Evidence for a contact between glutamine-18 of lac repressor and base pair 7 of *lac* operator

("loss-of-contact" mutations/repressor-operator interaction/helix-turn-helix motif/catabolite gene activator protein)

RICHARD H. EBRIGHT

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115

Communicated by Jonathon Beckwith, September 16, 1985

ABSTRACT Glutamine-18 of the *lac* repressor (*lacR*) has been substituted by glycine, by serine, and by leucine. The specificities of wild-type *lacR* and of the three substituted *lacR* variants have been analyzed with respect to base pairs 5, 6, 7, 8, 9, and 10 of the *lac* operator (*lacO*). The data indicate that [Gly¹⁸]*lacR*, [Ser¹⁸]*lacR*, and [Leu¹⁸]*lacR* lose the ability to distinguish between the *O*⁺ base pair G-C and the *O*^c base pairs T-A and A-T at position 7 of *lacO* ($K_{do^-}/K_{do^+} \approx 1$). In contrast, the three substituted variants retain the ability to discriminate *O*⁺ from *O*^c at each other position, by factors of 9 to 37. Therefore, I propose that glutamine-18 contacts base pair 7 of *lacO*. These data suggest that the interaction between the helix-turn-helix motif and DNA may be very similar or identical in *lacR* and the catabolite gene activator protein.

The structures of three sequence-specific transcriptional regulatory proteins have been determined by x-ray diffraction techniques; these are the catabolite gene activator protein (CAP), the amino-terminal fragment of the λ repressor, and λ -cro (reviewed in refs. 1 and 2). A prominent feature of each protein is an α -helix-turn- α -helix supersecondary structure. The 24 amino acids that comprise this motif in CAP, λ repressor, and λ -cro exhibit an identical three-dimensional structure and a homologous primary sequence (1, 2). It has been suggested that the two helix-turn-helix motifs of the DNA-binding protein dimer interact in a 2-fold symmetric fashion with successive major grooves of right-handed B-type DNA (2). The proposed interaction is such that amino acid side chains of the second α -helix of the helix-turn-helix motif are positioned to contact solvent-exposed atoms of DNA base pairs, thereby determining the specificity of binding.

In the case of CAP, we have shown elsewhere that substitution of glutamic acid-181—residue 2 of the second α -helix of the helix-turn-helix motif—by lysine, leucine, or valine alters the DNA-sequence specificity (3); the alteration observed is a change in specificity from G-C to A-T at base pairs 7 and 16 of the 2-fold symmetric DNA site. We have proposed, therefore, that glutamic acid-181 of CAP makes a direct contact to base pairs 7 and 16 (3). It has been possible to infer from this proposal, in conjunction with the three-dimensional structure of CAP, (i) the specific chemical interactions by which glutamic acid at position 181 interacts with G-C, whereas lysine, leucine, and valine interact with A-T, and (ii) the structure of the intermolecular CAP-DNA complex (4). The resulting model for the CAP-DNA complex is identical in its geometry to a model proposed independently by Steitz and co-worker (5) based on electrostatic and structural considerations.

Matthews, Sauer, Steitz, and co-workers (6–8) have pointed out that residues 5 to 26 of *lac* repressor (*lacR*) exhibit a significant amino acid sequence homology to the residues that comprise the helix-turn-helix motifs of CAP, λ repressor,

and λ -cro (Fig. 1*a*). Therefore, these workers have proposed that residues 5 to 26 of *lacR* adopt a helix-turn-helix conformation that is identical to that in the three known structures (6–8). This prediction has been verified by NMR data indicating that (i) the implicated region of *lacR* in fact folds as two α -helices, with the end-points of the helices at the positions predicted by the sequence homology (13), and (ii) the two α -helices are oriented with respect to each other as in CAP, λ repressor, and λ -cro (14, 15). Matthews, Sauer, Steitz, and co-workers (6–8) also have proposed that the implicated region and *lacR* mediates part or all of the sequence-specific interaction between *lacR* and *lacO*. This second proposal is consistent with a large body of genetic (16–18), chemical (19), and spectroscopic (20, 21) data accumulated over two decades.

The object of the experiments in this report was to identify which base pair of *lacO*—if any—is contacted by the amino acid of *lacR* equivalent in its position within the helix-turn-helix motif to glutamic acid-181 of CAP: i.e., glutamine-18 (Fig. 1*a*). To address this question, I have used an approach involving "loss of contact" amino acid substitutions (Fig. 2). A detailed description of this approach, as well as a preliminary account of part of the data, was presented at the 1985 Biomolecular Stereodynamics Conference (22).

OUTLINE OF EXPERIMENTAL PROCEDURE

The basis of the experimental approach is to eliminate the ability of glutamine-18 to make a contact to DNA, and then to determine at which base pair—if any—specificity is eliminated. To eliminate the contact by residue 18, I substitute glutamine-18 by glycine, which has no side chain. I have also performed parallel experiments in which glutamine-18 was substituted by serine and by leucine, the side chain of each of which is shorter and chemically distinct from that of glutamine. As outlined in Fig. 2, I reason that a "loss of contact" variant of *lacR* of this type will (i) lose the ability to discriminate between *O*⁺ and *O*^c at one position of *lacO*—this result defining the position contacted in wild-type *lacR* by glutamine-18—but (ii) retain the ability to discriminate between *O*⁺ and *O*^c at each other position.

In ref. 22, I have reported the construction of four sets of *Escherichia coli* K-12 strains, each set comprised of seven strains. These strains enable one to determine the profile of specificity with respect to *O*⁺ and to each *O*^c operator illustrated in Fig. 3*a*, of wild-type *lacR* and of *lacR* variants in which glutamine-18 is substituted by glycine, serine, and leucine. As an example, the set of seven strains used to investigate the profile of specificity of [Gly¹⁸]*lacR* is presented in Fig. 3*b*. Each strain in this set consists of three critical components as follows: (i) the *lacZ* gene, which encodes the assayable product β -galactosidase, placed under the control either of *lacO*⁺, or one of the six *O*^c operators illustrated in Fig. 3*a*—i.e., base changes at positions 5, 6, 7, 8, 9, and 10

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *lacR*, *lac* repressor protein; CAP, catabolite gene activator protein.

FIG. 1. Comparison of lacR and CAP. (a) Sequence homology between residues 5–26 of lacR and the residues that comprise the helix-turn-helix motif (residues 168–189) in CAP (8). The positions of α -helices E and F of CAP are indicated beneath the sequences. Glutamine-18 of lacR and glutamic acid-181 of CAP are underlined. (b) Consensus DNA half sites of lacR (9, 10) and CAP (3). The base-pairs postulated to contact glutamine-18 of lacR (this work) and glutamic acid-181 of CAP (3) are underlined. Positions at which ethylation of a DNA phosphate interferes with binding (11, 12) are indicated by primes ('); guanine residues protected against methylation by binding (11, 12) are circled.

of *lacO*; (ii) the *lacI^q* gene, which encodes lacR, containing an amber (UAG) mutation at codon 18; and (iii) mutations in two tRNA genes that result in high-efficiency chain elongation past an amber codon, with concomitant insertion at the amber codon of glycine.

To investigate the profiles of specificity of wild-type lacR, [Ser¹⁸]lacR, and [Leu¹⁸]lacR, three additional sets of strains were utilized, analogous to the set in Fig. 3b, except that component iii was a nonsense suppressor mutation in a *tRNA^{Gln}*, *tRNA^{Ser}*, or *tRNA^{Leu}* gene, respectively. Preliminary experiments with [Ser¹⁸]lacR and [Leu¹⁸]lacR indicated that—as a consequence of poor repression by these two lacR variants—it was not possible to obtain precise quantitative data using sets of strains in which the *lacI^qamA5* gene was present on the single-copy episome *F' lac pro* (22). Therefore, in the present work, I have utilized sets of strains for [Ser¹⁸]lacR and [Leu¹⁸]lacR in which the *lacI^qamA5* gene was introduced on the multicopy plasmid vector *placI^q-PUV5 Δ tetP* (a derivative of pBR322; for construction, see ref. 22). Calculations based on the differential rates of β -galactosidase synthesis indicate that use of the multicopy plasmid results in a 57-fold and a 23-fold

increase in the intracellular concentration of [Ser¹⁸]lacR and [Leu¹⁸]lacR, respectively.

For each strain, two measurements were taken (i) the differential rate of β -galactosidase synthesis in the absence of isopropyl 1-thio- β -D-galactoside (IPTG), dZ/dB , and (ii) the differential rate of β -galactosidase synthesis in the presence of IPTG (i.e., in the absence of active lacR), $(dZ/dB)_I$. The ratio of these two values is defined as P , and is equal to the fraction of *lacO* not occupied by lacR (23, 26).

$$P = \frac{(dZ/dB)}{(dZ/dB)_I} \quad [1]$$

where Z is β -galactosidase activity, and B is bacterial cell density. By rearrangement of the equilibrium relationship, it can be shown that

$$P = \frac{K_d}{K_d + R} \quad [2]$$

where K_d is the dissociation constant, and R is equal to the molar concentration of unbound lacR tetramers (26). Therefore, for two strains such that $R_1 = R_2$ (i.e., any two strains within the same set), the ratio of the dissociation constants K_{d1}/K_{d2} can be calculated as follows:

$$\frac{K_{d1}}{K_{d2}} = \frac{P_1(1 - P_2)}{P_2(1 - P_1)} \quad [3]$$

Note that within each of the four sets of strains, the seven strains that comprise the set differ only in the sequence of *lacO*. All calculations made in this report have been restricted to data from strains within the same set. Therefore, this analysis does not incur complications due to either (i) differing stabilities of the mutant vs. wild-type proteins or (ii) differing efficiencies of suppression by different nonsense suppressors. Note also that in all operators utilized in this investigation the right half of the approximately 2-fold symmetric operator is unaltered; therefore, I presume that the contacts made to the right half of *lacO* by a specific lacR variant will be equal in the cases of *lacO⁺* and *lacO^c*.

RESULTS

Effects of Substitution of Glutamine-18 of lacR. The data from this analysis are summarized in Table 1. Columns 2 and 3 present the data for wild-type lacR, in which amino acid 18 is glutamine. The data in column 3 (K_{dO^c}/K_{dO^+}) indicate that wild-type lacR discriminates between *O⁺* and *O^c* at each position of *lacO* by a factor of 24 to 130, the average factor of discrimination being 75. The values obtained are in excellent

FIG. 2. Logic of the experiment. The basis of this approach is to eliminate the ability of residue 18 to make a contact to DNA. I reason that such a "loss of contact" lacR will (i) lose the ability to distinguish between *O⁺* and *O^c* at one position of *lacO*—this result defining the position contacted in wild-type lacR by glutamine-18—but (ii) retain the ability to discriminate between *O⁺* and *O^c* at each other position. (a) Interaction of wild-type lacR and four base pairs of *lacO⁺* (4 contacts). The residue of lacR participating in each contact is indicated schematically as a vertical line; the presence of a favorable contact is indicated schematically by a small closed circle. (b–d) Interaction of wild-type lacR with *lacO^c* operators (3 contacts). (a') Interaction of the residue 18-substituted lacR with *lacO⁺* (3 contacts). (b') Interaction of the residue 18-substituted lacR with the *O^c* site containing a base change at the position contacted in wild-type lacR by glutamine-18 (3 contacts; identical to the number with *lacO⁺*). (c' and d') Interaction of the residue 18-substituted lacR with other *O^c* sites (2 contacts).

FIG. 3. (a) The *lacO*⁺ sequence. Base positions are numbered from the start point of *lac* transcription; using this convention, the left-most base pair of *lacO* is designed as position 4. The axis of approximate 2-fold sequence symmetry is located at position 11. The sequence changes in the six *O*^c mutations used in this study (23–25) are indicated beneath the *lacO*⁺ sequence. (b) The set of strains used to determine the specificity profile of [Gly¹⁸]*lacR*. Each strain contains in addition the following markers: F'(lacPL8Z⁺Y⁺proA⁺B⁺) Δ*lacproABXIII* *recA1 srl::Tn10* *argEam* *ara* *nalA* *rpoB* (22).

agreement with those determined by others *in vivo* (23, 26), in nitrocellulose filter-binding experiments with intact λ h80*dlac* DNA (26), and in DNase I protection experiments (27).

The data for the lacR variant in which glutamine-18 is substituted by glycine are given in columns 4 and 5 of Table 1. The data in the K_{dO^c}/K_{dO^+} column indicate that, in the case of the O^c mutation located at position 7 of *lacO*, [Gly¹⁸]lacR loses completely the ability to discriminate between O^+ and O^c ($K_{dO^c}/K_{dO^+} = 0.93$). In contrast, for the O^c mutations located at positions 5, 6, 8, 9, and 10, [Gly¹⁸]lacR retains the ability to discriminate between O^+ and O^c —in this case by factors of 13 to 23.

The quantitative data for [Ser¹⁸]lacR and [Leu¹⁸]lacR, given in columns 6–9 of Table 1, are remarkably similar to those obtained for [Gly¹⁸]lacR. Thus, for the *O*^c mutation at position 7, [Ser¹⁸]lacR loses the ability to discriminate between *O*⁺ and *O*^c ($K_{do^c}/K_{do^+} = 1.4$), and [Leu¹⁸]lacR actually exhibits a reversal of specificity, displaying a 3- to 4-fold preference for the *O*^c base pair T·A vs. the *O*⁺ base pair G·C ($K_{do^c}/K_{do^+} = 0.30$). In contrast, for the *O*^c mutations located at positions 5, 6, 8, 9, and 10, [Ser¹⁸]lacR and [Leu¹⁸]lacR retain the ability to discriminate between *O*⁺ and *O*^c by factors of 9 to 37 and 12 to 35, respectively. These data are in good agreement with the qualitative results reported in ref. 22, based on experiments with *lacI^qama5* present on *F'**lac pro*.

Recognition of O^c Base Pair A·T at Position 7. In conjunction

with the logic outlined in Fig. 2, the data with [Gly¹⁸]lacR, [Ser¹⁸]lacR and [Leu¹⁸]lacR constitute a preliminary indication that glutamine-18 of lacR interacts with base pair 7 of *lacO*. To test this hypothesis, the effects of a second *O*^c base change at position 7 of *lacO*—i.e., substitution of the *O*⁺ base pair G-C by A-T—were investigated. The G-C → A-T substitution was obtained from W. Reznikoff; the isolate made available contained in addition the *lacPUV5* and *Z361* mutations, these being, respectively, a CAP-cAMP-independent promoter mutation, and a 27-base-pair insertion at codon 9 of *lacZ* (28). It was, therefore, necessary to compare the data obtained with this isolate to a *lacPUV5-Z361 O*⁺ control (for construction of strains, see Table 2). Table 2, column 3, shows that wild-type lacR discriminates between the *O*⁺ base pair G-C and the *O*^c base pair A-T at position 7 by a factor of 12. In contrast, [Gly¹⁸]lacR and [Ser¹⁸]lacR lose essentially all of the ability to discriminate between *O*⁺ G-C and *O*^c A-T ($K_{dO^c}/K_{dO^+} = 1.6$; 1.4), and [Leu¹⁸]lacR exhibits a modest reversal of specificity ($K_{dO^c}/K_{dO^+} = 0.83$) (Table 2, columns 5, 7, and 9).

DISCUSSION

Contact by Glutamine-18 to Base Pair 7. The data indicate that substitution of residue 18 of lacR by glycine, serine, or leucine in each case (*i*) eliminates—or reverses—the ability to distinguish between the O^+ base pair G-C and the O^c base pairs T-A and A-T at position 7 of *lacO*, but (*ii*) does not

Table 1. Specificity profiles of wild-type lacR, [Gly¹⁸]lacR, [Ser¹⁸]lacR, and [Leu¹⁸]lacR

Operator	Wild-type lacR		[Gly ¹⁸]lacR		[Ser ¹⁸]lacR		[Leu ¹⁸]lacR	
	<i>P</i>	<i>K</i> _{dOC} / <i>K</i> _{dO+}	<i>P</i>	<i>K</i> _{dOC} / <i>K</i> _{dO+}	<i>P</i>	<i>K</i> _{dOC} / <i>K</i> _{dO+}	<i>P</i>	<i>K</i> _{dOC} / <i>K</i> _{dO+}
<i>O</i> ⁺	0.00067	[1]	0.046	[1]	0.012	[1]	0.16	[1]
5 A·T	0.080	130	0.51	22	0.21	21	0.70	12
6 C·G	0.056	88	0.50	21	0.10	9.0	0.81	22
7 T·A	0.035	54	0.043	0.93	0.017	1.4	0.054	0.30
8 G·C	0.043	67	0.39	13	0.22	23	0.71	13
9 T·A	0.016	24	0.42	15	0.21	21	0.87	35
10 T·A	0.056	88	0.53	23	0.32	37	0.85	30

Operators are identified by the position and sequence of the O^c base change (sequences in Fig. 3A). The strains utilized were as follows: XAcE201B.2-209.2 for wild-type lacR ; XAcE701B.2-709.2 for [Gly¹⁸] lacR ; XAcE101B.5-109.5 for [Ser¹⁸] lacR ; and XAcE601B.5-609.5 for [Leu¹⁸] lacR (22). The differential rates of β -galactosidase synthesis were measured and corrected as described in ref. 22, except that media for strains XAcE101B.5-109.5 and XAcE601B.5-609.5 contained 40 μg of L-proline/ml and 75 μg of ampicillin/ml. P was calculated using Eq. 1. The values shown are the means of at least two independent determinations [mean standard error = ± 0.050 (P)]. K_{dO^c}/K_{dO^+} was calculated using Eq. 3.

Table 2. Recognition of the O^c base pair A-T at position 7 of *lacO*

Operator	Wild-type <i>lacR</i>		[Gly ¹⁸] <i>lacR</i>		[Ser ¹⁸] <i>lacR</i>		[Leu ¹⁸] <i>lacR</i>	
	<i>P</i>	K_{do^-}/K_{do^+}	<i>P</i>	K_{do^-}/K_{do^+}	<i>P</i>	K_{do^-}/K_{do^+}	<i>P</i>	K_{do^-}/K_{do^+}
$O^+/PUV5\text{-}Z361$	0.018	[1]	0.12	[1]	0.074	[1]	0.49	[1]
$7A\text{-}T/PUV5\text{-}Z361$	0.18	12	0.18	1.6	0.10	1.4	0.44	0.83

$\lambda cI857plac5\text{-}PUV5\text{-}Z361$ and $\lambda cI857plac5\text{-}PUV5\text{-}O^c7\text{-}Z361$ (28) were converted to $i434 cI^+$ by a phage cross with $\lambda i434plac5\text{-}\Delta OZS20$ (22). Except for the identity of the $\lambda i434plac5$ derivative, the strains used are identical to those in Table 1. The strain designations are: XAcE210.2, 211.2 for [Gln¹⁸]*lacR*; XAcE710.2, 711.2 for [Gly¹⁸]*lacR*; XAcE110.5, 111.5 for [Ser¹⁸]*lacR*; XAcE610.5, 611.5 for [Leu¹⁸]*lacR*. Values for *P* and K_{do^-}/K_{do^+} were obtained as in Table 1.

eliminate the ability to distinguish between O^+ and O^c at each other position. Therefore, I propose that glutamine-18 of *lacR* makes a direct contact to base pair 7 of *lacO*. I note that a contact between glutamine-18 and base pair 7 was predicted in ref. 30, based on a systematic attempt to align amino acid sequences and DNA sequences.

Extent of Independence of Contacts. Fersht and co-workers (29) have described a simple test of independence of contacts in a protein-ligand interaction; they designate this test the "double mutant procedure." The basis of this procedure is to compare the effects of substitution of residue A—a residue participating in a contact *other than* the contact under investigation—in each of two contexts: (i) the wild-type protein and (ii) a variant substituted in addition at residue B—a residue participating in the contact under investigation. Departures from complete independence between contacts A and B manifest themselves as a difference in the quantitative effect of substitution of residue A in case i as compared to case ii.

The data in Table 1 permit five applications each of this test to [Gly¹⁸]*lacR*, [Ser¹⁸]*lacR*, and [Leu¹⁸]*lacR*: i.e., where residue A is, respectively, base pair 5, 6, 8, 9, or 10. The data exclude a strict independence of contacts. Thus, as compared to wild-type *lacR*, each of the three substituted *lacR* variants exhibits a reduction in the ratio K_{do^-}/K_{do^+} at base pairs 5, 6, 8, 9, and 10—the factors of reduction ranging from 0.69 to 10.8 (mean = 4.4). As it is postulated that there is no contact between amino acid 18 and these base pairs, this reduction in K_{do^-}/K_{do^+} must be attributed to indirect—i.e., nonindependent—effects. The results indicate that upon substitution of glutamine-18 by glycine, serine, and leucine there may occur slight changes in peptide backbone conformation and/or in local surface packing.

This approximately 4.4-fold reduction in K_{do^-}/K_{do^+} is to be contrasted to the 58-, 39-, and 180-fold reductions in K_{do^-}/K_{do^+} at base pair 7—the latter effects attributed to the loss of a direct contact. It remains to be established whether in other instances also the quantitative effects of second-order, indirect interactions will be significantly exceeded by the effects of loss of a direct contact (thereby satisfying the logic of Fig. 2), or whether the situations observed with glutamine-18 of *lacR* (this work) and glutamic acid-181 of CAP (3) were fortuitous. If approximate independence/additivity of contacts is not atypical, it should be possible to use the "loss of contact" procedure described in this report to identify residues involved in other amino acid-base pair contacts, both in *lacR* and in other proteins.

Comparison to CAP. Fig. 1b illustrates the consensus DNA half sites of *lacR* and CAP. The DNA sites of both proteins contain the identical 5-base-pair sequence T-G-T-G-A (12, 30). Note also that the points of protein-DNA contact identified in chemical protection and modification experiments (11, 12) occupy identical positions in the sites of *lacR* and CAP with respect to the homologous 5-base-pair sequence. It has been shown elsewhere that the sequence T-N-T-N-A-N is a highly conserved element in the DNA-half sites of thirteen proteins known or thought to contain the helix-turn-helix motif, and that, in four of the five examples for which these data are available, the points of protein-

phosphate contact identified in chemical experiments are located with respect to the sequence T-N-T-N-A-N exactly as in *lacR* and CAP (30).

In the case of CAP, data have been presented indicating that residue 2 of α -helix F—glutamine acid-181—makes a direct contact to base pairs 7 and 16 of the symmetric DNA site (3). These positions correspond to the G-C base pair at the fourth position within the homologous 5-base-pair sequence T-G-T-G-A. Here, I have presented data indicating that residue 2 of the helix-F homolog of *lacR*—glutamine-18 (Fig. 1a)—contacts base pair 7 of *lacO*; in this case also, base pair 7 corresponds to the G-C base pair at the fourth position within the sequence T-G-T-G-A. It is striking that the equivalently located amino acids within the helix-turn-helix motifs of the two proteins contact equivalently located base pairs within the respective DNA half sites. The postulated contact between glutamic acid-181 of CAP and the G-C base pair is an H bond between a carboxylate oxygen atom of the glutamic acid side chain and the N-4 atom of cytosine (4). It was noted in ref. 4 that glutamine, "using the amide oxygen atom," could make a contact to G-C using the identical geometry; I suggest that this is what occurs in the case of glutamine-18 of *lacR*. I propose, therefore, that the geometry of the interaction between the helix-turn-helix motif and DNA in *lacR* and CAP may be identical or extremely similar.

Detailed models of the contacts to DNA by amino acids of the helix-turn-helix motif of CAP have been presented elsewhere (Fig. 4a; refs. 4, 5, and 30). By analogy to these models, and using the data in this report, it is possible tentatively to predict additional contacts to DNA by amino acids of the helix-turn-helix motif of *lacR*. These predictions are illustrated schematically in Fig. 4b. Note that 6 of the 10 predicted contacts are made by residues that are identical in CAP and *lacR* and that are predicted to contact identical atoms in the respective DNA sites. Apart from the putative contact by glutamine-18, the three strongest predictions for sequence-specific contacts to DNA base pairs follow. (i) The phenol hydroxyl of either tyrosine-7 or -17 makes an H bond to the guanine N-7 atom at G-C base pair 5 of *lacO*. In NMR experiments, irradiation of the guanosine H-8 frequency region has been reported to elicit a nuclear Overhauser enhancement at the tyrosine-7/tyrosine-17 frequency (21). (ii) The methyl of threonine-19 makes a van der Waals interaction with the thymine 5-methyl at A-T base pair 8 of *lacO*. Experiments with the base analogs uracil and 5-methylcytosine indicate that the thymine 5-methyl moiety is the chief determinant of specificity at base pair 8 (31). (iii) The guanidinium side chain of arginine-22 makes an H bond to either the thymine O⁴ atom or the adenine N-7 atom at A-T base pair 8 of *lacO*. Due to the ability of arginine-22 to reach to either base of base pair 8, it is unlikely that this contact confers significant sequence specificity (cf. ref. 30).

Differences from CAP. There are two arguments against the hypothesis that the geometry of the interaction between the helix-turn-helix motif and DNA in *lacR* and CAP is identical. First, glutamic acid is a nonfunctional substitution at residue 18 of *lacR* (18). Second, whereas evidence has been presented that suggests that substitution of glutamic acid-181 of CAP

FIG. 4. Contacts between DNA site and helix-turn-helix motif. (a) Proposal for CAP (4, 5, 30). Lines 3–6, the amino acids proposed to contact DNA base-pair edges. Numbers give the position of the amino acid with respect to the start of helix F (Fig. 1a); the identity of the amino acid side chain is given in lines 4 and 6, using the one-letter code. Residue 2 (glutamic acid-181) and the contacted DNA base pair are boxed (3, 4). Lines 7–10, the amino acids proposed to contact, or to approach closely, the DNA phosphate backbone (MC denotes polypeptide main chain); the phosphates proposed to receive these contacts are marked in lines 1 and 2 by arrows. (b) Proposal for lacR, assuming that the interaction between the homologous 5-base-pair DNA sequence and the helix-turn-helix motif is identical in lacR and CAP. Residue 2 (glutamine-18) and the contacted DNA base pair are boxed. Residues -10 and 1 (tyrosine-17 and tyrosine-7) can each be modeled to contact either (i) base pair 5 of lacO (position "2" in the sequence illustrated) or (ii) the upper-strand phosphate 5' to base pair 5.

by leucine results in specificity for A·T (3), substitution of glutamine-18 of lacR by leucine results in a loss of specificity (Tables 1 and 2). The simplest interpretation of these two observations is that the α and β carbon atoms of the amino acid side chain at residue 2 of the helix-F homolog do not have exactly the same orientation with respect to the contacted DNA base pair in the cases of lacR and CAP. However, there is an alternative explanation for each of these two observations.

(i) There are three positively charged amino acids within 5 Å of residue 181 of CAP (arginine-180, arginine-185, and lysine-188) but only one positively charged amino acid within 5 Å of residue 18 of lacR (arginine-22). Thus, introduction of the negatively charged glutamic acid side chain at position 18 of lacR may interfere with folding and/or with the interaction with the negatively charged DNA molecule.

(ii) The postulated leucine-181–A·T contact in [Leu¹⁸¹]CAP is a van der Waals interaction with the thymine 5-methyl (4). To make this contact with the residue 181 α and β carbons in the orientation postulated for the glutamic acid-181–G·C contact, the leucine side chain must adopt a disfavored conformation (see figure 2 of ref. 4). Alternatively, the residue 181 α and β carbons may retract slightly from the DNA in the leucine-181–A·T contact. If so, the differing effects of substitution by leucine in lacR and CAP may reflect a lesser amount of flexibility in lacR with respect to move-

ment of the leucine-18 α and β carbons—without necessarily a difference in the geometry of the wild-type helix-turn-helix motif–DNA interaction.

Ultimately, it will be desirable to obtain a catalog of the interactions of all 20 amino acids at residue 181 of CAP with all 4 base pairs at the contacted base position, and the analogous catalog for residue 18 of lacR. A comparison of the 160 entries in these two catalogs may permit detailed conclusions regarding the actual extent of the similarity in the interaction with DNA of the two proteins.

I thank J. Beckwith for comments on the manuscript and for research support. I acknowledge also W. Reznikoff for phage isolates. This work was supported by National Institutes of Health Grant GM13017 to J. Beckwith.

1. Takeda, Y., Ohlendorf, D., Anderson, W. & Matthews, B. (1983) *Science* **221**, 1020–1026.
2. Pabo, C. & Sauer, R. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
3. Ebright, R., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984) *Nature (London)* **311**, 232–235.
4. Ebright, R., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7274–7278.
5. Weber, I. & Steitz, T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3973–3977.
6. Matthews, B., Ohlendorf, D., Anderson, W. & Takeda, Y. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1428–1432.
7. Sauer, R., Yocum, R., Doolittle, R., Lewis, M. & Pabo, C. (1982) *Nature (London)* **298**, 447–451.
8. Weber, I., McKay, D. & Steitz, T. (1982) *Nucleic Acids Res.* **10**, 5085–5102.
9. Sadler, J., Sasnor, H. & Betz, J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6785–6789.
10. Simons, A., Tils, D., von Wilcken-Bergmann, B. & Muller-Hill, B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1624–1628.
11. Barkley, M. & Bourgeois, S. (1978) in *The Operon*, eds. Miller, J. & Reznikoff, W. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 177–220.
12. Majors, J. (1975) Dissertation (Harvard Univ., Cambridge, MA).
13. Zuiderweg, E., Kaptein, R. & Wuthrich, K. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5837–5841.
14. Arndt, K., Boschelli, F., Lu, P. & Miller, J. (1981) *Biochemistry* **20**, 6109–6118.
15. Kaptein, R., Zuiderweg, E., Scheek, R., Boelens, R. & van Gunsteren, W. (1985) *J. Mol. Biol.* **182**, 179–182.
16. Adler, K., Beyreuther, K., Fanning, G., Geisler, N., Gronenborn, B., Klemm, A., Muller-Hill, B., Pfahl, M. & Schmitz, A. (1972) *Nature (London)* **237**, 322–327.
17. Miller, J., Coulondre, C., Hofer, M., Schmeissner, U., Sommer, H. & Schmitz, A. (1979) *J. Mol. Biol.* **131**, 191–222.
18. Miller, J. (1984) *J. Mol. Biol.* **180**, 205–212.
19. Fanning, T. (1975) *Biochemistry* **14**, 2512–2520.
20. Buck, F., Ruterjans, H., Kaptein, R. & Beyreuther, K. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5145–5148.
21. Nick, H., Arndt, K., Boschelli, F., Jarema, M., Lillis, M., Sadler, J., Caruthers, M. & Lu, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 218–222.
22. Ebright, R. (1985) *J. Biomol. Struct. Dyn.* **3**, 281–297.
23. Smith, T. & Sadler, J. (1971) *J. Mol. Biol.* **59**, 273–305.
24. Gilbert, W., Gralla, J., Majors, J. & Maxam, A. (1975) in *Protein-Ligand Interactions*, eds. Sund, D. & Blauer, G. (de Gruyter, Berlin), pp. 193–207.
25. Maquat, L., Thornton, K. & Reznikoff, W. (1980) *J. Mol. Biol.* **139**, 537–549.
26. Jobe, A., Sadler, J. & Bourgeois, S. (1974) *J. Mol. Biol.* **85**, 231–248.
27. Pfahl, M. & Hendricks, M. (1984) *J. Mol. Biol.* **172**, 405–416.
28. Munson, L., Stormo, G., Niece, R. & Reznikoff, W. (1984) *J. Mol. Biol.* **177**, 663–683.
29. Carter, P., Winter, G., Wilkinson, A. & Fersht, A. (1984) *Cell* **38**, 835–840.
30. Ebright, R. (1986) in *Protein Structure, Folding, and Design*, ed. Oxender, D. (Liss, New York), in press.
31. Fisher, E. & Caruthers, M. (1979) *Nucleic Acids Res.* **7**, 401–416.