PROTEINS: Structure, Function, and Genetics 1:302-311 (1986)

# Isolation and Analysis of Arc Repressor Mutants: Evidence for an Unusual Mechanism of DNA Binding

**Andrew K. Vershon, James U. Bowie, Theresa M. Karplus, and Robert T. Sauer**
*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

*ABSTRACT* **We have isolated 64 different missense mutations at 36 out of 53 residue positions in the Arc repressor of bacteriophage P22. Many of the mutant proteins with substitutions in the C-terminal 40 residues of Arc have reduced intracellular levels and probably have altered structures or stabilities. Mutations in the N-terminal ten residues of Arc cause large decreases in operator DNA binding affinity without affecting the ability of Arc to fold into a stable three-dimensional structure. We argue that these N-terminal residues are important for operator recognition but that they are not part of a conventional helix-turn-helix DNA binding structure. These results suggest that Arc may use a new mechanism for sequence specific DNA binding.**

**Key words: mutant proteins, protein stability, operator binding**

## INTRODUCTION

Individual residues in protein sequences can play many roles. Some residues are required to form and/or stabilize the correctly folded three-dimensional structure. Others contribute mainly to functions such as ligand binding or catalysis. Still other residues appear to play no active role in either folding or activity, and probably serve only to connect one determinant of the structure to another. One approach toward understanding the relationships among protein sequence, structure, and function is to use classical genetic methods to isolate missense mutations that block function and then to use biochemical and physical methods to determine how a given sequence change affects the folding or activity of the purified mutant protein. Here, we report on such an analysis of the Arc repressor of bacteriophage P22.

Arc is a small, dimeric DNA binding protein (53 residues/monomer) that represses transcription from the $P_{ant}$ promoter of bacteriophage P22.[1-3] Arc binds to a single operator site, which partially overlaps the $-10$ and $-35$ regions of $P_{ant}$,[4] and appears to repress transcription by preventing the isomerization of bound RNA polymerase from a closed to an open complex (S.-M. Liao and W.R. McClure, personal communication). The three-dimensional structure of Arc is not known, but crystals have been grown and work on the structure is in progress.[5]

In this paper we describe the isolation of randomly generated mutations in the *arc* gene, the purification of twelve mutant Arc proteins, and the stability and DNA binding properties of these mutant proteins. Our results suggest that many of the primary determinants of DNA binding are located within the first ten residues of Arc, since mutations at these residues cause large decreases in the affinity of operator binding without affecting Arc's ability to fold into a stable three-dimensional structure. It is extremely unlikely that these residues of Arc form part of a standard helix-turn-helix DNA binding structure, and it therefore appears that Arc is a member of a different class of site-specific DNA binding proteins.

## MATERIALS AND METHODS

### Strains

*Escherichia coli* strain UA2F is a derivative of strain US3 (*thi*$^-$his$^-$lacZ$^-$*lac*Y$^+$ *sup*$^\circ$ *rec*A$^-$; ref. 6) that contains an F′ episome (*lac*I$^Q$ *lac*Z::Tn5 [*kan*$^R$] *pro*$^+$) and a λAC201 prophage (*imm*$^{21}$ containing a $P_{ant}$ promoter fusion to *cat*; ref. 7). The other *E. coli* strains used in this work are X90 (*ara*$^-$Δ*lac-pro nal*A *arg*E*am rif*$^R$ *thi*1$^-$; F′$^-$*lac*$^+$ *lac*I$^{Q1}$ *pro*$^+$; ref. 8); AP401 (X90, *lon*:minitet (tet$^R$); A. Pakula, unpublished); and CAG456 (*lac*Z*am trp*a*m pho*a*m sup*C$^{ts}$ *mal rps*L *phe htp*R165; ref. 9).

Plasmid pTA200[7] is a pBR322 derivative containing a $P_{tac}$-*arc* fusion and the M13 origin of replication and packaging site from plasmid pZ150.[10]

### Mutagenesis and Mutant Selection

Hydroxylamine or *mut*T mutagenesis of plasmid pTA200 followed standard procedures.[6,11] Oligonucleotide directed mutagenesis was performed by synthesizing seven mutagenic primers (23–26 bases), which spanned the *arc* gene, with nucleoside triphosphate stocks containing 1% contaminants of each of the other bases (i.e., 97% G, 1% A, 1% C, 1% T; ref. 7). These primers were annealed to single-stranded pTA200; second-strand synthesis was carried out[12]; and the DNA was transformed into a strain defective in mismatch repair. Plasmid DNA was purified from this strain and used for mutant selection.

Spontaneous mutations in pTA200 were obtained by preparing a 50-fold dilution of an overnight culture of strain X90/pTA200 in 2 ml LB broth; infecting with $2 \times 10^7$ plaque-forming units (pfu) of phage M13

Received October 27, 1986; accepted November 26, 1986.

Address reprint requests to Robert T. Sauer, Dept. of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

RV1[13]; and growing for 6 h at 37°C. A transducing lysate, which contained approximately equal quantities of packaged single-stranded pTA200 and M13, was then prepared by pelleting the cells and sterilizing the supernatant for 30 min at 65°C. Such lysates contain pTA200 mutants at a frequency of about 0.1%. These mutations probably arise from depurination during the growth or handling of the single-stranded plasmid DNA.[14]

Arc$^-$ derivatives of pTA200 were obtained by transforming or transfecting mutagenized pTA200 DNA into strain UA2F and selecting for growth at 37°C on LB plates supplemented with ampicillin (100 μg/ml), kanamycin (50 μg/ml), chloramphenicol (100 μg/ml), and isopropyl B-D-thio-galactopyranoside (IPTG) (0.4 μg/ml). Candidate colonies were purified by restreaking on selection plates, and overnight cultures were grown in LB broth supplemented with ampicillin (100 μg/ml) and kanamycin (50 μg/ml). Single-stranded plasmid DNA was purified from a M13 RV1 transducing lysate of these cultures as described,[7] and the mutant DNA was sequenced by the dideoxy procedure[15] with a primer that hybridized to a region 3′ to the end of the *arc* gene.

### Measurement of Arc Levels In Vivo

The level of Arc protein in lysates of cells (X90, AP401, or CAG456) containing pTA200 or mutant derivatives was assayed by sodium dodecyl sulfate (SDS)-polyacrylamide electrophoresis. Overnight cultures were diluted into LB broth (or LB broth containing 3× NaCl) supplemented with ampicillin (100 μg/ml) and grown to late log at 30°C or 37°C. IPTG was added to a final concentration of 100 μg/ml and growth was continued for 2.5h. One ml of the culture was centrifuged for 5 min in a microfuge, and the cell pellet was suspended in 200 μl of sample buffer (3% SDS, 65 mM Tris-HCl (pH 7.5), 10% glycerol, 5% 2-mercaptoethanol, and 0.002% bromphenol blue dye), and heated for 2 min at 90°C. After centrifugation in a microfuge for 15 min, the viscous pellet was removed with a toothpick, and 20 μl of the remaining lysate was electrophoresed on a 15% SDS-polyacrylamide gel.[16] Proteins were visualized after staining with Coomassie blue, and Arc was identified by comparison with purified Arc protein, which had been run on the same gel.

### Protein Purification

Purification of the Arc mutants was performed as described for wild-type Arc by Vershon *et al.*[3] with the following modifications. Mutants in strain X90 or AP401 were grown at 37°C, in 10 liters of LB supplemented with 1 g ampicillin. When the culture reached an absorbance of 1.0 at 600 nm, 2 g of IPTG was added to induce expression from the $P_{tac}$ promoter, and growth was continued for another 2.5 h. The cells were harvested, lysed, and carried through the first ammonium sulfate precipitation as described previously.[3] Back extractions of the ammonium sulfate pellet were not performed. Instead, the ammonium sulfate pellet was suspended with 40–50 ml of PCB (50 mM Tris-HCl [pH 8.0], 0.1 mM EDTA, 5% glycerol, 1.4 mM 2-mercaptoethanol), the residual precipitate was removed by centrifugation at 10,000 rpm in a Sorval SA-600 rotor for 30 min, and the supernatant was chromatographed on a Sephadex G75 column equilibrated in PCB plus 200 mM KCl. Gel filtration and phosphocellulose chromatography were performed as described.[3] The pooled fractions from the phosphocellulose column were loaded on a CM-Sephadex (C-50) column, washed in 2 column volumes of SB buffer (50 mM Tris-HCl [pH 7.5], 0.1 mM EDTA, 5% glycerol, 1.4 mM 2-mercaptoethanol) with 50 mM KCl, and a gradient was developed from 50 mM to 1,000 mM KCl. Column fractions containing Arc protein were pooled, dialyzed into 10 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, and 50 mM KCl, and stored at −20°C. The concentrations of the mutant proteins were determined by using a molar extinction coefficient at 280 nm of 7800.

Some of the mutant proteins eluted from the cation exchange columns at different salt concentrations than wild-type protein. The KT2 and RH23 proteins should be less positively charged than wild-type Arc, and both proteins eluted from the cation exchange columns at a lower salt concentration than wild type. GR3, EA17, and EA43 should be more positively charged, and each protein eluted from the columns at higher salt concentrations than wild type.

### Protein Stability

Thermal unfolding of the mutant proteins was monitored by changes in the molar ellipticity at 222 nm, with an AVIV model 60DS CD polarimeter. For these studies, the protein samples were diluted to 50 μg/ml in a buffer containing 10 mM potassium acetate (pH 4.0) and 100 mM KCl. (These experiments were performed at low pH because denaturation is reversible at pH 4, but only partially reversible at pH 7. The $T_m$ for wild-type Arc is approximately 20°C higher at pH 7.5 than at pH 4.) Melting curves were generated from $\Theta_{222}$ measurements taken at 5° intervals between 5°C and 75°C. The RH23 mutant was found to be unfolded at the lowest temperature tested. After denaturation, the samples were slowly cooled to room temperature and the circular dichroism (CD) spectra were found to be identical to those recorded before denaturation.

Denaturation induced by guanidine-hydrochloride was monitored by changes in intrinsic tryptophan fluorescence, by means of a Perkin-Elmer MPF-3 Fluorescence Spectrophotometer. The emission spectra of native Arc has a maximum at 332 nm, and the intensity of this peak shows a linear dependence on Arc concentration. Upon denaturation in 6 M GuHCl, the emission maximum shifts to 348 nm and the emission intensity at 332 nm decreases almost threefold. Aliquots of 6 M GuHCl were added to protein samples (100 μg/ml in 10 mM Tris-HCl [pH 7.5], 50 mM KCl)

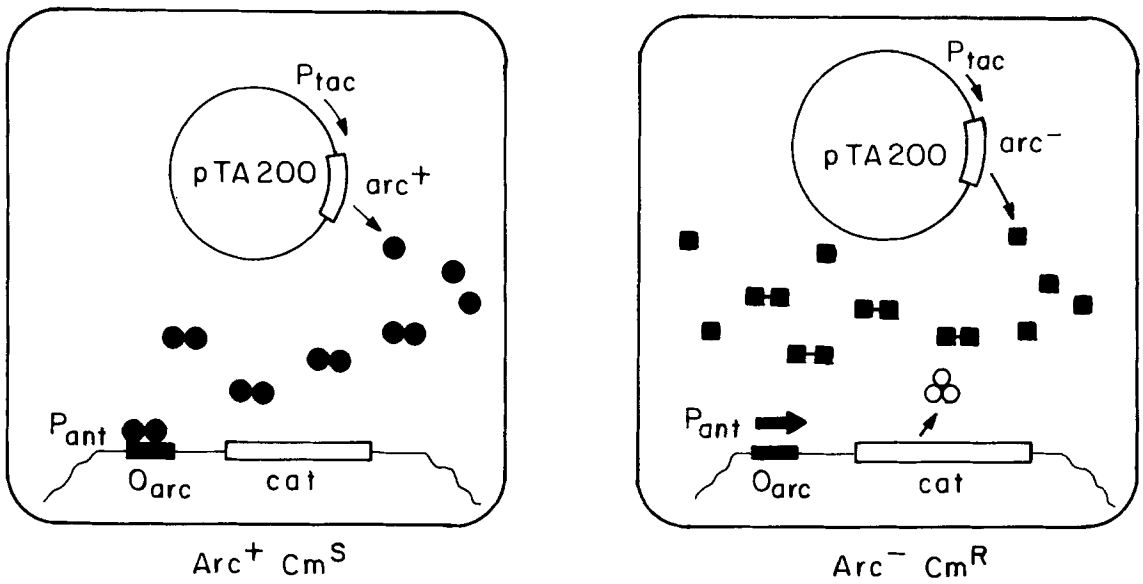


Fig. 1. Mutant selection strategy.

and the emission intensity at 332 nm was measured. Intensities were corrected to account for dilution of the protein sample.

### Operator DNA Binding Assays

The binding of mutant Arc proteins to operator DNA was monitored by the gel mobility shift assay or by pancreatic DNase I protection experiments, with the methods described previously for wild-type Arc.[4,17] All binding experiments were performed in a buffer containing 10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 0.1 mM EDTA, 100 mM KCl, and 100 μg/ml bovine serum albumin (BSA). The 250-base pair (bp) *Eco*RI/*Hin*dIII fragment from plasmid pI0101 was used as a source of operator DNA; the 752-bp *Eco*RI/*Pst*I fragment from pBR322 was used as a source of nonspecific DNA. These fragments were end-labeled with DNA polymerase I large fragment and deoxyadenosine 5′-[$\alpha$-$^{32}$P]triphosphate ($\alpha$-$^{32}$Pd-ATP) and deoxythymidine 5′-[$\alpha$-$^{32}$P]triphosphate ($\alpha$-$^{32}$P-dTTP).

## RESULTS

### Strategy for Mutant Isolation

The primary goal of this study was to generate single amino acid changes in the Arc repressor and to examine the effects of these changes on protein folding and DNA binding activity. Figure 1 outlines the strategy that we have used to obtain mutations in the *arc* gene. Arc protein is expressed from plasmid pTA200,[7] which contains the *arc* gene fused to the hybrid $P_{tac}$ promoter.[8] Following mutagenesis, plasmid DNA is transformed into a selection strain, which harbors a fusion of the $P_{ant}$ promoter to the gene encoding chloramphenicol acetyl transferase (*cat*). In cells transformed with $Arc^+$ plasmids, $P_{ant}$ is repressed by Arc and the cells are sensitive to chloramphenicol. Cells transformed with $Arc^-$ plasmids can be selected by their resistance to chloramphenicol since the *cat* gene is expressed. To obtain as broad a spectrum of mutations as possible, we used a variety of mutagenic methods, including hydroxylamine treatment *in vitro*, oligonucleotide mutagenesis with primers containing random base substitutions, growth in the *mut*T mutator strain, and propagation of the *arc* gene in a single-stranded form (see Methods).

### $Arc^-$ Mutations

DNA sequence changes were determined for 176 independent $Arc^-$ mutations. Of these, 125 are single missense mutations, 24 are small insertions or deletions in the *arc* gene or promoter region, 15 are nonsense mutations, and 9 are base changes in the promoter or ribosome binding site regions. In addition, two mutants contain double or triple missense changes, and two mutations alter the natural TAA termination codon. These latter mutations (termed +YS55) extend the *arc* reading frame by two codons and result in addition of a Tyr-Ser dipeptide as residues 54 and 55 of the mutant Arc protein.

Table I lists each different Arc missense mutation or nonsense mutation, the base changes, the method of mutagenesis employed, and the number of independent isolates. The missense mutations are designated by the one-letter code for the wild-type and mutant amino acids, and the residue position of the sequence change. For example, the KT2 mutation changes the wild-type lysine (K) at residue 2 of Arc to threonine (T).

The collection of $Arc^-$ mutations contains 64 different missense mutations. As shown in Figure 2, the missense changes are distributed throughout the Arc sequence and affect 36 of the possible 52 positions (missense mutations cannot be obtained in the initia-

**TABLE I. Arc – Mutations***

| Mutant | Position (bp) | Change | Mutagen | No. isolated | In vivo levels |
|---|---|---|---|---|---|
| KQ2 | 4 | A→C | MT | 1 | H |
| KT2 | 5 | A→C | MT | 1 | H |
| GK3 | 7, 8 | GG→AA | HA | 1 | H |
| GR3 | 7 | G→A | HA | 2 | H |
| ML4 | 10 | A→C | MP | 1 | ND |
| MR4 | 11 | T→G | MT | 1 | M |
| MI4 | 12 | G→A | HA | 1 | H |
| SR5 | 13 | A→C | MT | 1 | L |
| SC5 | 13 | A→T | MP | 1 | H |
| PL8 | 23 | C→T | HA | 2 | H |
| Qam9 | 25 | C→T | HA | 7 | ND |
| QL9 | 26 | A→T | MP | 1 | L |
| FV10 | 28 | T→G | MP | 1 | H |
| FC10 | 29 | T→G | MT | 2 | H |
| FS10 | 29 | T→C | MP | 1 | L |
| NK11 | 33 | T→G | MT | 1 | L |
| Lam12 | 35 | T→A | SP | 2 | ND |
| LW12 | 35 | T→G | MT | 5 | L |
| LF12 | 36 | G→T,C | SP, MP | 3 | L |
| RW13 | 37 | C→T | SP, MP | 6 | L |
| RQ13 | 38 | G→A | HA | 3 | L |
| WR14 | 40 | T→C | SP, MP | 4 | L |
| Woc14 | 41, 42 | GG→AA | HA | 1 | ND |
| Wop14 | 42 | G→A | HA | 1 | ND |
| PS15 | 43 | C→T | HA | 1 | L |
| PL15 | 44 | C→T | HA,SP,MP | 10 | L |
| EA17 | 50 | A→C | MT | 1 | M |
| VE18 | 53 | T→A | SP | 1 | ND |
| VG18 | 53 | T→G | MT | 1 | M |
| DN20† | 58 | G→A | HA | 1 | L |
| DY20 | 58 | G→T | MP | 1 | L |
| DA20 | 59 | A→C | MT, MP | 2 | M |
| LF21 | 63 | G→T | MP | 1 | M |
| RC23 | 67 | C→T | HA, SP | 7 | L |
| RH23‡ | 68 | G→A | HA | 1 | L |
| VI25 | 73 | G→A | HA | 1 | M |
| VG25 | 74 | T→G | MT | 2 | L |
| AT26 | 76 | G→A | HA | 1 | L |
| AV26 | 77 | C→T | SP | 1 | L |
| EK28 | 82 | G→A | HA | 3 | L |
| Eam28 | 82 | G→A | SP | 1 | ND |
| EA28 | 83 | A→C | MT | 1 | M |
| NY29 | 85 | A→T | MP | 1 | L |
| NT29 | 86 | A→C | MT, MP | 2 | L |
| NK29 | 87 | T→G | MT | 1 | L |
| RW31 | 91 | C→T | HA,SP,MP | 5 | L |
| RL31 | 92 | G→T | SP | 1 | L |
| SA32 | 94 | T→G | MT | 1 | M |
| SF32 | 95 | C→T | HA | 1 | L |
| VG33 | 98 | T→G | MP | 1 | L |
| NH34 | 100 | A→C | MT | 2 | L |
| NK34 | 102 | T→A | MP | 3 | L |
| EK36 | 106 | G→A | HA, MP | 2 | L |
| EA36 | 107 | A→C | MT | 1 | L |
| EG36 | 107 | A→G | MP | 2 | L |
| IM37 | 111 | T→G | MT | 3 | L |
| YD38 | 112 | T→G | MT | 2 | L |
| YS38 | 113 | A→C | MT | 4 | L |
| YC38 | 113 | A→G | MP | 1 | L |
| Qam39 | 115 | C→T | HA, SP | 4 | ND |
| QP39 | 116 | A→C | MP | 2 | L |

*(continued)*

**TABLE I. Arc−Mutations* (continued)**

| Mutant | Position (bp) | Change | Mutagen | No. isolated | In vivo levels |
|---|---|---|---|---|---|
| Rop40 | 118 | C→T | HA | 1 | ND |
| RQ40 | 119 | G→A | HA | 1 | L |
| VA41 | 122 | T→C | MP | 1 | L |
| Eoc43 | 127 | G→T | MP | 1 | ND |
| EG43 | 128 | A→G | MP | 1 | L |
| EA43 | 128 | A→C | MT | 1 | M |
| FS45 | 134 | T→C | MP | 2 | L |
| FC45 | 134 | T→G | MT | 4 | L |
| KT46 | 137 | A→C | MP | 1 | M |
| EK48 | 142 | G→A | HA | 2 | L |
| RP50 | 149 | G→C | MP | 1 | L |
| +YS55** | 162 | A→C | MT | 2 | H |

*am, amber mutation; oc, ochre mutation; op, opal mutation. The base-pair numbers refer to Figure 1, where the A of the initiator ATG is assigned as base 1. HA, hydroxylamine; MT, *mut*T; MP, oligonucleotide directed mutagenesis with mutant primers; SP, spontaneous mutations arising from single-stranded DNA. Levels refer to the steady-state intracellular concentrations of each mutant protein—H, high; M, moderate; L, low; or ND, not determined, as described in the text.

†DN20 contains an additional silent change (G→A) at base pair 57.

‡RH23 contains a silent mutation (G→A) at base pair 116.

**+YS55 arises from a mutation in the termination codon, which results in addition of Tyr54 and Ser55 to the wild-type Arc sequence.

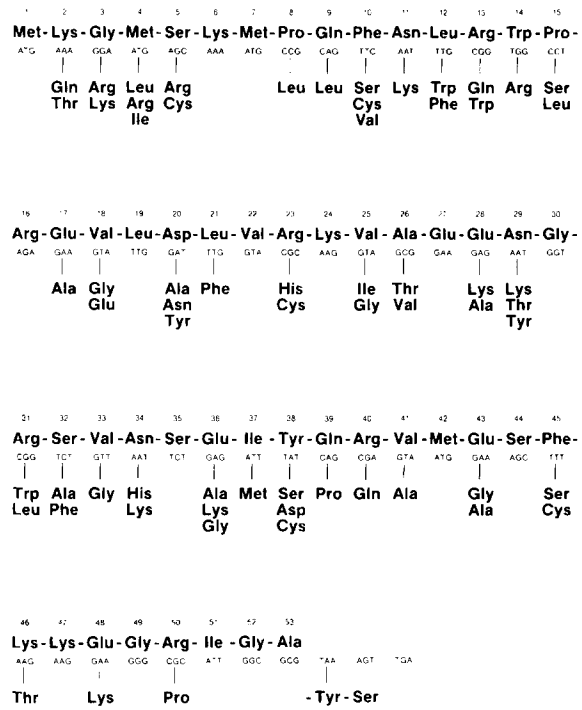


Fig. 2. Missense amino acid substitutions in Arc. The wild-type DNA sequence is shown below the protein sequence.

tor methionine). Thus, a large number of the residues appear to play some role in the structure or function of the Arc repressor.

## Intracellular Levels of the Mutant Proteins

The steady-state intracellular levels of wild-type Arc and each of the missense mutants were compared after SDS-polyacrylamide gel electrophoresis of crude cell lysates. Following induction of the *tac* promoter with isopropyl-β-D-thio-galactoside (IPTG), wild-type Arc is present as a major band on a Coomassie-stained gel. Under comparable conditions, the mutant Arc proteins show a range of levels. Table I lists the levels for each mutant protein. Forty-two mutants had low steady-state levels (less than 10% of the wild-type level); ten had moderate levels (10–50% of wild-type); and nine had high levels (greater than 50% of wild-type).

There are a number of possibilities that could account for the reduced levels of many of the mutant Arc proteins. Increased turnover caused by intracellular proteolysis probably contributes to the reduced levels. However, the mutant levels were not increased significantly by conditions that have been reported to stabilize some mutant proteins; these conditions include growth at 30°C, growth in high-salt medium,[18] and growth in *lon*$^-$ or *htp*R$^-$ strains.[9,19] Reduced transcription, mRNA stability, or translation may also contribute to the reduced levels of some of the mutant proteins. It should be noted that we have measured the steady-state levels of the mutants after IPTG induction of the *tac* promoter, whereas the mutants were selected under repressed conditions. It is possible that overexpression of the mutant proteins increases their stability, and that their relative levels would be even lower under the conditions of the mutant selection. Unfortunately, the mutant protein levels cannot be measured reliably when the *tac* promoter is repressed.

## Purification of the Mutant Proteins

Many of the mutant Arc proteins had cellular levels that were too low to allow purification but represen-

tative mutants at seventeen different positions appeared to be present in sufficient quantities to attempt purification. In seven cases (VG18, DA20, LF21, RC23, EA28, KT46, and EK48), the mutant proteins were either degraded or lost by precipitation during the purification. In twelve cases (KT2, GR3, MI4, SC5, PL8, FV10, EA17, DN20, VI25, RH23, EA43, and +YS55), we were able to purify significant quantities of the mutant proteins to greater than 95% homogeneity. With the exception of expected differences in elution from cation exchange columns, each of these mutant proteins purified in a manner similar to wild-type Arc protein.

Amino acid compositions were determined for each of the purified mutant proteins and with one exception were found to be in good agreement with those predicted from the DNA sequences (data not shown). In one case, the composition did not agree with the sequence originally predicted for the mutant protein. However, reinvestigation revealed an error in the analysis of the DNA sequence. Correction of this error brought the composition and predicted sequence into accord.

### Stabilities and Structural Properties of the Mutant Proteins

As a first step in characterizing the purified mutant proteins, we determined their stabilities and that of wild-type Arc to thermal and guanidine-HCl induced denaturation. These experiments monitor the equilibrium between the folded and unfolded structures of the mutant proteins. Since the folded form includes both monomer and dimer species, the overall equilibrium constant should be a function of both the intrinsic stability of the monomer and the strength of the dimer contacts. One would therefore expect an increase in stability with increasing protein concentration. This behavior was clearly observed for wild-type Arc protein (data not shown). As a result of this concentration dependence, the stabilities of the wild-type and mutant proteins were all determined at the same protein concentration. Under these conditions, if the stability of a mutant protein is equivalent to that of wild type, it suggests that the mutant protein is similar to wild type in both tertiary and quaternary structure. The oligomeric form of each mutant in the concentration range used for the folding studies was studied directly by chromatographing small zones of the purified proteins on a Sephadex G50 column. Wild-type and all the mutant proteins, including those with altered stability, eluted at the position expected for a dimer.

The thermal stability of each of the mutant proteins was measured by monitoring changes in circular dichroism. Figure 3A shows melting curves for wild-type Arc and the PL8 and DN20 mutant proteins, which have increased and decreased thermal stabilities, respectively. One additional mutant (RH23) was found to be less thermally stable than wild type, but the remaining nine mutant proteins have thermal denaturation profiles that are very similar to those of wild-type Arc. The temperatures of half-denaturation ($T_m$) and van't Hoff enthalpies ($\Delta H$) for wild-type and each of the mutant proteins are listed in Table II.

**TABLE II. Stabilities of Purified Mutant Proteins***

| Mutant | $D_{½}$ (M) | $T_m$ (C) | $\Delta H_{VH}$ (kcal/mole) |
|---|---|---|---|
| WT | 1.2 | 34 | 22 |
| KT2 | 1.1 | 34 | 24 |
| GR3 | 1.2 | 34 | 24 |
| MI4 | 1.2 | 33 | 25 |
| SC5 | 1.2 | 35 | 20 |
| PL8 | 1.9 | 56 | 45 |
| FV10 | 1.2 | 33 | 25 |
| EA17 | 1.2 | 36 | 24 |
| DN20 | 1.0 | 26 | 22 |
| RH23 | 0.8 | <5 | — |
| VI25 | 1.1 | 31 | 22 |
| EA43 | 1.1 | 33 | 24 |
| +YS55 | 1.2 | 38 | 24 |

*$D_{½}$ is the molar concentration of GuHCl required to unfold 50% of the protein (100 μg/ml) in 10 mM Tris-HCl (pH 7.5), 50 mM KCl at room temperature. $T_m$ is the temperature at which 50% of the protein (50 μg/ml) is denatured in 10 mM potassium acetate (pH 4.0), 100 mM KCl. $\Delta H_{VH}$ is the van't Hoff enthalpy determined from the thermal denaturation data.

Denaturation by guanidine-HCl was monitored by changes in the intrinsic fluorescence of the single tryptophan residue in Arc. Table II lists the GuHCl concentrations ($D_{½}$) at which wild-type Arc and each of the mutant proteins are 50% unfolded, and Fig. 3B shows complete curves for wild-type and three mutant proteins. Nine of the mutant proteins have $D_{½}$ values that are the same or very similar to wild-type; the RH23 mutant was significantly less stable, the DN20 mutant was slightly less stable, and the PL8 mutant was significantly more stable to GuHCl-induced denaturation.

The experiments described above measure the stability of the folded form relative to the unfolded form. Other evidence suggests that the folded structures are similar to wild-type. Since purification properties are determined by the solubility, structure, and oligomeric form of a protein, the finding that the mutants all purified like wild type suggests that the mutant Arc proteins form dimers that have structures similar to that of wild-type Arc. Finally, the similarities of the fluorescence, circular dichroism, and nuclear magnetic resonance (NMR) spectra all argue against major alterations of the mutant protein structures.

The increased stability of the PL8 mutant protein is striking. This substitution of proline by leucine results in a 22°C increase in the $T_m$ of the mutant protein and dramatically increases the stability of the protein to GuHCl denaturation. It will be of interest to determine whether this increased stability results

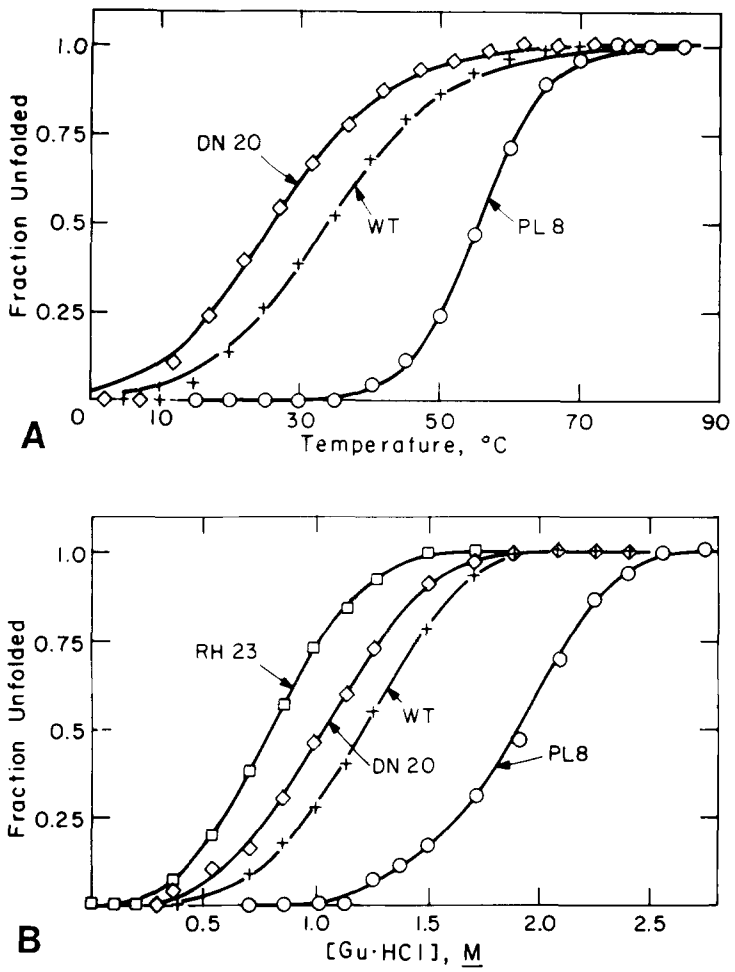


Fig. 3. Denaturation of wild-type and mutant proteins. **A:** CD melting curves for the wild-type, PL8, and DN20 Arc proteins. The fraction of protein unfolded at each temperature was calculated from measurements of the ellipticity at 222 nm. The curves are theoretical plots from the van't Hoff equation using values of the $T_m$ and $\Delta H_{VH}$ that best fit the data. **B:** Fluorescence denaturation curves for the wild-type, PL8, DN20, and RH23 Arc proteins as a function of guanidine-hydrochloride concentration. The fraction of protein unfolded at each concentration was calculated from the fluorescence emission intensity at 332 nm.

solely from local interactions or from more extensive, propagated changes. For example, if Pro8 terminates an α-helix, then this helix might be extended for several additional residues in the mutant structure.

### DNA Binding Affinity

The operator and non-operator binding of wild-type Arc and the mutant proteins was measured by gel electrophoretic shift assay. Figure 4A shows such an assay for wild-type Arc. As the concentration of wild-type Arc is raised, the operator-containing restriction fragment moves from its free position (band A) to a bound position (band B). At higher concentrations, Arc begins to bind to the non-operator sites on this fragment and the DNA migrates high on the gel (band D). A restriction fragment without an operator site moves from its free position (bands C) to a saturated position (band E) at approximately the same Arc concentration where the operator fragment shows nonspecific binding.

DNA binding assays for the SC5 and VI25 mutants are shown in Fig. 4B,C, respectively. SC5 shows very poor operator binding, whereas VI25 binds operator only slightly less well than wild-type Arc. Table III lists the relative concentrations of protein required for half-maximal binding of each of the mutant proteins to the operator and non-operator DNA fragments. Seven of the mutants (KT2, GR3, MI4, SC5, PL8, FV10, and RH23) show 60–250-fold decreases in binding to the fragment containing the *arc* operator site. The remaining mutants (EA17, DN20, VI25, EA43, and +YS55) have operator binding activities within fourfold of wild-type Arc. Wild-type Arc and most of the mutants bind to the non-operator fragment at similar concentrations (Table III). The SC5 mutant and VI25 mutants show slight decreases in non-operator binding, and the EA17 and +YS55 mutants show slight increases.

For the mutants with significantly decreased operator affinities, pancreatic DNase I protection experi-

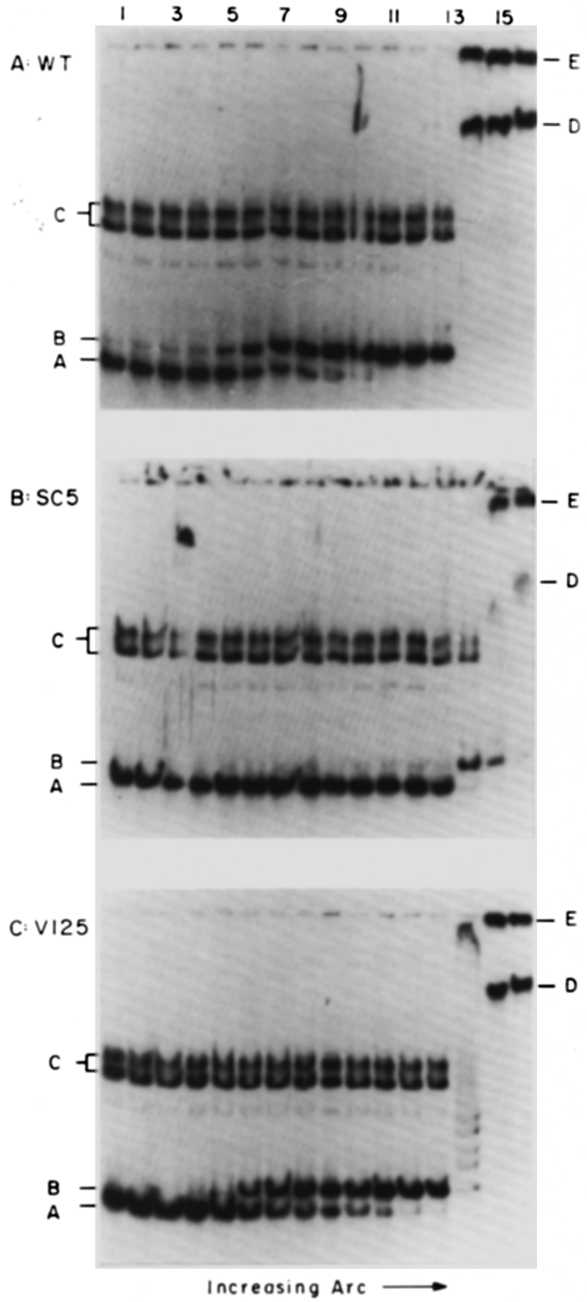


Fig. 4. Gel electrophoretic DNA binding assays for the wild-type, SC5, and VI25 Arc proteins. Bands at positions A and B correspond to unbound and specifically bound operator fragment, respectively. Bands at position C correspond to unbound nonspecific fragments. Bands D and E correspond to operator fragment and nonspecific fragments, respectively, that are bound nonspecifically by Arc protein. **A:** Wild-type protein. **B:** SC5. **C:** VI25. The same dilution series was used for all three proteins. **Lane 1:** $4 \times 10^{-6}$ M. **Lanes 2–14:** Twofold dilutions between $4 \times 10^{-6}$ M and $2.3 \times 10^{-10}$ M. **Lane 15:** $2.3 \times 10^{-10}$ M. **Lane 16:** No added Arc.

ments were performed to assay specific protection of the *arc* operator site. At sufficiently high concentrations of each mutant protein, the operator was protected from DNase digestion, while other sites in the fragment were not protected (data not shown).

**TABLE III. DNA Binding of the Purified Mutants Relative to Wild-Type Arc***

| Mutant | Specific binding | Nonspecific binding |
|---|---|---|
| WT | 1 | 1 |
| KT2 | 250 | 2 |
| GR3 | 200 | 2 |
| MI4 | 60 | 1 |
| SC5 | 500 | 4 |
| PL8 | 200 | 2 |
| FV10 | 120 | 2 |
| EA17 | 1 | 2 |
| DN20 | 4 | 1 |
| RH23 | 100 | 0.5 |
| VI25 | 4 | 5 |
| EA43 | 1 | 2 |
| + YS55 | 1 | 0.3 |

*The values listed are the protein concentrations (relative to wild-type Arc) required for half-maximal binding to operator or non-operator DNA. Under the conditions used [10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$, 0.1 mM EDTA, 100 mM KCl, 100 μg/ml BSA, room temperature], half-maximal binding of the operator fragment occurs at a wild-type Arc concentration of $1 \times 10^{-9}$ M, and half-maximal binding of the non-operator fragment occurs at a concentration of $1 \times 10^{-7}$ M.

## DISCUSSION

Our studies have identified 64 missense mutations, which as a group affect 36 of the 53 residues of Arc repressor. This implies that the identities of the side chains at 70% or more of the residues are important for some aspect of Arc structure or activity. Since the majority of the missense mutant proteins have very low intracellular levels when compared to wild-type Arc, it seems likely that many of the mutant substitutions interfere with Arc's ability to maintain a protease resistant structure in the cell. These findings are similar to those recently obtained for another small repressor, the λ Cro protein; missense mutations were isolated at half of the residue positions in Cro, and increased cellular degradation of the mutant proteins was shown to contribute to most of the mutant phenotypes.[20]

Approximately 20 of the mutant Arc proteins were present in levels sufficient to attempt purification, and 12 of these mutant proteins were purified. Two of these mutants, DN20 and RH23, were shown to be less stable than wild type to heat or GuHCl denaturation. For these mutants, instability of the folded structures is probably sufficient to account for their mutant phenotype in the cell, since the denatured protein would be inactive in operator binding and should be susceptible to rapid degradation by cellular proteases.

The EA17, VI25, EA43, and +YS55 mutants were found to be as stably folded as wild type and to have affinities for operator and non-operator DNA that were within fourfold of the wild-type value. The EA17, VI25, and EA43 proteins had reduced cellular levels,

which presumably accounts at least in part for their defective phenotypes. It is possible that these mutants have low cellular levels because they are expressed at a lower level than wild type or because the the mutations create protease sensitive sites. However, these mechanisms could not account for the defective phenotype of the +YS55 mutant, since this protein has a level in vivo that is the same as wild-type Arc. It is possible that the increased binding affinity of +YS55 for non-operator DNA (Table III) contributes to its defective phenotype.

The mechanism of Arc repression is unusual. Arc does not prevent polymerase binding to the promoter but does block the isomerization of bound RNA polymerase from a closed to open complex (S.-M. Liao and W.R. McClure, personal communication). It is therefore possible that there is a class of Arc mutants which retain operator binding activity but lose the ability to prevent the isomerization step. The EA17, DN20, VI25, EA43, and +YS55 mutants were tested for such a phenotype, but each of these proteins retained the ability to repress $P_{ant}$ transcription in vitro (S.-M. Liao and W.R. McClure, personal communication).

Crystallographic, biochemical, and genetic studies of λ Cro, catabolite activator protein, λ repressor, 434 repressor, and Trp repressor have provided strong evidence that each of these repressors uses side chains from a conserved helix-turn-helix structure for most contacts with their operator DNA's.[21–30] Many additional repressors, whose crystal structures are not known, also appear to utilize helix-turn-helix structures for operator recognition (for review, see ref. 31). Arc, however, does not share direct sequence homology with any of the helix-turn-helix repressors. In the remaining discussion we argue that Arc may use a different mechanism for operator recognition.

The KT2, GR3, MI4, SC5, PL8, and FV10 mutations are clustered in the first ten amino acids of Arc and appear to define major determinants of operator recognition. Each of these mutations causes a 60-fold to 250-fold reduction in apparent operator affinity (Table III). While it is formally possible that these large decreases in operator binding result from alterations in the structures of the mutant proteins, several results suggest that this is not the case. The KT2, GR3, MI4, SC5, PL8, and FV10 proteins are present at wild-type levels in vivo, and are as stable or more stable than wild type in vitro. Moreover, spectroscopic evidence indicates that these mutant proteins have structures like wild type. It is therefore likely that their operator binding affinities are reduced because favorable contacts with the DNA have been removed or unfavorable contacts have been introduced. Both mechanisms imply close approach of residues 2–5, 8, and 10 to the operator DNA. While we do not rule out the involvement of other residues in DNA binding, it seems clear that the ten amino terminal residues of Arc comprise a major part of the DNA binding surface of the protein.

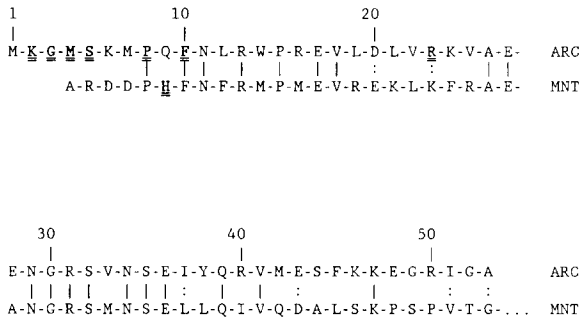


Fig. 5. Homology between the P22 Arc and the P22 Mnt repressors. The Arc sequence and the first 50 residues of Mnt are shown. The sites of known DNA binding mutations are underlined (ref. 32 and this work).

Arc repressor shares strong sequence homology with the Mnt repressor of bacteriophage P22 (see Fig. 5). This suggests that the two proteins have similar structures and use similar mechanisms for binding to their respective operators. Since the first ten residues of Arc appear to mediate its operator binding, then the aligned residues of Mnt (residues 1–7; see Fig. 5) would be expected to serve the same function. This seems to be the case, as altered specificity experiments suggest that residue 6 of Mnt makes major groove contacts with the 3rd and 15th bases of the 17-bp *mnt* operator.[32–33] In the helix-turn-helix repressors, these operator bases would be recognized by side chains from the N-terminal end of the DNA recognition α-helix. Thus, the N-terminal residues of Mnt and Arc appear to have functional roles analogous to residues in the DNA recognition helices of the helix-turn-helix proteins. However, the N-terminal 7–10 residues of Arc or Mnt could not be part of a conventional DNA recognition α-helix because there would be no residues to form the preceding helix and turn of the helix-turn-helix structure. Taken together, these results suggest that Arc and Mnt are not simple helix-turn-helix repressors and are probably members of a new family of sequence specific DNA binding proteins.

## NOTE ADDED IN PROOF

The *arc* and *mnt* operators differ at 14 of 21 basepairs and neither Arc nor Mnt binds to the other repressor's operator.[4] We have recently shown that a fusion protein, created by replacing the first 6 residues of Mnt repressor with the first 9 residues of Arc repressor, binds to the *arc* operator but not to the *mnt* operator (K. Knight and R. Sauer, in preparation). This result, taken with the results presented here, indicates that the N-terminal residues of Arc are responsible for the specificity of operator recognition as well as most of the affinity of operator binding.

## ACKNOWLEDGMENTS

We thank Karen Blackmer for help in mutant isolation and sequencing; Anna Jeitler-Nilsson for help

in purifying the MI4 and PL8 proteins; Sha-Mei Liao, Will McClure, and Michael Weiss for communication of unpublished results; and Peter Kim for use of the CD instrument. This work was supported by NIH grants #AI-16892 and 15706. A.K.V. and J.B. were supported by NIH training grant GM-072871.

## REFERENCES

1. Susskind, M.M. A new gene of bacteriophage P22 which regulates synthesis of antirepressor. J. Mol. Biol. 138:685–713, 1980.
2. Sauer, R.T., Krovatin, W., DeAnda, J., Youderian, P., Susskind, M.M. Primary structure of the *imm*I immunity region of bacteriophage P22. J. Mol. Biol. 168:699–713, 1983.
3. Vershon, A.K., Younderian, P., Susskind, M.M., Sauer, R.T. The bacteriophage P22 Arc and Mnt repressors: Overproduction, purification, and properties. J. Biol. Chem. 260: 12124–12129, 1985a.
4. Vershon, A.K., Liao, S.-M., McClure, W.R., Sauer, R.T. Interactions of the bacteriophage P22 Arc repressor with operator DNA. J. Mol. Biol. (in press).
5. Jordan, S.R., Pabo, C.O., Vershon, A.K., Sauer, R.T. Crystallization of the Arc repressor. J. Mol. Biol. 185:445–446, 1985.
6. Nelson, H.C.M., Hecht, M.H., Sauer, R.T. Mutations defining the operator-binding sites of bacteriophage λ repressor. Cold Spring Harbor Symp. Quant. Biol. 47:441–449, 1983.
7. Vershon, A.K., Blackmer, K., Sauer, R.T. Mutagenesis of the Arc repressor using synthetic primers with random nucleotide substitutions. In: "Protein Engineering: Applications in Science, Medicine, and Industry." Inouye, M., Sarma, R, eds. Orlando, FL: Academic Press. 1986: 243–256.
8. Amman, E., Brosius, J., Ptashne, M. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene 25:167–178, 1983.
9. Baker, T.A., Grossman, A.D., Gross, C.A. A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. USA 81:6779–6783, 1984.
10. Zagursky, R.J., Berman, M.L. Cloning vectors that yield high levels of single-stranded DNA for rapid DNA sequencing. Gene 27:183–191, 1984.
11. Nelson, H.C.M., Sauer R.T. Lambda repressor mutations that increase the affinity and specificity of operator binding. Cell 42:549–558, 1985.
12. Zoller, M.J., Smith, M. Oligonucleotide-directed mutagenesis: A simple method using two oligonucleotide primers and a single-stranded DNA template. DNA 3:479–488, 1984.
13. Levinson, A., Silver, D., Seed, B. Minimal size plasmids containing an M13 origin for production of single-stranded transducing particles. J. Mol. Appl. Genet. 2:507–517, 1984.
14. Kunkel, T.A. Mutational specificity of depurination. Proc. Natl. Acad. Sci. USA 81:1494–1498, 1984.
15. Sanger, F., Nicklen, S., Coulson, A.R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977.
16. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685, 1970.
17. Vershon, A.K., Liao, S.-M., McClure, W.R., Sauer, R.T. Bacteriophage P22 Mnt repressor: DNA binding and effects on transcription *in vitro*. J. Mol. Biol. (in press).
18. Kohno, T., Roth, J. Electrolyte effects on the activity of mutant enzymes *in vivo* and *in vivo*. Biochem. 18:1386–1392, 1979.
19. Gottesman, S., Zipser, D. Deg phenotype of *Escherichia coli lon* mutants. J. Bacteriol. 133:844–851, 1978.
20. Pakula, A., Young, V., Sauer, R.T. Effects of missense mutations on the activity and degradation of λ Cro. Proc. Natl. Acad. Sci. USA, in press, 1986.
21. Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA. Nature 290:754–758, 1981.
22. McKay, D.B., Steitz, T.A. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature 290:744–749, 1981.
23. Pabo, C.O., Lewis, M. The operator-binding domain of λ repressor: Structure and DNA recognition. Nature 298:443–497, 1982.
24. Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y., Matthews, B.W. The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature 298:718–723, 1982.
25. Hecht, M.H., Nelson, H.C.M., Sauer, R.T. Mutations in λ repressor's amino-terminal domain: Implications for protein stability and DNA binding. Proc. Natl. Acad. Sci. USA 80:2676–2680, 1983.
26. Ebright, R.H., Cossart, P., Cicquel-Sanzeym, B., Beckwith, J. Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. Nature 311:232–235, 1984.
27. Kelly, R.L., Yanofsky, C. Mutational studies with the *trp* repressor of *Escherichia coli* support the helix-turn model of repressor recognition of operator DNA. Proc. Natl. Acad. Sci. USA 82:483–487, 1985.
28. Anderson, J.E., Ptashne, M., Harrison, S.C. A phage repressor-operator complex at 7 Å resolution. Nature 316:596–601, 1985.
29. Schevitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L., Sigler, P.B. The three-dimensional structure of *trp* repressor. Nature 317:782–786, 1985.
30. Wharton, R.P., Brown, E.L., Ptashne, M. Substituting an α-helix switches the sequence-specific DNA interactions of a repressor. Cell 38:361–369, 1985.
31. Pabo, C.O., Sauer, R.T. Protein-DNA recognition. Annu. Rev. Biochem. 53:293–321, 1984.
32. Youderian, P., Vershon, A.K., Bouvier, S., Sauer, R.T., Susskind, M.M. Changing the DNA-binding specificity of a repressor. Cell 35:777–783, 1983.
33. Vershon, A.K., Youderian, P., Weiss, M.A., Susskind, M.M., Sauer, R.T. Mnt repressor-operator interactions: Altered specificity requires N-6 methylation of operator DNA. In: "Sequence Specificity in Transcription and Translation." Calendar, R., Gold, L., eds. New York: Alan R. Liss, Inc., 1985:209–218.