High-Resolution Structure of an Engineered Cro Monomer Shows Changes in Conformation Relative to the Native Dimer†

Ronald A. Albright,‡,§ Michael C. Mossing,*,|| and Brian W. Matthews*,‡,||

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

Received August 18, 1995; Revised Manuscript Received November 6, 1995⊗

ABSTRACT: A rationally designed, genetically engineered, monomeric form of the Cro protein from bacteriophage λ has been crystallized and its structure determined by isomorphous replacement and refined to a resolution of 1.54 Å. The structure confirms the rationale of the design but, at the same time, reveals 1−2 Å shifts throughout the monomer structure relative to the previously determined structure of the dimeric wild-type protein. These changes include a 1.6 Å main-chain shift in part of the β-sheet region of the molecule relative to the α-helical region and a 1.1 Å shift of a buried phenylalanine within the core as well as a correlated 2.2 Å shift in a solvent-exposed β-hairpin. The conformational adjustments appear to reflect an inherent flexibility of the protein that is associated with its DNA-binding function.

One approach to structure−function studies of large, complex, proteins is to adopt a strategy of "divide and conquer". Since larger proteins are usually made up of subunits or subdomains (Kyte, 1995), they can often be reduced to smaller units that are themselves stable. Analysis of such domains has, in some cases, been very informative, especially when the parent structure is too large for conventional analysis [e.g., Zuiderweg et al. (1983)]. This approach does, however, assume that the substructures faithfully maintain the wild-type conformation. The present analysis of an engineered monomeric variant of Cro protein from phage λ addresses this question.

The wild-type DNA-binding form of λ Cro repressor is a homodimer (Anderson et al., 1981) (Figure 1). As far as is known, the 66-residue wild-type Cro monomer is not stable. Folding occurs only upon dimerization (Takeda et al., 1977; Gitelson et al., 1991). The dimer interface consists of a short antiparallel β-ribbon, with Phe 58 from one monomer buried within the core of the other monomer and *vice versa*. As a result, the folding domain core of the monomer also serves a dual role as part of the dimer interface. The surface area of the interface is 1280 Å^2, or 640 Å^2 per half-dimer.

The logical engineering of stable Cro monomer mutants was demonstrated by Mossing and Sauer (1990). The strategy was to maintain all of the intermolecular interactions in an intramolecular fashion by engineering a β-turn into the antiparallel β-strands of the wild-type dimer interface, just beyond the molecular dyad. Formation of this β-hairpin allowed the monomer to donate Phe 58 to its own core, as well as maintain the hydrogen bonds along the antiparallel β-ribbon. In order to achieve this goal, the polypeptide chain had to be lengthened, and five amino acids were therefore inserted into the wild-type Cro sequence between Lys 56 and Pro 57, resulting in 71-residue molecules. One of the most stable of the resulting monomer mutants, formerly referred to as Cro.mDG (Mossing & Sauer, 1990) was Cro K56-[DGEVK].[1] The name [cf. Heinz et al. (1993)] indicates that the five amino acids Asp-Gly-Glu-Val-Lys have been inserted following Lys 56. This protein is monomeric and is substantially more stable than the wild-type dimer (Mossing & Sauer, 1990). In this paper, we present the high-resolution crystal structure of the Cro K56-[DGEVK] monomer mutant (Figure 2).

MATERIALS AND METHODS

Nomenclature. Wild-type Cro dimer residues are 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 5 and 6), α1 (7−14), α2 (16−23), α3 (27−35), β2 (40−45), and β3 (49−56). The β3β3′ antiparallel β-ribbon (residues 54−56 and 54′−56′) forms part of the dimer interface (Figure 1).

The Cro K56-[DGEVK] monomer mutant retains the numbering of the wild-type sequence, except for the five inserted residues which are identified as Asp 56a, Gly 56b, Glu 56c, Val 56d, and Lys 56e. The resulting overall sequence of this 71-residue protein is 1−56, 56a−56e, and 57−66. Residues 56a and 56b were expected to form a tight β-turn, while residues 56c−56e were designed to be analogous to 54′−56′ of the wild-type dimer. Accordingly, K56-[DGEVK] has an additional β-strand, β4, colored red in Figure 2, that replaces the part of β3′ involved in the wild-type dimer interface.

Crystallization and Crystal Characterization. Purified Cro K56-[DGEVK] protein was obtained as previously described (Mossing & Sauer, 1990), dissolved in 20 mM K_2HPO_4 (pH

† This work was supported in part by National Institutes of Health grants GM46513 (to M.C.M.) and GM20066 (to B.W.M.).

* To whom correspondence should be addressed (FAX: 503-346-5870).

‡ University of Oregon.

§ Present address: Department of Molecular Biophysics and Biochemistry, Yale University, Boyer Center for Molecular Medicine, Room 154, 295 Congress Ave., New Haven, CT 06510.

|| University of Notre Dame.

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; T_m, temperature at which half the protein is unfolded; HTH, the helix−turn−helix region of the molecule, i.e., residues 16−35; K56-[DGEVK], the engineered monomeric Cro monomer with Asp-Gly-Glu-Val-Lys inserted following Lys 56; PIP, di-μ-iodobis(ethylenediamine)diplatinum(II) nitrate; MIR, multiple isomorphous replacement.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1996.

0006-2960/96/0435-735$12.00/0

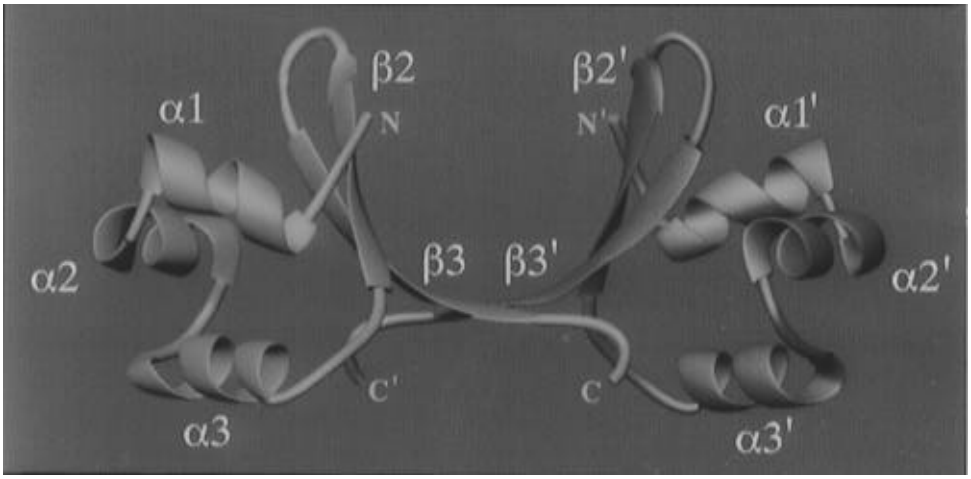

FIGURE 1: Ribbon representation of the structure of the wild-type Cro dimer (Anderson et al., 1981). The major secondary structure elements are indicated. For clarity, however, the short β1 strand (residues 5−6, immediately preceding helix α1) is shown as an extended chain and is not explicitly labeled. In this view the 2-fold axis of symmetry is vertical. The antiparallel pair of β-sheet strands, β3 and β3′, links one monomer with the other. Phe 58, near the C-terminus of each strand, is buried within the hydrophobic core of the partner subunit.

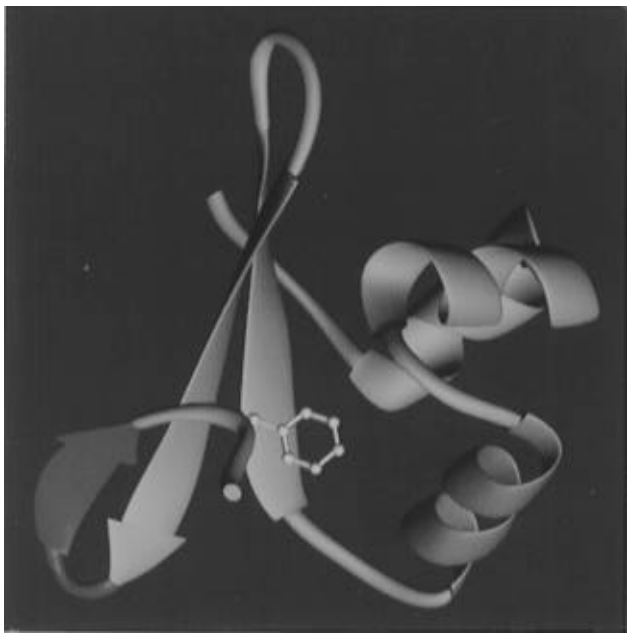

FIGURE 2: Structure of the engineered Cro monomer. The five-residue insertion following Lys 56 is colored red. The first two residues of this insertion form a β-turn, while the remaining three residues comprise the β4 strand, mimicking the β3′ strand near the center of the wild-type dimer interface (Figure 1). Phe 58, close to the carboxy terminus, is shown in yellow. The recognition helix (α3) is colored mauve. The orientation differs from that of Figure 1, allowing an unencumbered view of Phe 58 within the core.

7.0) and 0.1 mM EDTA, and then concentrated to 16 mg/mL using a 3000 MW cutoff Centricon filter and centrifugation. The best crystals were obtained by mixing this protein stock solution with an equal volume of precipitant buffer [4.2−4.8 M sodium formate (pH 7.8)] and equilibrating against the precipitant solution using the hanging drop method. Large crystals (1.0 mm × 0.6 mm × 0.5 mm) typically grow within 1 week at room temperature. These crystals belong to space group $P2_12_12_1$ with a = 34.8 Å, b = 39.2 Å, c = 48.3 Å. There is a single protein molecule per asymmetric unit with a V_M value of 2.1 Å^3/Da (Matthews, 1968). The crystals diffract isotropically to a limit of approximately 1.4 Å resolution under the conditions described below and are exceptionally resistant to radiation-induced damage.

Table 1: X-ray Data Collection Statistics[a]

	native	heavy-atom derivatives	
		K_2PtCl_4	PIP
cell dimensions			
a (Å)	34.8	34.8	35.3
b (Å)	39.2	39.2	39.2
c (Å)	48.3	48.1	48.3
resolution (Å)	1.54	1.60	1.70
reflections measured	45435	46137	44479
unique reflections	9834	8727	7616
completeness (%)	95.8	91.8	97.5
R_{sym} (%)	2.9	4.8	4.4
R_{iso} (%)		17.4	14.2

[a] Both the native and derivative crystals were in 200 mM sodium formate, 100 mM Tris (pH 7.8), and 30% PEG 3350 (see text). All data collected to the resolution at which the average of $I/\sigma(I)$ was equal to 2.0 were used. R_{sym} gives the average discrepancy between symmetry-related intensities and R_{iso} the average change in the structure amplitudes due to introduction of the heavy atom.

Data Collection and Reduction. All X-ray data were collected on a Xuong−Hamlin area detector (Hamlin, 1985; Howard et al., 1985; Zhang & Matthews, 1993) using graphite-monochromated Cu Kα radiation from a Rigaku rotating anode generator. Each data set was collected from a single crystal at room temperature. The durability of the crystals allowed 4−5-fold redundancy of the observations with little intervening degradation. Due to the strong intensity of some reflections, data were collected in two separate sweeps. All data sets were of high quality (Table 1).

Structure Determination and Refinement. Attempts to solve the structure by molecular replacement, using models based on the structure of the dimer (Anderson et al., 1981), proved unsuccessful. Searches using models based on the low-resolution electron density map of the Cro−DNA complex (Brennan et al., 1990) were also fruitless. In retrospect these searches failed for a combination of reasons. First, uncertainty regarding the locations of residues in the vicinity of the insertion, plus the loop region near residue 47, dictated that they were, in most cases, deleted. This reduced the search model to 65−75% or less of the actual structure. Second, the remaining "core" structure of the search model is now known to differ from the target by ∼1

Table 2: Statistics from Heavy-Atom Refinement

	$PtCl_4$	PIP
no. of sites	2	2
no. of platinum atoms	2	4
primary protein ligands	Met 12, Lys 8	Met 12
secondary protein ligands[a]		Arg 13, Gln 27
resolution range (Å)	20−1.7	20−1.7
no. of centric reflections	1185	1160
centric *R*-factor[b]	0.58	0.67
no. of acentric reflections	6259	6258
no. of Bijvoet differences used	5965	
phasing power[c]	1.60	1.03

[a] The two PIP molecules each contain two platinum atoms spaced approximately 2.5 Å apart, with one platinum interacting with Met 12. A platinum of one PIP contacts Met 12 in the same position as $PtCl_4$, with the remaining platinum interacting with Arg 13. The other PIP resides at a crystal contact, interacting with Met 12 from a different position, while the other platinum contacts Gln 27* of a symmetry-related molecule. [b] The centric *R*-factor is $R = \Sigma||F_{PH} - F_P| - F_H|/\Sigma|F_{PH} - F_P|$, where F_{PH}, F_P, and F_H are respectively the amplitudes of the reflections for the heavy-atom derivative, the native protein, and the heavy atom and the summations include only the centrosymmetric reflections. [c] The phasing power is $\langle F_H\rangle/\langle|F_{PH} - F_P| - F_P\rangle$, where F_{PH}, F_P, and F_H are as defined above and the brackets indicate the root-mean-square value.

Å (see below). Finally, with small proteins, as is the case here, the intramolecular vectors tend to overlap the intermolecular ones, confusing the search process.

Multiple isomorphous replacement (MIR) was therefore used to obtain initial phases. Careful transfer of the crystals into a different solvent [200 mM sodium formate, 100 mM Tris (pH 7.8), 30% PEG 3350] prior to soaking with heavy-atom compounds proved critical to the formation of useful derivatives (Sigler & Blow, 1965). Transfer often resulted in fine hairline fractures within the crystals but had little discernible effect on data quality. Native data were collected from crystals which had also been transferred to this new solvent to eliminate possible nonisomorphism arising from differing solvent environments.

Heavy-atom derivatives were obtained by soaking native crystals in 2 mM K_2PtCl_4 for 4 days or in 3 mM di-μ-iodobis-(ethylenediamine)diplatinum(II) nitrate (PIP) for 10 days (Table 1). Initial heavy-atom positions were determined manually from difference Patterson maps and checked using the program VERIFY written by S. Roderick (personal communication). Subsequent refinement of heavy-atom positions and occupancies (Table 2) was achieved using the program HEAVY (Terwilliger & Eisenberg, 1983). A series of MIR maps including the anomalous dispersion signal of the K_2PtCl_4 derivative were calculated and appeared to improve as successively higher resolution data were included. The map used for model building included all data between 20 and 1.7 Å resolution and had a figure of merit of 0.57.

The electron density was unambiguous for most of the protein main chain, as well as for many side chains (Figure 3). It was also immediately apparent that this density did not strictly correspond to the structure of the wild-type dimer (Anderson et al., 1981). The quality of the map allowed the *de novo* construction of the K56-[DGEVK] model directly into the MIR density, thereby avoiding potential bias. Residues 4−55 and 57−59, along with the inserted residues 56d and 56e, were built first using the graphics program FRODO (Jones, 1982), with some solvent-exposed side chains truncated. The *R*-factor of this initial model was 39% (using 20−2.0 Å data). Phases from this model were then combined with the MIR phases to yield an improved map into which the engineered β-turn residues (56, 56a, 56b, 56c) were built. All refinement was carried out using the TNT least-squares package (Tronrud et al., 1987; Tronrud, 1992). Additional positional and *B*-factor refinement, combined with model building and further expansion of the model, reduced the *R*-factor to 23%. Ordered water molecules were added as the resolution was gradually extended to 1.54 Å. To be retained in the model, water molecules had to have at least one hydrogen-bonding partner, a *B*-factor of less than 80 Å^2, corresponding positive density in "$F_o - F_c$" and "$2F_o - F_c$" electron density maps, and roughly spherical density in "omit" maps. Isotropic *B*-factors were allowed to adjust freely. The final model contains 559 non-hydrogen atoms, including 64 protein residues (3−56, 56a−56e, and 57−61) and 57 water molecules, two of which were modeled as having slightly different positions, each with 50% occupancy.

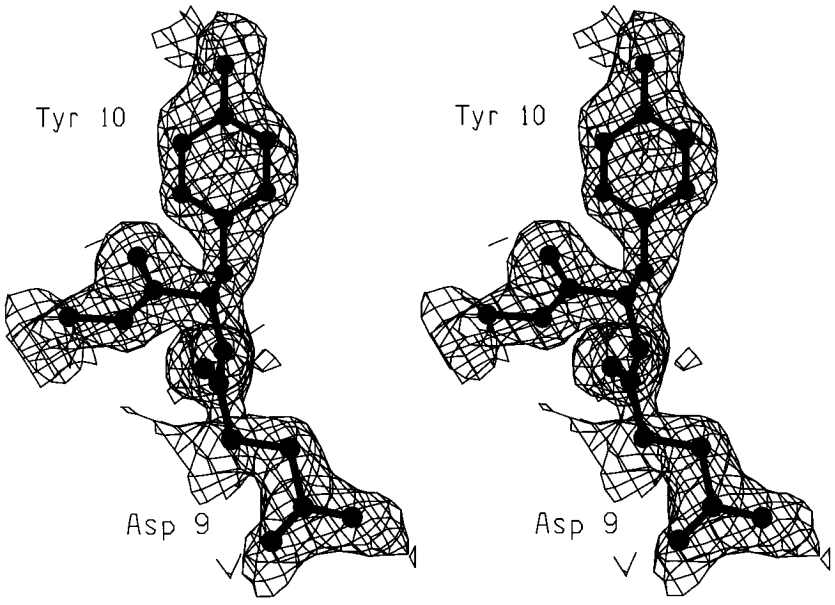

FIGURE 3: Multiple isomorphous replacement electron density map in the vicinity of Tyr 10. Resolution = 1.7 Å. The map is contoured at 1σ, where σ is the root-mean-square electron density throughout the unit cell.

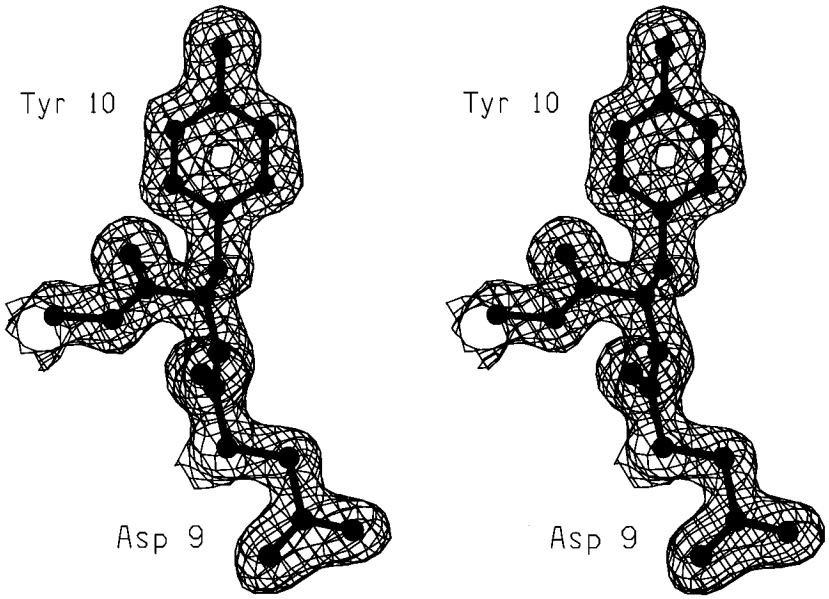

FIGURE 4: Representative region of the final electron map. Coefficients are $(2F_o - F_c)$, where F_o is the observed structure amplitude and F_c is that calculated from the final refined model. Resolution = 1.54 Å. The map is contoured at 1σ (*cf.* Figure 3).

At the amino terminus, electron density is not apparent for residues 1 and 2. At the carboxy terminus, the density is weak and ambiguous near residues 62 and 63 and essentially absent for residues 64−66. We assume all these amino acids to be disordered. The same carboxy-terminal residues were disordered in the dimer (Anderson et al., 1981). The Gln 27 side chain is modeled in two different orientations, each with 50% occupancy, as are two water molecules. The final *R*-factor is 17.8% for all data between 20 and 1.54 Å resolution (9834 reflections). The coordinates have been deposited in the Brookhaven Protein Data Bank (Access Code 1ORC).

Calculations of surface area and other parameters were carried out with the program EDPDB (Zhang & Matthews, 1995).

RESULTS

Quality of the Structure. Following refinement of the Cro K56-[DGEVK] structure the final electron density map is highly detailed, featuring well-defined carbonyl "bumps" along the backbone and holes through the aromatic rings (Figure 4). The average discrepancies of the bond lengths and bond angles from their expected values are 0.015 Å and 2.3°, respectively. A Luzzati (1952) plot (not shown) suggests a coordinate error of 0.18 Å. Fifty-four of the 55 non-glycine and non-proline residues (98.2%) are in the "most-favored" regions of the Ramachandran plot (Figure 5), compared with 87.9% for other structures of similar resolution, as determined by the program package PROCHECK (Laskowski et al., 1993). The sole outlier is the first residue of the engineered turn, Asp 56a ($\phi = 54°$, $\psi = 54°$), which lies in an additional "allowed" region with values common for a β-turn. The $\chi 1$ angles were also determined to be slightly better than average for the given resolution.

Main-chain thermal factors are shown in Figure 6, with an average of 21.6 Å^2 for the globular part of the protein (residues 4−59). The rapid increase in main-chain *B*-factors near both ends of the model reflects the intrinsic disorder of the termini. The engineered β-turn region (residues 56, 56a, 56b, and 56c) contains the highest main-chain thermal factors within the globular portion of the protein. These decrease substantially within the remainder of the insertion (residues 56d and 56e), and become average around Phe 58. As expected, slight increases in thermal factors are also observed at the other β-turn regions (near residues 15, 25, and 47), as well as the short loop near residue 37. The side-chain *B*-factors vary more widely and are much more dependent on the identity and solvent accessibility of each residue (data not shown).

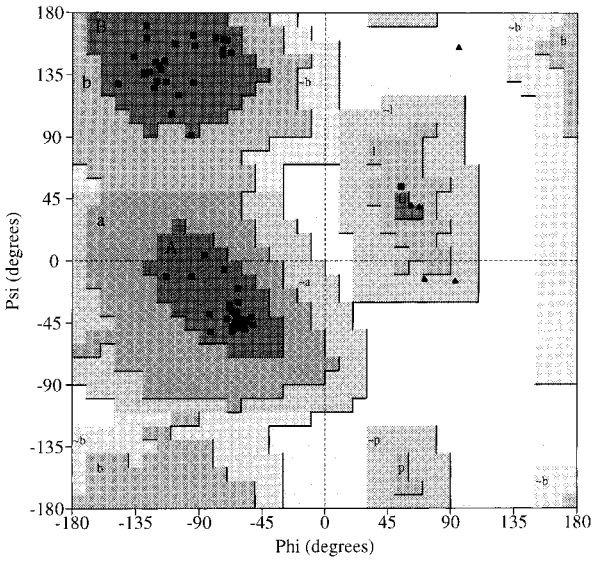

FIGURE 5: Ramachandran diagram, calculated using PROCHECK (Laskowski et al., 1993), for the final refined structure of the Cro monomer. Glycine residues are indicated by triangles and non-glycines by squares.

Overall Structure. The major structural features anticipated in the original design of the monomer (Mossing & Sauer, 1990) are, indeed, present in the structure of Cro K56-[DGEVK] (Figure 2). All of the intermolecular interactions observed within the wild-type dimer are satisfied in an intramolecular fashion within the monomer. The first two residues of the five-residue insertion (Asp 56a and Gly 56b) form a tight β-turn, while the remaining three residues (Glu

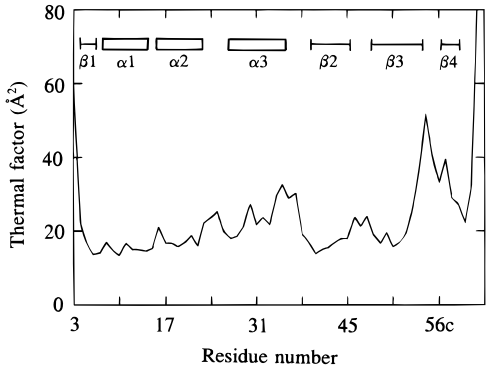

FIGURE 6: Main-chain thermal factors for the engineered Cro monomer. The value plotted for each residue is the average of the main-chain *B*-factors. The locations of the α-helices (α1−α3) and the β-sheets (β1−β4) are indicated.

56c, Val 56d, and Lys 56e) maintain the main-chain hydrogen-bonding interactions analogous to those observed within the antiparallel β-ribbon of the wild-type dimer interface (Glu 54′, Val 55′, and Lys 56′) (Figure 1). The creation of the β-hairpin allows the aromatic ring of Phe 58 to make hydrophobic interactions within the monomer core in a manner similar to Phe 58′ of the wild-type dimer. These interactions will now be reviewed in greater detail.

The backbone (ϕ,ψ) angles of turn residues Asp 56a and Gly 56b are (+54, +54) and (+71, −13), respectively. The main-chain carbonyl of the residue immediately preceding the turn (Lys 56) also accepts a hydrogen bond from the main-chain amide of the residue immediately following the turn (Glu 56c). This pattern is consistent with a type I′ β-turn, for which Asp-Gly is a "consensus" sequence (Sibanda et al., 1989).

All of the interactions along the short antiparallel β-ribbon interface of the wild-type dimer are maintained in the monomer mutant, right up to the engineered β-turn. Main-chain−main-chain hydrogen bonds occur between N54 and O56e (O56′), O54 and N56e (N56′), N56 and O56c (O54′), and O56 and N56c (N54′), respectively (with the corresponding residue numbers of the wild-type dimer shown in parentheses). ("N54" indicates the main-chain amide nitrogen of residue 54 and "O56e" the main-chain carbonyl oxygen of residue 56e, etc.) The Pro 57-Phe 58-Pro 59 segment serves to disrupt the β-ribbon. Pro 57 is in the *trans* conformation and Pro 59 is *cis*. As a result, Phe 58 is at the apex of a sharp bend from which it extends into the hydrophobic core (Figures 2 and 7). The benzyl ring is enclosed by Leu 7, Leu 23, Val 25, Ala 29, Ile 30, Ala 33, Ile 40, Leu 42, Ala 52, and Pro 59 and is covered by Arg 38 and Glu 54, which form a salt bridge. The tip of the aromatic ring points toward the center of the recognition helix (α3) near the main chain between Ala 29 and Ile 30.

Virtually all of the hydrogen-bonding interactions thought to occur throughout the entire wild-type dimer are maintained in K56-[DGEVK]. Of special note is the previously uncharacterized native β2β3-turn between residues 45 and 49. The sequence, main-chain hydrogen-bonding, and (ϕ,ψ) angles of this turn have now been determined to be consistent with a "type I [L1−L4] + G1 β-bulge" (Sibanda et al., 1989). This solvent-exposed β-hairpin appears to be made more rigid by an extensive internal network of hydrogen bonds, with side chains optimized for stabilizing interactions. The O^{γ} of Ser 49 accepts a hydrogen bond from the $N^{\delta 2}$ of Asn 45, while simultaneously donating a hydrogen bond to the $O^{\delta 1}$ of Asp 47. Several well-ordered water molecules are also observed to link different parts of this hairpin together.

Comparison of Monomeric and Dimeric Cro. In the previously determined structure of native Cro protein there were four molecules per asymmetric unit labeled O, A, B, and C, disposed with noncrystallographic 222 symmetry (Anderson et al., 1981), with the A-C and O-B pairs corresponding to physiological dimers. Comparisons of the "A-half" (residues A4−A55 plus C55−C60) of this A-C dimer with the monomeric Cro mutant structure are shown in Figures 8 and 9. In Figure 8, superposition is based on the main-chain atoms of residues 16 to 35, i.e., on the helix−turn−helix region. As can be seen, however, there are significant differences elsewhere in the respective structures. The nature of these differences can be seen in the "difference−distance" plot (Figure 9), which shows the changes in distance between all pairs of α-carbon atoms in the respective structures.

The monomer can be considered as made up of two parts, an α-helical region (7−35) and a β-sheet region (40−59), linked by a short loop. These regions have been shown to exhibit somewhat independent thermodynamics (Griko et al., 1992). The major structural changes correspond to an overall shifting of the β-sheet region relative to the α-helical part. The main chain of the β-sheet region undergoes a straightening and twisting, resulting in a lever-arm-like motion about an effective fulcrum point near residues 44 and 50, the position at which the β2β3-ribbon leaves the core and becomes solvent-exposed. The central regions of the β2,

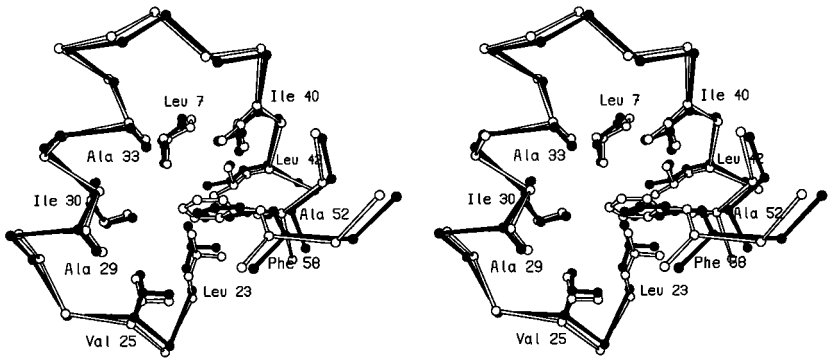

FIGURE 7: Stereo drawing showing how Phe 58 in the engineered monomer structure (solid bonds) moves away from the α3 helix (at left) relative to its position in the wild-type Cro dimer (open bonds). The side chain of Phe 58 is covered by Arg 38 and Glu 54, both of which have been omitted for clarity. The structures are superimposed on the basis of the main-chain atoms within the helix−turn−helix (residues 16−35).

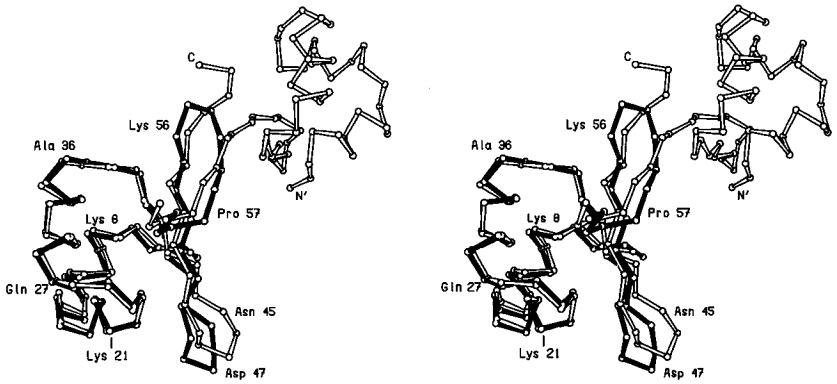

FIGURE 8: Superposition, based on the helix−turn−helix main-chain atoms (residues 16−35), of the engineered Cro monomer (solid bonds) on the dimer of wild-type Cro (open bonds).

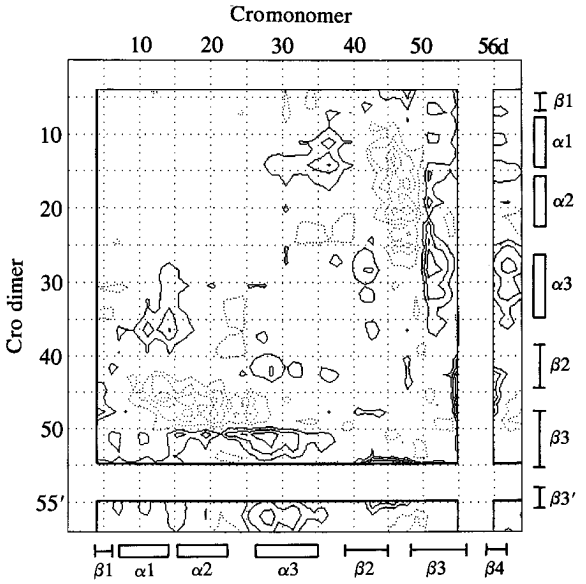

FIGURE 9: Difference−distance plot (Nishikawa et al., 1972) showing the change in distance between all pairs of C^{α} atoms in the mutant Cro structure relative to the wild-type dimer. The locations of the α-helices and β-strands are indicated. Positive contours (solid) and negative contours (dashed) are drawn at increments of 0.5 Å with the zero contour omitted. The residues that are outside the boxed regions cannot be compared because they correspond either to the disordered region at the amino terminus (residues 1−3) or to the inserted loop region in the monomer (residues 56−56c). The coordinates of the dimer are from crystallographic refinement at 2.2 Å resolution (D. E. Tronrud, D. H. Ohlendorf, and B. W. Matthews, unpublished) and will be described elsewhere, as will a description of the protein in complex with operator DNA at 3.0 Å resolution (R. A. Albright and B. W. Matthews, unpublished).

β3, and β4 strands (near residues 42, 52, and 56d), as well as Phe 58, all move away from the recognition helix (α3), with some of the shifts exceeding 1.6 Å (Figure 9). At the same time, the solvent-exposed β2β3-hairpin (residues 46−49) swings 2.2 Å away from residue 55, i.e., away from the engineered turn, and toward the α1α2-turn near residue 14.

The resulting hydrophobic core of the monomer is less tightly packed than that of the dimer, even though the identity of the residues in this region are not altered. This is particularly true in the region of the Phe 58 aromatic ring, which, although remaining predominantly buried, moves over 1.1 Å away from the center of the α3-helix (Figure 7). Extra space is thereby created in the core between Phe 58, Ala 29, and Ala 33. Subsequent random mutagenesis has identified mutants at these positions that exhibit substantial increases in thermal stability, presumably due to improved packing (Mollah et al., 1996).

DISCUSSION

Interconversion of Dimers and Monomers. The successful design of the monomeric form of Cro leads one to consider it as a general approach to engineer monomers from dimers. The feature that makes the design successful is the ability to transfer interactions that normally stabilize the dimer into interactions that occur within a single monomer. Given a polypeptide chain corresponding to wild-type Cro, the carboxy-terminal "tail", in particular Phe 58, is simply not long enough to bind within its "own" hydrophobic core. As a consequence, Phe 58 binds within the hydrophobic core of a second monomer, leading to the observed dimeric structure of the wild-type protein. The extension of the carboxy terminus, as in the engineered variant (Mossing & Sauer, 1990), permits Phe 58 to bind within its own hydrophobic core. In principle, the engineered variant could also form dimers, but the monomer is strongly favored because of the high effective concentration of Phe 58 (Creighton, 1983).

Recently, an "inverse" experiment in which a dimer was engineered from a monomer was reported by Green et al. (1995). In this case six amino acids in staphylococcal nuclease were deleted, preventing the carboxy-terminal α-helix from reaching its normal site within the monomeric structure of the enzyme and leading to its incorporation instead, in another molecule, resulting in a stable dimer.

The same principles of monomer−dimer interconversion are also seen in the comparison of monomeric and dimeric diphtheria toxin (Bennett et al., 1994) and in the interconversion of monomeric and dimeric forms of CD2 (Murray et al., 1995).

Structure of the Monomer Relative to the Dimer. The most surprising feature of the Cro monomer mutant is that its structure, while similar to that of wild type, displays significant perturbations. The largest changes involve the solvent-exposed β2β3-hairpin, but main-chain shifts in excess of 1.6 Å also occur within the core between portions of the β-sheet and α-helical regions. Substantial changes in hydrophobic packing occur near the tip of the Phe 58 aromatic ring, which has shifted 1.1 Å away from the α3-helix. This was unanticipated, since model building had indicated that a β-turn could be built into the antiparallel β-ribbon interface of the dimeric structure without distortion (Mossing & Sauer, 1990). Additionally, although torsional

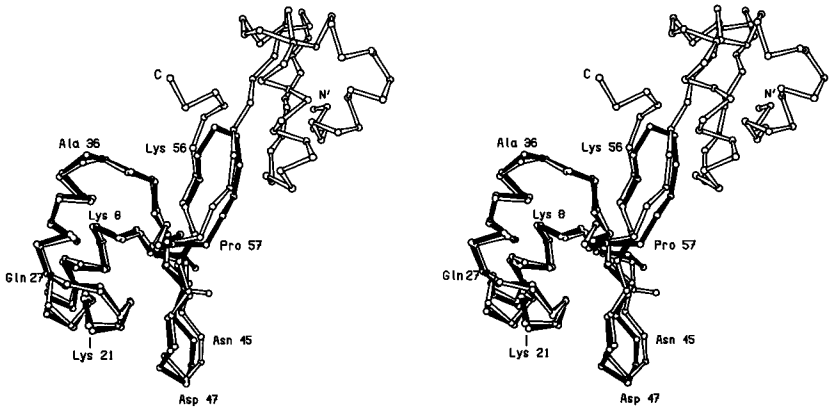

FIGURE 10: Superposition of the structure of the engineered monomer (solid bonds) on the dimeric structure of Cro (open bonds) as seen in a complex with operator DNA at 3.0 Å resolution (R. A. Albright and B. W. Matthews, unpublished). The superposition is based on residues 16−35, as in Figures 7 and 8.

flexibility at the center of the dimer interface was seen in a Cro−operator complex (Brennan et al., 1990), at the limited resolution available, no flexing of the core was noted.

Since the engineered insertion of the monomer mutant occurs in a solvent-exposed region of the structure, the core residues remain identical to those of the wild type. What, then, is the underlying reason for the structural changes? Since the hydrogen-bonding interactions that are present in the dimer are maintained in the monomer, the structural change does not include a distinct change from one set of interactions to another. Rather, the perturbations result from numerous small correlated changes throughout the folding domain. To some degree, such adjustments can be induced by differences in crystal packing environments but are usually limited in magnitude to about 0.5 Å (Zhang et al., 1995), substantially less than the changes in the present instance. On the other hand, the Cro structure is quite small and might be either more flexible or more susceptible to structural change than a larger protein. Yet another possibility is that the unusual changes in structure may be functionally relevant. This will be discussed in the following section.

Comparison of the Cro Monomer with Cro Bound to Operator DNA. The recent structure determination of a Cro−operator complex to 3.0 Å resolution (R. A. Albright and B. W. Matthews, unpublished) has allowed a substantially more detailed analysis of the structural changes that occur on binding DNA than was possible previously (Brennan et al., 1990). In Figure 10, the main-chain atoms of the monomer mutant are superimposed on those from the protein dimer as seen in the 3.0 Å resolution Cro−operator complex. As can be seen, the monomer is more similar to the structure of Cro bound to DNA (Figure 10) than it is to free Cro (Figure 6). For the backbone of residues 5−56 plus 56c−59, i.e., deleting just the two turn residues 56a and 56b, the root-mean-square discrepancy for the monomer *versus* wild-type Cro is 1.04 Å whereas it is 0.66 Å for the monomer *versus* DNA-bound Cro. The recent structure determination of the K56-[DGEVK] protein bound to DNA (R. A. Albright and B. W. Matthews, unpublished) shows that the monomer can adopt a conformation that is even closer to that of wild-type Cro complexed with operator DNA (data not shown). This suggests that the conformational changes seen in the engineered monomer structure relative to wild-type Cro are similar to the changes in the Cro structure that accompany DNA binding. It suggests that the Cro structure may be designed to readily permit this particular change in conformation.

CONCLUSIONS

In summary, the design of the Cro K56-[DGEVK] monomer has proven to be a remarkably successful example of protein engineering, while at the same time demonstrating the need for detailed structural information when analyzing the products of mutagenesis. The results also show that the structure of an isolated monomer, or a protein fragment, cannot be assumed to be identical with the structure *in toto*. It may be that the unanticipated structural changes within the Cro protein are in part due to changes in crystal contacts and to the presence of the engineered β-turn, but they seem, primarily, to reflect an inherent flexibility that may facilitate the large conformational changes in the protein that accompany DNA binding.

ACKNOWLEDGMENT

We thank Dr. X.-J. Zhang and Dr. R. H. Jacobson for helpful discussions.

REFERENCES

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 290*, 754−758.

Bennett, M. J., Choe, S., & Eisenberg, D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3127−3131.

Brennan, R. G., Roderick, S. L., Takeda, Y., & Matthews, B. W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8165−8169.

Creighton, T. E. (1983) *Biopolymers 22*, 49−58.

Gitelson, G. I., Griko, Y. V., Kurochkin, A. V., Rogov, V. V., Kutyshenko, V. P., Kirpichnikov, M. P., & Privalov, P. L. (1991) *FEBS Lett. 289*, 201−204.

Green, S. M., Gittis, A. G., Meeker, A. K., & Lattman, E. E. (1995) *Nature Struct. Biol. 2*, 746−751.

Griko, Y. V., Rogov, V. V., & Privalov, P. L. (1992) *Biochemistry 31*, 12701−12705.

Hamlin, R. (1985) *Methods Enzymol. 114*, 416−452.

Heinz, D. W., Baase, W. A., Dahlquist, F. W., & Matthews, B. W. (1993) *Nature 361*, 561−564.

Howard, A. J., Nielsen, C., & Xuong, N. H. (1985) *Methods Enzymol. 114*, 452−472.

Jones, T. A. (1982) in *Crystallographic Computing* (Sayre, D., Ed.) pp 303−317, Oxford University Press, Clarendon, Oxford.

Kyte, J. (1995) *Structure in Protein Chemistry*, Garland Publishing, New York.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
Luzzati, P. V. (1952) *Acta Crystallogr. 5*, 802−810.
Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
Mollah, A. K. M. M., Aleman, M., Albright, R. A., & Mossing, M. C. (1996) *Biochemistry 35*, 743−748.
Mossing, M. C., & Sauer, R. T. (1990) *Science 250*, 1712−1715.
Murray, A. J., Lewis, S. J., Barclay, A. N., & Brady, R. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7337−7341.
Nishikawa, K., Ooi, T., Isogai, Y., & Saito, N. (1972) *J. Phys. Soc. (Jpn.) 32*, 1331−1337.
Sibanda, B. L., Blundell, T. L., & Thornton, J. M. (1989) *J. Mol. Biol. 206*, 759−777.
Sigler, P. B., & Blow, D. M. (1965) *J. Mol. Biol. 14*, 640−644.
Takeda, Y., Folkmanis, A., & Echols, H. (1977) *J. Biol. Chem. 252*, 6177−6183.
Terwilliger, T. C., & Eisenberg, D. (1983) *Acta Crystallogr. A39*, 813−817.
Tronrud, D. E. (1992) *Acta Crystallogr. A48*, 912−916.
Tronrud, D. E., Ten Eyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489−503.
Zhang, X.-J., & Matthews, B. W. (1993) *J. Appl. Crystallogr. 26*, 1274−1284.
Zhang, X.-J., & Matthews, B. W. (1995) *J. Appl. Crystallogr. 28*, 624−630.
Zhang, X.-J., Wozniak, J. A., & Matthews, B. W. (1995) *J. Mol. Biol. 250*, 527−552.
Zuiderweg, E. R. P., Kaptein, R., & Wüthrich, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5837−3841.

BI951958N