tively thicken on entering a more viscous layer. These results are consistent with the focal mechanisms of earthquakes within slabs, with observed kinks in slabs from residual sphere analysis, with the shape of Benioff zones from both seismicity and tomography, and with the broadening of slabs from waveform analysis of shear waves. Taken together, these seismic observations support the hypothesis that the lower mantle is at least 10–30 times more viscous than the upper mantle, consistent with the picture emerging from observations and models of the geoid[22], from global seismic tomography[23], and from isotope geochemistry[24].

This work was supported by the National Science Foundation and the Ametek Corporation. Computing was carried out on the facilities of the San Diego Supercomputer Center; we thank Steve Lamont at SDSC for assistance in animating the simulations.

Received 20 April; accepted 14 July 1988.

1. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **84**, 1031–1048 (1979).
2. Hager, B. H., O'Connell, R. J. & Raefsky, A. *Tectonophys.* **99**, 165–189 (1983).
3. Davies, G. F. *EOS, Trans. Am. geophys. U.* **67**, 381 (1986).
4. Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **91**, 3573–3589 (1986).
5. Isacks, B. & Molnar, P. *Rev. Geophys.* **9**, 103–174 (1971).
6. Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505–509 (1984).
7. Zhou, H.-W. & Clayton, R. W. *EOS Trans. Am. geophys. U.* **68**, 1379 (1987).
8. Silver, P. G., Carlson, R. W. & Olson, P. *A. Rev. Earth planet. Sci.* **16**, 477–541 (1988).
9. Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **89**, 4389–4402 (1984).
10. Kincaid, C. & Olson, P. *J. geophys. Res.* **92**, 13832–13840 (1987).
11. Stevenson, D. J. & Turner, J. S. *Nature* **270**, 334–336 (1987).
12. Tovish, A., Schubert, G. & Luyendyk, B. P. *J. geophys. Res.* **83**, 5892–5898 (1978).
13. Richter, F. M. *J. geophys. Res.* **84**, 6783–6995 (1979).
14. Vassiliou, M. S., Hager, B. H. & Raefsky, A. *J. Geodyn.* **1**, 11–28 (1984).
15. Uyeda, S. & Kanamori, H. *J. geophys. Res.* **84**, 1049–1061 (1979).
16. Vassiliou, M. S. thesis, Calif. Inst. Technol. (1983).
17. Abe, K. & Kanamori, H. *J. geophys. Res.* **84**, 3589–3592 (1984).
18. Fischer, K. M., Jordan, J. H. & Creager, K. C. *J. geophys. Res.* **93**, 4773–4783 (1988).
19. Vidale, J. E. *Geophys. Res. Lett.* **14**, 542–545 (1987).
20. Vidale, J. E. & Garcia-Gonzalez, D. *Geophys. Res. Lett.* **15**, 369–372 (1988).
21. Tanimoto, T. *Geophys. J.* (in the press).
22. Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
23. Hager, B. H. & Clayton, R. W. in *Mantle Convection* (ed. Peltier, W. R.) (in the press).
24. Gurnis, M. & Davies, G. F. *Geophys. Res. Lett.* **13**, 541–544 (1986).
25. Jarrard, R. D. *Rev. Geophys. Space Phys.* 217–284 (1986).
26. Karig, D. E. *Geol. Soc. Am. Bull.* **82**, 323–344 (1971).

Crystal structure of *trp* repressor/operator complex at atomic resolution

Z. Otwinowski, R. W. Schevitz*, R.-G. Zhang†, C. L. Lawson*, A. Joachimiak*, R. Q. Marmorstein‡, B. F. Luisi & P. B. Sigler§

Department of Biochemistry and Molecular Biology, and ‡ Department of Chemistry, University of Chicago, 920 East 58 Street, Chicago, Illinois 60637, USA

The crystal structure of the trp *repressor/operator complex shows an extensive contact surface, including 24 direct and 6 solvent-mediated hydrogen bonds to the phosphate groups of the DNA. There are no direct hydrogen bonds or non-polar contacts to the bases that can explain the repressor's specificity for the operator sequence. Rather, the sequence seems to be recognized indirectly through its effects on the geometry of the phosphate backbone, which in turn permits the formation of a stable interface. Water-mediated polar contacts to the bases also appear to contribute part of the specificity.*

HERE we describe the atomic details of a specific protein/DNA interface observed in the crystal structure of the *trp* repressor/operator complex refined to 2.4-Å resolution. The formation of this complex depends on the binding of the corepressor ligand, L-tryptophan, and represents the main mechanism for the transcriptional control of L-tryptophan levels in enteric bacteria[1]. The crystal structure clearly shows the details of both direct and solvent-mediated contacts between the protein and DNA. These interactions and a substantial solvent-excluded contact surface are created by structural adjustments in both the DNA and protein which permit a far more intimate fit than anticipated from docking studies[2]. The crystal structure also shows that, in addition to the corepressor's positive allosteric effect on the conformation of the DNA-binding domains[3], the bound tryptophan itself and four of the surrounding side chains of its binding site are positioned to form hydrogen bonds to the operator to a degree not expected from modelling.

The most striking finding is that the crystal structure shows no direct hydrogen-bonded or non-polar contacts with the bases that could explain in simple chemical terms the repressor's preferential binding to the operator sequence. Instead, the direct hydrogen-bonded contacts are mainly to the operator's phosphate groups. In each complex there are water-mediated hydrogen bonds to six bases that have been shown *in vivo*[4] to be important in specifying the operator as the target sequence for the *trp* repressor. Specificity in the *trp* system does not rely on a simple complementarity between the repressor's amino-acid side chains and the operator's bases, but rather depends on a combination of solvent-mediated interactions with critical bases and on still poorly understood principles by which the base sequence restrains the conformation of B-form DNA.

Structure determination and refinement

The structure was determined from the monoclinic crystals described by Joachimiak *et al.*[5]; ($P2_1$, two complexes/asymmetric unit). The crystals required about a year to grow from 35% 2,4-dimethyl pentanediol, 50 mM NaCl, 11 mM $CaCl_2$, 10 mM cacodylate, *p*H 6.8. The operator was simulated by a symmetrical 18-base-pair duplex with overhanging 5′ T residues. Data were collected at room temperature with rotation photographs and area detectors. All photographic data were obtained from monochromatic synchrotron X-ray sources, either at the Stanford Synchrotron Radiation Laboratory (the parent data, $\lambda = 1.54$ Å) or at the Cornell High Energy Synchrotron Source (the $PbCl_2$ derivative data, $\lambda = 1.57$ Å). Synchrotron radiation was essential to obtain accurate data beyond 3.0 Å. Data from the area detector facility at the University of California, San

* Present addresses: Lilly Research Laboratories, Indianapolis, Indiana 46285, USA (R.W.S.); Department of Physical Chemistry, N. Rijksuniversitat, Groningen, The Netherlands (C.L.L.); Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland (A.J.).
† Permanent address: Chinese Academy of Sciences, Institute of Biochemistry, Shanghai, Republic of China.
§ To whom correspondence should be addressed.

Table 1 Statistics of data collection and refinement

Resmax	Predicted*	Refined†	% Coverage‡	⟨I⟩§	⟨Error⟩‖	% Weak F¶	R-sym*	R-factor**	Phasing power $PbCl_2$ derivative††
10.0-4.80	2,976	2,945	99.0	4,641	194	0.5	0.07	0.287	1.5
4.80-3.81	3,344	3,320	99.3	3,555	227	0.7	0.11	0.227	1.0
3.81-3.33	3,286	3,239	98.6	2,467	185	1.3	0.13	0.227	0.8
3.33-3.02	3,322	3,207	96.5	1,465	142	3.0	0.15	0.249	
3.02-2.81	3,291	3,001	91.2	583	83	7.2	0.20	0.270	
2.81-2.64	3,283	2,683	81.7	461	82	12.3	0.23	0.259	
2.64-2.51	3,296	2,448	74.3	331	78	18.3	0.29	0.243	
2.51-2.40	3,252	1,895	58.3	236	72	28.4	0.31	0.230	
Total	26,050	22,726	87.3	1,817	137	7.5	0.11	0.249	

* The number of possible reflections in each resolution shell.

† The number of reflections used in refinement.

‡ The extent of coverage of the data used in refinement, expressed as the percentage of the number predicted.

§ The average intensity, expressed in arbitrary units.

‖ The average error.

¶ The number of observations with $|F|$ less than two standard deviations.

* $\sum_h \langle|I(h') - \bar{I}(h)|\rangle / \sum_h \bar{I}(h)$ where $\langle|I(h') - \bar{I}(h)|\rangle$ is the average absolute deviation of equivalent (symmetry and/or Friedel) reflections $I(h')$ from the average $I(h)$. The largest component to R-sym was the crystal-to-crystal variation. On average, each reflection was measured 4.2 times. Eight crystals were used.

** $\sum_h ||F^o(h)| - |F^c(h)|| / \sum_h |F^o(h)|$, where $|F^o|$ and $|F^c|$ are observed and calculated structure factor amplitudes respectively.

†† There were 10,595 unique reflections to 2.8-Å resolution from a single crystal giving 60% coverage. The R-sym is 5%. The r.m.s. difference between the parent and $PbCl_2$ structure amplitude was 28%. Phasing power of the $PbCl_2$ derivative, expressed as $\sum_h |F^c_H|_h / \sum_h [||F^o_P| \exp(i\phi) + F^c_H| - |F^o_{PH}|]_h$, where $|F^o_P|$ and $|F^o_{PH}|$ are the observed amplitudes of the parent and heavy-atom derivative structure factors respectively, ϕ is the phase of the parent structure factor, and $|F^c_H|$ is the structure amplitude of the heavy atoms alone. The denominator combines the experimental and lack-of-isomorphism error. The latter was the larger component. Derivative data were extended to 2.6 Å, but useful phasing power fell sharply beyond 3 Å.

Diego supplemented the synchrotron parent data at lower resolution (to 3.2 Å). Cooling the crystals to as low as −15 °C provided no significant improvement in lifetime or resolution. A high quality data set was obtained for the parent structure to a resolution of 2.4 Å, although many films produced accurate data to beyond 2.2 Å. There was a modest anisotropy in the intensity distribution that reflected a strong underlying molecular transform in the direction of the DNA helical axis. Table 1 summarizes the data.

The structure was solved mainly by molecular replacement using a modified version of Fitzgerald's program package, MERLOT[6], applied to the parent data between 8.0-Å and 3.6-Å resolution. The most effective search structure was a variant of the *trp* repressor crystal structure, in which the potential for error in the search structure was diminished. The fact that the two crystal forms of the free *trp* repressor differ significantly in the conformation of their DNA-binding domains[7], as well as many solvent-exposed side chains, suggested that in the complex these elements could deviate from both crystal structures. The structure factors of the search structure were, therefore, calculated from a composite model where the atoms were placed midway between those of the trigonal and orthorhombic repressor-crystal forms. An atom's contribution to the calculated structure factor was weighted down in proportion to the difference ε_i between the atom's position in the two crystal structures. This was done in a resolution-dependent manner by applying to the i^{th} atomic scattering factor a pseudo-thermal term $\exp(-8\pi\,\varepsilon_i^2 \sin^2\theta/\lambda^2)$. Second, and more important, by subtracting the origin peak from the observed Patterson, we eliminated the effect of the origin term in the calculated component of the translation function. The origin term of the calculated Patterson and its contribution to the translation function increases linearly with the degree of molecular overlap. As correctly placed molecules will never overlap, and incorrectly placed molecules will usually overlap, failure to exclude the origin of the Patterson can significantly raise the noise level of the translation function and obscure the correct solution. With these modifications, the two solutions to the rotation function (first and fifth highest peaks) were 7.9 and 3.9 times the root-mean-square (r.m.s.) value of the map. The solutions to the translation function were 9.8, 8.1 and 9.6 (cross-translation) times the r.m.s. value. These were the highest peaks in their respective maps and were self-consistent.

A difference map comparing a $PbCl_2$ derivative and the parent, phased with the molecular replacement solution, showed heavy-atom sites consistent with the non-crystallographic symmetry elements. Phases calculated from the coordinates of the two *trp* repressor molecules, combined with the phases from the $PbCl_2$ derivative, produced a Fourier map that clearly showed an operator interfaced to each repressor. This initial map (3.0 Å, the limit of the isomorphous replacement phasing was 3.3 Å) was used to build the operator structures and to make the initial adjustments to the repressor model using the computer-graphics program FRODO[8].

The transformation relating the two independent repressor/operator complexes in the asymmetric unit was established by superimposing the two *trp* repressor molecules found by molecular replacement. We constrained this procedure to superimpose both repressor dyads. This provided the symmetry operations needed to average the electron density corresponding to the four crystallographically independent protomers, namely the four half operator/half repressors. After each stage of refinement the symmetry operators were reassessed with the same procedure using the backbone coordinates of the conformationally invariant 'central core' (residues 18-60)[7]. Remodelling proceeded in two modes; first the model was fitted to the unambiguous portions of the fourfold average electron density. The model was then adjusted to fit the unaveraged density so that local variations would be preserved. In the latter stages of the refinement, the model was fitted mainly to the unaveraged difference maps.

The protein and nucleic acid components of the model were simultaneously refined with a fast Fourier version of a restrained least-squares procedure that combined the programs of Hendrickson and Konnert[9], Agarwal[10], Finzel[11] and Westhof[12]. Bond lengths and inter-bond angles were restrained to be, on average, within one standard deviation of their literature values. There were no stereochemical restraints placed on the protein/DNA

interactions, on the relative orientation of the bases, or on any hydrogen bond in the entire complex. Table 1 summarizes the refinement.

Local symmetry implies validity

As the crystallographic asymmetric unit contains two potentially dyad-symmetrical repressor/operator complexes, the crystal structure presents four independent images of the half-repressor/half-operator complex. Except for the amino-terminal arms (residues 2 through 14) and the carboxyl-terminal three residues (106–108), these four protomers are superimposable with an r.m.s. deviation of backbone atoms from their average positions of 0.4 Å. The same degree of similarity applies to the operator structures, as well as to most of the side chains. Unless otherwise stated, the results presented are based on a model that represents the average of the four independent half-repressor/half-operator protomers. To the extent that the four protomers have the same conformation, the structures are free of fortuitous distortions imposed by crystal-packing and are therefore functionally valid representations defined by the intrinsic properties of the complex.

Structure of the repressor

When the repressor structures from the trigonal or orthorhombic crystal forms are superimposed on the repressor structure in the crystalline complex, the conformation in the central core is well conserved. However, the flexible reading heads of the complex showed significant deviations from both crystal forms of the repressor (Fig. 1). This was not unexpected because Lawson *et al.*[7] have pointed out that the reading heads of the *trp* repressor are flexible, that is, activation of aporepressor to repressor by the binding of L-tryptophan does not impart a rigidly unique structure to the DNA-binding domains of the repressor.

The side chains on the surface of the free repressors are typically mobile. In the repressor/operator complex, however, the temperature factors for the atoms in the side chains that contact the DNA are reduced to the values of the main chain or even lower. Thus, the mechanics of complex formation appear to include the meshing of flexible backbone elements and side chains to form a rigid complementary surface.

The N-terminal 'arms' of the repressor (residues 2 to 14) are poorly ordered and, to the extent that they can be modelled and refined, the four representations in the asymmetric unit assume unrelated conformations defined by interactions with neighbouring protein and DNA molecules and not with their cognate operators. This ill-defined conformation of the amino-terminal residues agrees with the observation made in the crystal structures of the *trp* aporepressor[3], pseudorepressor[13], and two different crystal forms of the repressor[2,7], that the conformations of the N-terminal 11 amino acids were largely determined by

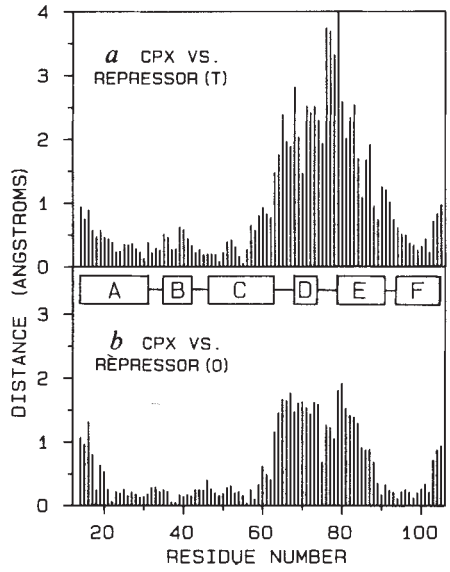

Fig. 1 The *trp* repressor in the complex compared to uncomplexed repressors. The model of the *trp* repressor in the complex was superimposed on models of the *trp* repressor in the trigonal crystal form[2] (top panel) and in the orthorhombic crystal form[7] (bottom panel) so as to minimize the r.m.s. deviation of the corresponding α-carbon atoms in the 'core' of *trp* repressor dimer. The ordinate is the deviation of corresponding α-carbon atoms after superimposition. The boxes *A* through *F* represent α-helical segments. Helices *D* and *E* comprise the helix-turn-helix motif (reading head).

fortuitous packing contacts. The behaviour of this amino-terminal segment reflects the fact that residues 2–15 project out from the surface of the repressor's globular fold.

Carey has shown that chymotryptic removal of amino-acid residues 2 to 7 of both subunits reduces affinity for both operator and nonspecific DNA by between 30- and 100-fold (private communication). A role in DNA binding has been attributed to the presumably flexible N-terminal and C-terminal arms of *cI* and *cro* repressors respectively[14,15]. The failure of the N-terminal segment to form part of the repressor/operator interface in the crystal may reflect an important difference between the natural interactions and those in the crystal. In the crystal, unlike the operator in its natural setting, DNA fragments are densely packed around the cognate operator offering alternative non-specific binding sites that can be easily reached by the flexible arms. The N-terminal arms might also be involved in protein/protein contacts which stabilize the postulated binding of tandem repressors to overlapping operator sequences in both the *trpEDCBA* and *aroH* operons[16].

Table 2 DNA orientation parameters

Base	Roll (degrees)	Tilt (degrees)	Slide (Å)	Helical twist (degrees)
G_{-9}/C_{+9}	3.5	−0.8	0.4	24.2
T_{-8}/A_{+8}	−1.5	−2.7	0.2	42.8
A_{-7}/T_{+7}	−2.4	−2.0	1.0	32.1
C_{-6}/G_{+6}	2.6	2.9	0.8	31.8
T_{-5}/A_{+5}	10.9	2.3	1.0	31.2
A_{-4}/T_{+4}	6.0	0.4	0.2	35.8
G_{-3}/C_{+3}	−1.4	2.5	0.9	35.3
T_{-2}/A_{+2}	1.6	2.9	−0.3	38.7
T_{-1}/A_{+1}	8.8	0.0	−0.8	31.2
A_{+1}/T_{-1}				

All parameters were calculated from the fourfold averaged operator with the programs BROLL and CYLIN (R. E. Dickerson, personal communication). The movements are defined in Fig. 3*c*.

Structure of the operator

Table 2 lists the orientation parameters (defined in Fig. 3*c*) for each base pair or each step between base pairs. The corresponding parameters deduced from the NMR study[17] of a similar symmetrical operator fragment in solution correlate poorly with those of the operator in the complex. The individual deviations of these local parameters from the average are modest in the crystalline complex—indeed, they are smaller than those seen in crystalline uncomplexed DNA. For example, the r.m.s. (and extreme) deviation of the roll and tilt angles for a symmetrical 12-base-pair B-DNA fragment are 5.1° (9.8°) and 1.7° (3.5°) respectively[18]. The corresponding values for the *trp* operator fragment in the complex are 4.3° (8.1°) and 2.2° (2.9°). The operator has an average of 10.6 base pairs per turn, with only modest deviations in twist angle (Table 2). The minor groove

is wider in the *trp* operator (5.5 to 8.1 Å) than in the crystalline dodecamer (3.1 to 7.2 Å) (ref. 18). Its widest point is opposite the parts of the major groove that interact with the reading heads of the repressor. There is also a slight widening of the minor groove about the dyad axis. The range of propeller twists (1.2° to 15.3°) is too small to produce non-Watson-Crick hydrogen bonds between strands. The small local deviations of the operator suggest that there are no focal points for forces distorting the operator, and that the strains of any distortion are distributed.

Figure 3*b* shows a polar projection of the unit vectors normal to the base pairs. The convex bend in the operator (Fig. 2*a*) is composed of three straight segments indicated by the three clusters in Fig. 3*b*. The three segments arise from two symmetrically disposed bends centred eight base pairs apart between position −5 and −4 and between position 4 and 5. This separation is less than the helical periodicity of the DNA; therefore, the molecule will bend in two planes. As a consequence, the DNA appears to be slightly S-shaped when viewed down the molecular dyad. When viewed along the perpendicular to this dyad as in Fig. 2*a*, the DNA appears C-shaped and follows the protein surface.

The repressor (as seen in the complex) was docked to helically uniform B-DNA. The canonical DNA of this modelled complex was used as a reference point to characterize the deviation of the operator in the complex (Fig. 2*b*). The atoms comprising bases −5 to +5 superimpose well, with an r.m.s. deviation of 1.4 Å. The phosphate oxygens associated with these bases and contacting the protein superimpose less well, with an r.m.s. deviation of 3.5 Å. In contrast, nucleotides −6 to −9 and +6 to +9 show substantially larger differences in which the phosphates that contact the protein, −7, −8 and −9, deviate by 6.7, 9.1 and 9.6 Å respectively.

The large deviations of the operator from uniform B-DNA are necessary to obtain the 24 phosphate-protein contacts seen in the complex (Fig. 2*a*). These shifts are due to the cumulative effect of modest local distortions. The largest values of slide and roll occur in the base-pair steps of the sequence ACTAG (from −7 to −3 and its symmetry equivalent), a region that is also the most sensitive to mutations (Fig. 3*a*). These parameters explain the salient structural features of the DNA. The positive roll in steps T-A and A-G bends the DNA. The steps A-C, C-T and T-A each has a slide of nearly 1 Å, and this run of high slide and roll causes the large displacement of the last three base pairs at both ends of the operator noted above. Thus the segment of operator that carries the strongest sequence determinants for repressor affinity is the segment that exhibits the largest deviations in local orientational parameters and produces the most distinct conformational variation in the repressor/operator complex.

Stabilizing interactions

Hydrogen bonds. There are 14 direct hydrogen bonds between the half-repressor/half-operator (Figs 2*a*, 4 and 5). All but two involve the unesterified oxygens of six phosphate groups. The exceptions are the two hydrogen bonds in the interaction of Arg 69 with G_{-9} which represent the only direct hydrogen bonds to the functional groups of a base. But these hydrogen bonds are relatively unimportant in specifying the target DNA sequence, for symmetrical mutations of GC_{+9}/CG_{-9} have a negligible effect on repressor binding *in vivo*[4]. Four of the protomer's 14 direct contacts to the half-operator are made by amino-acid residues that are not part of the helix-turn-helix motif. The absence of direct hydrogen bonds to mutationally sensitive base pairs implies that, in the *trp* system, specificity is not due to direct hydrogen-bonded contact of the major groove's polar groups by the hydrogen bonding groups of the protein[19].

Six additional solvent-mediated contacts are evident, three to the phosphates and three to the base pairs. One such contact probably involves a bridging calcium ion between the side chain of Asp 46 and phosphate 1. Two water molecules are positioned to form hydrogen bonds with the repressor and the two unesterified oxygens of phosphate 3; one involves Oγ of Thr 53 and the other Oγ of Thr 81. Phosphate 3 is thereby the focus of four hydrogen-bonded interactions, two mediated by the solvent and at least two directly from the protein (Arg 54 and the ring nitrogen of the bound tryptophan).

There are three well-ordered water molecules that mediate hydrogen-bonded interactions between the half-repressor and the base pairs of the half-operator. Figure 5*c–e* shows how two of the water molecules bridge the peptide nitrogens of Ile 79 and Ala 80 at the amino terminus of the E-helix to the hydrogen-bond acceptors of the purines A_5 and G_6. The water that bridges N79 to N7 of G_6 appears to be firmly positioned by an additional interaction with the N_ζ of Lys 72. This arrangement of hydrogen bonds channels the positive potential of the E-helical dipole to the bases that face the tip of the helix-turn-helix. The third water molecule links the γ-hydroxyl of Thr 83 to N6 and N7 of A_7. Figure 3 shows that the bases involved in these water-mediated interactions are among the most important in specifying the repressor's affinity for the operator sequence.

Non-polar contacts. The α-carbon of Gly 78, the fourth residue of the turn in the helix-turn-helix, is close enough to the hydrophobic surface of the DNA (C6 of T_4 and C8 of A_5; Fig. 5*e*) to exclude water; however, this contact has no discriminatory function because other bases can be accommodated without steric clashes. Furthermore, the region of the operator that opposes the α-C of Gly 78 is hydrophobic, irrespective of the DNA sequence. The contact is close enough, however, to explain why amino acids other than glycine should disrupt repressor binding[20]. The C_γ of Thr 83 forms a non-polar contact to the C2′ of A_{-7}. Interestingly, the non-polar side chains of Ile 79 and Ala 80, positions postulated to be important in DNA recognition[21], do not make extensive contact with non-polar surfaces of the operator. Instead, they face mainly the hydrogen-bonded functional groups of base pairs ±6 and ±7 (Fig. 5*c*). The $C_{\gamma 1}$ (methylene) of Ile 79 does contact the C5 methyl of T_8, but apparently does not serve a discriminatory function because this position is not conserved in the three operators of the *trp* system.

The role of bound L-tryptophan. By comparing the high-resolution crystal structures of the unliganded aporepressor with that of the liganded active repressor, Zhang *et al.*[3] have shown that the main effect of the corepressor, L-tryptophan, is to orient the flexible reading heads so that the helix-turn-helix motif can penetrate successive major grooves of the operator. Figures 2*a* and 5*f* and *g* show that, in addition to the 'allosteric' effect on the protein's conformation, bound L-tryptophan and the neighbouring residues of the pocket interact directly with DNA. The indole ring nitrogen of L-tryptophan forms a hydrogen bond to phosphate 3. The 'floor' of the indole ring's binding pocket is formed by the proximal methylene groups of Arg 54 whose guanidino group makes a hydrogen bond (possibly two) with phosphate 3. L-tryptophan's α-carboxylate makes a bifurcated hydrogen bond which positions the guanidino group of Arg 84 so that it interacts with phosphate 2. The corepressor's role in forming these contacts depends on an unusual binding mode, in which tryptophan's aliphatic side chain is maintained under torsional strain by strong contacts with the protein[13,22]. These contacts, combined with the direct interactions of the ligand and its binding site with the operator, cause the bound tryptophan to have the lowest thermal parameters in the entire structure; thus, the corepressor is the most firmly fixed element in the complex.

Solvent excluded interface. The flexibility of both the DNA-binding domains and the operator permit a moulding of the contact surface that makes 2,900 Å^2 inaccessible to solvent at the interface of the *trp* repressor/operator complex. This large solvent-excluded surface is typical of subunit interfaces in hetero-oligomeric proteins where structural adjustments are made on aggregation[23]. Nearly all the polar groups that become

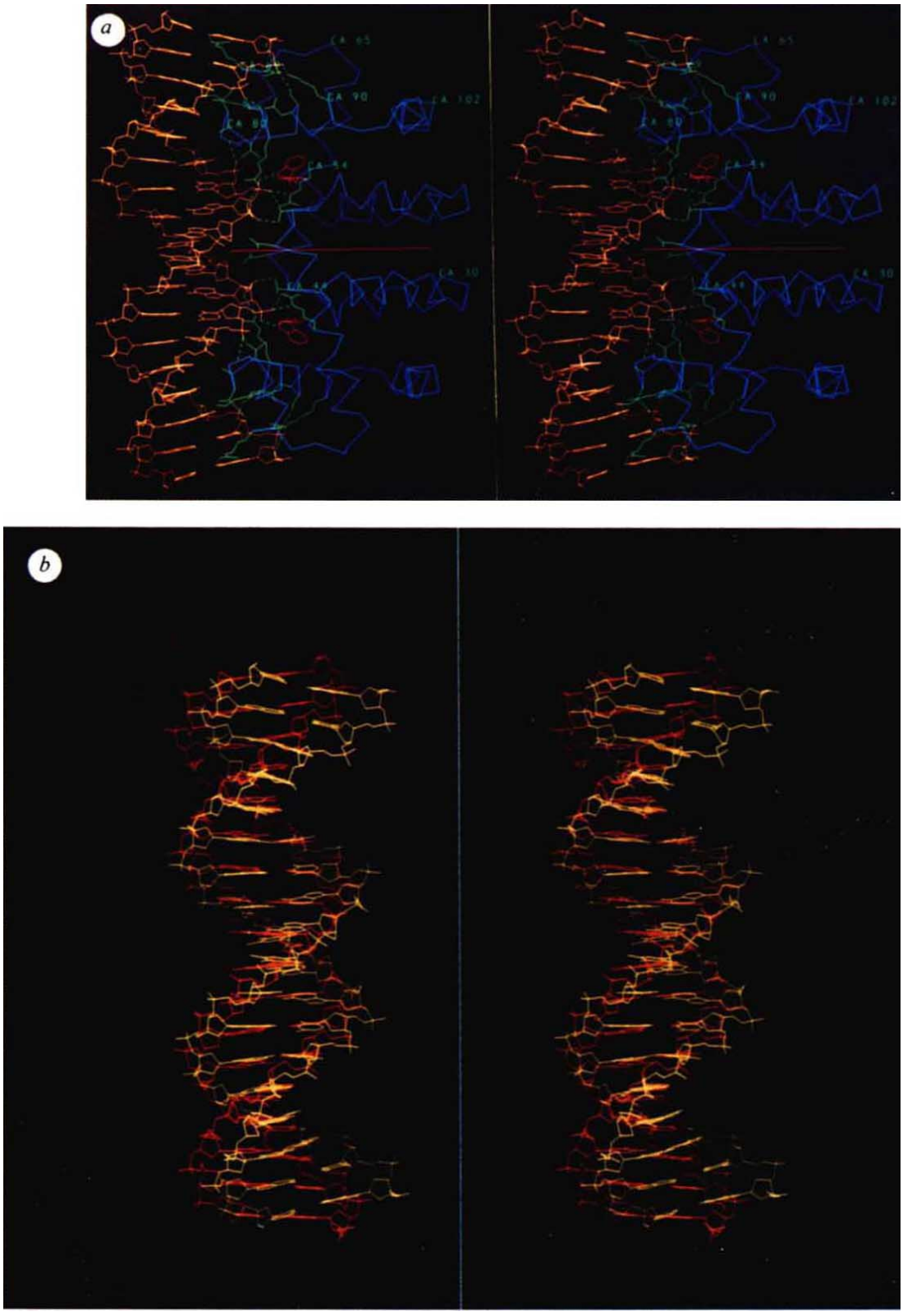

Fig. 2 The *trp* repressor/operator complex. *a*, The crystal structure. An α-carbon trace of the repressor backbone (blue) plus the side chains (green) that make direct hydrogen bonds (dashed line) to the operator (gold). *b*, The DNA structure. Helically uniform B-DNA (orange) is contrasted to the DNA as seen in the complex (gold).

solvent-inaccessible are hydrogen-bonded and no phosphate group is completely buried by protein. Assembly of protein subunits is favoured by the entropic effect of mobilizing water from the buried surfaces, which more than compensates for the restrictions imposed by stable tertiary and quaternary structures[24]. Record[25] has noted that mobilization of water is also one of the main contributors to the stabilization of specific protein/nucleic-acid complexes. The release of water would be expected to contribute a large entropic component to the stabilizing free energy of the *trp* repressor/operator interaction in view of the substantial solvent-excluded surface seen in the crystalline complex. The stereochemical adjustment of both the protein and nucleic-acid components of the complex appear designed to exploit this mode of stabilization.

***trpR* mutations.** Changes in amino-acid side chains can diminish operator affinity in three ways; by depriving the complex of an important stabilizing interaction, by introducing a side chain that sterically or electrostatically disturbs the contact surface,

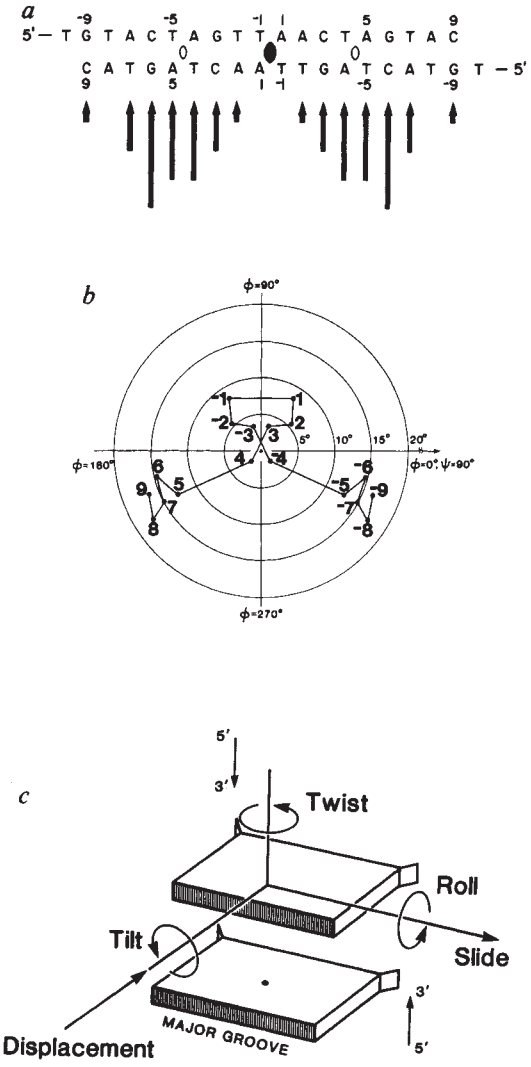

Fig. 3 Base-pair orientations. *a*, The sequence of the duplex DNA fragment in the co-crystals that simulates the operator. Symmetry was achieved by changing the *trpEDCBA* operator sequence from $(A\cdot T)_{-8}$ to $(T\cdot A)_{-8}$ to match $(A\cdot T)_{+8}$. This does not affect repressor binding[4]. Numbering of base pairs is from 5′ to 3′ on the top strand from an origin at the twofold rotational axis of symmetry, so the dyad axis is between −1 and +1. The numbering of nucleotides on the bottom strand retains dyad symmetry. The arrows indicate the nucleotides where mutations diminish repressor affinity. The size of the arrow is a rough measure of the base pair's importance in specifying repressor affinity, as deduced from the preservation of the nucleotides in the three operator sequences of *Escherichia coli* (*trp EDCBA*, *trpR* and *aroH*) and the directed mutational studies of Bass *et al.*[4] For example, any symmetrical substitution for $(G\cdot C)_6$ or $(C\cdot G)_{-6}$ totally disrupts operator binding in the assay used by Bass *et al.* *b*, A polar projection (on to the equatorial plane) of unit vectors normal to each base pair's average plane. Base pairs are designated by the number of the base in the top strand of *a*. Coordinates are the latitude (polar angle ψ) measured from the pole, and longitude ϕ measured counterclockwise from the complex's dyad (which has the coordinates $\phi = 0°$, $\psi = 90°$). The pole represents the overall axis of the DNA. A straight segment of B-DNA produces a cluster around the segment's axis. An abrupt bend forms a separate new cluster. *c*, Definition of relative base-pair orientational angles and 'slide' used in Table 2. The arrows indicate the positive sense. The angles describe the change in orientations proceeding from one base pair to the next (a 'step'). Slide and displacement refer to a movement in relation to the 'local' helical axis defined by a base pair and its immediate neighbour. The orthogonal coordinate basis is defined by a local normal to the average plane of the two adjacent base pairs of the step, and the bisector of the base pair to be rotated.

or by disrupting the structure of the repressor. Examination of the crystalline complex enables us to explain all the negative missense *trpR* mutations[20] in terms of one or more of these effects. Deleterious mutational changes have been shown to occur at most of the positions in contact with the operator. Figure 4*c* shows the correspondence between amino-acid sequence positions in the helix-turn-helix, where deleterious missense mutational changes occur, and the amino-acid side chains that contact the operator.

Electrostatic attraction. Warwicker (personal communication) calculated the electrostatic charge potential surrounding the repressor. The positive charge potential distribution clearly follows the contours of B-DNA, suggesting that electrostatic field effects may contribute significantly to interaction with the operator. This view is supported by the observation that mutations such as Glu 18 to Lys increase repression[20], even though the mutationally inserted lysine side chains cannot possibly contact the operator. We suggest that the tight fit of the specific complex strengthens this general electrostatic attraction by diminishing the electrostatic screening effect of the solvent.

Tandem binding. Kumamoto *et al.*[16] proposed, on the basis of methylation-protection and footprinting experiments, that one *trp* repressor dimer binds the *trp*R operator and that more than one repressor might bind to overlapping operator-like sequences in both the *trp EDCBA* and *aroH* operons. The relationship of the tandemly overlapping repressors can be best described as a strict twofold rotation of the entire repressor around an axis placed between base pairs 4 and 5, as shown by the open sysmbols in Fig. 3*a*. A 180° rotation of the entire repressor around this axis (with no further adjustments) places the 'second' repressor in a position that preserves the protein-phosphate contacts described above. In this model, no stereochemical clashes occur that could prevent this tandem binding.

Discussion

Different B-DNA sequences have characteristic distributions of hydrogen-bond donor and acceptor groups on the bases, but rather similar sugar-phosphate backbones. This fact has led to a postulate that specific affinity of proteins for their DNA targets arises from complementary protein-DNA hydrogen bonding through groups in the major and minor grooves of the nucleic acid[19]. This 'direct-readout' mechanism for recognition can be extended to include polar contacts mediated through ordered water molecules.

A second principle, called 'indirect readout', relies on the increasingly recognized fact that the B-DNA exhibits a large degree of sequence-dependent structural variation[26-29]. Arguments stemming from this observation have been used to explain the positioning of DNA in nucleosomes[30,31,33], the specificity of nucleases[33,34], the bending of DNA[35-41] and the preferential binding of *lac* repressor to poly(dA·dT) (ref. 42). In the indirect readout mechanism, the primary function of the DNA sequence is to permit a conformation that substantially enhances attractive interaction with its cognate protein. Direct and indirect readouts do not invoke mutually exclusive roles for the base sequence, and clearly both principles can confer specificity to various degrees in any particular complex.

Two unrefined crystal structures of specific protein/DNA complexes have been reported at resolutions that provide some insight into the stereochemistry of specific protein/DNA interactions. Rosenberg and his colleagues pointed out that significant protein-induced DNA-deformation, as well as potential hydrogen-bonding interactions with the functional groups of the major groove, are important in forming the complex between the *Eco*RI restriction endonuclease and its six-base-pair substrate[43]. Harrison and his colleagues suggested that in the crystalline complex of *p434* repressor and its 14-base-pair operator, the non-contacted bases at the centre of the operator may govern the degree of helical twist needed to accommodate both polar and

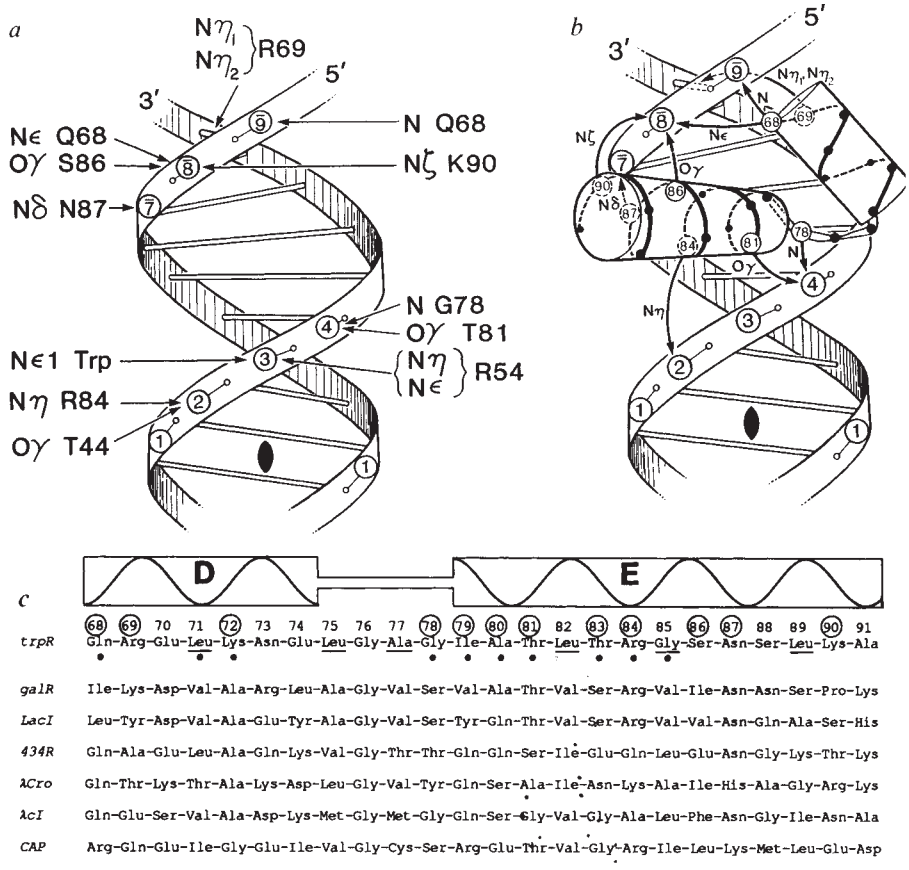

Fig. 4 Residues involved in direct contacts between *trp* repressor and operator. *a*, Schematic view of the upper half of *trp* operator as 'viewed' by the *trp* repressor showing operator's phosphates (circles with numbers) and the repressor's functional groups (arrows) that form direct hydrogen bonds. *b*, Schematic view of the amino-acid positions (circled amino-acid numbers) in the helix-turn-helix that make direct hydrogen bonds to the operator (shown by arrows) as seen in Figs 1 and 4*a*. The figure emphasizes that the second helix, or E-helix, of the bihelical motif is 'pointed into' the major groove with its α-helical axis almost perpendicular to the global DNA axis. The E-helix does not 'lie in' the major groove with its axis parallel to the groove. *c*, The amino-acid sequence of the helix-turn-helix. The circled residues make direct contact to the operator. The dots indicate residues which when altered by mutation significantly diminish repression[20]. The underlined residues are involved in stabilizing the conformation of the DNA-binding domain. The sequences of other bacterial regulatory proteins are shown for comparison[27].

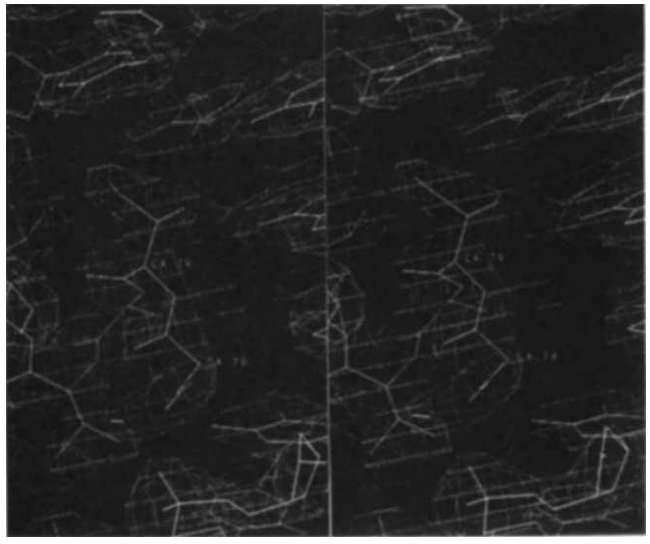

Fig. 6 Electron density map. A Fourier synthesis containing coefficients $[m(r|F_o| - (r-1)|F_c|)\exp(\phi)]_h$, where ϕ is the centroid phase and m the figure-of-merit of a joint-phase probability distribution with contributions from both the refined model and the $PbCl_2$ derivative. The scaling term r is ~2.5 and adjusts the relative contributions to minimize model bias in the maps. The map is averaged fourfold as described in the text. The model shows the amino-terminal end of the E-helix deep in the operator's major groove. Note the density for a water molecule near N80.

non-polar stabilizing contacts[44,45]. The graded affinity for tandem operators is at the heart of the biological role of the repressors in this family of temperate phages, and it is mainly the central base pairs that distinguish the operator sequences.

In this study of the *trp* repressor/operator complex, we have sufficient resolution, accuracy and redundancy to state that there are no direct hydrogen bonds between the functional groups of the bases and the protein that can account for the specific affinity of the *trp* repressor for the operator sequence. The only direct hydrogen bonds within the major groove are between Arg 69 and G_{-9}; however, for reasons pointed out earlier, this interaction contributes little to the specificity. Similarly, there are also no van der Waals' contacts between the non-polar side chains of the protein and the non-polar surfaces of the bases that can account for specificity. In fact, the contact surface does not appear to be able to exclude non-operator DNA sequences by steric interference, such as that produced by a snug contact with a C5 of cytosine that would clash with a 5-methyl group were it replaced by a T. It should be emphasized that our failure to observe direct protein contacts with the contours of the major groove does not arise from our inability to see the side chains in the electron density. With the exception of the N-terminal arms, all the amino acids that could conceivably contact the DNA bases have refined robustly with a high degree of positional certainty and local symmetry. Figure 6 illustrates the typically strong electron density for the side chains of the contact surface.

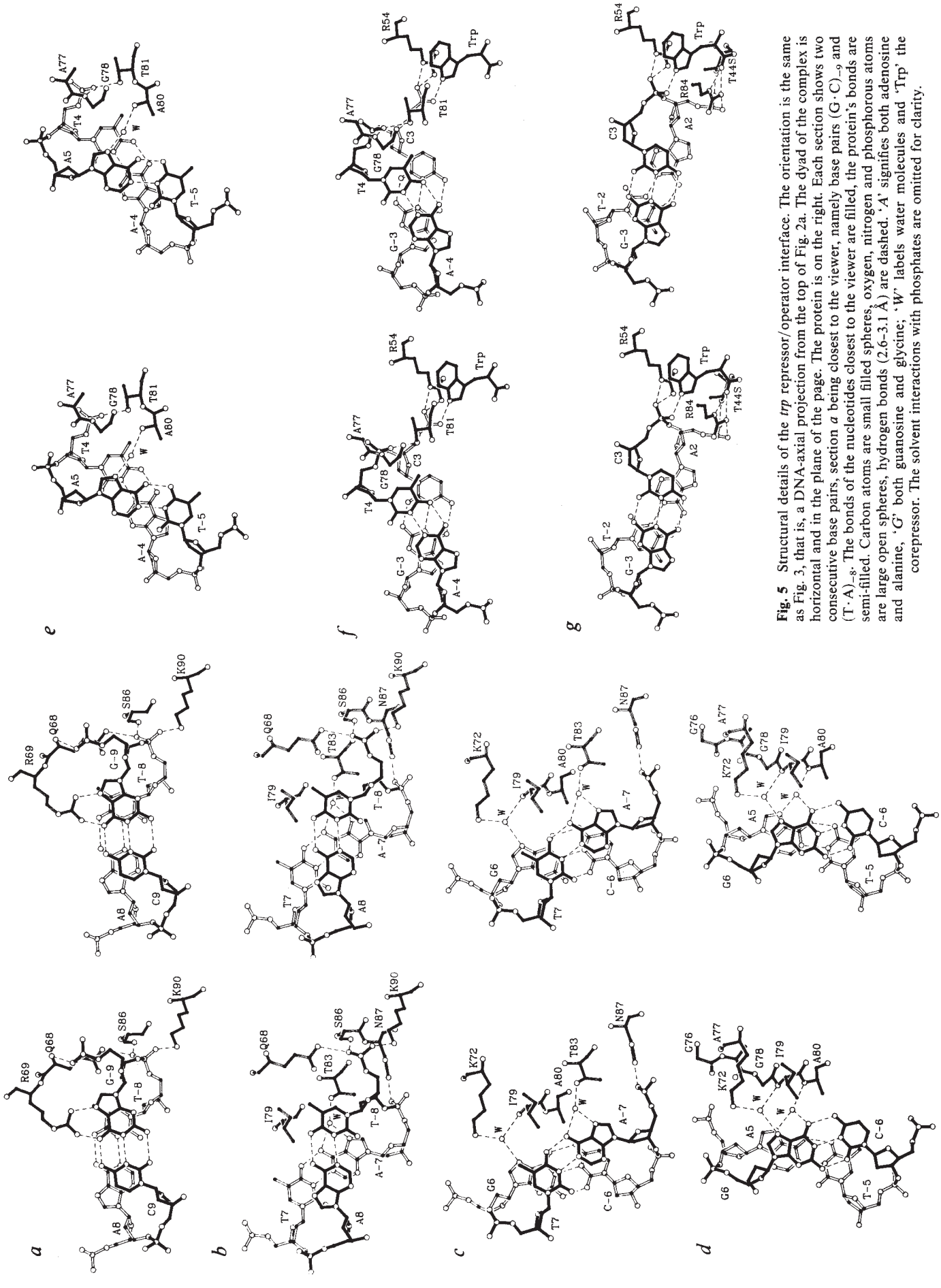

Fig. 5 Structural details of the *trp* repressor/operator interface. The orientation is the same as Fig. 3, that is, a DNA-axial projection from the top of Fig. 2a. The dyad of the complex is horizontal and in the plane of the page. The protein is on the right. Each section shows two consecutive base pairs, section *a* being closest to the viewer, namely base pairs $(G \cdot C)_{-9}$ and $(T \cdot A)_{-8}$. The bonds of the nucleotides closest to the viewer are filled, the protein's bonds are semi-filled. Carbon atoms are small filled spheres, oxygen, nitrogen and phosphorous atoms are large open spheres, hydrogen bonds (2.6–3.1 Å) are dashed. '*A*' signifies both adenosine and alanine, '*G*' both guanosine and glycine; '*W*' labels water molecules and 'Trp' the corepressor. The solvent interactions with phosphates are omitted for clarity.

We have, however, identified three polar contacts mediated by well ordered water molecules that bridge the functional groups of mutationally sensitive base pairs[4] to the protein and thereby contribute to 'direct readout' of the operator sequence (Fig. 5*b–d*). Two of the water molecules (contacting N79 and N80) are compatible only with a G at position 6 and a purine at position 5. These restrictions are completely consistent with the result of a directed mutational analysis[4], in which symmetrical substitution of any base for G at position 6 or of a pyrimidine at position 5 maximally reduced *in vivo* operator affinity. The pattern of mutational effects at position 7 can be similarly rationalized by the solvent-mediated contacts between A_7 and Thr 83. Moreover, repression is diminished by *trpR* mutations that change the residues whose side chains are involved in these water-mediated interactions, Lys 72 (S. Bass, V. Sorrells and P. Youderian, private communication) and Thr 83 (ref. 20). It should be noted that water-mediated contacts cannot explain the importance of base pairs 3 and 4. Moreover, there is no assurance that alternative patterns of water-mediated contacts could not serve to bridge non-operator sequences to the repressor. What then accounts for the specificity of the *trp* repressor/operator complex?

We believe that in the *trp* repressor/operator system a primary mode of sequence specific recognition is 'indirect readout'; that is, the operator sequence permits the DNA to assume a characteristic conformation that will make 24 direct hydrogen bonds to the protein through its phosphate groups and increase the solvent-excluded surface. It is obvious from docking experiments that straight uniform DNA with average conformational parameters can support only a small fraction of such a hydrogen-bonded network. The fact that four of the hydrogen-bonded phosphates in the protomer accept more than one hydrogen bond, and that two of these are caged by three direct hydrogen bonds, requires that the DNA backbone should assume a very precise geometry. Those conformational adjustments also enlarge the solvent-excluded contact surface to one that is comparable to those between subunits in stable oligomeric proteins. Moreover, exclusion of the solvent's screening effect is likely to augment both general and local electrostatic attractions.

We propose that DNA sequences other than an operator sequence, such as those containing operator constitutive mutations, might achieve this favourable conformation, but at too high a cost in internal energy to form a stable complex. *In vitro* measurements of the *trp* repressor specificity[46] show a 10^4-fold preference for the operator over a random non-operator site. This implies a difference of at least 5 to 6 kilocalories per mole in the free energy of binding. It is likely that the water-mediated hydrogen bonds involving base pairs 5 through 7 contribute part of the free-energy difference that targets *trp* repressor to the *trp* operator. Could the bulk of the free-energy difference responsible for operator-specificity be accounted for by sequence-dependent differences in the work required to achieve the conformation of the operator in the complex? Substantial differences in stacking energies have been calculated for small changes in the orientation of adjacent base pairs in B-DNA[28,47,48]. These variations in energy are highly dependent on the species of neighbouring base pairs, and might easily account for strong sequence-specific restraints on the required adjustments to the DNA's conformation.

We thank P. M. D. Fitzgerald for her program MERLOT; R. L. Fox, W. Mandecki, M. Hayden, R. Metzger and B. Powell (Abbott Laboratories) for synthesizing the oligonucleotides; P. Phizakerley and E. Merritt (Stanford Synchrotron Radiation Laboratory); W. Schildkampf and K. Moffat (Cornell High Energy Synchrotron Source), and N.-g. Xuong and others at the Area Detector Facility (UCSD) for providing instrumental resources for data collection; J. Carey, P. Youderian, S. Bass and R. Gunsalus for providing genetic and biochemical information before publication; C. Yanofsky, V. Horn and J. Paluh for overproducing strains; H. Drew, M. McCall, C. Calladine, R. Dickerson and W. Olson for help with the analyses of the DNA structure; and finally, G. Sechler, N. Freed, J. Cohen and C. Anderson for help in preparing the manuscript. This work was supported by grants from the USPHS and the ACS. C.L.L. was a predoctoral trainee of the USPHS, and B.F.L. is a Fellow of the Jane Coffin Childs Memorial Fund for Cancer Research.

Received 6 June; accepted 9 August 1988.

1. Somerville, R. L. *Amino Acid Biosynthesis and Regulation* (ed. Hermann, R. M. & Somerville, R. L.) 351-378 (Addison-Wesley, Reading, Massachusetts, 1983).
2. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. *Nature* **317**, 782-786 (1985).
3. Zhang, R. *et al. Nature* **327**, 591-597 (1987).
4. Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P. & Youderain, P. *Genes and Development* **1**, 565-572 (1987).
5. Joachimiak, A. *et al. J. biol. Chem.* **262**, 4917-4921 (1987).
6. Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21**, 273-278 (1988).
7. Lawson, C. L. *et al. Proteins* **3**, 18-31 (1988).
8. Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303-317 (Clarendon, Oxford, 1982).
9. Hendrickson, W. A. & Konnert, J. H. *Structure, Conformation, Function and Evolution* Vol. 1 (ed. Srinivasan, R.) 43-57 (1981).
10. Agarwal, R. C. *Acta Crystallogr.* **A34**, 791-809 (1978).
11. Finzel, B. C. *J. appl. Crystallogr.* **20**, 53-57 (1987).
12. Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119-145 (1985).
13. Lawson, C. L. & Sigler, P. B. *Nature* **233**, 869-871 (1988).
14. Pabo, C. O., Krouatin, W., Jeffrey, A. & Sauer, R. T. *Nature* **298**, 441-443 (1982).
15. Caruthers, M. H. *et al.* in *Protein Structure Folding and Design* (ed. Oxender, D. L.) 221-228 (Liss, New York, 1986).
16. Kumamoto, A. A., Miller, W. G. & Gunsalus, P. *Genes and Development* **1**, 556-564 (1987).
17. Lefevre, J. F., Lane, A. N. & Jardetzky, O. *Biochemistry* **26**, 5076-5090 (1987).
18. Dickerson, R. E., Kopka, M. L. & Pjura, P. *Biological Macromolecules and Assemblies* Vol. 2 (ed. McPherson, A. & Jurnak, F.) 38-126 (1985).
19. von Hippel, P. H. & Berg, O. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1608-1612 (1986).
20. Kelley, R. L. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 483-487 (1982).
21. Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7274-7278 (1984).
22. Marmorstein, R. Q., Joachimiak, A., Sprinzl, M. & Sigler, P. B. *J. biol. Chem.* **262**, 4922-4927 (1987).
23. Chothia, C., Wodak, S. & Janin, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3793-3797 (1976).
24. Janin, J. & Chothia, C. *FEBS: 12th Meeting, Dresden* **52** (ed. Hoffman, E., Pfeil, W. & Aurich, H.) 227-237 (Pergamon, Oxford, 1978).
25. Record, M. T. in *Unusual DNA Structures* (ed. Wells, R. D. & Harvey, S. C.) 237-252 (Springer, New York, 1988).
26. Dickerson, R. E. & Drew, H. R. *J. molec. Biol.* **149**, 761-786 (1981).
27. Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. *J. molec. Evol.* **19**, 109-114 (1983).
28. Haran, T. E., Berkovich-Yellin, Z. & Shakked, Z. J. *Biomolec. Struct. Dynam.* **2**, 397-412 (1984).
29. Calladine, C. R. & Drew, H. R. *J. molec. Biol.* **192**, 907-918 (1986).
30. Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191**, 659-675 (1986).
31. Drew, H. R. & Calladine, C. R. *J. molec. Biol.* **195**, 143-173 (1987).
32. Travers, A. A. & Klug, A. *Phil. Trans. R. Soc.* **B317**, 537-561 (1987).
33. Lomonossoff, G. P., Butler, P. J. G. & Klug, A. *J. molec. Biol.* **149**, 745-760 (1981).
34. Suck, D., Lahm, A. & Oefner, C. *Nature* **332**, 464-468 (1988).
35. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7664-7668 (1982).
36. Wu, H.-M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
37. Bossi, L. & Smith, D. M. *Cell* **39**, 643-652 (1984).
38. Snyder, M., Buchman, A. R. & David, R. W. *Nature* **324**, 87-89 (1986).
39. Zahn, K. & Blattner, F. R. *Science* **236**, 416-427 (1987).
40. Stenzel, T. T., Patel, P. & Bastia, D. *Cell* **49**, 709-717 (1987).
41. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. *Nature* **330**, 221-226 (1987).
42. Klug, A. *et al. J. molec. Biol.* **131**, 669-680 (1979).
43. Frederick, C. A. *et al. Nature* **309**, 327-331 (1984).
44. Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
45. Harrison, S. C. *et al. Bull. Inst. Pasteur* **86**, 55-64 (1988).
46. Carey, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 975-979 (1988).
47. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. *Biopolymers* **17**, 2341-2360 (1978).
48. Gotoh, O. & Tagashira, Y. *Biopolymers* **20**, 1033-1042 (1981).