PROTEINS: Structure, Function, and Genetics 3:18–31 (1988)

Flexibility of the DNA-Binding Domains of *trp* Repressor

Catherine L. Lawson, Rong-guang Zhang, Richard W. Schevitz, Zdzislaw Otwinowski, Andrzej Joachimiak, Paul B. Sigler

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637

ABSTRACT **An orthorhombic crystal form of *trp* repressor (aporepressor plus L-tryptophan ligand) was solved by molecular replacement, refined to 1.65 Å resolution, and compared to the structure of the repressor in trigonal crystals. Even though these two crystal forms of repressor were grown under identical conditions, the refined structures have distinctly different conformations of the DNA-binding domains. Unlike the repressor/aporepressor structural transition, the conformational shift is not caused by the binding or loss of the L-tryptophan ligand. We conclude that while L-tryptophan binding is essential for forming a specific complex with *trp* operator DNA, the corepressor ligand does not lock the repressor into a single conformation that is complementary to the operator. This flexibility may be required by the various binding modes proposed for *trp* repressor in its search for and adherence to its three different operator sites.**

Key words: flexibility, *trp* repressor, DNA-binding domains

INTRODUCTION

The *trp* repressor is a small dimeric regulatory protein that represses three operons in *Escherichia coli*: *trpEDCBA*, which encodes enzymes involved in L-tryptophan synthesis, *aroH*, which encodes one of three isozymes for the first step in aromatic amino acid biosynthesis, and *trpR*, which encodes the *trp* aporepressor protein.[1–3] Repression of these operons occurs by the selective binding of the repressor to an operator site, which prevents RNA polymerase from binding to the overlapping promoter site.[4–6] The three operator sites have slight variations on a consensus sequence which has two-fold symmetry, and it is proposed that the repressor recognizes the operator by forming nearly symmetrical contacts with critical base-pairs of this consensus sequence.[6]

Binding of the corepressor ligand, L-tryptophan, to the aporepressor protein is required for the recognition of the operator sequence. We have described the crystal structures of a liganded active repressor[7] and an unliganded inactive aporepressor.[8] From a comparison of the repressor and aporepressor models, two structural motifs within the repressor dimer were identified. Helices A, B, C, and F form an extensive interlocking structure, which has been termed the "central core." This central core contains all of the intersubunit contacts of the dimer, and its structure is identical in both the repressor and aporepressor. Helices D and E, on the other hand, form two independent "reading heads," which have distinctly different conformations in the repressor and aporepressor structures. The two symmetrically disposed and independent L-tryptophan binding sites can be defined in terms of these structural motifs: they are located in the interfaces between the central core and the two reading heads. Docking experiments with canonical B-form DNA[9] revealed that a major role of the L-trytophan ligand is to serve as a spacer to align the two symmetrically disposed reading heads of the dimer such that they can penetrate into two successive major grooves. In the absence of ligand, the reading heads collapse toward the central core and are too close together to penetrate successive major grooves.

The reading heads of the *trp* repressor dimer are composed of two α-helices connected by a short turn, or a helix-turn-helix motif. Four other dimeric sequence-specific DNA-binding proteins have also been shown by crystallographic structure determination to contain this helix-turn-helix motif. These are the *E. coli* catabolite gene activator protein[10,11] the repressor proteins *cI*[12] and *cro*[13] from bacteriophage λ, and the *cI* repressor of bacteriophage 434.[14,15] Additionally, the *E. coli lac* repressor has been shown by NMR to contain the helix-turn-helix motif.[16] Each of these proteins also recognizes either an exact or an approximate two-fold symmetric DNA sequence.

Abbreviations used: Repressor(t), trigonal crystal form of *trp* repressor; repressor(o), orthorhombic crystal form of *trp* repressor.

Received September 17, 1987; accepted October 15, 1987.

Address reprint requests to Paul B. Sigler, Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637.

Rong-guang Zhang is a visiting fellow from Shanghai Institute of Biochemistry, Chinese Academy of Sciences, Shanghai, People's Republic of China.

Richard W. Schevitz is now at the Physical Chemistry Research Division, Eli Lilly and Co., Indianapolis, IN 46206.

Andrzej Joachimiak is now at the Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61704 Poznan, Poland.

Catherine L. Lawson is now at the Laboratory for Chemical Physics, University of Groningen, 9747 AG Groningen, The Netherlands.

Crystal structure analysis of a complex between the amino-terminal fragment of the 434 repressor, which is closely related to λ*cI* repressor, and a fragment of DNA containing the 434 operator suggests that the specific contacts of the helix-turn-helix with the base-pairs of the operator are roughly as proposed in docking experiments.[14,15] From analyses of sequence homology, it has been proposed that several other genetic regulatory proteins would also recognize their cognate sequences using this structural scheme.[17–19]

Regulatory proteins of this class are known to be able to bind to DNA, which does not contain the cognate sequence, but with weaker affinity.[21] It has been proposed that nonspecific binding to DNA enables the regulatory protein to "search" for its cognate sequence by diffusing along the length of the DNA duplex.[20] In order to search for a cognate sequence, the protein must be able to adjust to the variation in DNA conformations and the wide range of contours presented in the major groove of nonoperator DNA. Upon recognition of the operator sequence, it must then form a specific interaction of very high affinity. For this reason it has been proposed that the region of the protein involved in binding to DNA is probably flexible.[21,22]

We report here that the reading heads of the active, liganded form of *trp* repressor can assume at least two distinct conformations. This was observed by comparing the structures of an orthorhombic crystal form of the active liganded repressor with the previously described trigonal crystal form.[7] Since the trigonal and orthorhombic forms of the repressor are grown under identical crystallization conditions with comparable ease, the two reading-head conformations produced in these crystal structures are of comparable stability. We presume that both of these stable conformational states and perhaps a distribution of intermediate conformations are compatible with nonoperator DNA-binding. This is in distinct contrast to the unliganded state, which can bind neither operator nor nonoperator DNA effectively.[23] By comparing the conformational states of two forms of repressor with those of the aporepressor we can sharpen our focus on structural features which are critical for DNA binding in general and operator binding in particular.

MATERIALS AND METHODS

The Crystal Structures

The structure determination and refinement of the trigonal crystal form of *trp* repressor (repressor(t)) and the orthorhombic crystal form of the *trp* aporepressor have been reported.[7,8] The structure determination and refinement of the orthorhombic crystal form of *trp* repressor (repressor(o)) and further refinement of the repressor(t) structure are reported below. Table I gives a descriptive summary of the crystallographic parameters for the three models used in the structural comparisons, and Table II outlines the refinement parameters for the orthorhombic and trigonal repressor model structures.

Crystallization and Data Collection

The *trp* aporepressor protein was purified as described.[24,25] Orthorhombic crystals of *trp* repressor were discovered as occasional contaminants in preparations of the trigonal crystal form,[26] but they can be grown reproducibly with seeding techniques.[27] The orthorhombic crystal form has space group $P2_12_12$, with unit cell dimensions a = 53.3 Å, b = 53.6 Å, c = 33.2 Å. The crystals have typical dimensions 0.4 × 0.4 × 0.5 mm, and they diffract to at least 1.5 Å resolution. Crystallization and exchange to an ammonium sulfate stabilizer were carried out in a manner identical with the method reported for the trigonal form.[7] A complete data set was collected to 1.65 Å on oscillation photographs using cylindrical cassettes and was processed by the program DENZO (Z. Otwinowski, unpublished). Film processing and all subsequent calculations described below were performed using either a VAX 11/750 or a Microvax II.

Molecular Replacement

The orthorhombic form of *trp* repressor was solved by the molecular replacement method,[28] using the

TABLE I. Crystal and Refinement Parameters of Structures Compared

	Repressor(T)	Repressor(O)	Aporepressor
Space group	$P3_221$	$P2_12_12$	$P2_12_12$
Unit cell	a = 50.6 Å	a = 53.3 Å	a = 44.5 Å
		b = 53.6 Å	b = 57.4 Å
	c = 73.6 Å	c = 33.2 Å	c = 34.2 Å
Growth conditions*	2.5 M sodium phosphate 0.6 M ammonium chloride	2.5 M sodium phosphate 0.6 M ammonium chloride	1.8 M ammonium sulfate 0.5 M sodium, potassium phosphate
Ligand	2.0 mM L-tryptophan	2.0 mM L-tryptophan	—
pH	5.25	5.25	7.0
R-factor†	20.4%	18.0%	20.4%
Resolution	2.2 Å	1.65 Å	1.8 Å

*Both repressor crystal forms were exchanged to a phosphate-free stabilizer (2.4 M ammonium sulfate, 0.4 M sodium chloride, 50 mM sodium acetate, and 2.4 mM L-tryptophan, pH = 5.4) prior to data collection.
†Calculated using data from d_{max} = 5.0 Å to resolution indicated, with $F \geq 2\sigma$.

TABLE II. Refinement Summary for Trigonal and Orthorhombic *trp* Repressor Models

	Trigonal	Orthorhombic
R-factor = $\Sigma\|F_o-F_c\|/\Sigma\|F_o\|$	20.4%*	18.0%†
No. of structure factors	4,292	10,699
No. of non-hydrogen atoms	888	920
No. of solvent molecules	59	84
r.m.s. deviation of bond lengths (σ = .020 Å) ‡	.012 Å	.012 Å
r.m.s. deviation of angle-related distances (σ = .040 Å)	.039 Å	.033 Å
r.m.s. shift in final cycle		
position	.007 Å	.008 Å
temperature factor**	.080 Å^2	.210 Å^2

*For $|F| > 2\sigma$ and 5.0 Å $\geqslant$ d $\geqslant$ 2.2 Å.
†For $|F| > 2\sigma$ and 5.0 Å $\geqslant$ d $\geqslant$ 1.65 Å.
‡r.m.s. deviation = $[\Sigma_{n=1}^{N}(x_n - x_o)^2/N]^{1/2}$ where x_n is the value of the nth parameter, and x_o is the accepted average value from the literature having a standard deviation of σ.
**Neighbor restraints on temperature factor values were tighter for the trigonal form than for the orthorhombic form (trigonal: bonded atoms σ = 0.3 Å^2, shared bond atoms σ = 0.6 Å^2, orthorhombic: bonded atoms σ = 1.0 Å^2, shared bond atoms σ = 1.5 Å^2).

forerunner of the program package MERLOT.[29] The crystal has unit cell dimensions that are too small to fit more than one subunit of the dimeric protein in the asymmetric unit (using $\bar{v}$ = 0.736 cm^3/g,[30] the calculated volume of one subunit is 62.9% of the asymmetric unit); therefore, the molecular two-fold axis of the protein must coincide with a crystallographic two-fold axis. Because of this constraint, the solution by molecular replacement was reduced to a simple two-dimensional problem. After aligning the dyad axis of the repressor with the two-fold axis of the crystal, it was only necessary to 1) orient the repressor by rotating about the crystallographic two-fold, then 2) position the oriented repressor by translating along the crystallographic two-fold.

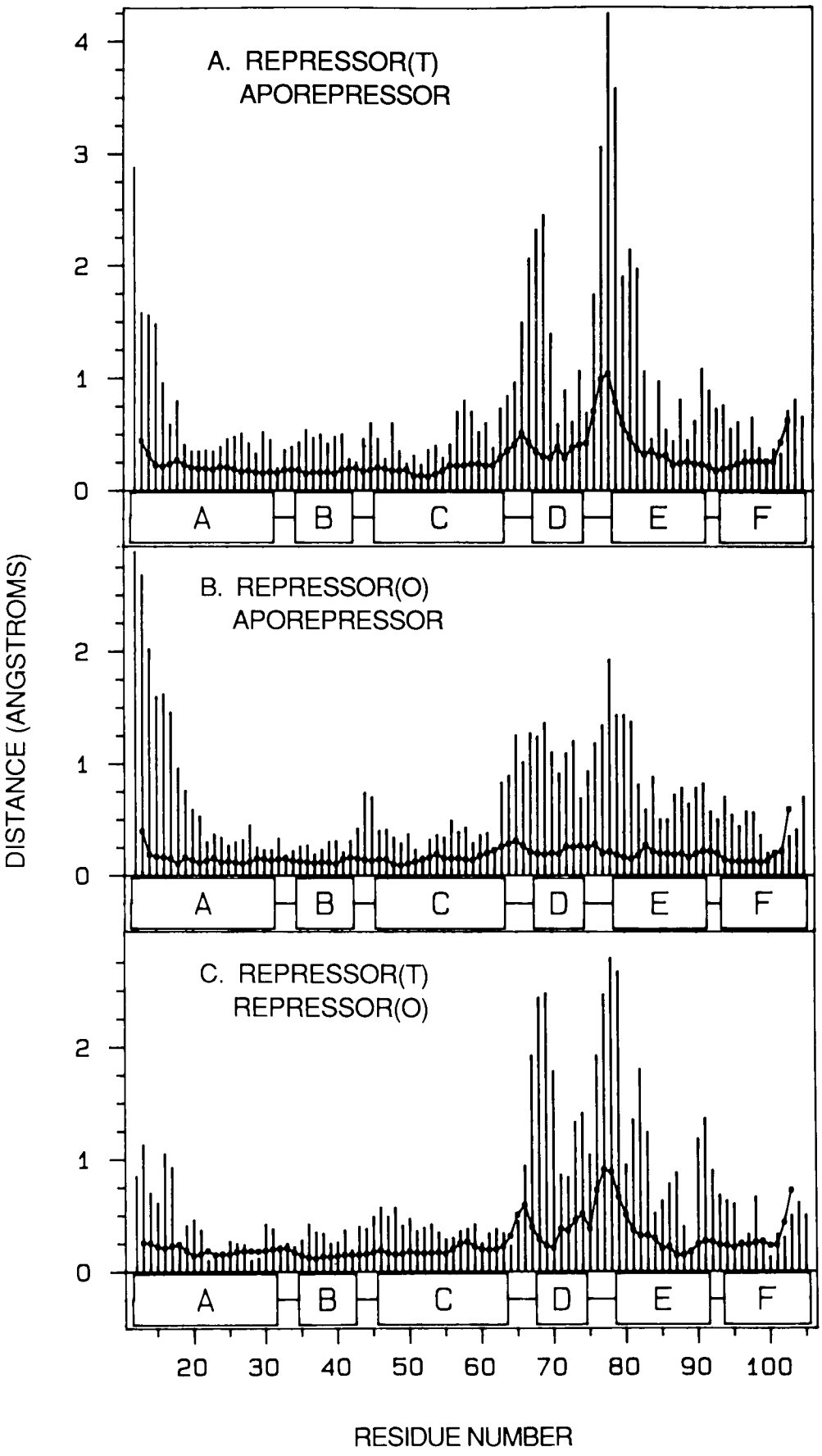
The search structure used to solve the orthorhombic *trp* repressor crystal form was identical with the one used to solve the orthorhombic *trp* aporepressor crystal form.[8] This consisted of a partially refined model of the trigonal crystal form of the *trp* repressor, which had been truncated to remove the disordered amino-terminal arm, the L-tryptophan ligand, and the carboxyl-terminal three residues, but included residues 12–105 of the 108 residues determined from the *trpR* gene sequence.[31] The search structure was placed so that the molecular dyad coincided with the x-axis of an artifical unit cell having dimensions of 120 Å in each of three orthogonal directions. The fast-rotation function of Crowther[32] was calculated using data from 10 to 4 Å, with a 25 Å radius of integration, and the resulting map gave a distinct peak of 10.1 standard deviations above the mean value of the map at α = 93°, β = 90°, γ = 0°. This solution placed the molecular twofold on the crystallographic z-axis as expected. The T(1) translation function of Crowther and Blow[33] was then calculated for this orientation, but using only one subunit of the dimer as the search structure. A cross-translation search between the subunit with general position x, y, z and the subunit with general position x + ½, $\bar{y}$ + ½, $\bar{z}$ using 10- to 5.5 Å data gave a distinct maximum at fractional unit cell translation vector values of TA = 0.50, TB = 0.50, and TC = 0.18, corresponding to a z translation of 0.41 (TC = 1–2z). The peak height of the translation function solution was 2.5 standard deviations above the mean value of the map.

Refinement

When the search structure was placed in the orthorhombic cell using the molecular replacement solution, the resulting residual (R factor) was 56.9%, including 10- to 5.5-Å data. The rotation and translation parameters of the solution were then refined by a whole-molecule rigid-body refinement using the program RVAMAP.[29] This procedure reduced the R factor to 55.8%. The program CORELS[34,35] was then used to refine the model in six rigid-body segments, defined by the six α-helical segments in the search structure, allowing only the ϕ,ψ angles of the interhelical segments to vary. After five cycles of refinement, using 9- to 5-Å data, the R factor converged to 47.2%. At this point the α-helical constraints were removed, each amino acid residue was constrained as a rigid body domain, and the resolution was extended from 5 to 3.5 Å in three steps over 29 cycles of refinement. This resulted in an R factor of 37.4%.

The refinement program of Konnert and Hendrickson[36] was then used for further stereochemically re-

Fig. 1. Pairwise comparison of the crystal structures-fitting methods. Using the program PUBFIT,[41] pairs of the corresponding α-carbon atoms of the three crystal structures were aligned by least-squares distance minimization. The vertical bars show the resulting minimized α-carbon distances. The r.m.s. deviations per atom for each comparison (see titles on figure) are **A**) 1.11 Å, **B**) 0.87 Å, and **C**) 0.91 Å, including all α-carbons between residues 12 and 105. PUBFIT was also used to compare five residue fragments of main chain atoms (N,C$_\alpha$,C,O) centered on each residue, in order to assess local conformational differences. The r.m.s. deviation per atom for each five residue fragment is shown by the points connected by a solid line. Boxes labeled A–F represent the α-helical segments of the polypeptide chain.


A. REPRESSOR(T)
APOREPRESSOR
B. REPRESSOR(O)
APOREPRESSOR
C. REPRESSOR(T)
REPRESSOR(O)
A
B
C
D
E
F
DISTANCE (ANGSTROMS)
RESIDUE NUMBER
0
1
2
3
4
20
30
40
50
60
70
80
90
100

strained refinement of the model against the observed data. Two versions of this program were used: the refinement initially employed a standard structure factor calculation algorithm but was later switched to the fast Fourier version PROFFT.[37] Refinement was also aided by the inclusion of symmetry-related contact restraints.[38] The structure was first refined at low resolution (9–3.5 Å) with tight stereochemical restraints. After 10 cycles the 9- to 6-Å data was dropped from refinement, the resolution was gradually increased, and the stereochemical restraints were loosened. At appropriate points in the refinement, difference maps were calculated, and the model was adjusted using the program FRODO.[39] "Omit maps" were also calculated in which particularly troublesome regions of the structure were omitted from the structure-factor calculation, in order to minimize bias from incorrectly placed atoms. The bound L-tryptophan, which had been omitted from the initial molecular replacement model and from early stages of refinement, was clearly seen in the first $(2F_{obs} - F_{calc}),\alpha_{calc}$ map. On the other hand, the interpretation of electron density for regions containing large structural shifts such as the D-helix and the C-D and D-E turns was not immediately evident, and omit maps at several stages of refinement were required to rebuild these regions.

In early stages of refinement the scale and overall temperature factor that gave the best fit between F_{obs} and F_{calc} in Wilson plots for the data from 5- to 1.65-Å resolution were adopted for refinement. These values were redetermined when the number of atoms in the model changed. In the 10- to 5-Å resolution zone, the average F_{obs} deviated significantly from the average F_{calc}. Modeling of bulk solvent was not included in the refinement; therefore, to avoid errors resulting from incorrectly represented solvent/molecule contrast, all data with resolution below 5 Å were dropped from refinement. In later stages of the refinement, when high-resolution data were included ($d_{min} \leqslant 2.2$ Å), individual isotropic temperature factors were refined.

During the process of refinement 84 water molecules and an ordered sulfate anion were added to the model, and additionally, residues 5–11 and 106–108 were fitted to difference maps.

Repressor(t) Refinement

The repressor(t) structure[7] was subjected to further refinement using the program PROFFT.[37] Aside from adjustment of torsion angles in several side chains, the main improvement in the structure was the addition of 59 water molecules. Manual adjustment of the repressor(t) model was often aided by comparison with the repressor(o) model, particularly in correcting regions with poor stereochemistry.

In the last cycles of refinement, individual isotropic temperature factors were refined, but because the intensity data set used for refinement was incomplete in the 2.4- to 2.2-Å resolution shell, tight neighbor restraints were imposed (see Table II).

RESULTS

Comparison of the Crystal Structures

The structures of the trigonal and orthorhombic crystal forms of the repressor (repressor(t) and repressor(o)) and the orthorhombic crystal form of the aporepressor were compared using both fitting and autocorrelation methods. Fitting methods require that two models be placed in a common coordinate system, and analysis is therefore biased by the method used for the superimposition. They are nevertheless useful for initial comparison of structural models, and they are essential for visual comparison. Autocorrelation methods, on the other hand, use information that is intrinsic to the structure irrespective of its position. As Kundrot and Richards have pointed out, they are particularly useful for characterizing structural transitions within a molecule.[40] Both of these types of analysis are required for understanding the intramolecular shifts that occur between the three crystal structures.

It is important to note that in all the crystal structures compared in this study, the solvent environment is very similar (Table I). Differences in conformation can therefore be attributed to the presence or absence of bound L-tryptophan and crystal contacts. In particular, solvent conditions such as pH, ionic strength, or specific ion effects cannot be invoked as the basis for conformational change.

The comparisons included the residues 12–105 and excluded the poorly defined flexible amino-terminal arm (residues 2–11) and final three residues (106–108). The latter two segments extend beyond the protein's globular envelope and appear to conform to different crystal packing requirements in each of the three crystal forms. For each method of analysis three separate pairwise comparisons were made: repressor(t) vs. aporepressor, repressor(o) vs. aporepressor, and repressor(t) vs. repressor(o).

Least-Squares Superimposition

Alpha-carbon atom coordinates for each model were fitted to each of the other models using the program PUBFIT.[41] The resulting distances between corresponding atoms are represented by the vertical bars in Figure 1. The set of α-carbon atoms that have the largest deviations in position are identical in both of the repressor vs. aporepressor comparisons (Fig. 1A,B). The regions that fit poorly are the regions of the protein that are also least well ordered in the crystalline lattices (Fig. 2): the amino-terminal half of the A helix (12–17), the helix-turn-helix region (65–92), and the carboxyl-terminus (104–105). However, there are qualitative and quantitative differences between the shift profiles that indicate that the two repressor structures are not fully equivalent.

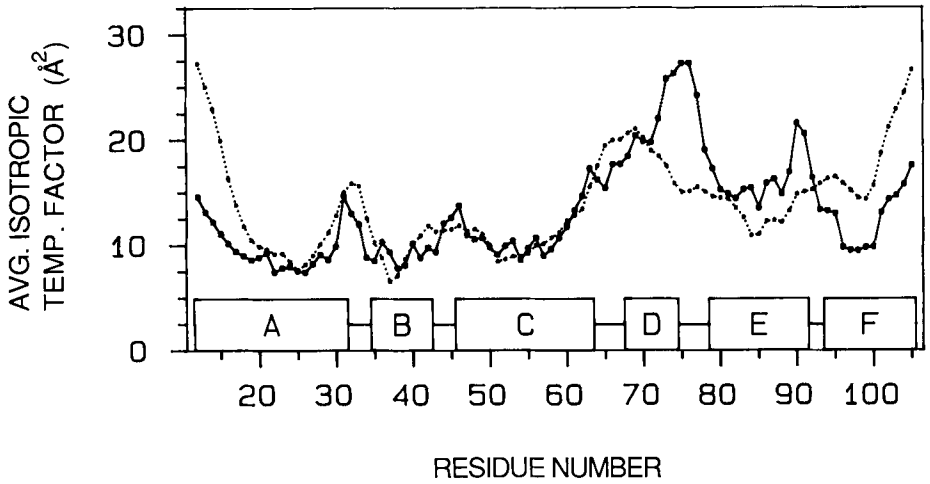


Fig. 2. Main chain average isotropic temperature factor profiles. The average temperature factor for main-chain atoms (N,C_α,C_β,C,O) was calculated for residues 12–105 in both the repressor(o) structure (points connected by solid line) and the aporepressor structure (points connected by dotted line). The temperature factors used in the calculation were refined values from the Konnert-Hendrickson least-squares refinement program.[26] Neighbor restraints for isotropic temperature factor refinement of $\sigma = 1.0$ Å^2 for main chain atoms that share a bond and $\sigma = 1.5$ Å^2 for main chain atoms that bond to a common atom were applied in both repressor(o) and aporepressor refinements. Boxes labeled A–F represent the α-helical segments of the polypeptide chain.

Surprisingly, in the least-squares superimposition of the two repressor structures (Fig. 1C), the deviations in position, particularly in the region of the helix-turn-helix, are of the same magnitude as the deviations observed between either of the repressors and the aporepressor. One must conclude from this analysis that the repressor's reading heads possess an inherent flexibility that, unlike the repressor/aporepressor transition, is not caused by the binding or loss of the corepressor ligand. Therefore, a more careful analysis of the three crystal models is clearly required to define 1) the apparent modes of flexibility in the active form of the repressor and 2) the structural role of L-tryptophan ligand binding in activating the repressor to bind to operator and nonoperator DNA.

Local Conformational Changes

In order to determine the differences in local conformation of the three model structures, a series of least-squares distance minimizations were calculated between the main chain atoms (N, Cα, C, O) of five residue fragments centered on each residue of the protein for each model pair (Fig. 1, points superim-

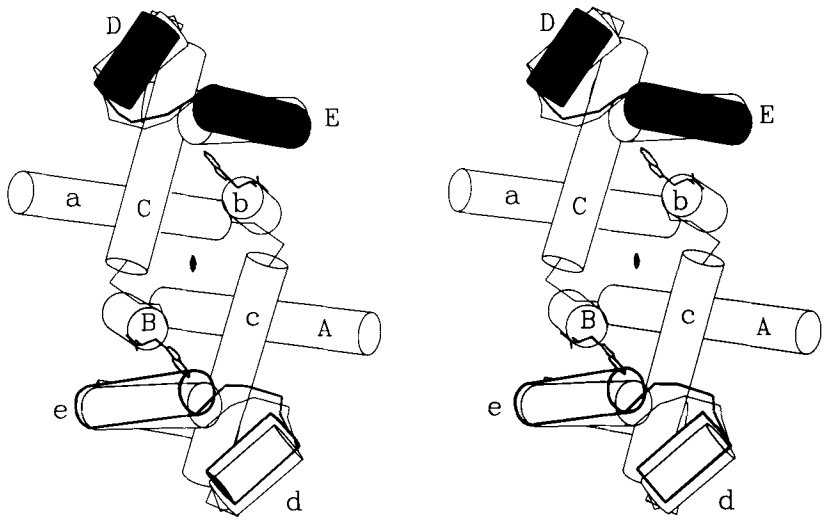


Fig. 3. "Dynamite model" showing the helical shifts of the *trp* repressor. The central core residues of the three structures were superimposed using PUBFIT.[41] The resulting coordinate sets were used as input in the drawing package of A. Lesk and K. Hardman.[51] Cylinders labeled A–E represent the helices of one subunit in the repressor(o) dimer structure; those labeled in lower case letters represent the other subunit. The F-helices are hidden behind the E-helices in this view. Superimposed on the repressor(o) D and E helices are the D and E helices of the trigonal repressor (above, dark shading) and the aporepressor (below, light shading).

posed on bar graph). The result of this analysis shows that the largest local conformational changes between the two repressor forms are located in the C-D and D-E turns (Fig. 1C). There is also a minor conformational shift in the second turn of the D-helix. Since conformational change is concentrated in the turns between helices, the structural integrity of the individual helices is largely maintained. Thus, the two crystal forms of repressor have adjusted to different crystal packing forces by shifting the positions of their D- and E-helices.

In contrast, the atomic shifts between repressor(o) and aporepressor are due to much more subtle local conformational changes that are also located in the region of the reading heads (Fig. 1B). Large conformational adjustments in the main chain of the reading heads therefore do not seem to be required for the binding of L-tryptophan, even though there is a large overall shift of the entire helix-turn-helix motif associated with ligand binding. The profile of local conformational change between the repressor(t) and aporepressor (Fig. 1A), unlike the essentially flat profile between the repressor(o) and aporepressor, closely follows the profile between the two repressor models, with large conformational changes in the C-D and D-E turns. The shifts observed between repressor(t) and aporepressor are clearly not caused by L-tryptophan binding alone but are a result of L-tryptophan binding plus shifts of the D- and E-helices.

Helical Shifts

The program PUBFIT[41] was used to determine the magnitude of the shifts of the D- and E-helices by the two-step fitting procedure described by Lesk and Chothia.[42] First, the α-carbon atoms of the central core (residues 13–62 and 94–105 of both subunits) were superimposesd. Next, the α-carbon atoms of test segments (either the D-helix, residues 68–74, or the E-helix, residues 80–91) were superimposed. PUBFIT calculates the translation distance required to superimpose the centers of gravity of the test segments and the rotation about a single axis required to superimpose the two sets of atoms.

The relative positions of the D- and E-helices in the three models after superimposition of the central core are displayed schematically in Figure 3. For simplicity, only one structure is represented in its entirety (repressor(o)), with one aporepressor reading head (light shading) and one repressor(t) reading head (dark shading) superimposed on the repressor(o) reading heads. Table III lists the translations and rotations required for independent superimposition of the D- and E-helices.

The translational components of helical movement are in each case no more than 1.4 Å, similar to values observed for helical shifts in other proteins such as insulin, citrate synthetase, and hemoglobin.[42–45] The helical shifts between the two repressor structures have large rotational components, 21° for the D-helix, and 11° for the E-helix, while the helical shifts between the repressor(o) and aporepressor structure have smaller rotational components, 10° and 5°, respectively. The 21° rotation of the D-helix is approximately twice the maximum rotation observed for helical shifts in other proteins;[42–45] however, the D-helix is also quite short, containing only two α-helical turns.

Difference Distance Matrices

In order to determine whether the results observed by least-squares fitting analysis were indeed due to concerted shifts of segments of the protein, diagonal difference plots[46] using α-carbon atom positions were calculated for each of the three pairwise comparisons. Each position (i,j) in the difference distance plot represents the shift in the distance between atom i and atom j for the crystal structures compared. Since the distance shifts are ordered by residue number, intramolecular shifts that involve concerted movement of a region of the peptide chain can be more easily recognized and interpreted. Also, since the shifts are signed, the direction as well as the magnitude of the movement can be determined.

Difference distance plots of intersubunit and intrasubunit shifts for [repressor(t)-aporepressor], [repressor(o)-aporepressor], and [repressor(t)-repressor(o)] are shown in Figure 4A–C. In each of the three comparisons, the largest interatomic distance shifts are found in the region of the plot representing the reading head domain of one subunit (helices D and E) versus the reading head domain of the opposing subunit (helices D′ and E′). Significant shifts are also found in the regions of the plot representing the reading head domain (helices D and E) vs. the central core (helices A, B, C, F and A′, B′, C′, F′).

The two [repressor-aporepressor] plots (Fig. 4A,B) are similar in the sense that they both indicate a shift of the reading heads away from the central core upon L-tryptophan binding. The [repressor(t)-repressor(o)] plot (Fig. 4C), on the other hand, indicates no overall direction of shift of the reading head domain; the shift pattern is best described in terms of independent

TABLE III. Translations and Rotations of the D and E Helices

Models compared	D-helix	E-helix
Repressor(t)-aporepressor	1.02 Å, 21.3°	0.75 Å, 11.2°
Repressor(o)-aporepressor	1.44 Å, 10.2°	0.95 Å, 4.9°
Repressor(t)-repressor(o)	1.09 Å, 21.3°	0.84 Å, 11.3°

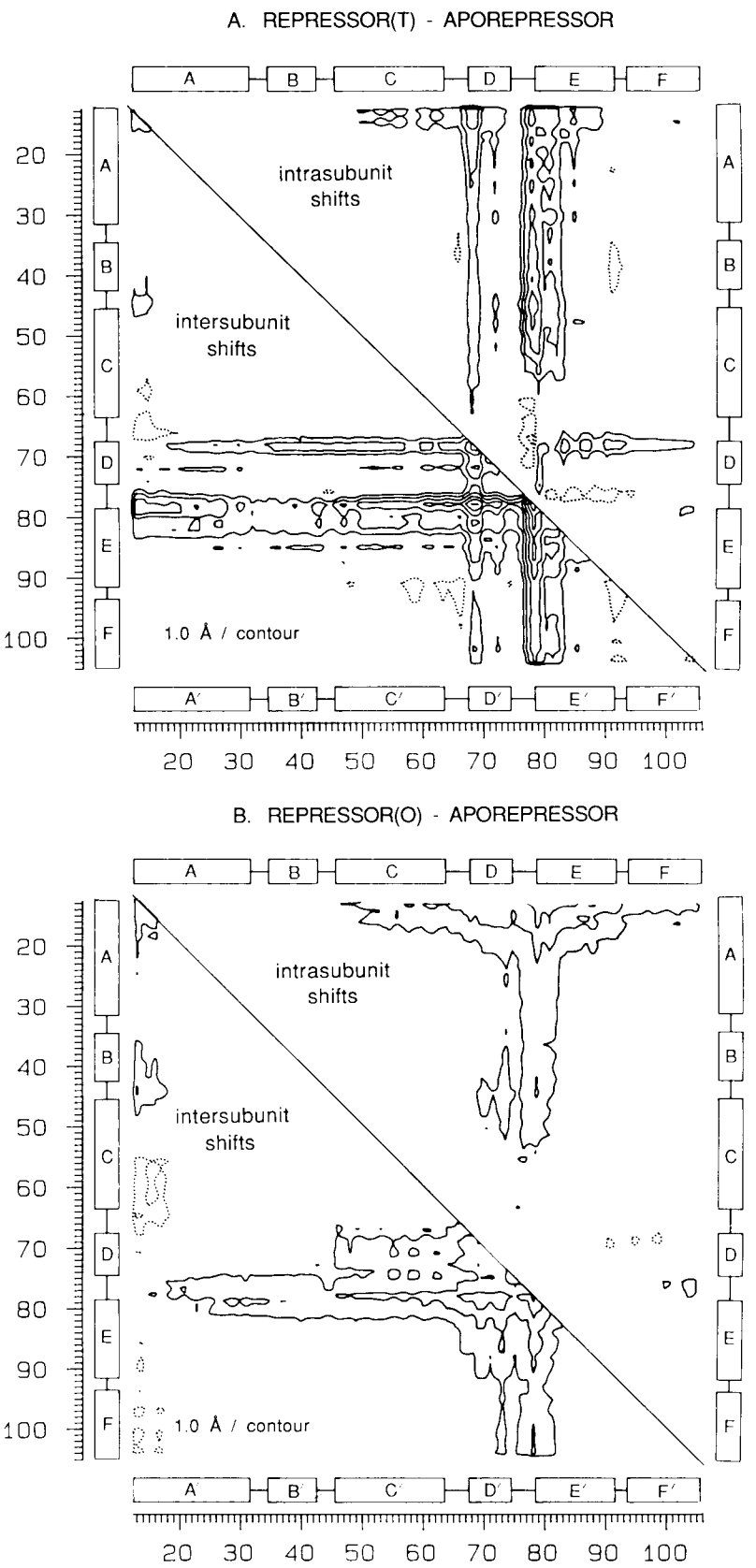


Fig. 4. Difference distance matrices. Model pairs are as indicated in the headings of **A–C** (model 1-model 2). Positive shifts are contoured with solid lines, negative shifts with dotted lines. Boxes labeled A–F represent helices of one subunit, and those labeled A′–F′ represent helices of the dyad-related subunit. "Intrasubunit shifts:" each matrix element *i,j* represents the distance between α-carbon atoms *i* and *j* within one subunit of model 1 minus the distance between atoms *i* and *j* within one subunit of model 2. "Intersubunit shifts:" as above, except that atom *i* and atom *j* are on opposing subunits of the protein dimer.

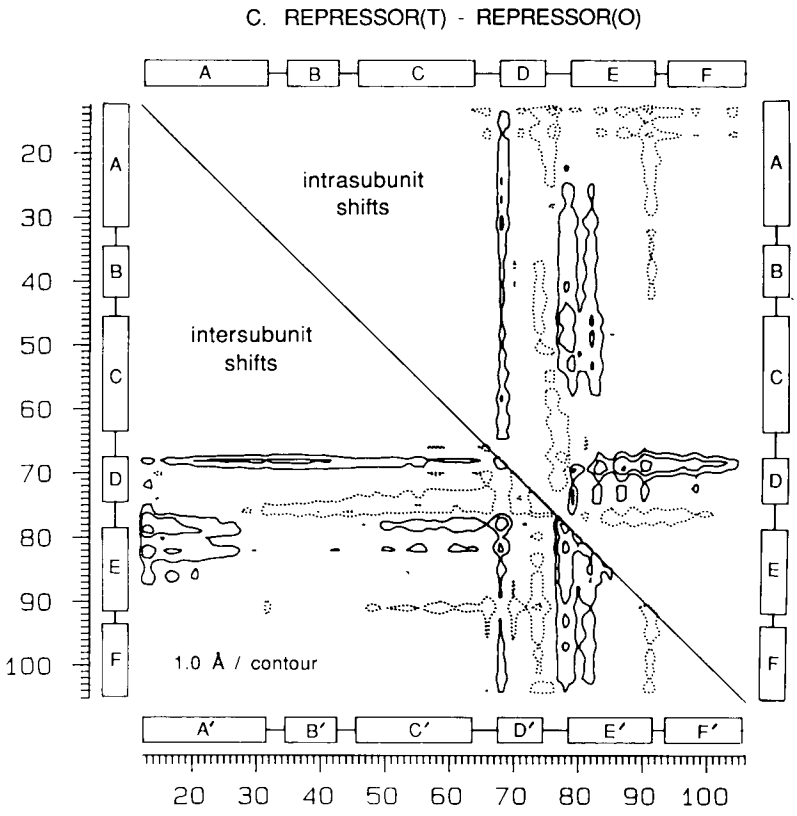


Fig. 4. Continued.

movement of the two α-helices within the reading head. Specifically, there are two shift "gradients," which correspond exactly with helices D and E: the amino-terminal ends of the D and E helices are shifted outward in the repressor(t) structure with respect to the repressor(o) structure, while the carboxyl terminal ends of both helices are shifted inwards.

The [repressor(o)-aporepressor] plot indicates an "en bloc" shift of the entire reading head with respect to the central core. The [repressor(t)-aporepressor] plot, which has the same helical shift "gradients" observed in the [repressor(t)-repressor(o)] plot, indicates individual movements of the D and E helices superimposed on an "en bloc" shift of the reading head.

Distance Shifts Between Symmetry-Related Atoms

The distance shift between each α-carbon atom and its respective symmetry mate across the molecular dyad axis is plotted for each residue in the three standard pairwise comparisons in Figure 5. This is an alternative representation of the "diagonal" in the intersubunit shifts of Figure 4, but since the interaction of the repressor with the operator depends on symmetrical contacts on either side of the common twofold, replotting the diagonal data allows one to highlight structural differences that are likely to be critical for operator recognition.

The difference plot between repressor(o) and aporepressor (Fig. 5B) clearly indicates an "en bloc" shift of the entire reading head (helices D and E) away from the twofold axis upon L-tryptophan binding, providing the correct separation to interact with DNA. The difference plot between repressor(o) and repressor(t) (Fig. 5C), on the other hand, does not indicate an overall shift either toward or away from the twofold axis but rather a large change in the conformation of the reading head. The difference plot between repressor(t) and aporepressor (Fig. 5A) can be most easily explained as a combination of "en bloc" outward shift of the reading heads that permits interaction with DNA, plus a conformational change within the reading head.

Flexibility of the Interhelical Turn

In Figure 6, the conformations of the reading heads in the three crystal structures are shown with the E-helices superimposed. It is clear upon visual inspection that the D-E turn of the repressor(t) structure has a different conformation than the turns of the repressor(o) and aporepressor structures, which are similar to each other. The observed differences in conformation are due to small adjustments in the ϕ, ψ angles of the turn ($\leqslant 30°$) rather than radical changes to different local energy minima. The flexibility of the D-E turn is probably aided by its unusual sequence: [D-helix], Leu^{75}, Gly^{76}, Ala^{77}, Gly^{78}, [E-helix]. The first glycine in this turn is highly conserved in other helix-turn-helix containing proteins, the position occupied by alanine is usually taken by

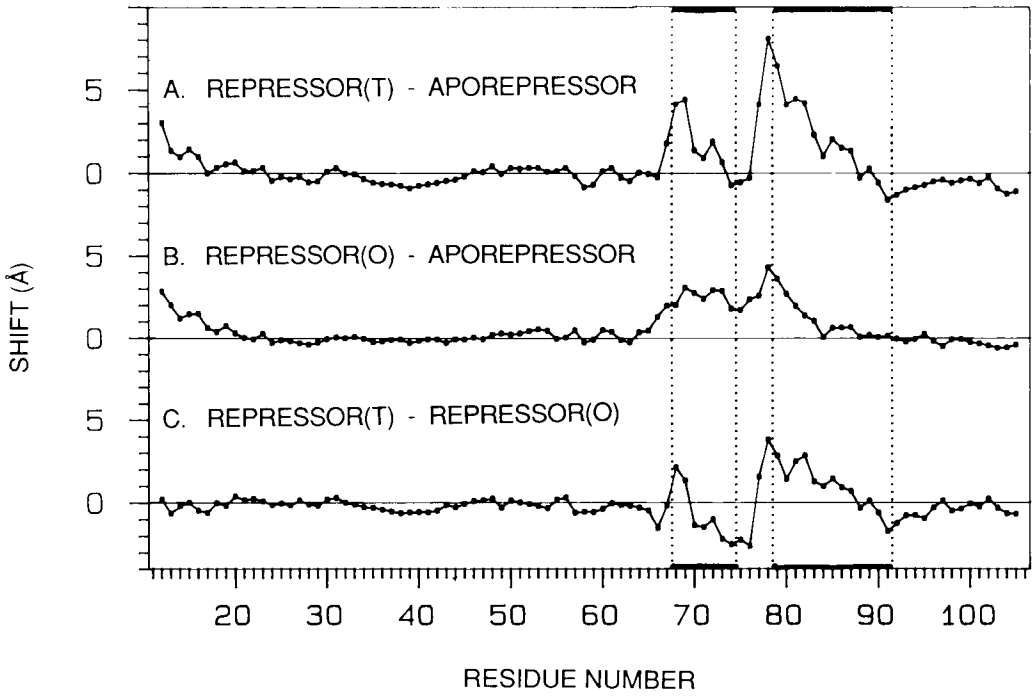


Fig. 5. Symmetry-related α-carbon shifts. Model pairs are as indicated in the headings of **A–C** (model 1-model 2). The plotted values represent the distance between each α-carbon atom and its symmetry mate across the molecular dyad axis of model 1 minus the corresponding distance in model 2. Boxes labeled A–F represent the α-helices in one protein subunit.

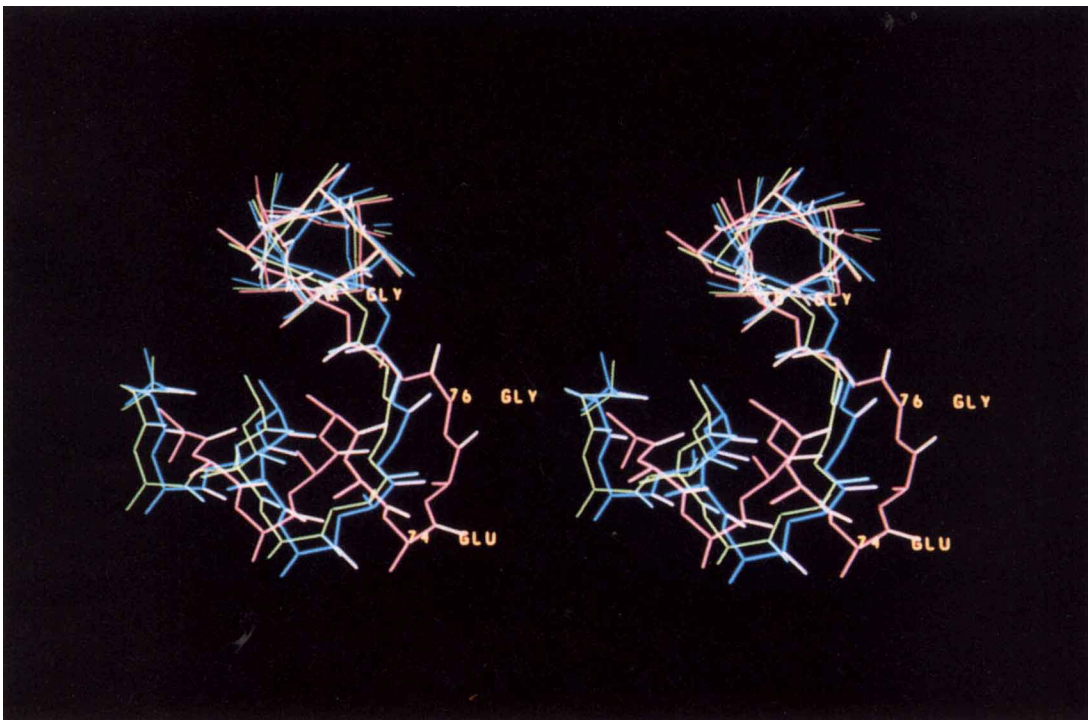


Fig. 6. Flexibility of the interhelical turn. α-carbon atoms of the E-helix in the three models were superimposed. The resulting positions of main-chain atoms are shown for the D-helix, D-E turn, and E-helix. Blue: aporepressor; pink: repressor(t); green; repressor(o). α-carbon atoms of the repressor(t) structure are labeled, and carbonyl bonds of each model are highlighted with white in order to distinguish them from C_α–C_β bonds.

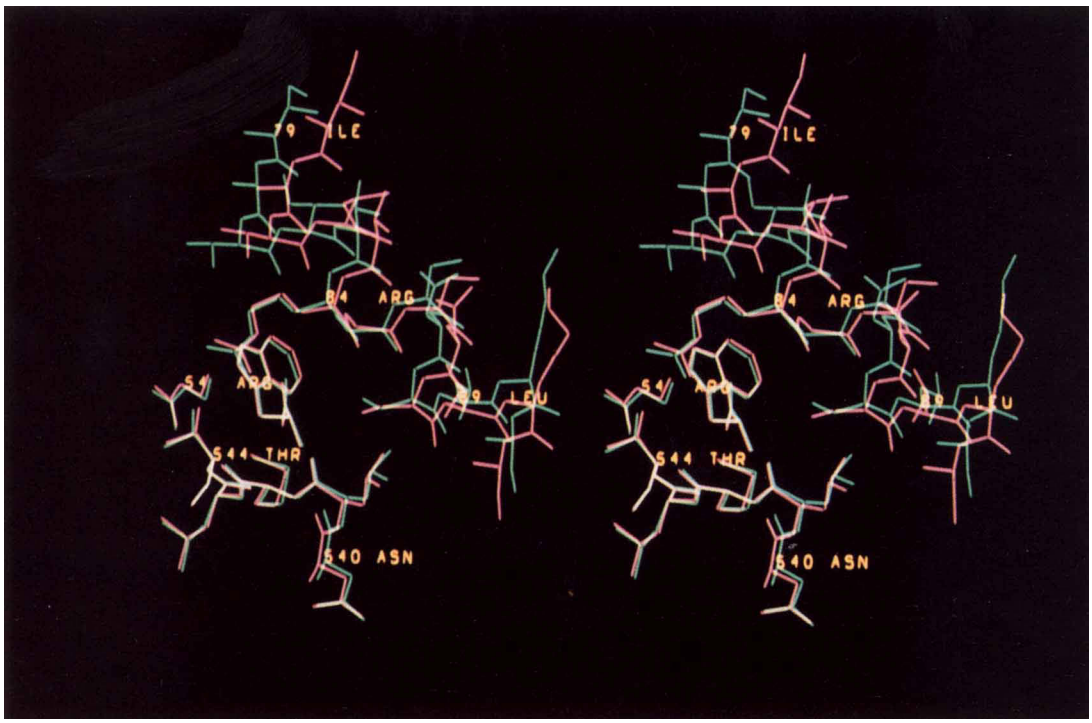


Fig. 7. Superimposition of the two L-tryptophan-binding sites. α-carbon atoms of central core residues in the two repressor structures were superimposed. The resulting atomic positions in the region of the L-tryptophan binding site are shown. Pink: repressor(t); green: repressor(o). Atoms of the repressor(o) structure are labeled. Residues from the symmetry-related subunit of the repressor dimer are numbered (N + 500).

an amino acid with a bulkier hydrophobic group, and the second glycine is unusual in its position.[17,18] The unusual sequence of the *trp* repressor turn appears to impart a great deal of torsion angle flexibility, thereby allowing the two helices it connects to move relative to each other.

Interestingly, the interhelical turn appears to be less well ordered in both repressor structures than in the aporepressor structure, as assessed by individual isotropic temperature factors obtained in the last stages of refinement (Fig. 2, data for repressor(t) not shown). Since there are no direct crystal contacts in either structure in the D-E turn, these data suggest that the flexibility of the reading head is more limited in the aporepressor structure. If this is true, one important effect of L-tryptophan binding may be to increase the allowed flexible motion of the turn, in addition to alignment of the reading heads.

Ligand Binding Site in Repressor

The two distinct conformations of the reading head seen in the repressor(t) and repressor(o) structures surprisingly do not alter the conformation of the L-tryptophan binding pocket significantly (Fig. 7). The L-tryptophan ligand is positioned identically in the two repressor structures with respect to the central core. Even though the E-helix is rotated by 11° between the two structures, the residues from the E-helix that contribute to the L-tryptophan binding pocket (Arg^{84}, Gly^{85}, Ser^{88}) are positioned nearly identically. Only the positions of residues that reside at either the amino- or carboxyl-terminal ends of the E-helix are significantly shifted. One might therefore describe the L-tryptophan binding site as the pivot point for rotation of the E-helix.

DISCUSSION

Stable and Flexible Domains in *trp* Repressor

Comparison of the repressor(o) model structure with the repressor(t) and aporepressor model structures clearly supports our view of the repressor dimer as being composed of three structural motifs: a stable central core formed by helices A, B, C, and F of both subunits, and two flexible reading heads, formed by the D- and E-helices. The central core, which contains all of the intersubunit contacts of the dimer, is essentially identical in the three structures. On the other hand, the reading heads, which have essentially no intersubunit contacts, have different conformations in each of the three structures. It is clear from comparison of the two repressor structures with the aporepressor structure that the relative position of each reading head with respect to the central core is influenced by the state of ligation. However, from comparison of the two repressor structures, it is also clear that when the L-tryptophan ligand is bound, the position of the reading head is not fixed, but can assume at least two different conformations. From this information, we can define two distinct types of flexibility

of the reading heads: "en bloc" domain movement, which is associated with the binding or loss of L-tryptophan, and intradomain flexibility, which is distinct from the ligand-induced transition.

"En Bloc" Domain Movement Associated With Ligand Binding

Even though there are only subtle local conformational differences in the main chain between the repressor(o) and aporepressor model structures, there is a large "en bloc" shift in the position of the reading heads with respect to the central core of the dimer. This overall shift can be largely attributed to a hinge-like motion distributed evenly along the torsional angles in residues at the two joints between the reading head and the central core: the C-D and E-F turns. The conformational differences observed between the repressor(t) and the aporepressor are due to the superimposition of this overall shift and independent hèlical shifts. As described previously,[8] the reading heads of the aporepressor structure are too close together to fit into two successive major grooves of canonical B-form DNA.[9] Our results here are consistent with the idea that the binding of L-tryptophan to the aporepressor dimer results in a repositioning of the two reading heads such that they can penetrate successive major grooves of B-DNA. This "en bloc" repositioning of the reading heads requires only small but possibly crucial adjustments in the conformation of the main chain within the reading head.

A Variable DNA-Binding Surface

In contrast to the "en bloc" domain movement that is associated with the binding or loss of L-tryptophan, there is an intradomain flexibility of the reading heads that is distinct from the ligand-induced transition. This flexibility can be characterized as movements of the two helices in the reading head motif relative to each other accompanied by conformational adjustments in the main chain of both the interdomain hinge turns (C-D and E-F) and the interhelical turn (D-E). Helical shifts have been observed in comparisons of crystal structures of several proteins, e.g., insulin, citrate synthetase, and hemoglobin,[42-45] and they are usually correlated with a function, such as state of ligation. In the case of *trp* repressor, we must correlate the helical shifts observed with an inherent flexibility of the protein, since both observed conformations are in the same liganded state. Since the trigonal and orthorhombic forms were grown under identical crystallization conditions, a difference in solvent environment can be excluded as the cause for the conformational changes. We must conclude instead that the two observed conformations of the repressor reading heads were induced by alternative crystal packing schemes of comparable stability. It is noteworthy that Ile^{79}, whose α-carbon atom has the second largest distance shift, forms an intermolecular contact in both trigonal and orthorhombic repressor crystals.

The helical movements observed between the two repressor models cause the presumed DNA-binding surface of the repressor to be altered significantly. When the repressor(t) model is compared to the repressor(o) model, the D-E turn appears "stretched out" so that the amino-terminus of the E-helix protrudes out further from the surface of the molecule, while the carboxyl-terminus of the D-helix is pulled in closer to the globular fold. It is not likely that both of these repressor models represent *the* specific operator-binding conformation, and it is possible, or even probable, that neither of them do. What is clear from observation of the two repressor structures is that the binding of L-tryptophan does not lock the repressor into a single conformational state.

Flexibility in Genetic Regulatory Proteins

It has been suggested by several authors[21,22] that conformational flexibility is an important property of sequence-specific DNA-binding proteins, particularly those that have a helix-turn-helix binding motif like the *trp* repressor. Flexibility might be important in accommodating the many and varied surfaces encountered in DNA binding, and in its "one-dimension" search for a target-binding sequence.

There is solid evidence for conformational flexibility in just one other protein of this class, the catabolite activator protein (CAP). There are two different orientations of the small DNA-binding domains of CAP with respect to the larger cyclic AMP-binding domains to which they are linked by a short flexible span of the peptide chain.[10] The *lac*, λcI, 434*cI*, and *cro* repressors are the only other proteins which have been shown by either X-ray crystallography or nuclear magnetic resonance (NMR) to contain the helix-turn-helix DNA-binding motif.[12-16] Based on proteolysis experiments, it appears that subunits of the *lac* and λ*cI* repressors, like CAP, also have a small relatively independent DNA-binding domain connected by a single span of peptide chain to a larger domain, which forms intersubunit contacts.[47-49] NMR results also suggest independent movement of the DNA-binding domains of *lac* repressor.[50] For *cro* repressor, it has been suggested that the dimer subunit interface is flexible enough to allow relative movement of the two helix-turn-helix motifs.[22]

The first *trp* repressor structure that we reported[7] appeared to be different from the other helix-turn-helix motif containing proteins in the sense that no mode of flexible motion for the helix-turn-helix was apparent. However, the results we report here from comparison of two crystal structures indicate that the active liganded form of *trp* repressor is capable of independent positioning and flexible movement of its reading heads, via conformational adjustments in the unusual sequence of its interhelical turn.

ACKNOWLEDGMENTS

We thank Professor Charles Yanofsky, Dr. Richard Kelley, and Ms. Virginia Horn of Stanford University

for supplying the overproducing strains of *trp* repressor, Dr. Paula M.D. Fitzgerald of the University of Alberta for her molecular replacement program package and her advice in its use, Dr. Barry Finzel of E.I. Du Pont De Nemours & Co. for his fast Fourier version of the Konnert-Hendrickson refinement program (PROFFT), Dr. Arthur Lesk of the MRC Cambridge for his program PUBFIT, and Professor Florante Quiocho and his colleagues at Rice University for PS300 FRODO 6.1. The research was supported by grants from the USPHS (GM 15225) and the American Cancer Society (NP-444). C.L.L. was a predoctoral trainee of the USPHS (GM 07183), and R.-g.Z. was a Fellow of the Cancer Research Campaign administered by the International Union Against Cancer (Geneva).

NOTE ADDED IN PROOF

The crystal coordinate sets used in this comparative study have been deposited in the Protein Data Bank, Upton, N.Y. 11973, from which copies will soon be available.

REFERENCES

1. Rose, J.K., Squires, C.L., Yanofsky, C., Yang, H.-L., Zubay, G. Regulation of *in vitro* transcription of the tryptophan operon by purified RNA polymerase in the presence of partially purified repressor and tryptophan. Nature 245:133–137, 1973.
2. Zurawski, G., Gunsalus, R.P., Brown, K.D., Yanofsky, C. Structure of *aroH*, the structural gene for the tryptophan-repressible 3-deoxy-D-arabino-heptulosonic acid-7-phospate synthetase of *E. coli*. J. Mol. Biol. 145:47–73, 1981.
3. Kelley, R.L., Yanofsky, C. *trp* aporepressor production is controlled by autogenous regulation and inefficient translation. Proc. Natl. Acad. Sci. U.S.A. 79:3120–3124, 1982.
4. Squires, C.L., Lee, F.D., Yanofsky, C. Interaction of the *trp* repressor and RNA polymerase with the *trp* operon. J. Mol Biol. 92:93–111, 1975.
5. Brown, K.D., Bennett, G.N., Lee, F., Schweingruber, M.E., Yanofsky, C. RNA polymerase interaction at the promoter-operator region of the tryptophan operon of *E. coli* and *S. typhimurium*. J. Mol. Biol. 121:153–177, 1978.
6. Oppenheim, D.S., Bennett, G.N., Yanofsky, C. *E. coli* RNA polymerase and *trp* repressor interaction with the promoter-operator region of the tryptophan operon of *S. typhimurium*. J. Mol. Biol. 144:133–142, 1980.
7. Schevitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L., Sigler, P.B. The three-dimensional structure of *trp* repressor. Nature 317:782–786, 1985.
8. Zhang, R.-G., Joachimiak, A., Lawson, C.L., Schevitz, R.W., Otwinowski, Z.O., Sigler, P.B. The crystal structure of the *trp* aporepressor at 1.8 A shows how tryptophan enhances DNA affinity. Nature 327:591–597, 1987.
9. Arnott, S., Hukins, D.W.L. Optimized parameters for A-DNA and B-DNA. Biochem. Biophys. Res. Commun. 47:1504–1509, 1972.
10. McKay, D.B., Steitz, T.A. Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. Nature 290:744–749, 1981.
11. McKay, D.B., Weber, I.T., Steitz, T.A. Structure of catabolite gene activator protein at 2.9 A. Incorporation of amino acid sequence and interactions with cyclic AMP. J. Biol. Chem. 257:9518–9524, 1982.
12. Pabo, C.O., Lewis, M. The operator-binding domain of λ repressor: Structure and DNA recognition. Nature 298:443–447, 1982.
13. Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA. Nature 290:754–758, 1981.
14. Anderson, J.E., Ptashne, M., Harrison, S.C. A phage repressor-operator complex at 7 Å resolution. Nature 316:596–605, 1985.
15. Anderson, J.E., Ptashne, M., Harrison, S.C. Structure of the repressor-operator complex of bacteriophage 434. Nature 326:846–852, 1987.
16. Zuiderweg, E.R.P., Kaptein, R., Wüthrich, K. Sequence-specific resonance assignments in the ^{1}H nuclear-magnetic-resonance spectrum of the Lac repressor DNA-binding domain 1-51 from *E. coli* by two-dimensional spectroscopy. Eur. J. Biochem. 137:279–292, 1983.
17. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M., Pabo, C.O. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature 298:447–451, 1982.
18. Ohlendorf, D.H., Anderson, W.F., Matthews, B.W. Many gene-regulatory proteins appear to have evolved from a common precursor. J. Mol. Evol. 1:109–114, 1983.
19. Weber, I.T., Steitz, T.A. A model for the non-specific binding of catabolite gene activator protein to DNA. Nucleic Acids Res. 12:8475–8487, 1984.
20. Berg, O., Winter, R.B., Von Hippel, H. Diffusion-driven mechanisms of protein translocation on nucleic acids 1. Models and Theory. Biochemistry 20:6929–6948, 1981.
21. Steitz, T.A., Harrison, R., Weber, I.T., Leahy, M. Ligand-induced conformational changes in proteins. In: "Mobility and Function in Proteins and Nucleic Acids," Ciba Foundation Symposium 93. Porter, R., O'Connor, M., Whelan, J., eds. London: Pitman Books, Ltd. 1983:25–46.
22. Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y., Matthews, B.W. The molecular basis of DNA-protein recognition inferred from the structure of *cro* repressor. Nature 298:718–723, 1982.
23. Wagner, T., Marmorstein, R.Q., Joachimiak, A., Sigler, P.B. Manuscript in preparation.
24. Joachimiak, A., Kelley, R.L., Gunsalus, R.P., Yanofsky, C., Sigler, P.B. Purification and characterization of *trp* aporepressor. Proc. Natl. Acad. Sci. U.S.A. 80:668–672, 1983.
25. Joachimiak, A., Marmorstein, R., Schevitz, R.W., Mandecki, W., Fox, J.L., Sigler, P.B. Crystals of the *trp* repressor-operator complex suitable for x-ray diffraction analysis. J. Biol. Chem. 262:4917–4921, 1987.
26. Joachimiak, A., Schevitz, R.W., Kelley, R.L., Yanofsky, C., Sigler, P.B. Functional inferences from crystals of *E. coli trp* repressor. J. Biol. Chem. 258:12641–12643, 1983.
27. McPherson, A. The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. Methods Biochem. Anal. 23:249–345, 1976.
28. Rossmann, M.G., (ed.): "The Molecular Replacement Method: A Collection of Papers on the Use of Non-Cyrstallographic Symmetry." New York: Gordon and Breach, 1972.
29. Fitzgerald, P.M.D., MERLOT: A package of computer programs for the determination of phases using the molecular replacement method. J. Appl. Cryst., submitted.
30. Schwert, G.W.R., Kaufman, S. The molecular size and shape of the pancreatic proteases. III. α-chymotrypsin. J. Biol. Chem. 190:807–816, 1951.
31. Gunsalus, R.P., Yanofsky, C. Nucleotide sequence and expression of *E. coli trpR*, the structural gene for the *trp* aporepressor. Proc. Natl. Acad. Sci. U.S.A. 77:7117–7121, 1980.
32. Crowther, R.A. The fast rotation function. In: "The Molecular Replacement Method: A Collection of Papers on the Use of Non-Crystallographic Symmetry." Rossmann, M.G., ed. New York: Gordon and Breach. 1972:173–178.
33. Crowther, R.A., Blow, D.W. A method of positioning a known molecule in an unknown crystal structure. Acta Cryst. 23:544–548, 1967.
34. Sussman, J.L., Holbrook, S.R., Church, G.M., Kim, S.-H. A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. Acta Cryst. A33:800–804, 1977.
35. Hoard, L.G., Nordman, C.E. A Gauss-Siedel least-squares refinement procedure with rigid-group and parameter restraint capabilities. Acta Cryst. A35:1010–1015, 1979.
36. Hendrickson, W.A., Konnert, J.H. Stereochemically restrained crystallographic least-square refinement of macromolecule structures. In: Biomolecular Structure, Conformation, Function and Evolution," Vol. 1. Srinivasan, R., ed. Oxford: Pergamon Press. 1981:43–57.
37. Finzel, B.C. Incorporation of fast Fourier transforms to speed restrained least-squares refinement of protein structures. J. Appl. Cryst. 20:53–55, 1987.

38. Sheriff, S. Addition of symmetry-related contact restraints to PROTIN and PROLSQ. J. Appl. Cryst. 20:55–57, 1987.
39. Jones, T.A. Interactive computer graphics: FRODO. In: "Methods in Enzymology," Vol. 115, part B. Wyckoff, H.W., Hirs, C.H.W., Timasheff, S.N., eds. Orlando: Academic Press. 1985:157–171.
40. Kundrot, C.E., Richards, F.M. Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1,000 atmospheres. J. Mol. Biol. 193:157–170, 1987.
41. Lesk, A. A toolkit for computational molecular biology II. On the optimal superimpositon of two sets of coordinates. Acta. Cryst. A42:110–113, 1986.
42. Lesk, A.M., Chothia, C. Mechanisms of domain closures in proteins. J. Mol. Biol. 174:175–191, 1984.
43. Baldwin, J., Chothia, C. Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism. J. Mol. Biol. 129:175–220, 1979.
44. Chothia, C., Lesk, A., Dodson, G.G., Hodgkin, D.C. Transmission of conformational change in insulin. Nature 302:500–505, 1983.
45. Chothia, C., Lesk, A.M. Helix movements in proteins. Trends Biochem. Sci. 10:116–118, 1985.
46. Nishikawa, K., Ooi, T., Isogai, Y., Saito, N. Tertiary structure of proteins I. Representation and computation of the conformations. J. Phys. Soc. Japan 32:1331–1337, 1972.
47. Geisler, N., Weber, K. Isolation of the amino-terminal fragment of lactose repressor necessary for DNA binding. Biochemistry 16:938–943, 1977.
48. Jovin, T.M., Geisler, N., Weber, K. Amino-terminal fragments of *E. coli lac* repressor bind to DNA. Nature 269:668–672, 1977.
49. Pabo, C.O., Sauer, R.T., Sturtevant, J.M., Ptashne, M. The λ repressor contains two domains. Proc. Natl. Acad. Sci. U.S.A. 76:1608–1612, 1979.
50. Wade-Jardetzky, N.G., Bray, R.P., Conover, W.W., Jardetzky, O., Giesler, N., Weber, K. (1979). Differential mobility of the N-terminal headpiece in the *lac* repressor protein. J. Mol. Biol. 128:259–264, 1979.
51. Lesk, A., Hardman, C. Computer generated schematic diagrams of protein structures. Science 216:539–540, 1982.