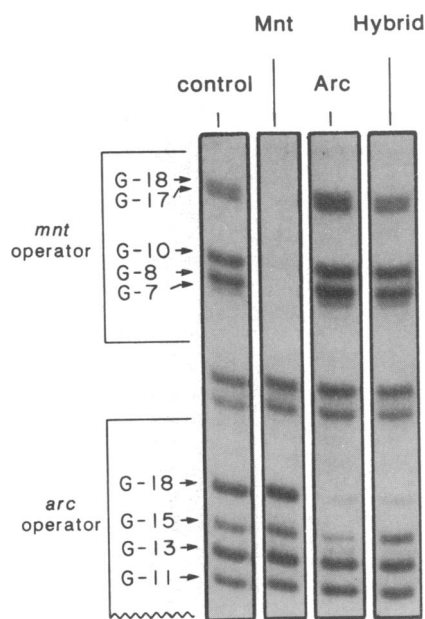


FIG. 4. Methylation protection experiments. End-labeled DNA containing adjacent *arc* and *mnt* operators was incubated in the absence of protein (control) or in the presence of Mnt (40 nM), Arc (400 nM), or hybrid protein (2 nM). After methylation and DNA cleavage, products were electrophoresed on an 8% acrylamide sequencing gel and visualized by autoradiography.

this question, we expressed Arc, Mnt, and the hybrid protein from plasmids and assayed their activities in the cell. To measure Arc activity, we used a strain in which Arc can repress expression of the *cat* gene and thereby render the cell sensitive to chloramphenicol (10). Such strains were chloramphenicol-sensitive when they contained either Arc or the hybrid protein (Fig. 5 Upper). Thus, the hybrid protein has Arc activity *in vivo*. In the same assay, the strain containing Mnt was chloramphenicol resistant. Hence, the hybrid protein behaved *in vivo* as expected on the basis of its binding to the *arc* operator *in vitro*.

Mnt activity can be assayed in a phage λ lysogen of *E. coli* by the resistance of the cell to superinfection by λ DA2, a phage carrying the *imm1* region of bacteriophage P22 (11). In this assay, Mnt prevents the synthesis of P22 anti-repressor, which, if expressed, inactivates the prophage repressor and permits growth of the superinfecting phage (1, 17). Mnt prevented growth of λ DA2, whereas Arc and the hybrid protein allowed the superinfecting phage to grow (Fig. 5 Lower). Thus, in this system Arc and the hybrid protein behaved similarly in that neither showed the ability to negatively regulate anti-repressor synthesis. Again, the operator binding specificities observed for each of these proteins *in vitro* also pertained *in vivo*.

DISCUSSION

We have shown that the operator-binding specificity of the Mnt repressor can be changed to that of the Arc repressor by simple exchange of a few N-terminal amino acids. Specifically, when the six N-terminal residues of Mnt are replaced by the corresponding nine residues of Arc, the resulting hybrid protein acquires the binding specificity of Arc. These results show that residues in the N-terminal regions of both Arc and Mnt are the primary determinants of operator-binding specificity for these related repressors. That the N-terminal residues of Arc can function in the context of a hybrid protein composed mainly of Mnt residues further suggests that Arc and Mnt present their DNA-binding regions

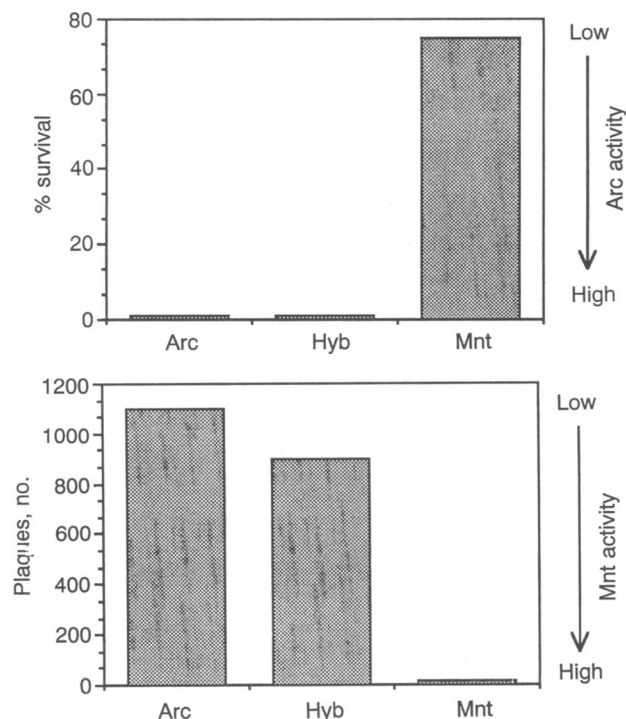


FIG. 5. Operator-binding activities *in vivo*. (Upper) Ability of Arc, Mnt, and the hybrid protein to confer chloramphenicol sensitivity to strain UA2F. The ordinate represents the fraction of transformants that survive in the presence of chloramphenicol at 100 μ g/ml. Low survival values correspond to high Arc activity. (Lower) Ability of the proteins to prevent lytic growth of λ DA2 in strain Y90. The ordinate represents the number of progeny phage produced 60 min after infection. Low numbers of phage correspond to high Mnt activity.

to their respective operators in the same basic manner. This, in turn, indicates that the two proteins must have similar tertiary and quaternary structures.

The *arc* and *mnt* operators show some homology (6 of 21 base pairs), but studies of mutant operators indicate that the base pairs that are most important for the recognition of each operator by its cognate repressor are not conserved (3, 18, 21). These functionally important base pairs are boxed in the operator sequences below.



Recognition of these bases probably occurs via major groove interaction, as each of the N7 positions at the functionally important guanine bases in the *mnt* operator is protected against methylation by bound Mnt (5). Our results indicate that the ability of each repressor to discriminate between these two operators is determined entirely by residues at the N terminus. We suggest, therefore, that this selectivity is accomplished by the interaction of N-terminal residues of Arc and Mnt with the functionally important base positions in the major groove of their respective operators.

The altered specificity experiments of Youderian *et al.* (6) originally suggested that His-6 of Mnt played a key role in recognition of the *mnt* operator. They found that an Mnt mutant bearing a His-6 \rightarrow Pro substitution bound with wild-type affinity to an *mnt* operator bearing symmetric

operator mutations at base pairs 5 and 17. Moreover, the strong binding of the Pro-6 mutant protein to the mutant operator was dependent upon N-6-methylation at base pairs 5, 6, 16, and 17 (7). As the N-6 position is in the major groove of the DNA, this result provides additional evidence for major-groove recognition of the operator at this site.

The finding that the N-terminal residues of Arc and Mnt are critical determinants of operator binding specificity does not mean that these residues are the only ones involved in operator binding. First, the observed patterns of contact between Arc or Mnt and their respective operators extend over 65–70 Å (4, 5), and it is not possible that all of these contacts could be mediated by a few N-terminal residues. There are undoubtedly other residues in the C-terminal regions of Arc and Mnt that make contacts with the DNA and thus contribute to the affinity of operator binding (8, 12). For example, Lys-79 in Mnt appears to be a critical residue for maintaining high-affinity operator binding (12). Second, the N-terminal residues of Arc and Mnt almost certainly need to be presented to their operators in the proper tertiary and quaternary contexts. We know, for example, that mutations throughout Arc are able to cause a repressor-defective phenotype, apparently by destabilizing the protein (8). We also have found that a synthetic peptide corresponding to the 13 N-terminal residues of Mnt shows no detectable operator binding *in vitro* (J. Bowie, K.L.K., and C. Pabo, unpublished results) and that a protein containing the N-terminal 13 residues of Arc fused to the C-terminal dimerization domain of phage λ repressor does not bind to the *arc* operator *in vivo* (unpublished results).

The ability of Arc and Mnt to switch operator-binding specificities as a consequence of exchange of only a few amino acid residues is reminiscent of the results of Wharton and Ptashne (19). They showed that the operator specificity of the phage 434 repressor, a helix-turn-helix protein, could be changed to that of the related P22 repressor by changing 4 residues in phage 434-repressor to the corresponding residues from phage P22 repressor. Each of the four residue changes altered side-chains in the "recognition" α -helix of phage 434 repressor. Moreover, the crystal structure of the 434 repressor-operator complex showed that two of the altered side chains are directly involved in recognition of the 434 operator by the 434 repressor (20). In our studies, the N-terminal regions of Arc and Mnt appear to play roles analogous to the recognition α -helices of the 434 and P22 repressors. In the 434/P22 case, the identity of residues in the recognition helix of 434 repressor determines the protein's ability to distinguish between the 434 and P22 operators. In the Arc-Mnt case, the identity of N-terminal residues determines whether the *arc* or *mnt* operators will be recognized. It is important to note, however, that the Arc and Mnt repressors do not appear to be helix-turn-helix repressors (2, 8). In fact, NMR studies suggest that at least a portion of the

N-terminal operator-binding region of Arc assumes an extended conformation in solution (M. Zagorsky and D. Patel, personal communication).

We thank Jim Bowie, Jim Hu, Robin Kelley, Wendell Lim, Dawn Parsell, and Drew Vershon for advice, information, help with the figures, and comments on the manuscript. This work was supported by National Institutes of Health Grant AI-16892. K.L.K. was a fellow of the Charles A. King Trust.

1. Susskind, M. M. & Youderian, P. (1983) in *Lambda II*, eds. Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 347–366.
2. Sauer, R. T., Krovatin, W., DeAnda, J., Youderian, P. & Susskind, M. M. (1983) *J. Mol. Biol.* **168**, 699–713.
3. Vershon, A. K. (1986) Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge, MA).
4. Vershon, A. K., Liao, S.-M., McClure, W. R. & Sauer, R. T. (1987) *J. Mol. Biol.* **195**, 323–331.
5. Vershon, A. K., Liao, S.-M., McClure, W. R. & Sauer, R. T. (1987) *J. Mol. Biol.* **195**, 311–322.
6. Youderian, P., Vershon, A. K., Bouvier, S., Sauer, R. T. & Susskind, M. M. (1983) *Cell* **35**, 777–783.
7. Vershon, A. K., Youderian, P., Weiss, M. A., Susskind, M. M. & Sauer, R. T. (1985) in *Specificity in Transcription and Translation*, eds. Calendar, R. & Gold, L. (Liss, New York), pp. 209–218.
8. Vershon, A. K., Bowie, J. U., Karplus, T. M. & Sauer, R. T. (1986) *Proteins: Struct. Funct. Genet.* **1**, 302–311.
9. Amman, E., Brosius, J. & Ptashne, M. (1983) *Gene* **25**, 167–178.
10. Vershon, A. K., Blackmer, K. & Sauer, R. T. (1986) in *Protein Engineering: Applications in Science, Medicine, and Industry*, eds. Inouye, M. & Sarma, R. (Academic, Orlando, FL), pp. 243–256.
11. De Anda, J. L. (1985) Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge, MA).
12. Knight, K. L. & Sauer, R. T. (1988) *Biochemistry* **27**, 2088–2093.
13. Vershon, A. K., Youderian, P., Susskind, M. M. & Sauer, R. T. (1985) *J. Biol. Chem.* **260**, 12124–12129.
14. Fried, M. & Crothers, D. M. (1981) *Nucleic Acids Res.* **9**, 6505–6525.
15. Garner, M. M. & Revzin, A. (1981) *Nucleic Acids Res.* **9**, 3047–3060.
16. Tullius, T. D. & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5469–5473.
17. Botstein, D., Lew, K. K., Jarvik, V. & Swanson, C. A. (1975) *J. Mol. Biol.* **91**, 439–462.
18. Youderian, P., Moyle, H. & Susskind, M. M. (1988) *Genetics*, in press.
19. Wharton, R. P. & Ptashne, M. (1985) *Nature (London)* **316**, 601–605.
20. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326**, 846–852.
21. Vershon, A. K., Kelley, R. D. & Sauer, R. T. (1989) *J. Biol. Chem.*, in press.