Protein Science (1998), 7:1485–1494. Cambridge University Press. Printed in the USA.

Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: Possible implications for nonspecific binding by the wild-type protein

RONALD A. ALBRIGHT,[1,3] MICHAEL C. MOSSING,[2] AND BRIAN W. MATTHEWS[1]

[1]Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403
[2]Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556

(RECEIVED January 26, 1998; ACCEPTED April 10, 1998)

Abstract

The structure has been determined at 3.0 Å resolution of a complex of engineered monomeric Cro repressor with a seven-base pair DNA fragment. Although the sequence of the DNA corresponds to the consensus half-operator that is recognized by each subunit of the wild-type Cro dimer, the complex that is formed in the crystals by the isolated monomer appears to correspond to a sequence-independent mode of association. The overall orientation of the protein relative to the DNA is markedly different from that observed for Cro dimer bound to a consensus operator. The recognition helix is rotated 48° further out of the major groove, while the turn region of the helix-turn-helix remains in contact with the DNA backbone. All of the direct base-specific interactions seen in the wild-type Cro-operator complex are lost. Virtually all of the ionic interactions with the DNA backbone, however, are maintained, as is the subset of contacts between the DNA backbone and a channel on the protein surface. Overall, 25% less surface area is buried at the protein–DNA interface than for half of the wild-type Cro-operator complex, and the contacts are more ionic in character due to a reduction of hydrogen bonding and van der Waals interactions. Based on this crystal structure, model building was used to develop a possible model for the sequence–nonspecific interaction of the wild-type Cro dimer with DNA. In the sequence-specific complex, the DNA is bent, the protein dimer undergoes a large hinge-bending motion relative to the uncomplexed form, and the complex is twofold symmetric. In contrast, in the proposed nonspecific complex the DNA is straight, the protein retains a conformation similar to the apo form, and the complex lacks twofold symmetry. The model is consistent with thermodynamic, chemical, and mutagenic studies, and suggests that hinge bending of the Cro dimer may be critical in permitting the transition from the binding of protein at generic sites on the DNA to binding at high affinity operator sites.

Keywords: Cro; helix-turn-helix; λ repressor; monomer–dimer; protein–DNA interaction

In common with a number of DNA-binding proteins (Berg et al., 1981), Cro protein from phage λ (Anderson et al., 1981) interacts with DNA in two distinctly different fashions (Takeda et al., 1986). Following initial contact with the DNA at some random noncognate site, Cro diffuses to its operator sites where the character of the protein–DNA interactions change (Kim et al., 1987). The nonspecific complex is stabilized predominantly by ionic interactions. Following a conformational change that buries substantially more surface area, additional van der Waals contacts, hydrogen bonds, and hydrophobic interactions are established with the operator (Takeda et al., 1986, 1992). In parallel with the two types of Cro–DNA complexes, in the absence of DNA, wild-type Cro might exist as a mixture of monomers and dimers (Jana et al., 1997). At submicromolar concentrations in vitro, monomers predominate. It is not clear which species is responsible for nonspecific binding.

The recent structure determination of a Cro-operator complex to 3.0 Å resolution (Albright & Matthews, 1998a, 1998b) has allowed the sequence-specific interactions between the protein and the DNA to be seen for the first time. The structure of the complex with noncognate DNA has, however, remained a mystery.

In this paper, we describe the crystal structure at 3.0 Å resolution of an engineered monomeric form of Cro (Mossing & Sauer, 1990) bound to a seven-base pair DNA duplex. This structure is distinctly different from that observed for wild-type Cro-operator complexes (Brennan et al., 1990; Albright & Matthews, 1998b). The protein binds the DNA in a sequence–nonspecific manner, and suggests how wild-type Cro might interact with noncognate DNA.

Reprint requests to: Brian W. Matthews, Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403.

[3]Present address: Howard Hughes Medical Institute, Department of Molecular Biophysics and Biochemistry, Yale University, Whitney Avenue/JWG 423, P.O. Box 208114, New Haven, Connecticut 06520-8114.

To date, there is relatively little information on the structure of such complexes (Luisi et al., 1991; Winkler et al., 1993; Gewirth & Sigler, 1995).

Results

Nomenclature

Wild-type Cro is a dimer of 66 residues numbered 1–66 for one monomer and 1′–66′ for the other. Secondary structural elements include three α-helices and three β-strands: β1 (residues 3–6), α1 (7–14), α2 (16–23), α3 (27–36), β2 (39–45), and β3 (49–56) (Anderson et al., 1981). Helices α2 and α3 form the helix-turn-helix motif, with α3 commonly referred to as the "recognition" helix. The $\beta 3\beta 3'$ antiparallel β-ribbon (residues 54–56 and 54′–56′) forms part of the dimer interface. By inserting five amino acids following Lys56, Mossing and Sauer (1990) generated a stable, monomeric form of Cro that is designated either Cro K56-[DGEVK] or simply "Cro monomer." The sequence numbering of wild type is retained for the monomer, except for the five inserted residues, which are identified as Asp56a, Gly56b, Glu56c, Val56d, and Lys56e. The resulting overall sequence of this 71-residue protein is thus 1–56, 56a–56e, and 57–66. The crystal and solution structures of Cro K56-[DGEVK] have been determined (Albright et al., 1996; Mossing, 1998). Residues 56a and 56b form a tight β-turn such that residues 56c–56e and 57–66 make interactions analogous to residues 54′–56′ and 57′–66′ respectively, of the wild-type dimer. Accordingly, the Cro monomer has an additional β-strand, β4, which replaces part of β3′ of the wild-type dimer.

Quality of the structure

The engineered Cro monomer was crystallized with a seven-base pair duplex (Fig. 1) and the structure determined and refined to 3.0 Å resolution as described under Materials and methods. Although the resolution of the structure determination is limited, the quality of the electron density map (Fig. 2) is sufficient to resolve base pairs and side chains, except for the disordered terminal residues 1 and 62–66. These residues are also disordered in the structure of the native protein (Anderson et al., 1981; Ohlendorf et al., 1998) and its complex with operator DNA (Albright & Matthews, 1998b). All nonglycine main-chain torsion angles fall within allowed regions of the Ramachandran plot (not shown), with 73% in the "most-favored" regions and the remaining 27% in the "additional allowed" regions as defined by PROCHECK (Laskowski et al., 1993). Subject to the limitations of the crystallographic resolution, the main-chain thermal factors observed in the complex are qualitatively similar to those observed in wild-type Cro (Anderson et al., 1981; Ohlendorf et al., 1998), in the Cro-operator complex (Albright & Matthews, 1998b) and in the noncomplexed Cro monomer (Albright et al., 1996).

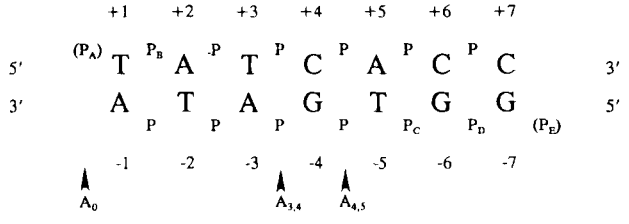

Fig. 1. DNA duplex used for the structure determination. The sequence shown was used for the crystallization, and corresponds to part of the consensus half-operator of Cro and λ-repressor proteins (Ptashne, 1986). Phosphates labeled $P_A \ldots P_E$ follow the convention used to describe the crystal structures of the operator complexes of λ-repressor (Jordan & Pabo, 1988) and Cro (Albright & Matthews, 1998b). The terminal phosphates, including P_A and P_E in parentheses, are absent. Because the binding of this DNA fragment to the Cro monomer is completely out of register with the binding of operator DNA to wild-type Cro, the phosphates $P_A \ldots P_E$ do not bind at analogous sites in the respective complexes. A_0, $A_{3,4}$, and $A_{4,5}$ indicate possible locations that might be considered for the crystallographic twofold axis (see Materials and methods).

The complex is sequence nonspecific

The overall complex is shown in Figure 3. Most strikingly, the orientation of the protein relative to the DNA is rotated about 48° from its position in the wild-type Cro-operator complex. In this alternative mode of interaction, no base-specific contacts are made by the helix-turn-helix (HTH) region, the motif that makes all of the specific interactions in the wild-type complex. The turn region (labeled $N_{\alpha 3}$ in Fig. 3A,B) remains in contact with the sugar–phosphate backbone, but both the beginning and end of the HTH (near residues Gln16 and His35, respectively) are now located more than 10 Å from the DNA backbone. In contrast, both of these ends directly contact phosphate groups in the wild-type Cro-operator complex (Albright & Matthews, 1998b). In terms of the recognition helix, the N-terminus remains near the DNA but does not make direct interactions with base pairs, while the C-terminus is well away.

Another feature accentuates the nonspecific nature of the complex. As discussed in Materials and methods, a crystallographic twofold axis passes through the middle region of the DNA fragment, between two base pairs, such that two Cro monomers contact the same fragment from opposite sides. Because the DNA is not palindromic (Fig. 1), these regions of contact necessarily have different sequences. The DNA is statistically disordered throughout the crystal, able to bind in either orientation. Notwithstanding this disorder, there are distinct peaks in the electron density map at the phosphate positions, and the planes of the base pairs are clearly resolved. This preservation of the phosphate positions permits conservation of the protein contacts with the DNA backbone (Fig. 4).

Interactions with the DNA

The Cro monomer interacts with the DNA predominantly through the sugar–phosphate backbone (Fig. 4). In the wild-type Cro-operator complex, two distinct regions of DNA backbone contacts were observed (Albright & Matthews, 1998b). In the first or "innermost" region, the DNA backbone passed through a channel and made salt-bridge contacts to positively charged side chains. The second or "outermost" region was more solvent exposed and consisted primarily of noncharged side chains. In the monomer complex, the innermost set of contacts are maintained, while the outermost ones are lost.

The aromatic ring of Tyr26 makes van der Waals contacts (3.2 Å) with the DNA backbone. However, unlike the wild-type Cro-operator complex, the hydroxyl group of this residue extends beyond the sugar–phosphate backbone into solvent and does not hydrogen bond with the DNA. The main-chain amide of Tyr26 donates a hydrogen bond (2.7 Å) to a phosphate group. The posi-

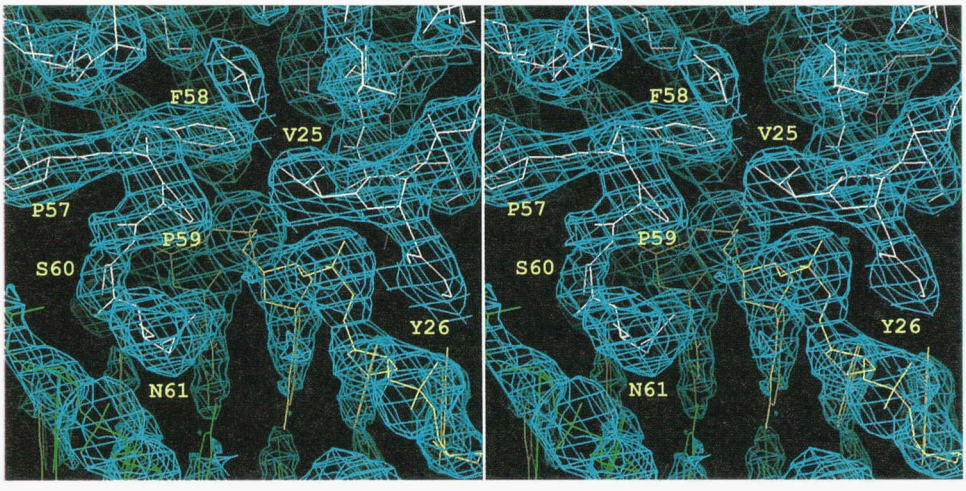

Fig. 2. Stereo figure showing the electron density in the region where the sugar–phosphate backbone of the DNA (yellow) passes through the channel on the surface of the Cro monomer (white). The part of the protein shown includes Phe58, which penetrates into the hydrophobic core, and the C-terminal residues (to Asn61), which occupy the minor groove of the DNA. Coefficients are $2F_o - F_c$ and phases are from the refined model. The map is contoured at 1σ.

tion of this phosphate relative to the protein is analogous to the "P_C" phosphate of the wild-type complex, lying at the beginning of the protein channel, partially buried by Val25, Tyr26, and Ala29. The next phosphate, analogous to "P_D" of the wild-type complex, is completely buried in the deepest part of the channel, where it is surrounded by Ala29, Ile30, Ala33, Arg38, Phe58, Pro59, and Ser60. Arg38 salt bridges (2.7 Å) to one of the phosphate oxygens. As in wild type, Arg38 is also buttressed by a salt bridge with Glu54. The

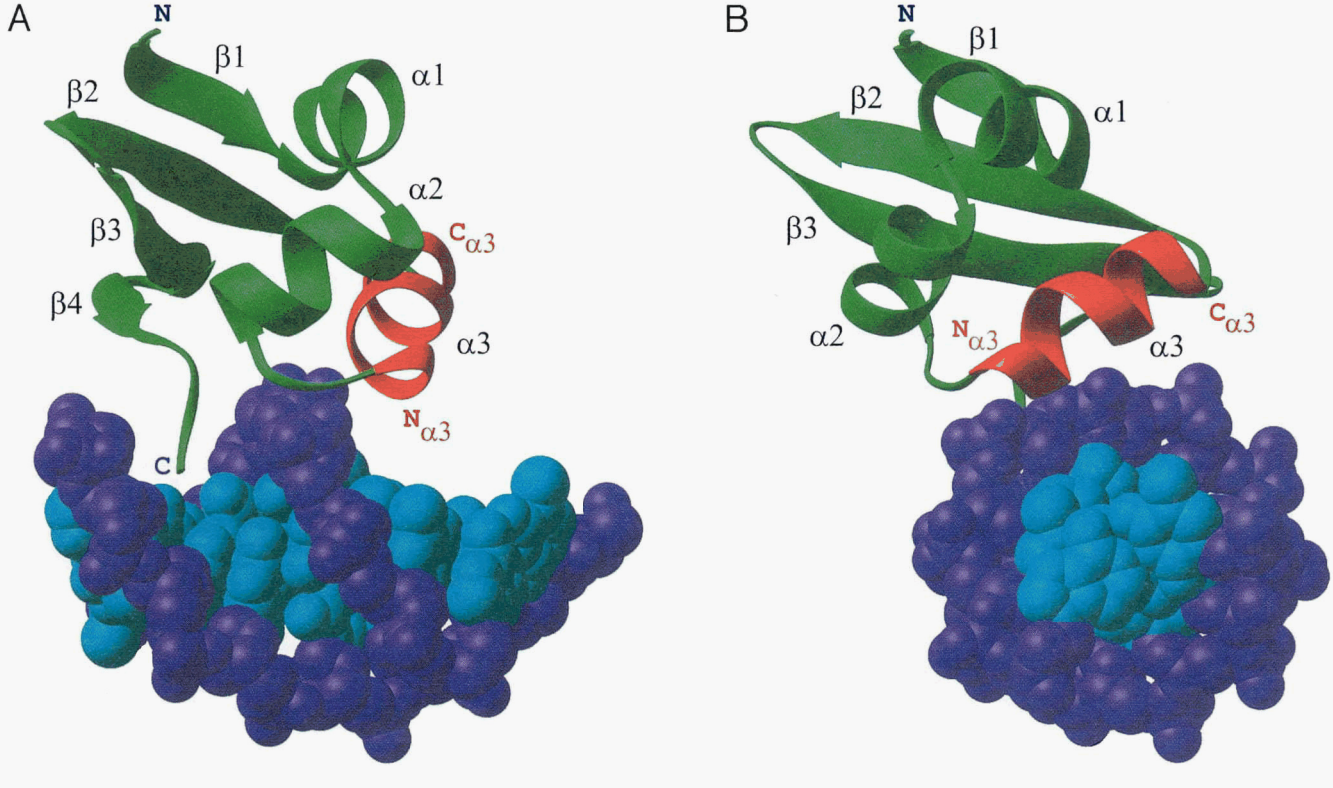

Fig. 3. Drawings showing the overall arrangement of the engineered Cro monomer bound to the 7-mer DNA duplex, as seen in the crystal structure. The "recognition" helix is shown in red, and the polypeptide chain as shown includes residues 2–61. Secondary structure elements are identified. $N_{\alpha 3}$ and $C_{\alpha 3}$ show, respectively, the N- and C-termini of the recognition helix. **A:** View parallel to the grooves of the DNA. **B:** View along the axis of the DNA.

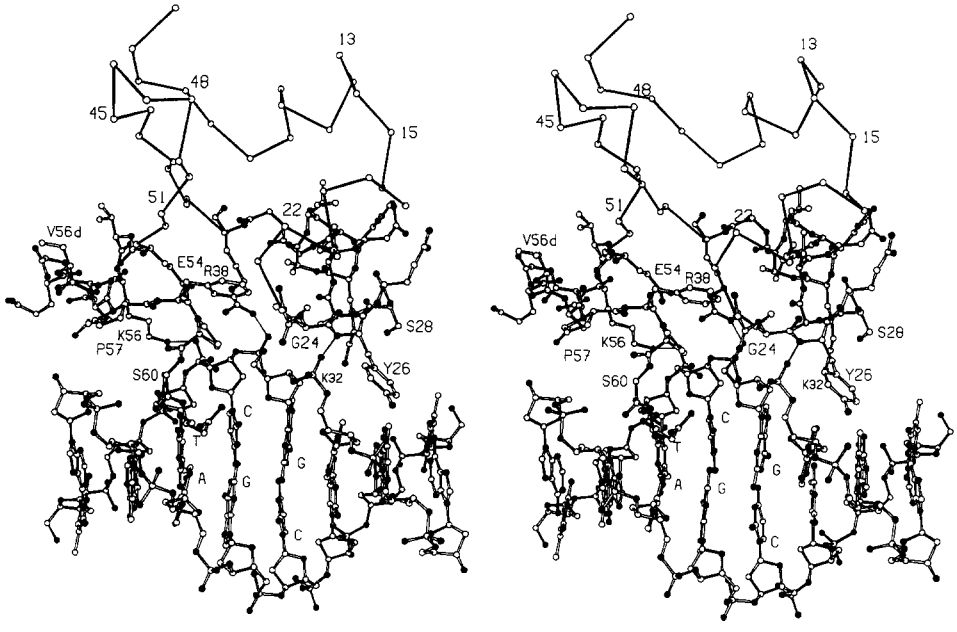

Fig. 4. Stereo figure showing the details of the protein–DNA interface. For clarity, protein side chains are shown if close to the DNA, but elsewhere only the C^{α}–C^{α} trace is included. Oxygen and nitrogen atoms are drawn solid; carbon atoms are shown as open circles. Hydrogen bonds between the protein and the DNA are shown as thin lines.

next phosphate group, analogous to "P_E" of the wild-type complex, emerges from the other end of the channel, where it forms a salt bridge (3.1 Å) with Lys56. Interactions with each of these phosphate groups are also seen in the wild-type Cro-operator complex.

The only positions where a direct sequence-specific interaction might occur are, surprisingly, at sites that are not involved in direct contacts in the wild-type operator complex. The C-terminus of the protein main chain crosses over the DNA backbone and into the minor groove at Ser60, whose main chain is close enough to potentially contact a base (Fig. 4). The side chain of Asn61 also appears to be located close enough (3.3 Å) to directly interact with a base in the minor groove. Unambiguous verification of possible sequence-specific contacts, however, is precluded by the effective degeneracy of the bases due to the statistical disorder in the crystal packing.

Protein conformation

A difference–distance plot comparing the C^{α}–C^{α} distances in the monomer complex with those in the apo Cro structure (Anderson et al., 1981; Ohlendorf et al., 1998) is shown in Figure 5. The recognition helix (27–36) shifts away from the core region of the β-sheet (residues 40–44, 50–53) by 0.9 Å, and from Phe58 by 0.5 Å. A slight straightening of the solvent-exposed $\beta2\beta3$-hairpin (45–48) moves it 1.6 Å away from residue 55. Qualitatively similar shifts relative to apo Cro occur when Cro binds operator DNA (Albright & Matthews, 1998b) and in the Cro monomer structure itself (Albright et al., 1996). The apo Cro monomer structure, however, exhibited a greater distortion of the core than is observed here. The engineered β-turn in the present complex (between $\beta3$ and $\beta4$) adopts a somewhat different conformation than in the apo monomer, seemingly in association with adjustments in the preceding residue, Lys56, which now interacts with the DNA. Structural heterogeneity is also seen in this region in the solution structure (Mossing, 1998). Additionally, Pro57 adopts a somewhat different conformation to avoid a steric clash with the DNA (Fig. 4). As a result of these adjustments, the engineered region of the monomer

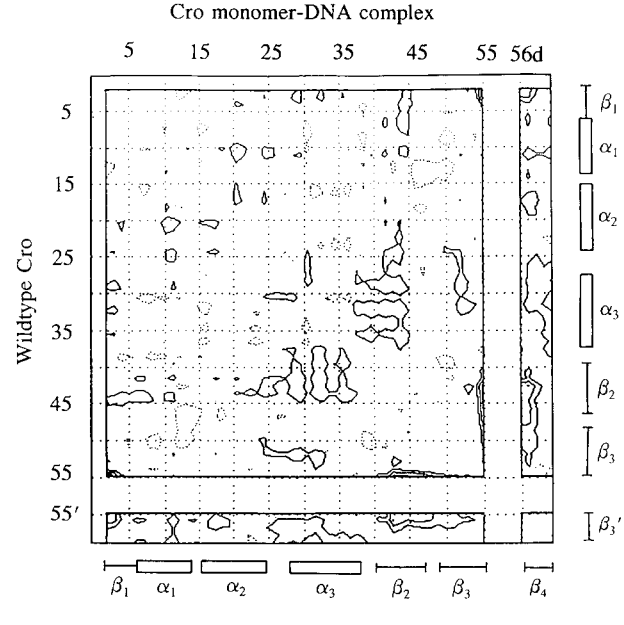

Fig. 5. Plot showing the change in distance between all pairs of α-carbon atoms in the Cro monomer DNA complex relative to the crystal structure of the uncomplexed wild-type Cro dimer (Ohlendorf et al., 1998). Contours drawn at increments of 0.5 Å, with the zero contour omitted. The solid contours show pairs of α-carbon atoms that are further apart in the monomer complex than in the wild-type dimer (e.g., on complex formation the $\beta2\beta3$ hairpin loop (45–48) moves 1.6 Å further away from residue 55). Locations of α-helices and β-strands are also shown.

adopts a conformation more similar to that observed in the apo Cro dimer. Consistent with these various structural changes, it has also been shown that the core of the Cro monomer is imperfectly packed, allowing adjustments that may be important in DNA recognition (Mollah et al., 1996).

DNA conformation

The DNA in the complex appears to be essentially straight B-form with some local distortions (Table 1). It should be remembered, however, that the crystal structure represents the statistical average of two or more oligomers in different orientations straddling a crystallographic twofold axis. The statistical disordering maintains the location of the phosphate backbone. The situation is different, however, for the base pairs, where the statistical disordering averages different base pairs at a given position. The overall result is equivalent to a DNA fragment with a unique backbone and composite base pairs. For this reason, the overall position of the DNA and its orientation relative to the protein are well established, but the conformations of the individual nucleotides are somewhat uncertain.

Discussion

Comparison with the wild-type Cro-operator complex

Because the DNA used in the Cro monomer complex corresponds to essentially one-half of the consensus Cro operator, it might be anticipated that the spatial relationship between the DNA and the protein would resemble one-half of the wild-type Cro-operator complex (Brennan et al., 1990; Albright & Matthews, 1998b). The two complexes, however, are strikingly different (Fig. 6). In the wild-type Cro-operator complex (Albright & Matthews, 1998b), the DNA is bent by about 40°, and numerous sequence-specific interactions occur between the HTH region of the protein and the base pairs within the major groove. In contrast, the monomer complex contains essentially straight B-form DNA and the relative orientation of the DNA with respect to the protein differs by 48° (Fig. 6). All sequence-specific interactions of the wild-type complex are lost, with only the turn of the HTH remaining close to the DNA.

Table 1. *DNA conformation*[a]

Base pair	Propellor twist (°)	Buckle (°)	Rise (Å)	Tilt (°)	Roll (°)	Helical twist (°)
1	0	−27				
			2.8	10	−2	43
2	−7	7				
			3.6	3	8	28
3	−18	−3				
			3.8	−7	−2	35
4	−7	−13				
			2.9	0	−11	45

[a]As described in the text, the DNA is statistically disordered about a crystallographic twofold axis. The parameters in the table correspond to four base pairs on one side of the twofold axis, with the outermost base pair (#1) having 50% occupancy. The remaining four base pairs are related by symmetry. Parameters were calculated using the program CURVES (Lavery & Sklenar, 1988).

A subset of the interactions between the protein and the DNA backbone, however, are retained in a region where the backbone passes through a channel on the surface of the protein. These include interactions with the backbone amide of Tyr26 and the side chains of Arg38 and Lys56 (Figs. 2, 4). The Cro–monomer complex buries 1,030 $Å^2$ of the surface area and is, therefore, 25% less intimate than the Cro-operator complex, which buries 1,376 $Å^2$ per half-complex. While all of the ionic interactions with the sugar–phosphate backbone observed in the wild-type operator complex are retained in the monomer complex, there is a substantial reduction in the number of hydrogen bonding interactions and van der Waals contacts due to loss of direct interactions within the major groove, as well as loss of the outermost region of DNA backbone contacts. The overall nature of the interactions in the monomer complex are, therefore, substantially more ionic in character than in the Cro-operator complex.

Rationale for nonspecific binding

Why the present complex is nonspecific remains an open question. One possible reason is that the intrinsically weak binding of the engineered Cro monomer is further reduced by the absence of phosphate contacts in the short DNA fragment used for crystallization. Although the seven-base pair fragment contains all of the bases that are contacted in the wild-type Cro-operator complex (Albright & Matthews, 1998b), it lacks two of the contacted phosphate groups (P_A and P_E), located at the 5′-terminal positions of each strand (Fig. 1). The P_A phosphate group, however, was also absent in the DNA fragment used in the low-resolution Cro-operator complex described by Brennan et al. (1990), and so would seem to be nonessential for specific complex formation. The binding of the engineered Cro monomer to a pseudo-symmetrized consensus–operator has been assayed using DNaseI footprinting and shown to be weak (Mossing & Sauer, 1990). In contrast to wild-type Cro, operator binding by the monomer could only be detected using low salt at 0 °C, with a reduction in binding affinity of at least 2,000-fold relative to wild type. Under these conditions, the Cro monomer exhibited a half-site preference based solely on the identity of the central base pair, the only position of asymmetry in the operator used. This is consistent with the present complex, which shows that base-specific interactions might be possible within the minor groove. The degree of specificity with which the engineered monomer binds DNA has not been established conclusively. Also, NMR studies that assayed the binding of the wild-type Cro dimer to operator half-sites showed the binding to be weak and potentially nonspecific (Baleja et al., 1991). As such, the original assumption that a monomer bound to an operator half-site would necessarily resemble one-half of the wild-type Cro-operator complex may be in question.

Kinetic studies (Kim et al., 1987) have shown Cro binding to be consistent with a two-step mechanism in which Cro first contacts the DNA at an arbitrary site and then travels along the DNA to its operator sites via a process consistent with a sliding mechanism (Berg et al., 1981). Cro interacts with noncognate and cognate DNA in distinctly different ways (Takeda et al., 1986, 1992). Affinity for its 17-base pair nonspecific sites (10^7 M^{-1}) is substantially less than for the specific operator sites (10^{11} M^{-1}). Takeda et al. (1992) have determined the thermodynamic parameters of both types of complexes using pulsed-flow microcalorimetry. Their data are consistent with the view that the interactions of Cro with noncognate DNA are more ionic in character, and that the overall

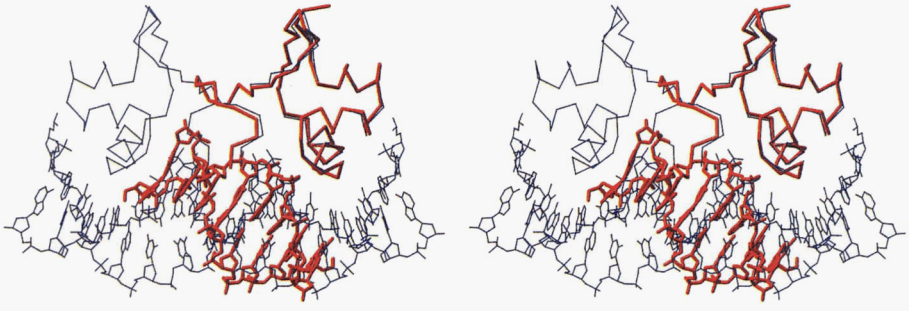

Fig. 6. Stereo drawing showing the superposition of the monomer–DNA complex described herein (red) on the sequence-specific complex of the wild-type Cro dimer with operator DNA (blue) (Albright & Matthews, 1998b). The superposition is based on the main-chain atoms of the helical region of the protein (residues 7–36).

complex is looser, burying less surface area and containing fewer hydrogen bonds and van der Waals interactions than in the specific complex. The formation of either type of complex is, in part, entropy driven, reflecting the displacement of counter ions and water molecules from the DNA surface. The increased ionic nature of the nonspecific complex is reflected in its greater salt sensitivity, with half-dissociation occurring around 70 mM KCl, while specific complexes appear to remain stable even at 500 mM KCl (Takeda et al., 1986).

Takeda et al. (1986) showed that the binding of wild-type Cro to sequence nonspecific DNA most strongly protects Lys56 from alkylation. This is the lysine that contacts the DNA in the present complex (Fig. 4). The second most protected lysine is Lys32, which is not in direct contact but could also be protected by the presence of the DNA (Fig. 4).

The disordered C-terminal tail of Cro (residues 62–66) plays a critical role in DNA binding, as determined by mutagenesis (Hubbard et al., 1990). The wild-type protein contains lysines at positions 62 and 63. Deletion of either of these impairs both specific and nonspecific binding. Furthermore, adding an extra positive charge at positions 64 or 65 increases nonspecific affinity almost 10-fold, but has no impact on specific binding. This indicates that the C-terminal region of Cro remains in the vicinity of the DNA backbone in the wild-type nonspecific complex, and that the interactions differ somewhat in the nonspecific and specific complexes.

In summary, the structure of the engineered monomer in complex with DNA described here appears to be consistent with all of the attributes expected for a nonspecific Cro–DNA complex. It might also serve as a model for the binding of wild-type Cro monomer to DNA (Jana et al., 1997).

Proposed model for the binding of wild-type Cro to noncognate DNA

Based on the observed complex of monomeric Cro with DNA, a tentative model can be proposed for the interaction of native, dimeric Cro with noncognate DNA. The preponderance of evidence (Kirpichnikov et al., 1985; Lee et al., 1987; Torigoe et al., 1991) suggests that the DNA in the nonspecific complex is essentially straight. [Atomic force microscopy, however, has suggested that wild-type Cro bends the DNA by 62 ± 23° when bound at nonoperator sites (Erie et al., 1994).] Thus, in the first stage of model building, the DNA in the observed structure of the monomer complex was simply extended. Then, to extrapolate from the known position of the monomer two options were considered to construct a dimer of Cro, either (1) the Cro dimer might have a conformation resembling that in the sequence-specific complex (Albright & Matthews, 1998b) or, (2) the structure of the dimer might correspond to that in the unbound form (Anderson et al., 1981; Ohlendorf et al., 1998). Tests of the first option by model building resulted in major steric clashes between the protein and the DNA (Fig. 6), which could only be rectified by substantially bending the DNA away from the protein. This arrangement, therefore, seemed unlikely. On testing the second option, however, the second Cro subunit was found to be close to the DNA, but in such a way that the part of the protein surface adjacent to the DNA was different from that in the first subunit. This arrangement, which might be considered as a possible model for the nonspecific interaction of Cro with DNA, is shown in Figures 7 and 8.

The contacts made by the first subunit with the DNA are, by definition, the same as those in the Cro monomer–DNA complex (Figs. 2, 4) and include primarily residues Val25, Tyr26, Arg38, Lys56, Phe58, Pro59, Ser60, and Asn61. Residues at the amino-terminus of the recognition helix (Gln27, Ser28) are close to the DNA but do not make direct contact with the bases (Figs. 4, 7A). In contrast, in the second subunit, as placed by model building, it is the C-terminus of the recognition helix (His35′) that contacts the DNA backbone, while the N-terminus of the helix lies far away (Fig. 7A,B). Arg38′ still maintains a phosphate contact, but due to the different relative orientation of the protein the DNA backbone lies along the edge of the protein channel rather than within it. Asn61′ and the main chain of Ser60′ appear able to directly contact bases in the minor groove, while the corresponding region of the other monomer is greater than 8 Å from the DNA. Lys32′ appears able to directly contact bases, and Lys39′ makes ionic interactions with the DNA backbone.

This model buries 41% less surface area at the protein–DNA interface than in the wild-type Cro-operator complex (1,622 Å^2 versus 2,751 Å^2) yet maintains virtually all of the ionic interactions. Essentially all of the base-specific contacts observed in the operator complex are lost, as are many of the DNA backbone contacts. These features are consistent with the data from thermodynamic, NMR, and mutant studies. On the other hand, the model does not match the twofold symmetry of the protein with the approximate twofold symmetry of the DNA backbone (Fig. 7B). In most cases, macromolecules tend to associate in a symmetrical manner (Matthews & Bernhard, 1973) but there is precedent for nonsymmetrical association (Steitz et al., 1976).

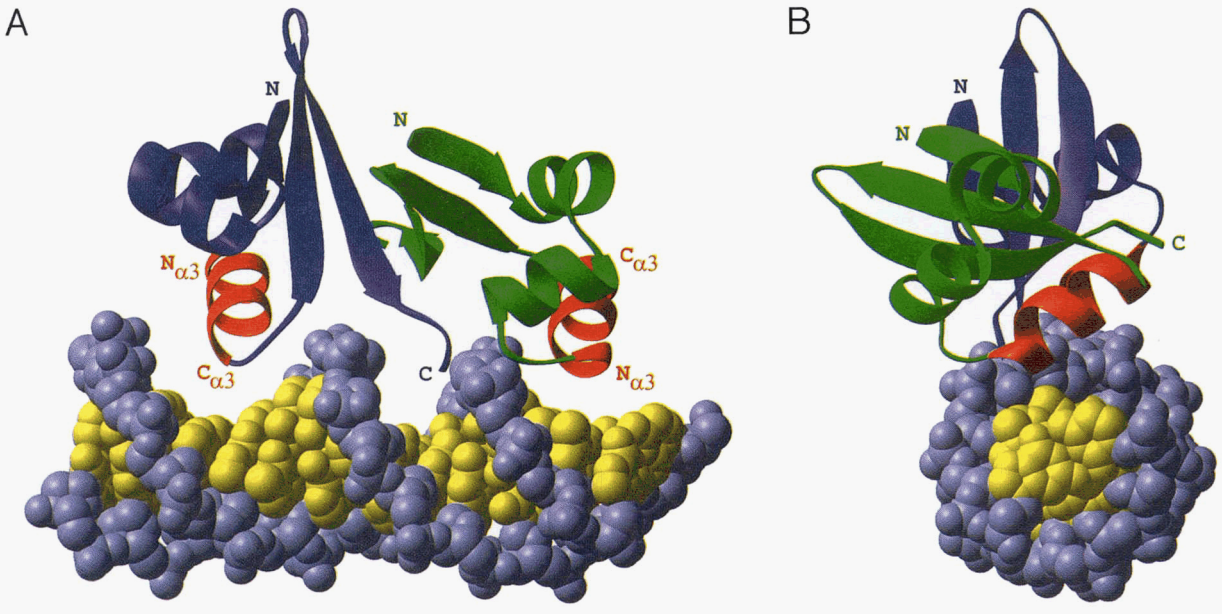

Fig. 7. Drawings showing a tentative model for a complex of a dimer of wild-type Cro with noncognate DNA. The model was constructed by superimposing one subunit of the wild-type Cro dimer (Ohlendorf et al., 1998) on to the engineered monomer of Cro in complex with nonspecific DNA (this work). **A:** The model viewed parallel to the grooves of the DNA. Because of the inherently asymmetric alignment of the dimer relative to the DNA (see text), on the right-hand side it is the amino-terminus of the recognition helix ($N_{\alpha3}$) that is close to the DNA, whereas on the left-hand side it is the carboxy-terminus of the recognition helix ($C_{\alpha3}$) that is close to the DNA. **B:** View of the model as seen parallel to the axis of the DNA. Note that in this view the two recognition helices (in red) are directly behind each other such that the second is largely obscured by the first.

The proposed model encompasses many of the features anticipated for proteins bound nonspecifically to DNA (e.g., see Berg et al., 1982; Pendergrast et al., 1994; Sidorova & Rau, 1996). In particular, the protein is separated from the DNA, relative to the specific complex, allowing water to intervene (Fig. 8). Also, there are fewer direct contacts between the protein and the DNA, and the DNA lacks major protein-induced changes in conformation. In the present case, the protein is much further removed from the DNA than is the case with either the glucocorticoid receptor or EcoRV endonuclease bound to noncognate DNA (Luisi et al., 1991; Winkler et al., 1993; Gewirth & Sigler, 1995). In this respect, the postulated Cro complex is more similar to that proposed for the nonspecific binding of CAP (Weber & Steitz, 1984), except that the protein is arranged asymmetrically with respect to the DNA.

Assuming, for the moment, that Cro moves along the DNA in a manner generally similar to that suggested by the present model, how then might sequence-specific recognition be achieved? At noncognate sites binding energy would be provided predominantly by electrostatic interactions, including those between the phosphate backbone and the channel of one Cro subunit. The initial stage of recognition of a specific site might include favorable hydrogen bonding interactions between Lys32′ and bases within the major groove as well as between Asn61′ and the minor groove. Thermal fluctuations could then allow the recognition helices to rotate into the major groove, bringing Gln27 and Ser28 close to base pairs 2 and 4, the only two positions that are invariant in all natural operator half-sites. Concomitantly, the channel of the second subunit would also enclose the DNA backbone. These induced fit adjustments would result in a displacement of the middle region of the dimer away from the center of the DNA, eliminating the interactions with the minor groove that are presumed to occur in the nonspecific complex. The resultant displacement of solvent and burial of additional surface area would explain the entropy-driven formation of the complex (Takeda et al., 1992).

Finally, the present crystal structure also provides a new perspective on cocrystals of Cro that were obtained with a six-base pair duplex from the region of the operator directly contacted by Cro (5′-ATCACC-3′), as well as with a nine-base pair duplex from the central region of the operator predominantly not contacted by Cro (5′-ACCGCAAGG-3′) (Anderson et al., 1983). Surprisingly, the same unit cell and space group was obtained in both cases, even though the DNA fragments were different in length and in sequence. A closely related unit cell was also obtained with another hexanucleotide (Malinina et al., 1985). As judged by the diffraction patterns, the DNA appeared to be essentially straight but disordered. In light of the present structure, it could be that both complexes were, in fact, nonspecific. This would allow the DNA to bind in different registry relative to the protein, explaining the straightness of the DNA, the ability of the crystals to accommodate different sequences, and also the disorder of the DNA.

Materials and methods

Crystallization and crystal characterization

Cro K56-[DGEVK] protein was purified as previously described (Mossing & Sauer, 1990). The protein was dissolved in 20 mM K_2HPO_4 (pH 7.0), 0.1 mM EDTA, and concentrated to 16 mg/mL by centrifugation using a 3,000 MW cutoff Centricon filter. Oligomers of 5′-TATCACC-3′ and its complement were purified by reverse-phase HPLC using a Hamilton PRP-1 column at 58 °C, then annealed by slow cooling from 85 to 4 °C. This blunt-end

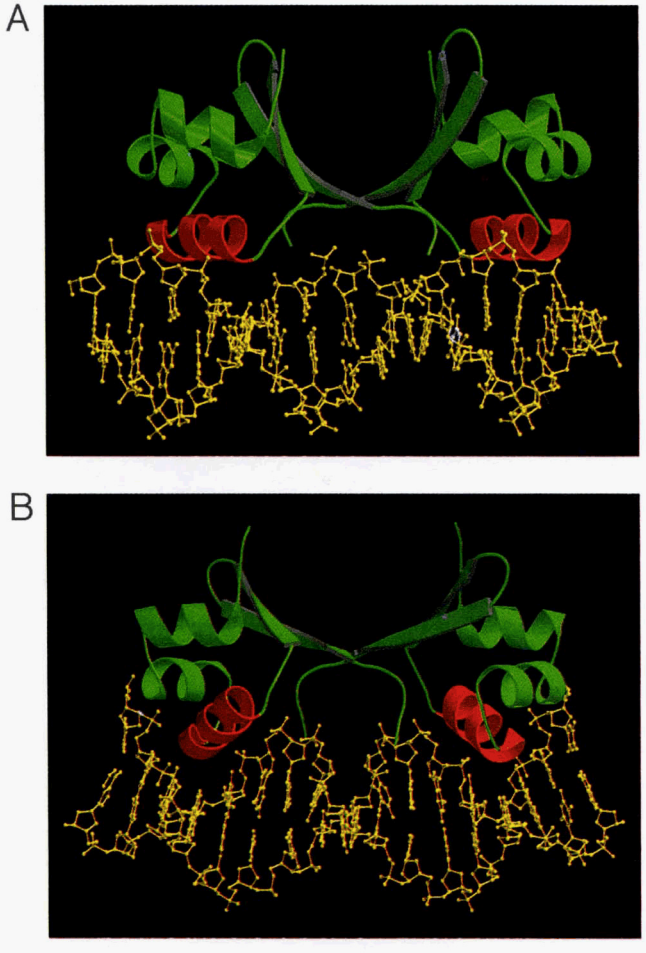

Fig. 8. Comparison of the binding of Cro to operator DNA with the tentative model for the binding to noncognate DNA. **A:** The model for nonspecific binding viewed perpendicular to the DNA (c.f. Fig. 7A). The right-hand monomer is aligned on the DNA as in the complex of the engineered Cro monomer. The remainder of the dimer was built assuming the conformation of wild-type Cro (Anderson et al., 1981; Ohlendorf et al., 1998). The recognition helices of both monomers, shown in red, are close to the DNA, but the DNA contacts are made by opposite ends of the helices, and are not equivalent. **B:** Binding of wild-type Cro to operator DNA (from Albright & Matthews, 1998).

seven-base pair DNA fragment corresponds to the part of the consensus half-operator that includes all of the base-specific interactions observed in the wild-type Cro-operator complex (Albright & Matthews, 1998b). The best cocrystals were obtained by mixing a 50% molar excess of the seven-base pair DNA duplex (13 mg/mL) with a protein monomer and combining with an equal volume of precipitant buffer [80 to 200 mM ammonium acetate, 29 to 33% PEG 3350, and 100 mM acetate buffer (pH 4.6)]. Equilibrating against the precipitant solution using the hanging drop method, plate-like crystals measuring 0.5 × 0.5 × 0.1 mm grew within 10 days at room temperature. These cocrystals belong to space group C2 with $a = 45.7$ Å, $b = 60.7$ Å, $c = 45.7$ Å, and $\beta = 112.7°$. As described below, a single protein molecule plus an averaged DNA half-fragment occupy the asymmetric unit, resulting in a solvent content parameter, V_M, of 2.9 Å^3/Da (Matthews, 1968). The crystals diffract to a limit of about 2.8 Å under the conditions described below.

Structure determination and refinement

X-ray data were collected on a Xuong-Hamlin area detector (Hamlin, 1985; Howard et al., 1985; Zhang & Matthews, 1993) using graphite-monochromated CuKα radiation from a Rigaku rotating anode generator. A complete native data set to 3.0 Å resolution was collected from a single crystal at room temperature (Table 2). No intensity cutoff restrictions of any kind were imposed.

Molecular replacement searches were carried out using the program package ROTFUN (Zhang & Matthews, 1994). The best search model proved to be the apo Cro monomer structure (Albright et al., 1996). With this model the rotation function peak was 6.4σ above average (1.5σ above the next-highest peak), and the translation search solution was 5.3σ above average (2.7σ above the next-highest peak). The protein accounts for about 77% of the molecular mass of the asymmetric unit. A $2F_o - F_c$ map phased on the molecular replacement solution revealed contiguous electron density for virtually the entire protein main chain. Density corresponding to parts of the DNA fragment was also apparent, even though the DNA was not included in the phase calculation. Unexpectedly, the orientation of the DNA density relative to the protein was substantially different from that in the complex with wild-type Cro (Albright & Matthews, 1998). Consistent with this observation, a second search model in which a DNA fragment was added to Cro monomer in the same relative position as observed in the wild-type Cro-operator complex (Albright & Matthews, 1998b) resulted in substantially weaker peaks in the molecular replacement calculations, but placed the protein in the same general position. Maps phased on this second model revealed good density for the protein but very poor density for much of the DNA, even though the DNA had been included in the phasing. This strongly suggested that the placement of the DNA in this model was not correct.

The "protein-only" molecular replacement solution served as the starting model for refinement, with an initial *R*-factor of 46% (20 to 4.0 Å data). All refinement was carried out using the TNT package (Tronrud et al., 1987; Tronrud, 1992), interspersed with rounds of model building using the graphics program FRODO (Jones, 1982). Rigid-body refinement of the protein reduced the

Table 2. *Crystallographic data collection and refinement statistics*[a]

Resolution range	20–3.0 Å
Total number of measured reflections	13,041
Number of unique reflections observed	2,452
Completeness	99.7%
Agreement between symmetry-related intensities	5.3%
Number of atoms refined	666
R (all reflections)	22.4%
Δ_{bonds}	0.015 Å
Δ_{angles}	2.7°
ΔB	2.8 Å^2

[a]The final refined model includes the engineered Cro monomer (residues 2–61, plus the five-residue insertion), four DNA base pairs, one at half occupancy, and four water molecules. *R* is the crystallographic *R*-factor. Δ_{bonds} and Δ_{angles} are the root-mean-square discrepancies of the refined bond lengths and angles from "ideal" values. ΔB is the discrepancy of the thermal factors from values expected from well-refined high-resolution structures (Tronrud, 1996).

R-factor to 40%. *B*-factors were fixed at 35 Å^2. Three nucleotides of the DNA strand nearest to the protein were then built into the electron density, with positional refinement reducing the *R*-factor to 34%. The three partner nucleotides of the complimentary strand were next built into the improved electron density, reducing the *R*-factor to 28%. These three base pairs terminated close to a crystallographic twofold axis such that three additional base pairs followed by symmetry. This, however, raised an ambiguity regarding the position of the DNA fragments in the crystal lattice. The simplest alternative was that two DNA fragments might be related by symmetry with their blunt ends stacked at the twofold axis. In terms of Figure 1, the twofold symmetry axis would be placed at A_0. Another possibility, however, was that the DNA fragment straddled the crystallographic twofold axis such that the dyad was positioned at, for example, $A_{3,4}$ or $A_{4,5}$ of Figure 1. Because the DNA has an odd number of base pairs, and also is nonpalindromic, it would have to be statistically disordered. Thus, the twofold axis of symmetry would not apply to the individual base pairs, but would apply to a statistically averaged DNA molecule.

The ambiguity was resolved by inspection of the electron density at the site where the DNA backbone would pass closest to the twofold axis. If the DNA was straddling the twofold axis, an internal phosphate group would occupy this site. Alternatively, if the DNA fragments were stacked end to end, this site would lack a phosphate group (Fig. 1). Inspection of an $F_o - F_c$ map in which the phosphate in question had been omitted from phase calculations revealed a clear positive peak 13 standard deviations above background (not shown), confirming that the DNA fragment was statistically disordered and straddled the crystallographic twofold axis. The clarity of this result also provided confirmation for the molecular replacement structure determination.

The crystallographic refinement suggests that the DNA fragment is randomly distributed between two orientations, related to each other by a 180° rotation about the symmetry axis. Clear electron density could be seen for six successive base pairs, three on each side of the twofold axis. Weaker flanking density suggested an additional base pair of reduced occupancy at each end, as would be expected for a statistically disordered seven-base pair fragment. Beyond these flanking base pairs, no additional electron density could be seen. This suggests that the crystallographic twofold axis passes through the DNA predominantly at sites $A_{3,4}$ and/or $A_{4,5}$ in Figure 1. The special position of the DNA fragment in the crystal lattice combined with its lack of sequence symmetry result in diffraction corresponding to a Cro monomer and what is effectively a four-base pair "averaged" DNA half-fragment, consisting of a composite of two or more base pair identities at each position and with the outermost base pair at half-occupancy. The other half of the fragment is generated by the twofold axis. Using such a model during refinement, however, would have increased the number of parameters and introduced a number of complications. In any event, it was unclear that refining composite bases would offer any analytical advantage over a single sequence. Therefore, in subsequent refinement the DNA was represented by a half-fragment model of the sequence 5′-TATC-3′, with the left-most base pair corresponding to the outermost base pair of the expanded fragment, at 50% occupancy. This fragment, which was not averaged in any way, accounts for the total of seven base pairs in the crystal, but it is understood that the model is only an approximation for a statistically averaged DNA structure. As the resolution was gradually increased, model building was alternated with refinement of atomic positions and correlated *B*-factors (Tronrud, 1996). Care was taken to maintain geometric and *B*-factor restraints between the DNA half-fragment and its symmetry mate. Both $F_o - F_c$ and $2F_o - F_c$ "omit" maps were used to check the entire structure. Final statistics are summarized in Table 2. A representative portion of the final "$2F_o - F_c$" electron density map is shown in Figure 2.

Structure comparisons and superpositions, as well as solvent accessibility and other routine calculations, were carried out with the program package EDPDB (Zhang & Matthews, 1995).

The coordinates of the refined complex have been deposited in the Brookhaven Protein Data Bank for immediate release (Access code 3ORC).

Acknowledgments

The help of Larry Weaver in preparing the figures is greatly appreciated. This work was supported in part by NIH Grants, GM46514 to M.C.M. and GM20066 to B.W.M.

References

Albright RA, Matthews BW. 1998a. How Cro and λ-repressor distinguish between operators: The structural basis underlying a genetic switch. *Proc Natl Acad Sci USA 95*:3431–3436.

Albright RA, Matthews BW. 1998b. Crystal structure of λ-Cro bound to a consensus operator at 3.0 Å resolution. *J Mol Biol*. Forthcoming.

Albright RA, Mossing MC, Matthews BW. 1996. High-resolution structure of an engineered Cro monomer shows changes in conformation relative to the native dimer. *Biochemistry 35*:735–742.

Anderson WF, Cygler M, Vandonselaar M, Ohlendorf DH, Matthews BW, Kim J, Takeda Y. 1983. Crystallographic data for complexes of the Cro repressor with DNA. *J Mol Biol 168*:903–906.

Anderson WF, Ohlendorf DH, Takeda Y, Matthews BW. 1981. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature 290*:754–758.

Baleja JD, Anderson WF, Sykes BD. 1991. Different interactions of Cro repressor dimer with the left and right halves of O_R3 operator DNA. *J Biol Chem 266*:22115–22124.

Berg OG, Winter RB, von Hippel PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry 20*:6929–6948.

Berg OG, Winter RB, von Hippel PH. 1982. How do genome-regulatory proteins locate their DNA target sites? *Trends Biochem Sci 7*:52–55.

Brennan RG, Roderick SL, Takeda Y, Matthews BW. 1990. Protein–DNA conformational changes in the crystal structure of a lambda Cro-operator complex. *Proc Natl Acad Sci USA 87*:8165–8169.

Erie DA, Yang G, Schultz HC, Bustamante C. 1994. DNA bending by Cro protein in specific and nonspecific complexes: Implications for protein site recognition and specificity. *Science 266*:1562–1566.

Gewirth DT, Sigler PB. 1995. The basis for half-site specificity explored through a non-cognate steroid receptor–DNA complex. *Struct Biol 2*:386–394.

Hamlin R. 1985. Multiwire area X-ray diffractometers. *Methods Enzymol 114*:416–452.

Howard AJ, Nielsen C, Xuong NH. 1985. Software for a diffractometer with multiwire area detector. *Methods Enzymol 114*:452–471.

Hubbard AJ, Bracco LP, Eisenbeis SJ, Gayle RB, Beaton G, Coruthers MH. 1990. Role of the Cro repressor carboxy-terminal domain and flexible dimer linkage in operator and nonspecific DNA binding. *Biochemistry 29*:9241–9249.

Jana R, Hazbun TR, Mollah AKMM, Mossing MC. 1997. A folded monomeric intermediate in the formation of lambda Cro dimer–DNA complexes. *J Mol Biol 273*:402–416.

Jones TA. 1982. FRODO: A graphics fitting program for macromolecules. In: Sayre D, ed. *Crystallographic computing*. Oxford: Oxford University Press. pp 303–317.

Jordan SR, Pabo CO. 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor–operator interactions. *Science 242*:893–899.

Kim JG, Takeda Y, Matthews BW, Anderson WF. 1987. Kinetic studies of Cro repressor–operator DNA interaction. *J Mol Biol 196*:149–158.

Kirpichnikov MP, Yartzev AP, Minchenkova LE, Chernov BK, Ivanov VI. 1985. The absence of non-local conformational changes in OR3 operator DNA on complexing with the *Cro* repressor. *J Biomol Struct Dynam 3*:529–536.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK:

A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Lavery R, Sklenar H. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J Biomol Struct Dynam 6*:63–91.

Lee SJ, Shirakawa M, Akutsu H, Kyogoku Y, Shiraishi M, Kitano K, Shin M, Ohtsuka E, Ikehara M. 1987. Base sequence-specific interactions of operator DNA fragments with the λ-cro repressor coupled with changes in their conformations. *EMBO J 6*:1129–1135.

Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, Sigler PB. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature 352*:497–505.

Malinina LV, Makhaldiani VV, Vainshtein BK, Kirpichnikov MP, Skryabin KG, Baev AA, Ivanova EM, Zarytova VF. 1985. Crystals of a nonspecific complex of cro-repressor with DNA. *Doklady Akademii Nauk SSSR 284*: 229–232.

Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol 33*:491–497.

Matthews BW, Bernhard SA. 1973. Structure and symmetry of oligomeric enzymes. *Annu Rev Biophys Bioeng 2*:257–317.

Mollah AKMM, Aleman MA, Albright RA, Mossing MC. 1996. Core packing defects in an engineered Cro monomer corrected by combinatorial mutagenesis. *Biochemistry 35*:743–748.

Mossing MC. 1998. Solution structure and dynamics of a designed monomeric variant of the lambda Cro repressor. *Protein Sci 7*:983–993.

Mossing MC, Sauer RT. 1990. Stable, monomeric variants of λ-Cro obtained by insertion of a designed beta-hairpin sequence. *Science 250*:1712–1715.

Ohlendorf DH, Tronrud DE, Matthews BW. 1998. Refined structure of Cro repressor protein from bacteriophage λ. *J Mol Biol*. Forthcoming.

Pendergrast PS, Ebright YW, Ebright RH. 1994. High-specificity DNA cleavage agent: Design and application to kilobase and megabase DNA substrates. *Science 265*:959–962.

Ptashne M. 1986. *A genetic switch. Gene control and phage λ*. Oxford, UK: Blackwell.

Sidorova NY, Rau DC. 1996. Differences in water release for the binding of *Eco*RI to specific and nonspecific DNA sequences. *Proc Natl Acad Sci USA 93*:12272–12277.

Steitz TA, Fletterick RJ, Anderson WF, Anderson CM. 1976. High resolution X-ray structure of yeast hexokinase, an allosteric protein exhibiting a non-symmetric arrangement of subunits. *J Mol Biol 104*:197–222.

Takeda Y, Kim J, Caday CG, Steers E Jr, Ohlendorf DH, Anderson WF, Matthews BW. 1986. Different interactions used by Cro repressor in specific and nonspecific DNA binding. *J Biol Chem 261*:8608–8616.

Takeda Y, Ross PD, Mudd CP. 1992. Thermodynamics of Cro protein–DNA interactions. *Proc Natl Acad Sci USA 89*:8180–8184.

Torigoe C, Kidokoro S, Takimoto M, Kyogoku Y, Wada A. 1991. Spectroscopic studies on λ Cro protein–DNA interactions. *J Mol Biol 219*:733–746.

Tronrud DE. 1992. Conjugate-direction minimization: An improved method for the refinement of macromolecules. *Acta Crystallogr A48*:912–916.

Tronrud DE. 1996. Knowledge-based B-factor restraints for the refinement of proteins. *J Appl Crystallogr 29*:100–104.

Tronrud DE, Ten Eyck LF, Matthews BW. 1987. An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr A43*:489–503.

Weber IT, Steitz TA. 1984. A model for the non-specific binding of catabolite gene activator protein to DNA. *Nucleic Acids Res 12*:8475–8487.

Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, Martin PD, Petratos K, Wilson KS. 1993. The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J 12*:1781–1795.

Zhang X-J, Matthews BW. 1993. STRAT: A program to optimize X-ray data collection on an area detector system. *J Appl Crystallogr 26*:457–462.

Zhang X-J, Matthews BW. 1994. Enhancement of the method of molecular replacement by incorporation of known structural information. *Acta Crystallogr D50*:675–686.

Zhang X-J, Matthews BW. 1995. EDPDB: A multi-functional tool for protein structure analysis. *J Appl Crystallogr 28*:624–630.