

Crystal Structure of *lac* Repressor Core Tetramer and Its Implications for DNA Looping

Alan M. Friedman, Thierry O. Fischmann, Thomas A. Steitz*

The crystal structure of the tryptic core fragment of the *lac* repressor of *Escherichia coli* (LacR) complexed with the inducer isopropyl- β -D-thiogalactoside was determined at 2.6 Å resolution. The quaternary structure consists of two dyad-symmetric dimers that are nearly parallel to each other. This structure places all four DNA binding domains of intact LacR on the same side of the tetramer, and results in a deep, V-shaped cleft between the two dimers. Each monomer contributes a carboxyl-terminal helix to an antiparallel four-helix bundle that functions as a tetramerization domain. Some of the side chains whose mutation reduce DNA binding form clusters on a surface near the amino terminus. Placing the structure of the DNA binding domain complexed with operator previously determined by nuclear magnetic resonance onto this surface results in two operators being adjacent and nearly parallel to each other. Structural considerations suggest that the two dimers of LacR may flexibly alter their relative orientation in order to bind to the known varied spacings between two operators.

The regulation of transcription is a central mechanism of metabolic control and development in all organisms. Positive and negative regulation are mediated by regulatory proteins that bind DNA either near to or at a distance from the start of transcription (1). The *lac* operon of *E. coli* and the *lac* repressor (2) are examples of negative transcriptional control acting both nearby and at a distance. In the absence of inducers, which signal a requirement for the enzymes encoded by the *lac* operon, the repressor binds a high-affinity *lac* operator DNA site (O1) centered at the +9 base pair (bp) from the start of transcription. This site overlaps the binding site for RNA polymerase (3). The bound repressor either inhibits the binding of RNA polymerase (4) or stops its entry into the processive elongation phase (5), preventing effective transcription of the genes for lactose metabolism. In the presence of inducers, the repressor's affinity for operator drops, and transcription ensues.

The *lac* repressor also mediates DNA looping that results from interaction with multiple operator sites (6, 7), a common theme in prokaryotic and eukaryotic transcription regulation (1). Numerous examples of DNA looping have been observed in negative regulation in prokaryotes, including phage λ repressor (8), the *ara* (9), *deo*

(10), *gal* (11), and *lac* operons (6, 7, 11), in positive regulation in prokaryotes, for ex-

ample NtrC (12), and in positive regulation in eukaryotes, such as the progesterone receptor (13) and Sp1 (14). Unlike these transcription regulators, LacR is a tetramer that can bind two DNA operator sites and form bidentate interactions with the *lac* promoter (6, 7, 11). These bidentate interactions are required for maintaining the repression efficiency found in the *lac* operon (15). In LacR, separated DNA binding sites are brought together by a single protein molecule, creating a favorable system for studying the structural organization of proteins bound to separated regulatory sites.

LacR is organized into an architecture common to many prokaryotic and eukaryotic transcriptional regulators that have a small DNA binding domain (headpiece) and a larger regulatory domain (core). Although about two dozen structures of DNA binding domains from transcriptional regulators complexed with DNA have been determined, few structures of multidomain transcription factors or of isolated regulatory domains are known. The intact 38-kilodalton repressor forms a homotetramer of 152 kD. Trypsin cleaves intact LacR into four 6-kD monomeric headpiece domains

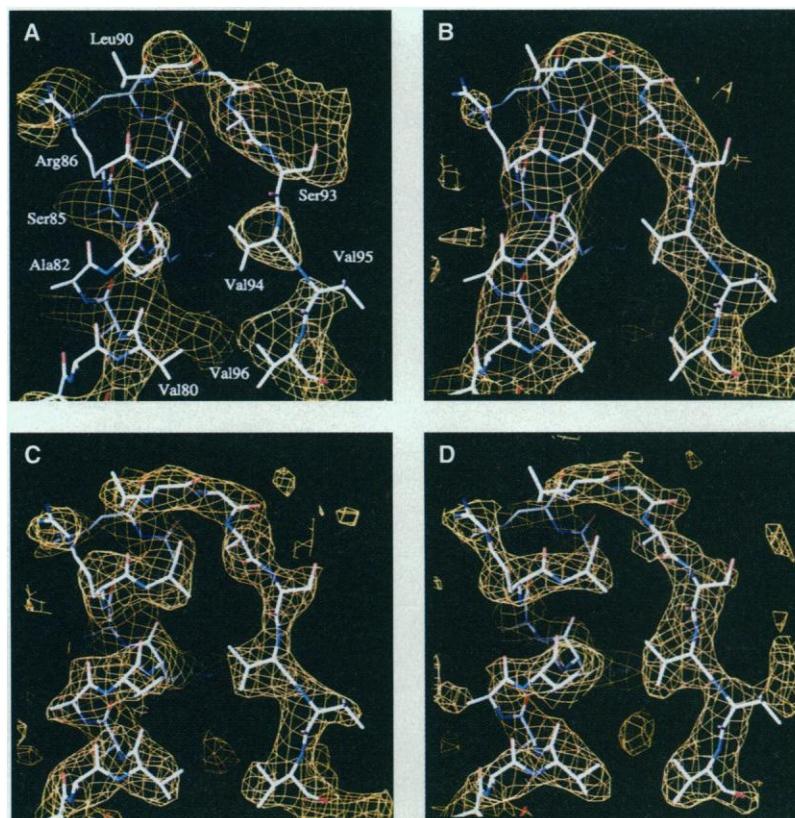


Fig. 1. A section through the electron density map at several stages of structure determination (33). The present atomic model is shown and all maps are contoured at 1.5σ . **(A)** Experimentally phased MAD map at 4.5 Å resolution. **(B)** Map after solvent flattening and fourfold symmetry averaging at 4.5 Å resolution. **(C)** Map after phase extension to 2.6 Å with solvent flattening and fourfold symmetry averaging restraints. **(D)** Simulated annealing omit map (34) in which the region shown was omitted and the remaining model subjected to simulated annealing refinement to remove phase bias. The map shown was calculated with $F_o - F_c$ amplitudes and phases calculated from the model after simulated annealing.

A. M. Friedman is in the Department of Molecular Biophysics and Biochemistry, T. O. Fischmann is in the Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, and T. A. Steitz is in the Department of Molecular Biophysics and Biochemistry, the Department of Chemistry, and the Howard Hughes Medical Institute, Yale University, New Haven, CT 06520-8114, USA.

*To whom correspondence should be addressed.

Table 1. Structure determination and refinement. The observed Bijvoet (shown on diagonal) and dispersive (shown off-diagonal) merging R factors on intensity between measurements in the MAD data set are shown. Anomalous scatterer parameters for the edge wavelength were refined while those for the remote wavelength were held fixed at their theoretical values (56). The completeness of the data for phase determination is detailed, as is the agreement between independent, redundant determinations of the MAD fundamental parameters, F_a , F_z , and $\Delta\phi$, where F_a is the magnitude of the anomalous structure factor, F_z is the magnitude of the wavelength-invariant structure factor, and $\Delta\phi$ is the difference in phase angle between the anomalous and the total structure factors (29). The quality of both the MAD fitting procedure and the subsequent anomalous scatterer refinement was confirmed by the agreement between F_a and F_c calculated from the refined parameters. Finally, the correctness of the statistical procedures is indicated by the close agreement between the σ of $\Delta\phi$ and the difference between independent determinations of $\Delta\phi$ ($<\Delta(\Delta\phi)>$) as well as by the agreement between the σ of F_a and the mean residual after anomalous scatterer refinement ($<|F_a - F_c|>$).

Data collection						
Wavelength (Å)	R^* for matched measurements in a phasing set				Scattering factors	
	Low resolution, 23–7.5 Å		Overall, 23–3.5 Å		f' (e ⁻)	f'' (e ⁻)
	Edge 1.0088 Å	Remote 0.9840 Å	Edge 1.0088 Å	Remote 0.9840 Å		
1.0088, edge	ND	0.016	ND	0.024	-23.4	1.9
0.9840, remote		0.031		0.046	-10.2	9.7
Centric Bijvoet R factor		0.018		0.032		
Phase determination						
Data type	Resolution range (Å)					
	23.0–6.0	6.0–3.5	23.0–3.5			
MAD phase determination sets measured (N)	4568	13055	17623			
Total possible unique reflections (N)	3504	13929	17433			
Unique reflections phased by MAD (N)	2801	10036	12837			
Unique reflections phased by MAD (%)	79.9	72.1	73.6			
Mean σ of $\Delta\phi$, $<\sigma(\Delta\phi)>$	49.1°	61.9°	57.2°			
Mean figure of merit $<m^\dagger>$	0.65	0.47	0.54			
Agreement between independent phase determinations						
Redundant phase det (No.)	1767	3019	4786			
$R(F_z)$ between redundant det‡	0.029	0.039	0.035			
$R(F_a)$ between redundant det‡	0.465	0.499	0.486			
Mean phase difference, $<\Delta(\Delta\phi)>$	50.0°	68.9°	61.9°			
Anomalous scatterer refinement						
Reflections in refinement (N)	442	298	740			
R factor between F_a and F_c §	0.231	0.414	0.311			
Mean residual, $< F_a - F_c >$	49.7	102.3	70.9			
Mean σ of F_a , $<\sigma(F_a)>$	38.0	49.9	42.8			
Symmetry averaging						
Data type	Resolution range (Å)					
	23.0–5.4	5.4–3.4	3.4–2.6	23.0–2.6		
Data used for averaging and refinement						
Total possible unique reflections (N)	4815	14056	23246	42117		
Unique reflections measured (N)	4808	14013	22259	41080		
Unique reflections measured (%)	99.9	99.7	95.8	97.5		
Merging R factor	0.048	0.079	0.253	0.095		
Symmetry averaging						
R factor on Fourier inversion of final map¶	0.122	0.152	0.284	0.188		

* $R = \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_i I_i(h)$, summation taken over a pair of matched Bijvoet mates or a cross-wavelength dispersive pair. †Figure of merit; defined as $\cos \langle \sigma(\Delta\phi) \rangle$ (29). ‡ $R = \sum_i |F_i(h) - \langle F(h) \rangle| / \sum_i F_i(h)$, summation taken over independent phase determinations. § $R = \sum_i |F_a(h) - F_c(h)| / \sum_i F_a(h)$, where F_a is the estimate of F_a fit from the observed data and F_c is the value calculated from the refined heavy atom parameters. The summation is taken over all reflections used for refinement, which were those meeting the quality criteria of $F_a > 4 \cdot \sigma F_a$, $F_a < 800$ (on absolute scale), and $\sigma(\Delta\phi) < 80^\circ$. ¶ $R = \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_i I_i(h)$, summation taken over all Bijvoet mates and all wavelengths. ¶¶ $R = \sum_i |F_i(h) - F_c(h)| / \sum_i F_i(h)$, where F_c is the observed structure factor magnitude and F_c is the structure factor magnitude calculated from Fourier inversion of the final cycle of averaging. The summation is taken over all measured reflections.

(residues 1 to 59) and the 128-kD tetrameric core fragment (16). An isolated headpiece domain binds specifically to one-half of the dyad symmetric operator and does not respond to inducer (17). The nuclear magnetic resonance (NMR) structures of the headpiece (18) and a headpiece-DNA half-operator complex (19) reveal a helix-turn-helix motif and protein-DNA contacts.

The LacR core retains the inducer binding and oligomerization properties of the intact molecule (20). Sequence similarity (21) and three-dimensional structure profiles (22) showed that the core is homologous to the periplasmic binding proteins (23), and allowed building an approximate model of the LacR monomer (24). A structure of the dimeric purine repressor complexed with DNA is now available (25), confirming the structural homology between the transport and regulator sugar binding protein families.

Each of the LacR core dimers orients the two attached headpiece domains for interaction with operator, enhancing their affinity (17, 26), since monomeric headpieces bind DNA weakly. Core also regulates DNA affinity by transmitting the allosteric signal from the inducer binding sites in the core to the DNA binding domains. The binding of isopropyl- β -D-thiogalactoside (IPTG) to lacR lowers lacR's affinity for operator DNA 1000 times (27), presumably through reorientation of the headpieces (17). Since the core retains the tetrameric structure of LacR, it not only positions a pair of headpieces to bind a single operator, but also positions the second pair of headpieces relative to the first, defining the global structure of bidentate complexes with DNA.

Structure determination. The tryptic core fragment (residues 60 to 360) was prepared (16) and reacted with ethylmercury phosphate before crystallization (28). Experimental phases were determined by the multiwavelength anomalous diffraction (MAD) method (29) with the incorporated mercury as the anomalous scatterer (Table 1) (30). The determination of this large structure by MAD and symmetry averaging extends the usefulness of this combination (31) to significantly smaller anomalous diffraction signals. Scaling of the MAD data and phase determination (Table 1) were done with a new suite of programs MADPRB (32) derived from the original package (29).

The MAD electron density map calculated at 4.5 Å resolution (Fig. 1A) was the starting point for phase improvement (Table 1) (33). Phases were improved first by solvent flattening and fourfold symmetry averaging at 4.5 Å resolution (Fig. 1B), and subsequent phase extension to 2.6 Å. After the COOH-terminal residues were located and included in the solvent flattening

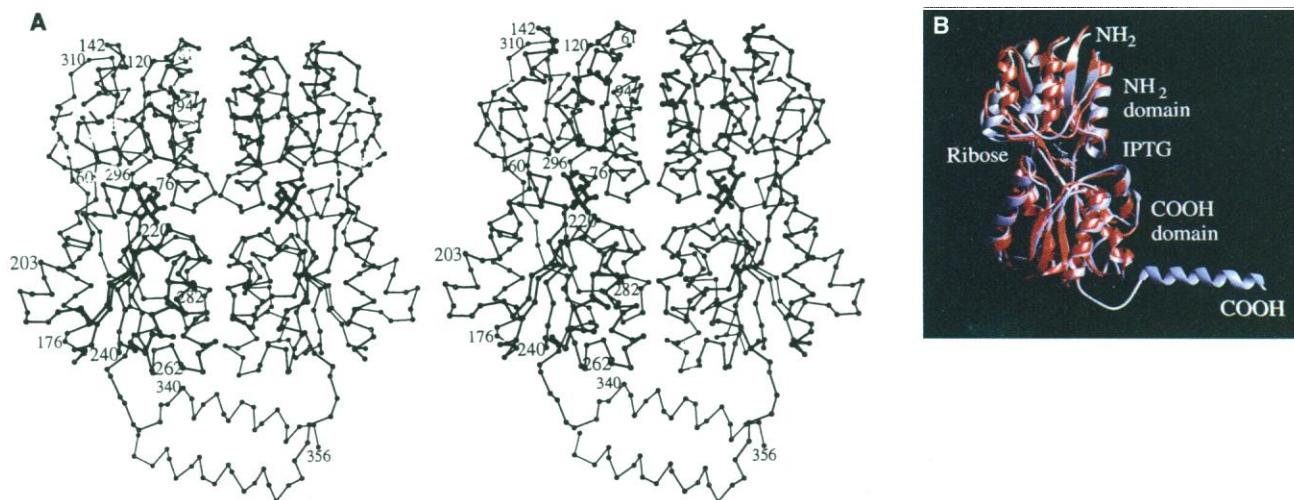


Fig. 2. Tertiary structure of the *lac* repressor core fragment. **(A)** Stereo drawing (57) of $C\alpha$ backbone of one of the two dimers of the core fragment. **(B)** Comparison of a periplasmic receptor protein and the *lac* repressor core fragment monomer. The $C\alpha$ coordinates of the ribose binding protein complexed with ribose (37) (in red) were superimposed on those of the *lac*

repressor core fragment monomer (in silver) using the program O (58). Residual rms differences after alignment were 1.6 Å for 242 $C\alpha$ atom pairs closer than 3.8 Å. Ribose transformed with the same alignment matrix and IPTG are also shown. [Figs. 2B and 3 through 6 were made with the program RIBBONS (59).]

mask, another round of phase improvement gave the final map (Fig. 1C). This map was used to build a model of residues 61 through 356, IPTG and ethylmercury. The structure was refined (34) to a crystallographic *R* factor of 22.2 percent for all data ($|F| > 2\sigma |F|$) between 8 and 2.6 Å resolution (Table 1). The root-mean-square (rms) deviations of bond length and angles from ideality were 0.012 Å and 1.8 degrees, respectively. The refined structure was confirmed by simulated annealing omit maps (Fig. 1D).

Each monomer has two α/β open-sheet domains and a COOH-terminal α helix (Fig. 2, A and B). The NH_2 -terminus of the core and most of the NH_2 -terminal residues are located in the NH_2 -domain. Likewise, most of the COOH-terminal residues are located in the COOH-domain. An extended coil segment (residues 331 to 337) connects the COOH-domain to a COOH-terminal α helix (residues 338 to 356) bearing two leucine heptad repeats. The importance of the leucine heptad repeats for tetramerization has been demonstrated (35, 36). This helix along with the corresponding helices of the other three monomers forms a four-helix bundle oligomerization domain, which knits together the tetramer.

The body of the monomer complexed with IPTG has a tertiary structure similar to the structures of the glucose-galactose-arabinose-ribose family of periplasmic receptor proteins complexed to their respective sugars (23). As predicted (22), the ribose binding protein (37) is the member of this family with the greatest structural similarity to LacR (Fig. 2B).

A cleft between the two domains encloses the inducer IPTG. Pyranose IPTG binds in a position similar to the pyranose isomer

of ribose in the periplasmic receptor structure (Fig. 2B). Important hydrogen bonding interactions are made with IPTG by residues including Ser⁶⁹, Arg¹⁰¹, Asp¹⁴⁹, Arg¹⁹⁷, Asn²⁴⁶, and Asp²⁷⁴. Hydrophobic contacts are made by residues including Leu⁷³, Ala⁷⁵, Pro⁷⁶, Ile⁷⁹, Trp²²⁰, Phe²⁹³, and Leu²⁹⁶.

Unusual, nonsymmetric LacR quaternary structure. The overall symmetry of the core tetramer differs from other known tetramer structures, which have either D2 symmetry (three mutually perpendicular twofolds) or C4 symmetry (a single four-fold). Tetramers having D2 and C4 symmetries cannot form larger aggregates by build-

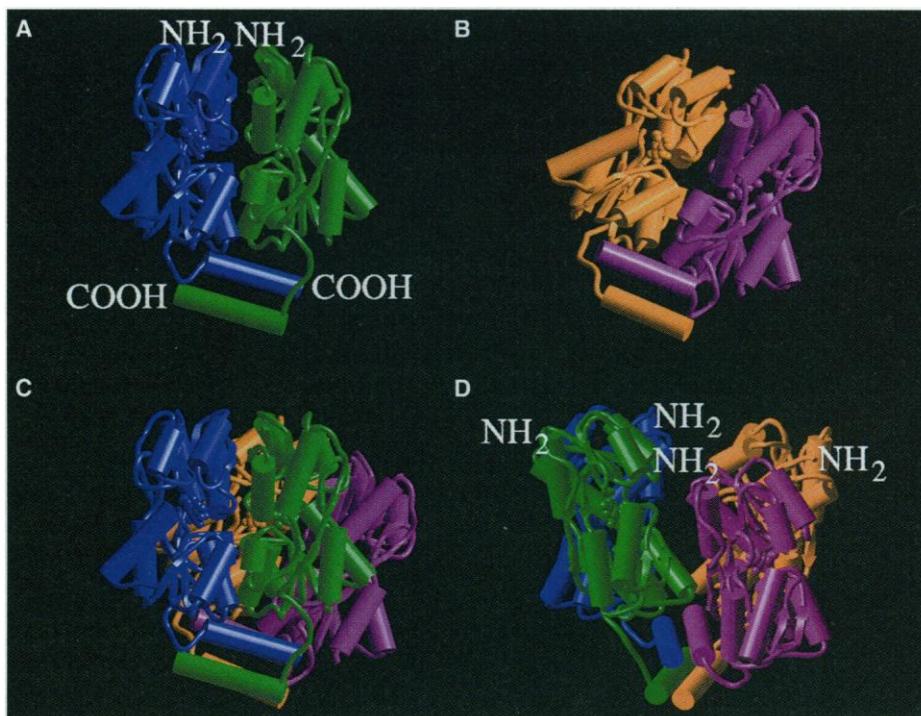


Fig. 3. Quaternary structure of the *lac* repressor core fragment. The coloring of subunits is constant throughout the panels while the same viewing orientation (approximately that of Fig. 2) is maintained in the first three. **(A)** One of the two dyad symmetric dimers. **(B)** The second dyad symmetric dimer, which lies behind the one shown in (A). **(C)** The complete tetramer comprising two dimers connected by the four-helix bundle oligomerization domain, which is located at the bottom of the panel. **(D)** The structure in (C) rotated approximately 90 degrees around the vertical axis.

ing on the tetramerization interfaces (38). In LacR, the bodies of each core monomer form two C₂-symmetric dimers (Fig. 3, A and B). These two dimers are arranged head-to-head and nearly parallel to one another with their NH₂-termini on the same face of the tetramer (Fig. 3, C and D, tops of both). The dyad axes of the dimers diverge from parallel by only 28°. The translation between the dimers and the small divergence of their dyad axes create a deep, V-shaped, solvent-filled cleft between them (Fig. 3D).

Although the overall symmetry of the core tetramer is not D₂, the four COOH-terminal helices form an oligomerization domain, an antiparallel four-helix bundle with D₂ symmetry (Fig. 3, C and D, lower part of each, and Fig. 4A) as predicted (35). Because the surface that forms the dimer-dimer interface is confined to the four-helix bundle, the D₂ symmetry of this domain prevents infinite aggregation (38). Because the four-helix bundle and the two dimers of the body of the subunits have different symmetries, the coil segment connecting the

terminal helix and the rest of the molecule adopts a different conformation in each monomer (Fig. 4B).

Independent evidence supports the V-shaped LacR core structure observed in the crystal being the predominant conformation in solution for both core and intact LacR. X-ray (39) and neutron (40) scattering studies on solutions of LacR and core are consistent with the quaternary structure in the crystal. Using the refined crystal structure atomic coordinates, we have calculated radius of gyration (r_g) and maximum dimension (D_{max}) values ($r_g = 32.2$ Å; $D_{max} = 94.5$ Å) that are close to experimental values derived from solution scattering ($r_g = 34.2$ Å; $D_{max} = 100$ Å). In contrast, a model-built extended D₂ symmetric structure gives calculated values ($r_g = 43.9$ Å; $D_{max} = 138.3$ Å) that are significantly larger than the observed values. Thus, the extended conformation cannot exist as a significant species in solution. Neutron scattering also predicts a solvent-accessible inhomogeneity, presumably the V-shaped cleft, within the scattering particle (40).

Negatively stained electron micrographs of intact LacR show a molecule of approximately the dimensions seen here with a deep cleft that fills with stain (41). These micrographs were interpreted as two pairs of cylindrical subunits diverging by an acute angle, as in the core structure. Finally, both LacR and core have a strong, permanent electric dipole (42), a finding inconsistent with D₂ symmetry. We conclude that the unusual quaternary structure observed in the crystal is not a crystal packing artifact but is the predominant species in solution for both core and intact LacR.

Although various techniques confirm that the crystal structure is the predominant structure in solution, the forces maintaining it are only moderate. The interaction between monomers forming a dyad-symmetric dimer buries 3000 Å² of accessible surface outside of the oligomerization domain. The association of the four α helices to form the oligomerization domain buries 3900 Å² of accessible surface and provides a stable anchor for the formation of the tetramer. The two dimers are oriented into the V-shaped quaternary structure by interactions between each dimer and the oligomerization domain and by a direct interaction across the V-shaped cleft between a subunit of each dimer. The total accessible surface buried in these three interactions is 1600 Å². Although significant, this is much less than the 3200 Å² typically buried in an interface between stable dimers (43), and suggests that this is the weakest interaction among the various forces determining the quaternary structure.

Although LacR clearly does not share sequence or structural similarity with the tumor suppressor p53, the only other tetrameric DNA binding protein for which there is significant structural information, the two proteins share a number of analogous features; p53 also has an oligomerization domain with a four-helix bundle motif (44). However, the p53 oligomerization domain has both parallel and antiparallel packing of the helices and contains both helices and β strands. The DNA binding site of p53 consists of a pair of dyad symmetric sites separated by 0 to 13 bp (45), and a model of a p53 tetramer bound to DNA has been proposed on the basis of the structure of the p53 core domain-DNA complex (44, 46). The arrangement of binding sites and of monomers in the model is not D₂ symmetric. Yet, like LacR, the oligomerization domain of p53 has D₂ symmetry. The combination of an asymmetric body with a symmetric oligomerization domain may prove to be a common motif in tetrameric DNA binding proteins.

The nature of the conformational change produced by the binding of IPTG to LacR is suggested by that produced by li-

Fig. 4. (A) The structure of the four-helix bundle oligomerization domain. (B) Superposition of the four helices showing the variation in the connecting loop.

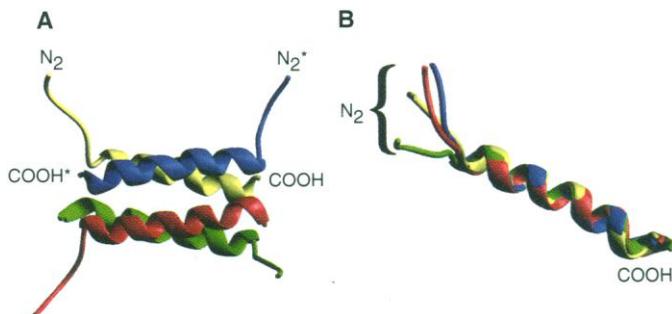
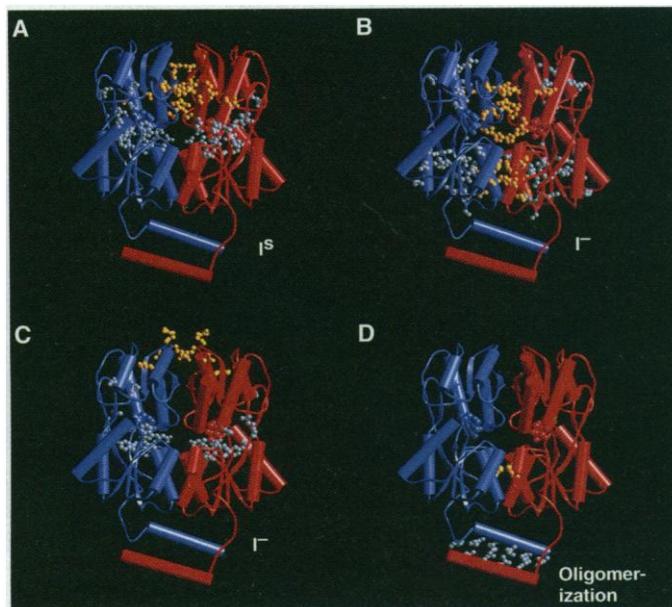


Fig. 5. Mutations in LacR core (24, 60) divided into eight classes on the basis of their phenotype and location. Alpha carbons and side chains of the wild-type residues whose mutation results in the various phenotypes are shown. (A) I^s mutations located in the inducer binding cleft are silver while those located at the dimer interface are gold. (B and C) I^r mutations, divided into those located within the hydrophobic core (B, in silver), those located at the dimer interfaces (B, in gold), those located within the IPTG binding cleft (C, in silver), and those located at the NH₂-terminal surface, presumed to form the interface with the headpiece domain (C, in gold). (D) Mutations that have been shown by biochemical means to alter the oligomerization of LacR. Side chains whose mutation results in dimers are shown in silver, while those which result in monomers are in gold.



gand binding to the homologous periplasmic sugar binding proteins. Ligand binding to these proteins changes the relative orientation of the α/β domains and closes the interdomain cleft (47, 48). Both the apo (open) and liganded (closed) structures of the leucine-isoleucine-valine binding protein (LIVBP) have been determined (48). Superimposing the COOH-domains of LIVBP and core shows that the dimer interface of LacR is roughly perpendicular to the axis of the hinge-bending motion induced by ligand binding to LIVBP. The analogous hinge-bending motion in LacR would slide the NH₂-domain dimer interfaces past each other when IPTG is bound. Such a motion would also alter the relative orientation of the DNA binding domains. The extent of hinge-bending motion between inducer-bound and DNA-bound forms of LacR must be less than that generally seen in the periplasmic binding proteins. Replicating the open structure of LIVBP by rotating the NH₂-domains of LacR core places the NH₂-termini of the core monomers 43.5 Å from the dyad axis, too far apart to connect to the COOH-termini of headpiece domains bound to operator DNA.

Structural basis of mutant LacR phenotypes. Mutations in LacR resulting in the inability to respond to inducer are called I^s (noninducible) mutations. The I^s mutant side chains cluster in the IPTG-binding cleft and at the dimer interface between NH₂-domains (Fig. 5A). I^s mutations within the IPTG-binding cleft could be either reducing the affinity of LacR for inducer or interfering with the transduction of the allosteric transition. Mutations in the NH₂-domain dimer interface presumably interfere with the allosteric transition (49) since IPTG-induced cleft closure would alter the relative orientation of the two NH₂-domains.

The inability to bind operator and repressor results in the I⁻ phenotype (non-DNA binding and nonrepressing). Excluding those found in the protein interior, non-DNA binding (I⁻) mutant side chains that are located in the core occur along surfaces that indirectly control the orientation of the headpiece domains such as the interfaces between monomers, the IPTG-binding cleft, and the NH₂-terminal surface. Mutations occurring at the interface between monomers that would prevent or alter dimer formation show the I⁻ phenotype because the high-affinity DNA binding unit is a dimer (Fig. 5B).

The cluster of I⁻ mutant side chains around the IPTG-binding cleft (Fig. 5C) may achieve that phenotype by altering the relative orientation of the NH₂- and COOH-domains of core and thus of the headpieces. The small group of I⁻ mutations on the surface of the NH₂-domain

near the NH₂-terminus (Fig. 5C) define a surface with which that NH₂-terminal headpiece may interact. Changes in these residues may directly alter the orientation of the headpiece domain.

Mutations resulting in repressor dimers lie within the oligomerization domain, whereas mutations promoting dissociation into monomers lie in the dimer interface between COOH-domains. Although mutations in the leucine heptad repeat result in dimers (15, 35, 36) (Fig. 5D), the leucine side chain of one monomer actually points toward the COOH-terminal helix of the other monomer in a dimeric subunit. The dimer-dimer interaction is mediated through the global hydrophobic packing of this domain, rather than by direct leucine-leucine contact. Mutations in the leucines disrupt the global hydrophobic packing of this domain. The separated dimers are still stabilized, however, by the extensive monomer-monomer interface between the NH₂- and

COOH-domains.

Classical genetic studies (50) have identified several regions (residues 65 to 75, 221 to 227, and 270 to 290) whose mutation yield monomeric mutants. These regions are in or near the NH₂- and COOH-domain dimer interfaces. Because of the imprecision of classical genetic mapping, we show (Fig. 5D) only those mutations that have been sequenced (51).

Models of intact repressor complex with operator and of DNA looping. We have constructed an approximate model of an intact repressor complexed to DNA by docking the structure of a LacR headpiece monomer complexed to a DNA half-operator site determined by NMR (18) onto the LacR core. The headpiece-half-operator complex structure (18) was converted into a dimer with the dyad axis passing through the central base pair of the operator. We have assumed that the dyad axis relating the two halves of the operator DNA and

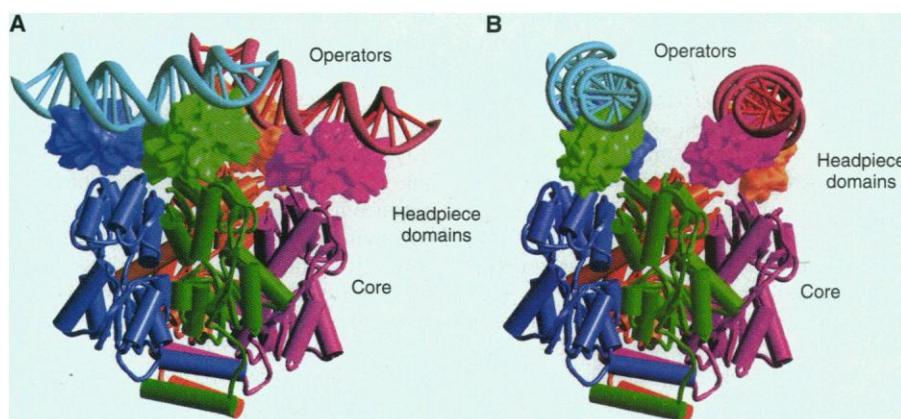


Fig. 6. Two possible models of a bidentate repressor-DNA interaction. The same viewing orientation as in Fig. 3 is maintained in both (A) and (B) but the orientation of the headpiece domains and DNA relative to the core is different in the two models. In the model in (A) the headpiece domains make contact with the NH₂-terminal surface of the core where I⁻ mutations are known to lie. The model in (B) approximately replicates the orientation seen in the homologous purR-DNA structure (25). In both models the headpiece domains have been portrayed as space filling solvent accessible surface.

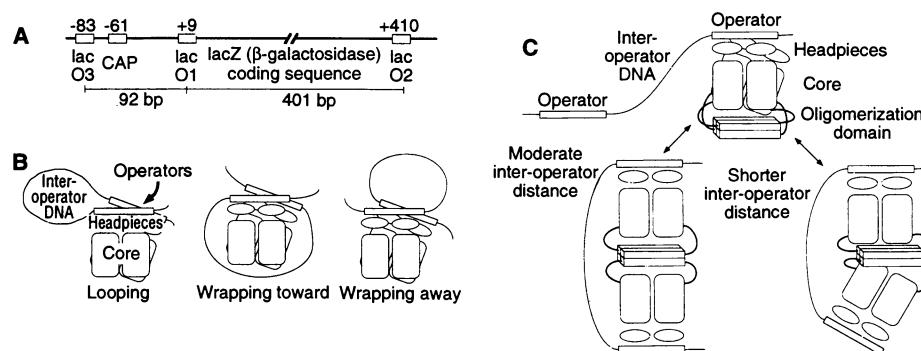


Fig. 7. Schematic diagrams of higher order interactions between LacR and the *lac* operon. (A) The control sites of the *lac* promoter. (B) Possible modes of bidentate interaction between two *lac* operators and LacR. (C) Postulated partial unfolding of *lac* repressor and subsequent segmental flexibility in binding two *lac* operators separated by distances shorter than can be accommodated by the repressor in its crystal conformation without severe strain. The extended coil connecting the COOH-domain and the oligomerization domain, which may form a flexible tether after the interaction of the two domains is disrupted, is shown as a heavy black line.

bound headpieces is coincident with the one relating the two subunits of the core dimer. Accordingly, only the relative rotation of the headpieces and core dimers about this axis and the translation along it require adjustment in docking these two domains. One possible orientation of the docked headpieces places them in contact with the I⁻ mutations (Lys¹⁰⁸, Leu¹¹⁴, Leu¹¹⁵, Glu¹¹⁷, Arg¹¹⁸, and Val¹⁷⁹) located on the core NH₂-surface (Fig. 6A) while another orientation (Fig. 6B) nearly orthogonal to the first is similar to that seen in the PurR-DNA complex structure (25). It is of no consequence to the subsequent discussion which of these orientations (or any in between) proves to be correct. In either of the extreme orientations the two operator DNAs bound to the two dimers are approximately parallel and on the same side of the molecule (Fig. 6, A and B). Our model has essentially straight DNA, in accord with solution studies that suggest LacR does not bend wild-type operator significantly (52).

A priori there are three ways that two operators in the *lac* promoter (Fig. 7A) can form a bidentate complex with the two sites in a LacR tetramer; these are looping, wrapping toward the protein, and wrapping away from it (Fig. 7B). The term "DNA looping" has been used (1, 6–14) to provide a general description of these interactions. Two cases should be considered. With lengths of DNA between operators that are longer than the persistence length of DNA, such as the 401 bp between operator O1 and the downstream operator O2, none of the models present steric problems. Electron micrographs of LacR bound to two operators spaced 535 bp apart show examples of both loops and wraps (6).

Shorter lengths of DNA between operators, such as the 92 bp between O1 and the upstream O3, restrict the possible models. Bidentate binding of LacR to DNAs containing two operators has been observed with spacings as small as 52 bp between operator centers (6). A characteristic pattern of alternating deoxyribonuclease I protection and hypersensitivity was generated and shows the direction of the DNA bend. For LacR this suggests that the DNA bends toward the protein (53), ruling out the wrapping away model.

Examination of our models of *lac* repressor (Fig. 6) and attempts to connect the two bound operators with intervening DNA suggest that formation of a complex between LacR and two operators spaced 52 bp apart would introduce severe or impossible strain when bound either as a loop or wrapped toward the protein; one way to achieve the bidentate interaction with two such closely spaced operators with a smaller energetic cost would be to disrupt the relatively modest interface between the two

LacR dimers and the oligomerization domain (Fig. 7C). Bidentate binding of LacR to DNA's with a range of interoperator distances from 52 through 84 bp has been experimentally observed (6). Two LacR dimers tethered to the tetramerization domain through the flexible segment could accommodate to this range of interoperator distances (Fig. 7C). This flexible arrangement of LacR may be the energetically most favorable one for the wild-type 92-bp spacing between operators O1 and O3. For the 401-bp spacing between O1 and O2, the repressor might or might not retain the crystal quaternary structure.

This general design consisting of two ligand-binding dimers tethered to a tetramerization domain through a flexible hinge region is analogous to the design of antibodies, which have two antigen-binding F_{ab} domains connected through a hinge to its F_c domain (54). This feature of antibody structure has long been recognized as important for crosslinking antigens with various relative orientations. Likewise, flexible connections between DNA binding domains of transcription regulatory proteins may increase the versatility of the protein in binding DNA sites separated along the DNA by varying distances. The tetrameric p53 molecule is also thought to have a flexible region between its DNA binding dimer units and its oligomerization domain (44, 46). Flexible connections may play a role in those dimeric transcriptional regulators that self-associate into larger complexes to loop DNA (8, 11, 13, 14). A model for DNA looping involving flexible connections between domains of *araC*, a dimeric transcriptional regulator that does not self-associate, has also been proposed (9, 55). Thus, multiple DNA binding domains flexibly attached to an oligomerization may prove to be one general structural principle in the architecture of prokaryotic and eukaryotic transcription factors.

REFERENCES AND NOTES

1. M. Ptashne, *Nature* **322**, 697 (1986); H. Echols, *J. Biol. Chem.* **265**, 14697 (1990); R. Schleif, *Annu. Rev. Biochem.* **61**, 199 (1992).
2. A. B. Pardee, F. Jacob, J. Monod, *J. Mol. Biol.* **1**, 165 (1959); W. Gilbert and B. Müller-Hill, *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1891 (1966).
3. W. Gilbert and A. Maxam, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3581 (1973).
4. J. Majors, *ibid.* **72**, 4394 (1975); H. Nick and W. Gilbert, *Nature* **313**, 795 (1985).
5. B. Chen, B. DeCrombrugge, W. B. Anderson, M. E. Gottesman, I. Pastan, *Nature New Biol.* **233**, 67 (1971); S. B. Straney and D. M. Crothers, *Cell* **51**, 699 (1987).
6. H. Krämer *et al.*, *EMBO J.* **6**, 1481 (1987).
7. J. A. Borowiec, L. Zhang, S. Sasse-Dwight, J. D. Gralla, *J. Mol. Biol.* **196**, 101 (1987); P. A. Whitson, W.-T. Hsieh, R. D. Wells, K. S. Matthews, *J. Biol. Chem.* **262**, 14592 (1987).
8. J. Griffith, A. Hochschild, M. Ptashne, *Nature* **322**, 750 (1986).
9. T. Dunn, S. Hahn, S. Ogden, R. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5017 (1984).
10. G. Dandaneil, P. Valentin-Hansen, J. E. L. Larsen, K.

- Hammer, *Nature* **325**, 823 (1987).
11. A. Majumdar and S. Adhya, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6100 (1984); N. Mandal, W. Su, R. Haber, S. Adhya, H. Echols, *Genes Dev.* **4**, 410 (1990).
12. W. Su, S. Porter, S. Kustu, H. Echols, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5504 (1990).
13. B. Théveny *et al.*, *Nature* **329**, 79 (1987).
14. W. Su, S. Jackson, R. Tjian, H. Echols, *Genes Dev.* **5**, 820 (1991).
15. S. Oehler, E. R. Eismann, H. Krämer, B. Müller-Hill, *EMBO J.* **9**, 973 (1990).
16. N. Geisler and K. Weber, *Biochemistry* **16**, 938 (1977).
17. R. T. Ogata and W. Gilbert, *J. Mol. Biol.* **132**, 709 (1979).
18. V. P. Chuprina *et al.*, *ibid.* **234**, 446 (1993).
19. R. Kaptein, E. R. P. Zuiderweg, R. M. Scheek, R. Boelens, W. F. van Gunsteren, *ibid.* **182**, 179 (1985).
20. T. Platt, J. G. Files, K. Weber, *J. Biol. Chem.* **248**, 110 (1973).
21. B. Müller-Hill, *Nature* **302**, 163 (1983).
22. J. U. Bowie, R. Lüthy, D. Eisenberg, *Science* **253**, 164 (1991).
23. N. K. Vyas, M. N. Vyas, F. A. Quijoco, *J. Biol. Chem.* **266**, 5226 (1991); F. A. Quijoco, *Curr. Opin. Struct. Biol.* **1**, 922 (1991).
24. J. C. Nichols, N. K. Vyas, F. A. Quijoco, K. S. Matthews, *J. Biol. Chem.* **268**, 17602 (1993).
25. M. A. Schumacher, K. Y. Choi, H. Zalkin, R. G. Brennan, *Science* **266**, 763 (1994).
26. J. Kania and D. T. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3529 (1976).
27. M. D. Barkley, A. D. Riggs, A. Jobe, S. Bourgeois, *Biochemistry* **14**, 1700 (1975).
28. LacR was purified [J. M. Rosenberg, O. B. Khallai, M. L. Kopka, R. E. Dickerson, A. D. Riggs, *Nucleic Acids Res.* **4**, 567 (1977)] from a T7 RNA polymerase expression clone. The tryptic core fragment in 10 mM sodium 2-N-morpholineethane sulfonic acid (NaMES), pH 7.0, 0.1 mM IPTG, 0.02 percent (w/v) Na₂S₂O₈, and 10⁻⁵ M NaEDTA was reacted with ethylmercury phosphate, concentrated to 17 mg/ml, and then dialyzed against 100 mM NaMES, pH 6.4, 4 percent (w/v) polyethylene glycol (PEG) 6K, 0.1 mM IPTG, 0.02 percent (w/v) Na₂S₂O₈. Crystals grew in space group P2₁, with cell dimensions *a* = 90.2 Å, *b* = 64.7 Å, *c* = 117.9 Å, and β = 91.8°. Crystals were harvested into 30 mM NaMES, pH 6.4, 8 percent (w/v) PEG 6K, 1 mM IPTG, 0.02 percent (w/v) Na₂S₂O₈ and stabilized in the same solution with 10 percent PEG 6K. After serial transfer into the latter solution with increasing concentrations of ethylene glycol to 25 percent (v/v) (final), crystals were shock-frozen in liquid propane and data were collected at -165°C [H. Hope *et al.*, *Acta Crystallogr.* **B45**, 190 (1989)].
29. W. A. Hendrickson, *Science* **254**, 51 (1991).
30. MAD data were collected on Fuji imaging plates at the F2 beamline at CHESS. Because of limited beam time, we measured only one Bijvoet difference and one dispersive difference. The Bijvoet difference was measured by inverse beam geometry (collecting oscillation films at ϕ and $\phi + \pi$) at a nominal $\lambda = 0.9840$ Å, which is on the high energy (large *f'*) side of the Hg L_{III} absorption edge. The dispersive difference was measured by collecting oscillation films at ϕ at a nominal $\lambda = 1.0088$ Å, the inflection point (most negative *f'*) of the measured fluorescence spectrum. Each set of data used for MAD phase determination thus consisted of three measurements of the same reflection matched in time and absorption path, a single measurement at the absorption edge, and a Bijvoet pair on the high energy side of the edge. This choice of measurements efficiently records the anomalous signal for elements such as mercury that lack strong peaks in their x-ray fluorescence spectra.
31. J. L. Smith *et al.*, *Science* **264**, 1427 (1994).
32. A. M. Friedman *et al.*, in preparation. MADPRB incorporates local block scaling algorithms for both Bijvoet and cross-wavelength scaling, a reformulation of the structure factor algebra with the use of *F* (rather than *f*), and a χ^2 and Bayesian-based analysis of phase probabilities.
33. In addition to the matched data collected for MAD phased determination, additional unmatched data were collected on the same frozen crystal to provide a complete set of observed amplitudes for phase

improvement and structure refinement. All the data from 23 to 2.6 Å were scaled and merged, treating as equivalent all observations related by point group symmetry or Bijvoet symmetry and collected at either wavelength, yielding a merging *R* on intensity of 9.5 percent (Table 1). The process of phase improvement was based on the RAVE suite [T. A. Jones, in *Molecular Replacement*, E. J. Dodson, S. Glover, W. Wolf, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1992), p. 91]. An initial mask for the refinement of noncrystallographic symmetry operators was constructed around a single monomer by orienting the homologous monomeric ribose binding protein (37) with a cross-rotation function and a phased translation function with the use of the experimental MAD phases in the program X-PLOR [A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988)]. Subsequent masks were constructed around partial models of LacR. Because of the absence of point group symmetry, initial operators relating this monomer to the other three were determined by pairwise superposition of the heavy atom sites followed by real space correlation refinement against the MAD map. Multiple sets of superpositions were screened until operators with correlation coefficients greater than 0.4 were found. The monomer mask and noncrystallographic symmetry (NCS) operators were then used to build a tetramer mask, and the MAD phases to 4.5 Å resolution were improved through histogram matching and solvent flattening with the tetramer mask and the program SQUASH [K. Y. J. Zhang, *Acta Crystallogr.* **D49**, 213 (1993)]. Maps calculated with $F_o - F_c$ amplitudes and combined phases were subjected to iterative cycles of symmetry averaging and solvent flattening in RAVE followed by Fourier inversion and phase combination at 4.5 Å (Fig. 1B). Symmetry averaging was repeated at 4.0 Å, 3.7 Å, and 3.5 Å, incorporating initial phases from the MAD experiment at each step, with subsequent extension from 3.5 Å to the limit of resolution, 2.6 Å, in 24 steps without experimental phases. A preliminary model was fitted into the 2.6 Å resolution map. Nearly every residue between 61 and 332 was readily visible. The NCS operators of the missing COOH-terminal residues were subsequently found to deviate signifi-

cantly from the NCS operators of residues 61 through 332. The missing residues were visible but poorly defined in a map calculated with $3F_o - 2F_c$ amplitudes and averaged phases. The averaging process was completed by averaging maps calculated with $2F_o - F_c$ amplitudes and combined phases at a fixed resolution of 2.6 Å. Fourfold averaging was applied within the region around residues 61 to 332, while electron density within the COOH-terminal region was not averaged but was protected from solvent flattening (Fig. 1C). The *R* factor on Fourier inversion of this final map was 18.8 percent (Table 1).

34. Crystallographic refinement consisted of two cycles of simulated annealing with both phase and NCS restraints. Each cycle consisted of slow-cooling simulated annealing followed by Powell minimization and manual rebuilding. These were followed by one cycle of simulated annealing without restraint, Powell minimization and conventional refinement of group B factors for main and side chain atoms of each residue. The refined coordinates have been deposited with the Brookhaven Protein Data Bank, accession code 1TLF.

35. S. Alberti, S. Oehler, B. v. Wilcken-Bergmann, H. Krämer, B. Müller-Hill, *New Biol.* **3**, 57 (1991).

36. A. E. Chakerian *et al.*, *J. Biol. Chem.* **266**, 1371 (1991).

37. S. L. Mowbray and L. B. Cole, *J. Mol. Biol.* **225**, 155 (1993).

38. J. Monod, J. Wyman, J. P. Changeux, *ibid.* **12**, 88 (1965).

39. D. B. McKay, C. A. Pickover, T. A. Steitz, *ibid.* **156**, 175 (1982).

40. M. Charlier, J. C. Maurizot, G. Zaccai, *Nature* **286**, 423 (1980).

41. Y. Ohshima, T. Horiuchi, M. Yanagida, *J. Mol. Biol.* **91**, 515 (1975).

42. D. Pörschke, *Biophys. Chem.* **28**, 137 (1987).

43. J. Janin, S. Miller, C. Chothia, *J. Mol. Biol.* **204**, 155 (1988).

44. W. Lee *et al.*, *Nature Struct. Biol.* **265**, 386 (1994).

45. W. S. El-Deiry *et al.*, *Nature Genet.* **1**, 45 (1992).

46. Y. Cho, S. Gorina, P. D. Jeffrey, N. P. Pavletich, *Science* **265**, 346 (1994).

47. J. S. Sack, M. A. Saper, F. A. Quiocho, *J. Mol. Biol.*

206, 171 (1989); B.-H. Oh *et al.*, *J. Biol. Chem.* **268**, 11348 (1993).

48. A. J. Sharff, L. E. Rodseth, J. C. Spurlino, F. A. Quiocho, *Biochemistry* **31**, 10657 (1992).

49. Interference with the allosteric transition has been suggested for substitutions of Lys⁹⁴ [W.-I. Chang, J. S. Olson, K. S. Matthews, *J. Biol. Chem.* **268**, 17613 (1993)].

50. A. Schmitz, U. Schmeissner, J. H. Miller, P. Lu, *ibid.* **251**, 3359 (1976).

51. A. E. Chakerian and K. S. Matthews, *ibid.* **266**, 22206 (1991).

52. S. Zinkel and D. M. Crothers, *Nature* **328**, 178 (1987); H.-M. Wu and D. M. Crothers, *ibid.* **308**, 509 (1984).

53. Deoxyribonuclease I cleavage of DNA containing two closely spaced operators bound to LacR shows hypersensitivity 15 bp from the center of the operators, suggesting an opening of the minor groove on the side away from the protein and thus bending toward the protein.

54. L. J. Harris *et al.*, *Nature* **360**, 369 (1992).

55. J. H. Carra and R. F. Schleif, *EMBO J.* **12**, 35-44.

56. D. T. Cromer and D. Liberman, *J. Chem. Phys.* **53**, 1891 (1970).

57. Program MOLSCRIPT [P. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991)].

58. T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).

59. M. Carson, *J. Appl. Crystallogr.* **24**, 958 (1991).

60. L. G. Kleina and J. H. Miller, *J. Mol. Biol.* **212**, 295 (1990); A. J. E. Gordon *et al.*, *ibid.* **200**, 239 (1988).

61. We thank W. Hendrickson for instructing us in the use of MAD; D. Crothers, R. Schleif, Z. Otwinowski, and J. Friedman for helpful discussions; S. Ealick, Q. Shen, P. Doing, T. Irving, and the staff of CHESS for assistance in data collection; P. Raccuia and S. J. Smerdon for constructing the LacR expression vector; V. Rath for comments on the manuscript; and D. Jeruzalmski, C. Correll, W. Yang, Y. Shamoo, L. Feinberg, J. Tormo, C. Brautigam, P. Klosterman, C. Scafe, and S. Bellon for their help in synchrotron data collection. Supported by NIH grant GM22778 to T.A.S.

10 November 1994; accepted 8 May 1995

AAAS–Newcomb Cleveland Prize

To Be Awarded for a Report, Research Article, or an Article Published in *Science*

The AAAS–Newcomb Cleveland Prize is awarded to the author of an outstanding paper published in *Science*. The value of the prize is \$5000; the winner also receives a bronze medal. The current competition period began with the 3 June 1994 issue and ends with the issue of 26 May 1995.

Reports, Research Articles, and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration for the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the competition period, readers are invited to nominate papers appearing in the Reports, Research Articles, or Articles sections. Nominations must be typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to the AAAS–Newcomb Cleveland Prize, AAAS, Room 924, 1333 H Street, NW, Washington, DC 20005, and **must be received on or before 30 June 1995**. Final selection will rest with a panel of distinguished scientists appointed by the editor-in-chief of *Science*.

The award will be presented at the 1996 AAAS annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.