The solution structure of Lac repressor headpiece 62 complexed to a symmetrical *lac* operator

Christian AEM Spronk, Alexandre MJJ Bonvin, Plachikkat K Radha†, Giuseppe Melacini‡, Rolf Boelens and Robert Kaptein*

Background: Lactose repressor protein (Lac) controls the expression of the lactose metabolic genes in *Escherichia coli* by binding to an operator sequence in the promoter of the *lac* operon. Binding of inducer molecules to the Lac core domain induces changes in tertiary structure that are propagated to the DNA-binding domain through the connecting hinge region, thereby reducing the affinity for the operator. Protein–protein and protein–DNA interactions involving the hinge region play a crucial role in the allosteric changes occurring upon induction, but have not, as yet, been analyzed in atomic detail.

Results: We have used nuclear magnetic resonance (NMR) spectroscopy and restrained molecular dynamics (rMD) to determine the structure of the Lac repressor DNA-binding domain (headpeice 62; HP62) in complex with a symmetrized *lac* operator. Analysis of the structures reveals specific interactions between Lac repressor and DNA that were not found in previously investigated Lac repressor–DNA complexes. Important differences with the previously reported structures of the HP56–DNA complex were found in the loop following the helix-turn-helix (HTH) motif. The protein–protein and protein–DNA interactions involving the hinge region and the deformations in the DNA structure could be delineated in atomic detail. The structures were also used for comparison with the available crystallographic data on the Lac and Pur repressor–DNA complexes.

Conclusions: The structures of the HP62–DNA complex provide the basis for a better understanding of the specific recognition in the Lac repressor–operator complex. In addition, the structural features of the hinge region provide detailed insight into the protein–protein and protein–DNA interactions responsible for the high affinity of the repressor for operator DNA.

Address: Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Present addresses: †Department of Molecular Biology, The Scripps Research Institute, San Diego, California 92037, USA and ‡Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093-0343, USA.

*Corresponding author.
E-mail: kaptein@nmr.chem.uu.nl

Key words: DNA-induced α helix, Lac repressor, molecular dynamics, NMR, protein–DNA complex

Received: 14 June 1999
Revisions requested: 22 July 1999
Revisions received: 23 August 1999
Accepted: 15 September 1999

Published: 23 November 1999

Structure November 1999, 7:1483–1492

0969-2126/99/$ – see front matter

Introduction

The crystal and solution structures of different forms of the *Escherichia coli* lactose (Lac) and purine (Pur) repressor proteins [1–9] have provided a basis for the interpretation of a wealth of biochemical and genetic data on the Lac repressor. The solution structures of the DNA-binding domain (or headpiece) and its complexes with *lac* operator sequences of varying length have provided a detailed understanding of the interactions that the repressor has with the *lac* operator [4–6]. The headpiece is a small globular domain with a hydrophobic core formed by three α helices. The first two helices span the helix-turn-helix motif of which the second helix is responsible for specific interactions with the *lac* operator. The structures of the full Lac repressor and its complexes with DNA and the gratuitous inducer isopropyl-β-D-1-thiogalactopyranoside (IPTG) have been solved by X-ray crystallography [1]. The crystallographic data of the apoprotein repressor and the IPTG-bound repressor did not show any electron density for the headpieces and the hinge region that connects the headpieces to the core domain. In the DNA-bound form the headpieces as well as the hinge regions, which form the so-called hinge helices in the complex, were observed. The hinge helices, through their interactions with the minor groove, cause a distinct bend in the operator DNA of about 45°. Nuclear magnetic resonance (NMR) studies have shown that the hinge region is unfolded when the headpieces are not bound to DNA and folds into an α helix when the headpieces bind as a dimer to a symmetrical operator sequence [9]. This folding transition in the hinge region plays an important role in the model for the allosteric transitions occurring in Lac repressor during induction. The crystal structures have revealed the effect of inducer binding and DNA-binding on the conformation of the Lac repressor core domain. The conformational changes in the core domain upon induction are translated to the DNA-binding domain through the connecting hinge region. Upon induction, the first residues of the hinge region are pushed apart, thereby disrupting protein–protein interactions between the hinge helices and causing the

hinge helices to unfold. The interactions of the hinge helices with the minor groove of the DNA are then lost and the affinity for the operator is reduced dramatically [1,9].

Although the hinge helices have been observed both in X-ray and NMR experiments, no high-resolution structural detail is, as yet, available for this functionally important domain. The X-ray structure of the repressor has a low resolution of 4.8 Å and provides, therefore, only little information on the hinge-helix–operator and hinge-helix–hinge-helix interactions. The NMR structures of headpiece 56 and its complex with an 11 base-pair operator half site (HP56–DNA complex) did not provide this information because the headpiece construct was too short to form the hinge helix. Also, protein–protein interactions, that could stabilise the hinge helix, are necessarily absent in the 1:1 HP56–DNA complex. Further, the bending of the *lac* operator DNA, as observed in the X-ray structures of the Pur and Lac repressor complexes and derived from biochemical data [10], changes the protein DNA interface not only in the minor groove but also in the major groove. This might cause differences in the major groove protein–DNA interactions between the complexes in which the hinge helices are present and the HP56–DNA complex.

In this paper we describe the solution structure of HP62 bound to the symmetrical *lac* operator sequence:

```
–1 1     5       10      15      20
G A A T T G T G A G C ◊ G C T C A C A A T T C
C T T A A C A C T C G ◊ C G A G T G T T A A G
```

The structure was determined by NMR spectroscopy and restrained molecular dynamics (rMD). HP62 binds as a dimer to this symmetrized *lac* operator and is large enough to form the hinge helices [9]. A full description of the structural features of the complex is given. In particular protein–protein and protein–DNA interactions occurring in the hinge region, which play a crucial role in the repressed state and are disrupted upon induction, are discussed in detail. A comparison is made between the complexes of HP62, HP56, intact Lac repressor and the Pur repressor with their operators. The structures shown here have some distinctly different protein–DNA interactions than those found in the previous NMR studies and provide a more complete structural basis for the understanding of biochemical and genetic data on Lac repressor.

Results and discussion

Structure calculations

The structure calculation protocol used for the HP62–DNA complex consisted of calculation of the structure of the HP62 monomer in the complex, docking of two duplicated headpieces onto DNA and a final refinement step (see the Materials and methods section). In the docking step a B-DNA template structure was used that was allowed to bend in order to accommodate the two HP62 molecules. The nuclear Overhauser enhancements (NOEs) observed for *lac* operator alone are not sufficient to allow a good definition of the DNA structure without protein bound to it. Although the structure of the headpieces and their contacts to the operator do provide additional long-range distance information, we used rMD simulations in water to obtain reasonable DNA structures. Table 1 shows the structural statistics and quality checks of the final 11 HP62–DNA structures. The structures (Figure 1) are of good quality and are well determined by the NMR data, as can be seen from the average pairwise root mean square deviations (rmsds).

The overall structure of HP62 in the complex

The global structure of the complex is very similar to the X-ray structures of the DNA-binding domains of the Lac and Pur repressors in complex with DNA (Figure 2). Two HP62 monomers, consisting of four α helices running from residues 6–13, 17–24, 32–45 and 51–56 (residues 50, 57 and 58 form helix extensions), bind symmetrically to *lac* operator with their helix-turn-helix motifs contacting bases in the major groove. The interface of the two HP62 monomers is formed by the two antiparallel hinge helices in the center of the operator, responsible for extensive protein–protein interactions and protein–DNA contacts to the minor groove. Because of these interactions the DNA

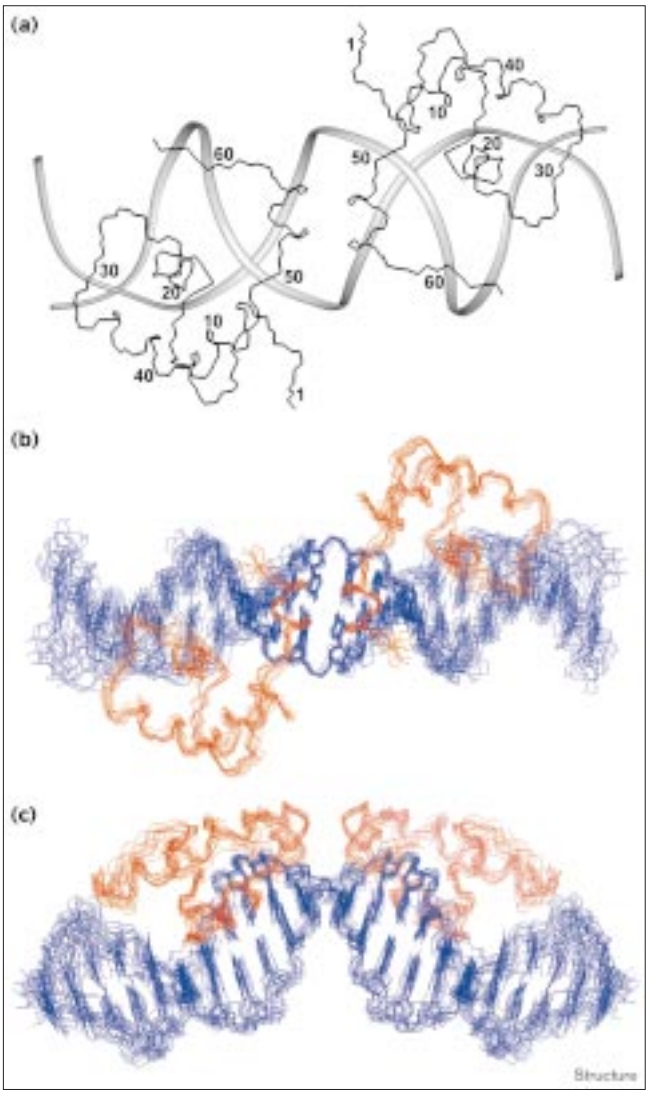
Table 1

Structural statistics.

Experimental input restraints for the complex	
HP62 intramonomer	1322
HP62 intermonomer	16
HP62 intra/intermonomer	30
Intra-DNA	1342
HP62–DNA	132
Average rmsds from distance restraints (Å)	0.053 ± 0.005
Average rmsds from idealized covalent geometry	
Bonds (Å)	0.00042 ± 0.00002
Angles (°)	2.6 ± 0.2
Impropers (°)	2.7 ± 0.5
Rmsds from average structure (Å)	
HP62 monomer (backbone/all heavy atoms)	0.35 ± 0.13/0.94 ± 0.31
HP62 dimer (backbone/all heavy atoms)	0.60 ± 0.24/1.09 ± 0.37
lac operator (backbone/all heavy atoms)	1.06 ± 0.45/0.95 ± 0.39
Complex (backbone of HP62 and Lac operator)	0.95 ± 0.35
Ramachandran plot	
Residues in favored regions (%)	91.4
Residues in allowed regions (%)	7.2
Residues in accepted regions (%)	1.0
Residues in disallowed regions (%)	0.4

The number of distance restraints are listed for the full dimeric complex. Rmsd calculations of HP62 were restricted to residues 4–25 and 32–57. Rmsd calculations of *lac* operator were restricted to the central 18 base pairs. Backbone atoms for the protein included the N, C′ and C^{α} nuclei. Backbone atoms of the DNA included the P, O3′, O5′, C3′, C4′ and C5′ nuclei.

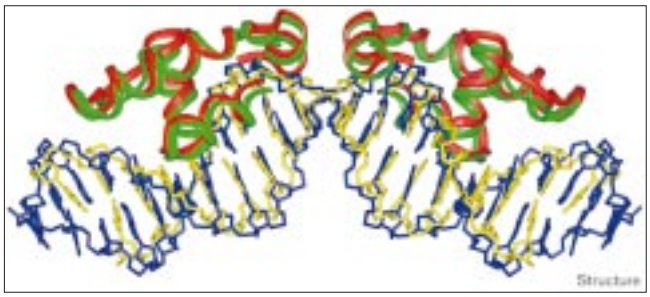
Figure 1



Ensemble of NMR structures of the HP62–DNA complex. (a) Backbone trace of the full HP62–DNA complex. Numbers indicate the residues of HP62. The DNA backbone is shown as a ribbon. (b) Overlay of the 11 final structures of the HP62–DNA complex. Superimposed are all atoms of residues 4–25, 32–58 of both HP62 monomers (red) and the central 18 base pairs of the *lac* operator (blue). Residues 4–59 of the two monomers and the full *lac* operator sequence are shown. (c) View perpendicular to (b). This figure was generated using the program Biosym InsightII.

structure shows an opening of the minor groove and a global bending of approximately 45°, which is similar to the deformations found in the X-ray structures of the complete Lac and Pur repressors in complex with DNA. Overlays of the NMR structures with the X-ray structures of the DNA complexes for residues 6–25, 34–57 of Lac repressor and residues 4–23, 32–55 in the Pur repressor on the C^{α} carbon atoms yield average pairwise rmsds of 0.94 and 0.72, respectively.

Figure 2



Superposition of one structure of the HP62–DNA ensemble with the X-ray structure of the Pur repressor–operator complex. The headpieces of the purine repressor are shown in green and its operator in yellow; the headpieces of the lactose repressor are shown in red and its operator is in blue. This figure was generated using Biosym InsightII.

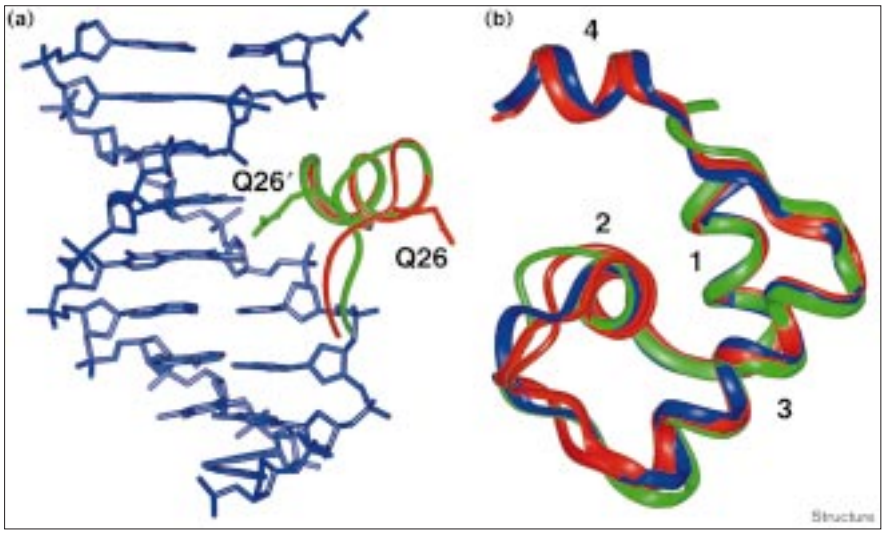
Comparison to the HP56–DNA structure

The major difference between the HP56–DNA and HP62–DNA complexes is the presence of the C-terminal hinge helix in the latter complex. The HP62–DNA and HP56–DNA structures, therefore, allow only a comparison for the residues spanning the first three helices and the loop between the recognition helix and the third helix. An overlay of the backbone atoms of the first three helices of HP62 with those of HP56 yields an average pairwise rmsd of 0.83 Å, whereas the overlay of the DNA sequences is poor. The largest difference in protein structure is found in the loop following the recognition helix of HP62, which is involved in contacts to the major groove of the DNA (Figure 3a). Previously, a discrepancy existed at Gln26, which in relaxation studies of free and complexed HP56 was hardly affected by DNA binding [11]. In contrast, in the structure of the half-site complex the Gln26 sidechain is involved in a water-bridged contact to the phosphate of Cyt7 causing the loop to point towards the DNA backbone. This discrepancy is resolved in the current studies of the HP62–DNA complex in which the observation of new intraprotein NOEs results in different loop conformations. In three of the HP62–DNA structures Gln26 points towards the DNA backbone in a similar though less pronounced fashion as seen in the HP56–DNA complex. The majority of the structures adopt a different conformation with the backbone pointing away from the DNA and the Gln26 sidechain being solvent-exposed. These results are in close agreement with our experimental NMR data, free MD simulations of the HP56–DNA complex (A Bonvin, personal communication) and genetic data [12,13], which indicate a disorded Gln26 sidechain that is not important for interaction with the *lac* operator.

The dimer interface

The formation of a protein–protein interface between the two hinge helices is necessary for the tight binding of the operator and occurs in the absence of inducer molecules.

Figure 3



Comparison of loop conformations in different repressor–DNA complexes. (a) The conformation of the loop connecting the second and third helix in the HP56–DNA and HP62–DNA complexes. The backbones of residues 15–29 and the sidechain of Gln26 of HP56 (green) and HP62 (red) are shown. The DNA is taken from the HP62–DNA structure. (b) Overlay of HP56, HP62 (two conformations) and the Pur repressor DNA-binding domain. The Pur repressor is shown in blue, HP62 in red and HP56 in green. The four α helices present in the HP62 are numbered 1–4. This figure was generated using the program Biosym InsightII.

At the interface extensive hydrophobic contacts are made between the sidechains of Val52, Ala53, Leu56 and, to a lesser extent, Gln55 (Figure 4; see Supplementary material). The arrangement of the sidechains at the interface is determined by the symmetry in the system and the NOEs observed between the two headpieces and between the headpieces and the DNA. The four methyl groups of the two Leu56 residues are in a more or less linear arrangement, with the $C^{\delta}2$ methyls in the center, and are involved in somewhat closer intermonomer contacts than Leu54 and Leu54′ in the Pur repressor–operator complex. The conformations and positions of the Val52 and Val52′ sidechains were determined to be in the favorable conformation of a valine residue in α helices [14]. This arrangement positions the two $C^{\gamma}1$ methyl groups of Val52 and Val52′ in close contact with each other in the center of the complex, whereas the corresponding valines in the Pur repressor complex are in contact with their $C^{\gamma}2$ groups.

Several mutations of Val52 are known to result in tight-binding Lac repressors, which indicates the importance of this residue [12,13]. Furthermore, the dimer interface can be stabilized by replacing Val52 by cysteine. Crosslinking the two hinge helices of this mutant by oxidation yields a repressor that binds Lac operator with an approximately sixfold higher affinity than the wild-type repressor does and the disulfide crosslink distorts the allosteric linkage between operator and inducer binding. As a consequence this complex is rendered non-inducible [15].

DNA binding and structure

An extensive discussion of the interactions between LacHP56 and the *lac* operator in the major groove and the comparison with genetic data has been given previously [6]. We will focus here on the differences found between the HP56–DNA and HP62–DNA complexes and discuss their possible consequences for the specific recognition of *lac* operator by repressor.

The interactions between HP62 and the *lac* operator can be divided into major groove and minor groove binding regions, the first being formed by the HTH motif and the loop following the recognition helix and the latter by the C-terminal hinge-helix (see Figures 4 and 5). In addition to the directly observed NOEs between HP62 and *lac* operator (see Supplementary material), information on protein–DNA interactions is obtained from an analysis of hydrogen bonding and apolar contacts of the structures collected during the last part of the MD simulation (see Supplementary material).

Interactions in the major groove

The residues of the first helix of the HTH motif that contact the *lac* operator in the HP62–DNA complex are Thr5, Leu6 and Tyr7. The main difference with the half-site complex is the conformation of the sidechain of Leu6. The conformation is well defined in the HP62–DNA complex by the observation of 12 NOEs to the DNA and is similar to that of Ile4 in the Pur repressor–DNA complex. In this new conformation the interaction of Leu6 with the DNA includes nonpolar interactions to Gua10, Cyt9 and Thy8, whereas such interactions were confined to Cyt9 in the half-site complex. In contrast, the interactions of the backbone of Leu6 with the operator are similar to those found in the HP56–DNA and the corresponding residue in the Pur repressor–operator complex.

The interaction of the recognition helix with the bases of the operator show some interesting differences in the HP62–DNA complex as compared with the half-site

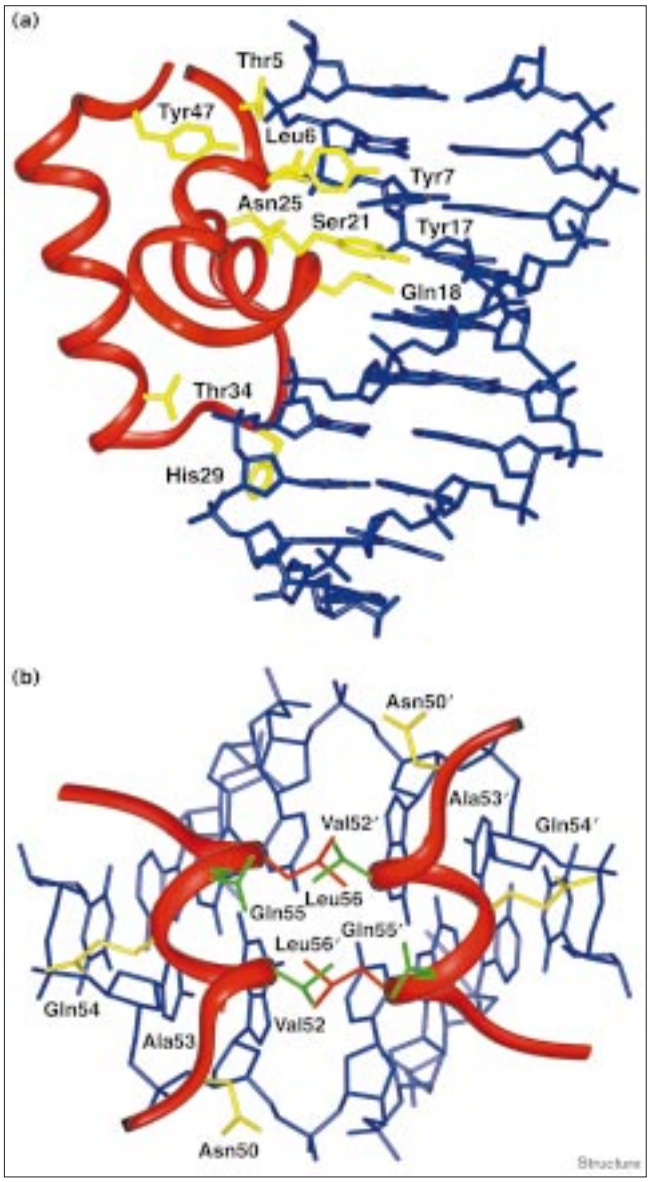
complex. The most important difference is that the interaction of the first two residues, Tyr17 and Gln18, with the DNA appears to be shifted away from the center of the operator in the HP62–DNA complex. These two residues are the most important residues for specific recognition of the *lac* operator and can be changed to obtain Lac repressors that recognize different operator sequences.

In the new structures the interactions of Tyr17 with the DNA involve Thy6, Gua7 and Thy8. The close contacts to Cyt9, present in the HP56–DNA complex, are not present, although weak NOEs between Tyr17 and Cyt9 have been observed. Further, we have observed NOEs that are probably those between the Tyr17 ring protons and the Cyt7 NH_2 group and between the Tyr17 ring protons and the Thy6 methyl protons, although they could not be assigned unambiguously. The shift in interaction of Gln18 is suggested by the contacts to the base-NH_2 of Ade6. Further, Gln18 interacts with Cyt7 through both hydrogen bonds and hydrophobic contacts [5,6,16]. These results from the structural analysis are in close agreement with the findings of Sartorius *et al.* [17,18] who, on the basis of genetic experiments, concluded that the specific binding of Tyr17 and Gln18 is directed mainly towards base-pair 7 and to a lesser extent to base-pair 6 of the *lac* operator. In the genetic experiments interactions to base-pair 8 were not found, probably because no repressor mutants were identified with a non-negligible affinity for any *lac* operator variant substituted in base-pair 8, which indicates the crucial role of this base pair in repressor–operator recognition. Interestingly, the complex of Pur repressor with a palindromic *pur* operator shows that the second residue of the recognition helix, Thr16, forms hydrogen bonds with base-pairs 5 and 6 of the operator simultaneously, corresponding to base-pairs 6 and 7 in the *lac* operator.

The three other residues in the recognition helix that are thought to be important for interaction with DNA are Ser21, Arg22 and Asn25. Of these, Ser21 and Asn25 appear to have similar interactions with the backbone of the DNA as seen in the HP56–DNA complex. No conclusions can be drawn for Arg22 because of structural disorder of its sidechain, caused by a lack of assignments of the sidechain protons. Previous genetic and structural data indicate, however, that Arg22 is important for the headpiece to interact with the DNA [6,17]. It could be interacting with a number of different base-pairs of *lac* operator in a dynamic process causing the NMR lines to broaden to the extent that structural studies become very difficult.

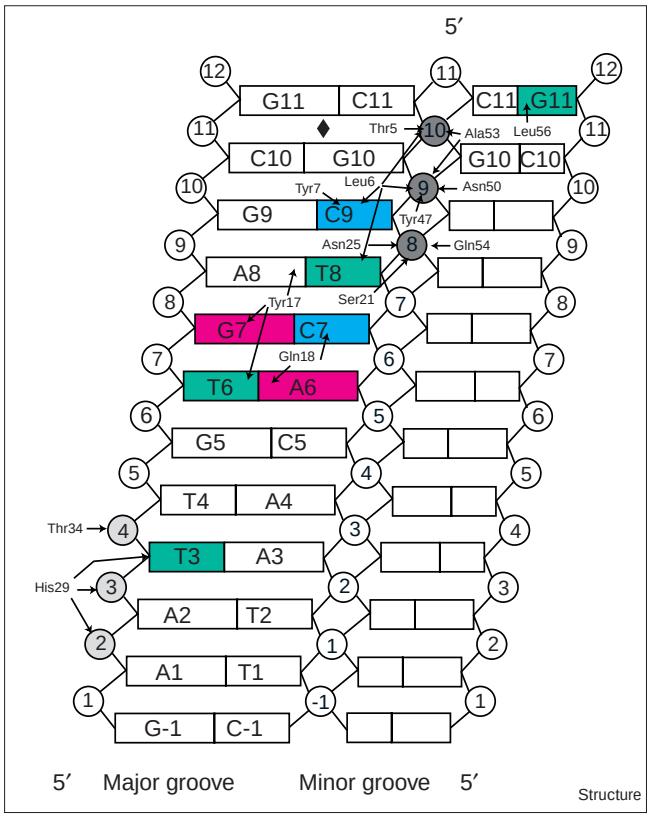
The differences in the interactions between the HTH motif of the headpieces and the half-site and full *lac* operator sequences can be largely explained by the distortions in the DNA structure upon binding to the full *lac* operator. Opening of the minor groove causes changes at the surface of the major groove and, as a consequence, changes in the protein–DNA interactions. For example, because of a steric clash between Tyr17, Leu6 and the DNA bases in the HP62–DNA complex Tyr17 adopts a different conformation than in the HP56–DNA complex. The nice aromatic-ring-stacking of Tyr7 and Tyr17, as seen in the

Figure 4



Intermolecular interactions in the HP62–DNA complex. (a) View of the interactions of the globular subdomain of HP62 (residues 1–49; red) with a half-site of the *lac* operator (blue). Residues involved in protein–DNA interactions are shown in yellow. (b) Protein–protein and protein–DNA interactions at the dimer interface of the HP62–DNA complex (residues 50–59 and the central six base pairs). Residues involved in protein–protein interactions are colored green, residues involved in protein–DNA interactions are colored yellow, and residues involved in both types of interactions are colored red. This figure was generated using the program Biosym InsightII.

Figure 5



Schematic drawing of all the protein–DNA interactions observed in the HP62–DNA complex. A cylindrical projection of the DNA is made, showing both the major and minor grooves of the *lac* operator. Bases involved in hydrophobic interactions are shown in green, bases involved in hydrogen-bonding interactions are shown in red, bases involved in both hydrophobic and hydrogen-bonding interactions are colored blue. The phosphate backbone groups and sugars contacted from the major groove are shaded light grey, the phosphate backbone groups and sugars contacted from both major and minor grooves are shaded dark grey.

HP56–DNA complex, is no longer possible in the HP62–DNA complex; this is in agreement with the clear differences in chemical shifts of the Tyr17 ring protons in the two complexes. Another contribution to differences in the structures of the HP56–DNA and HP62–DNA complexes might come from differences in the methods used to determine the structures. In the structure determination of the HP56–DNA complex, the conformation of the DNA was restricted using position restraining of the DNA backbone, thereby indirectly restricting the adaptability of the major groove to allow higher complementarity of the protein and DNA surfaces.

The loop following the recognition helix is, as shown by several studies [4,9,11], involved in interaction with the DNA. As mentioned previously, the interactions of Gln26 with the DNA are absent in the HP62–DNA complex. The important interactions between the loop and the DNA involve only residues 29–31, which is in agreement with genetic data that show low mutation sensitivity for residues 26–28, whereas residues 29–31 are sensitive to mutations [13]. His29 directly contacts the nucleotides Ade2 and Thy3 through apolar and hydrogen-bonding interactions. The cluster His29–Val30–Ser31 is, similar to the Pur repressor–operator complex, involved in a network of water-mediated hydrogen bonds to the phosphates of Ade2, Thy3 and Thy4. We observe these interactions in 21–24% of the analyzed coordinate sets of the rMD simulations. The presence of bound water molecules is supported by observations of NOE data, but could not be proven unambiguously because of the presence of the fast exchanging OH and NH groups in or near residues 28–31 [19].

It is important to note that the loop has some flexibility in the free state that provides the protein with some additional adaptability for binding DNA [11]. This flexibility, which is decreased upon DNA binding, appears to be present also in the purine repressor in which the structures of the headpieces show a slight disorder in the loop region [7]. In contrast, the NMR structure of the fructose repressor has a completely rigid connecting region, which was suggested to be required for the recognition of several pseudo-palindromic *fru* (fructose) operators [20].

In the third helix, which is important mainly for the stability of the headpiece, only Thr34 shows interactions with the DNA. Just beyond this helix important and stable hydrogen bonds to the DNA backbone are formed by the Tyr47 sidechain; this is similar to the Pur repressor–operator complex. For Lac repressor this hydrogen bonding was also present in rMD simulations of the HP56–DNA complex and is consistent with the clear observation of the OH proton signal of Tyr47 and its NOEs to the DNA in the NMR spectra of the HP62–DNA complex. The importance of this tyrosine residue is further reflected in its intolerance to mutations and its highly conserved nature in the LacI family of repressors [13,21,22]. Tyr47, therefore, has a double role; as well as being essential for the stability of the globular structure of the headpiece it is involved in protein–DNA interactions. Figure 4a shows a view of all the interactions between the DNA and the globular subdomain of HP62 (residues 1–49) as discussed above.

Interactions in the minor groove

The interactions of HP62 with the minor groove of the *lac* operator are all formed by residues of the hinge helices. Genetic data show that all residues in the hinge region are highly sensitive to amino acid replacements [13]. The conserved residues are Asn50, Ala53, Gln54 and Leu56 [22]. The importance of these residues becomes clear from the analysis of the structures of the HP62–DNA complex. The minor groove of the DNA is entered at the start of the hinge helix by Asn50, which anchors the helix to the

DNA backbone by hydrogen bonds involving both its backbone amide and sidechain NH_2 groups (Figure 4b). Ala53 and Leu56 are both involved in strong protein–protein and protein–DNA hydrophobic interactions that are similar to those seen in the Pur repressor–DNA complex. Leu56 intercalates in the central base-pair step of the operator and is largely responsible for the distortions of the DNA structure (Figure 4 and see the Supplementary material). In some members of the LacI-GalR family, TreR and CscR, this leucine is replaced by methionine. A mutant purine repressor, in which the intercalating leucine was replaced by a methionine, shows a similar binding mode and distortion of the *pur* operator as in the wild-type complex [23]. The Leu→Met mutation was in fact the only mutation found to retain repressor function *in vivo*. The only other residue that is found at the same position in the hinge helix in the LacI-GalR family is a valine in *Cyt*R. Although this variation can be considered conservative, a mutant Leu→Val purine repressor fails in its *in vivo* repressor function. Molecular modeling studies indicate that it is the inability of a valine sidechain to insert into the central base-pair step of the *pur* operator that abolishes the repressor function. The tolerance for the valine residue in CytR can be explained by the markedly different DNA-binding mode that is found in this repressor, in which the need for hinge helix formation and its minor groove insertion is absent. In contrast, a flexible hinge region is required for recognition of the different CytR operators that have half-site separations varying from 1–13 base-pairs [23,24].

Interestingly, Gln54 appears to be involved in hydrogen bonding with its sidechain NH_2 group to the DNA backbone at Thy8, very similar to the interaction formed by the sidechain of Arg52 (consensus residue) in the Pur repressor–operator complex. Although no NOEs were found between the Gln54 sidechain and the DNA, the NMR linewidths and relaxation parameters do indicate a decreased mobility of this sidechain upon DNA binding. Furthermore, the proton and nitrogen chemical-shift changes are indicative of hydrogen-bond formation, in agreement with the results from the structure calculations. In addition to the protein–protein interactions at the dimer interface, the electrostatic anchoring interactions between Gln54 and the DNA backbone might provide extra stability for the hinge helices. As described previously the induction of Lac repressor causes changes at the hinge region, thereby disrupting the helix-stabilizing interactions. The subsequent unfolding of the hinge helices decreases the affinity for the operator and the *lac* genes to be transcribed, demonstrating that structurally unstable domains can be functionally important in gene regulation.

Biological implications

The Lac repressor protein system is an important example of how gene regulation at the transcriptional level can be achieved by inducible switching between molecular conformations. In the present study the high-resolution structure of the complete DNA-binding domain of the Lac repressor in complex with a symmetrical *lac* operator sequence has been determined. In this dimeric complex the two DNA-binding domains induce a distinct bending of the DNA, resulting in significant differences in protein–DNA interactions as compared with the previously studied half-site complex. Analysis of the dimeric complex provides a better structural basis for understanding the genetic data on the specificity of the Lac repressor–operator interaction. Further, the structure provides detailed information on the so-called hinge helices, which connect the DNA-binding domain to the inducer-binding core domain, form the interface between the two DNA-binding domains, and induce the observed bending of the operator. Protein–protein and protein–DNA interactions that stabilize the interface and therefore the repressed state are delineated in atomic detail. The hinge helices are essential for the high-affinity binding of the repressor to the operator and play a crucial role in the induction of the *lac* operon. Upon induction, the interface of the DNA-binding domains is disrupted and as a result the hinge helices unfold and lose the tight interaction with the DNA. As discussed elsewhere, the emerging picture can be extrapolated to the interaction of the Lac repressor with the wild-type operator and provides a basic understanding of molecular interactions that determine the affinity and DNA bending observed in binding of Lac repressor to operator mutants. Most importantly, the induced DNA bending and the interaction between hinge helices in the wild-type complex is comparable with that found in the symmetrical complex [25].

Materials and methods

NMR sample preparation

Cloning, expression and purification of Lac HP62 was performed analogously to the method described by Slijper for Lac HP56 [26]. Unlabeled proline was added to all media and therefore Pro3 and Pro49 of HP62 were always present as $^{12}C/^{14}N$ amino acids in all samples. The palindromic 22 bp (5′-GAATTGTGAGCGCTCACAATTC-3′) *lac* operator DNA fragment was purchased at Carl Roth GmbH (Germany) and further purified on a Q-sepharose (Pharmacia) column. The following samples were used for NMR measurements: 3 mM ^{15}N-labeled HP62, 1.5 mM 22 bp duplex *lac* operator DNA; 2.8 mM $^{15}N/^{13}C$-labeled HP62, 1.4 mM 22 bp duplex *lac* operator DNA; 3 mM 10% ^{13}C-labeled HP62, 1.5 mM 22 bp duplex *lac* operator DNA; 2.8 mM of mixed $^{15}N/^{13}C$ HP62 and $^{14}N/^{12}C$ HP62 (stoichiometry = 1:1) and 1.4 mM 22 bp duplex *lac* operator DNA. All NMR samples contained 0.02 M KCl, 0.01 M potassium phosphate buffer pH 6.1 in 95% H_2O/5% D_2O. The ^{15}N-labeled complex was dissolved in either 95% H_2O/5% D_2O or in 100% D_2O. Trace amounts of NaN_3 were added to all NMR samples as preservative.

NMR spectroscopy

NMR spectra were recorded at 315K on Varian Unity+ 750 MHz, Varian Unity Inova 500 MHz, Bruker AMXT-600 and Bruker AMX-500 spectrometers equipped with triple-resonance gradient probes. All NMR spectra were processed using the NMRPipe software package

[27] and analyzed with the NMR analysis program REGINE on Silicon Graphics workstations. NMR experiments performed to obtain backbone and sidechain resonance assignments of HP62 were 3D-HNCA, 3D-HN(CO)CA, 3D-CBCA(CO)NNH, 3D-H(C)CH-DIPSY, 3D-(H)CCH-DIPSY, 3D-HC(C)H-DIPSY, 3D-(^{1}H-^{15}N)-NOESY-HSQC, 3D-(^{1}H-^{13}C)-NOESY-HSQC, 3D-(^{1}H-^{15}N)-TOCSY-HSQC, 3D-(^{1}H-^{15}N)-HMQC-NOESY-HSQC, 2D-NOE, 2D-TOCSY and were performed essentially as described in [28]. Stereospecific assignments of the methyl protons in the pro-chiral centers of all valine and leucine residues were obtained using the method described by Neri *et al.* [29]. In addition, 2D simultaneous ^{13}C-^{15}N double-half filter NOE experiments [30] were used for resonance assignments the unlabeled prolines at positions 3 and 49.

DNA assignments were obtained from 2D-NOE and the abovementioned simultaneous ^{13}C-^{15}N double-half filter NOE experiments. The sequential resonance assignment of the *lac* operator DNA, based on base-H6/H8 to sugar-H1′ NOEs [31], was nearly complete. The only missing NOE was between the C10-H1′ and G10-H8 protons because of an increased distance caused by the base-pair roll in the center of the operator upon HP62 binding. Most H2′, H2″, H3′ and some of the H4′, H5′ and H5″ resonances could be assigned on the basis of intraresidue and sequential NOEs. Additional assignments of some of the ribose protons could be made after initial structure calculations based on NOEs observed to neighboring assigned protons.

Protein–DNA interactions were assigned from 2D time-shared ^{13}C-^{15}N double-half filter-, 2D-NOE- and 3D-NOESY-HSQC experiments. The NMR spectra showed numerous NOEs between the hinge-helices and the central part of the *lac* operator that could be assigned unambiguously. In addition, most NOEs previously found in the half-site complex of Lac HP56 and an 11 base-pair *lac* operator sequence [5] were also found in the NMR spectra of the HP62–DNA complex. A number of additional protein–DNA NOEs could be assigned after the first structure calculations. Based on the calculated structures, protein–DNA NOEs between residues in the hinge-helices and the center of the *lac* operator could be assigned unambiguously, except for NOEs originating from Leu56 and Leu56′ methyl groups. Table S4 in the Supplementary material summarizes all the interactions observed between HP62 and the *lac* operator and which were used in the final stage of the structure calculations.

Protein–protein interactions were identified in a 3D-^{13}C-filtered-NOESY-HSQC as described by Zwahlen *et al.* [32] on a sample containing 50% unlabeled/50% ^{15}N-^{13}C labeled HP62 in complex with *lac* operator. The protein–protein interactions are listed in Table S5.

Distance restraints

Experimental distance restraints were derived from 2D-NOE spectra in H_2O and D_2O (50 ms, 80 ms and 100 ms mixing times), 3D-(^{1}H-^{15}N)-NOESY-HSQC (100 ms mixing time), 3D-(^{1}H-^{13}C)-NOESY-HSQC (40 ms mixing time), 3D-(^{1}H-^{15}N)-HMQC-NOESY-HSQC (100 ms mixing time) and 2D-double-half filtered NOE (75 ms mixing time) spectra. The restraints were calibrated using known distances in proteins and DNA (α-helical N-N, α-N, α-N(i,i + 3), α-N(i,i + 4) and Cytosine H5–H6 distances) and divided into three classes having upper bounds of 2.8 Å, 3.5 Å or 4.5 Å. In the structure calculations the NOEs were treated as $\Sigma(r^{-6})^{-1/6}$ sums. Where appropriate, upper bounds for NOEs involving diastereotopic groups were corrected as described by Fletcher *et al.* [33]. No lower bounds were used for experimentally derived distance restraints except for lower bound restraints on DNA-ribose H5′/H5″–H2′/H2″ distances and ribose H5′/H5″ to base H8/H6 distances obtained by the method of Kim *et al.* [34]. Hydrogen-bonding restraints for α-helical regions in HP62 were implemented as donor–acceptor distances of 2.30–3.50 Å and proton–acceptor distances of 1.70–2.50 Å. Hydrogen-bonding restraints for the DNA were implemented as donor–acceptor distances of 2.60–3.30 Å and proton-acceptor distances of 1.60–2.20 Å. An artificial distance restraints set, used in the docking of HP62 on the DNA, was obtained for the *lac* operator by generating a list of all proton–proton distances smaller than 5 Å in the operator B-DNA form and defining lower and upper bounds as ± 5% of the distances.

The protein–DNA distance restraints include two hydrogen-bonding restraints between HP62 and operator. The observation of the Tyr47 OH and its NOE to the H3′ proton of Cyt9 indicate that it is hydrogen bonded to the phosphate-oxygen atoms of the DNA backbone. The resonances of the sidechain NH_2 of Asn25 show a striking shift in the ^{1}H-^{15}N-HSQC spectrum of the complex, indicating hydrogen bonding of this group. Our initial structure calculations and the rMD study of Chuprina *et al.* [6] show that this hydrogen bonding occurs to the backbone phosphate oxygen atoms of Thy8 and was therefore included in the final structure calculations. These hydrogen-bonding restraints for the Asn25 and Tyr47 sidechain NH_2 and OH groups to the phosphate backbone of the DNA were defined as donor–acceptor upper bounds of 3.50 Å and proton–acceptor upper bounds of 2.50 Å, where the acceptor could be either one of the two available phosphate oxygen atoms. The protein–DNA distance restraints involving the methyl groups of Leu56 and Leu56′ were treated as ambiguous restraints between protons of one HP62 monomer and protons of both DNA strands.

All protein–protein NOEs were initially treated as ambiguous intra/interprotein restraints. Note that the interactions observed between the Val52 methyl groups of one monomer to the same group of the other monomer were not used as distance restraints in the structure calculations since these interactions occur as diagonal peaks in the 2D-NOE spectra. Furthermore, NOEs originating from sidechains at the interface (Val52, Ala53 and Leu56) that were not found in the filter-experiments were also used as ambiguous restraints in the initial calculations. After iteratively improving the structures, NOEs were assigned as either truly intermonomer, intramonomer or having both contributions.

Structure determination of the HP62 monomer

All structure calculations were performed with X-PLOR 3.851 [35] on Silicon Graphics workstations using the standard X-PLOR parallhdg.pro parameter set. A simulated annealing protocol [36] followed by an extra SA refinement was used to calculate the structure of the monomeric HP62 based on 676 experimental distance restraints. There were 15 restraints derived from NOEs that were identified as both intra- and intermonomer NOEs treated as intramonomer restraints in the calculation of the HP62 monomer structure. An initial ensemble of 250 structures was generated from a covalent template structure with randomized backbone φ and ψ angles. In this step hydrogen-bonding restraints were included for all four helices of HP62. In the additional SA refinement step hydrogen-bonding restraints were only used for the C-terminal hinge-helix.

Docking of HP62 onto lac operator DNA

Of the 250 calculated structures, 46 were selected based on good stereochemistry, good Ramachandran plot qualities for the structured regions and no restraint violations larger than 0.5 Å. The selected structures were duplicated, separated in space, rotated by 180° and docked onto the 22 bp *lac* operator B-DNA using simulated annealing. The temperature of the system was decreased in 10K steps from 300K to 100K, using 40,000 cooling steps of 0.005 ps. NOE force constants for protein–protein restraints, ambiguous inter/intraprotein restraints and protein–DNA restraints were increased from 2 kcal/mol^{-1} Å^{-2} to 50 kcal/mol^{-1} Å^{-2}. All other NOE force constants were held constant at 50 kcal/mol^{-1} Å^{-2} except those for the intra-HP62 NOEs, which were set to 100 kcal/mol^{-1} Å^{-2}. In addition to experimental distance restraints, noncrystallographic symmetry restraints were incorporated for both the HP62 monomers and the *lac* operator DNA sequence. Artificial distance and planarity restraints for the DNA were incorporated in the docking calculations in order to keep the DNA close to the B-form but allowing a slight bend necessary to accommodate the two headpiece molecules on the DNA. No experimental distance restraints for the DNA were used during the docking. The protein–DNA distance restraints were all set to upper

bound restraints of 7 Å during the docking, again in order to allow a proper accommodation of the headpieces on the DNA. Hydrogen-bonding restraints for the C-terminal α helix were used to keep this helix stable during the docking calculation. Selection of structures was done on the basis of good stereochemistry, no restraint violations larger than 0.4 Å and proper positioning of the His29 sidechain in the major groove of the DNA.

Restrained molecular dynamics

The 14 best structures were selected for rMD simulations in water using the CHARMM22 force-field for proteins and nucleic acids [37]. The structures were solvated in a rectangular water box with a minimum solute to wall distance of 10 Å. This was done by translating an equilibrated box of TIP3P water molecules [38] around the protein–DNA complex and deleting all water molecules within 2.3 Å of any solute heavy atom. The systems were then energy minimized first by 100 steps of conjugate gradient minimization with the solute fixed and subsequently for 250 steps using harmonic position restraints on the solute with force constants of 100 kcal/mol^{-1}Å^{-2}. After minimization, the system was neutralized by randomly replacing water molecules at least 5 Å away from the solute by 36 sodium atoms, followed by the same energy minimization protocol as before.

The structures were then refined by 24 ps of rMD at constant volume and temperature (NVT) under periodic boundary conditions. Covalent bond lengths were constrained with the procedure SHAKE [39] with a relative geometric tolerance of 0.0001. The non-bonded pair list was calculated with a cutoff distance of 13 Å and updated every time an atom moved by more than 0.5 Å. Electrostatic interactions were truncated at 12 Å by force shifting [40] and a potential-switching function between 10 and 12 Å was used for van der Waals interactions. The integration time was 0.002 ps. Initial velocities were taken from a Maxwellian distribution at 300K and the temperature was maintained by coupling the system to an external bath at 300K with a coupling constant of 100 ps^{-1}. In order to keep the peptide omega angles within two times the standard deviation from the ideal value the force-constants for dihedral angles dealing with the peptide planarity (groups O, C, N, HN and O, C, N, CA) were modified in the force-field to 50 kcal/mol^{-1} Å^{-2}. During the last 4 ps of the simulation the force-constant for the NOE-potential was increased from 10–20 kcal/mol^{-1} Å^{-2}. The resulting structures were finally energy minimized using 200 steps of minimization. The distance restraints used in the rMD simulation and the final energy minimization included all experimentally derived distance restraints, additional hydrogen-bonding restraints for DNA base-pairing and noncrystallographic symmetry restraints for the HP62 and *lac* operator.

Structural analysis

The quality of the protein structures was analyzed using PROCHECK_NMR [41]. Three of the 14 rMD structures were rejected on the basis of the occurrence of a D-amino acid or bad stereochemistry at residue 50 where the protein structure crosses the backbone of the DNA. The DNA helical parameters were calculated using the program SCHNAaP [42]. Analysis of the protein–protein and protein–DNA interactions was done on the last 3 ps of the trajectories of the rMD simulations of the 11 final structures, which comprised 77 sets of coordinates. Apolar interactions were analyzed by counting the occurrence of C–C distances smaller than 4 Å. The geometric criteria for hydrogen-bond analysis were an angle between donor, hydrogen and acceptor larger than 125° and a donor to acceptor cutoff distance of 3.5 Å.

Accession numbers

The coordinates have been deposited in the PDB with the accession code 1cjg. The PDB NMR restraint entry code is r1cjgmr.

Supplementary material

Supplementary material including tables listing all intermolecular interactions observed in the HP62–DNA complex and DNA helical parameters, is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

This work was financially supported by the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (CW-NWO). Giuseppe Melacini was supported by the European Community (EU Grant ERBFMGECT950002).

References

1. Lewis, M., *et al.*, & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254.
2. Friedman, A.M., Fischmann, T.O. & Steitz, T.A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* **268**, 1721-1727.
3. Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α-helices. *Science* **266**, 763-770.
4. Lamerichs, R.M.J.N., *et al.*, & Rüterjans, H. (1989). ^{1}H NMR study of a complex between the lac repressor headpiece and a 22 base pair symmetric *lac* operator. *Biochemistry* **28**, 2985-2991.
5. Lamerichs, R.M.J.M., Boelens, R., van der Marel, G.A., van Boom, J.H. & Kaptein, R. (1990). Assignment of the ^{1}H-NMR spectrum of a lac repressor headpiece–operator complex in H_2O and identification of NOEs. Consequences for protein–DNA interaction. *Eur. J. Biochem.* **194**, 629-637.
6. Chuprina, V.P., Rullman, J.A.C., Lamerichs, R.M.J.N., van Boom, J.H., Boelens, R. & Kaptein, R. (1993). Structure of the complex of lac repressor headpiece and an 11 base-pair half operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* **234**, 446-462.
7. Nagadoi, A., *et al.*, & Nichimura, Y. (1995). Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. *Structure* **3**, 1217-1224.
8. Slijper, M., Bonvin, A.M.J.J., Boelens, R. & Kaptein, R. (1996). Refined structure of lac repressor headpiece (1–56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to *lac* operator. *J. Mol. Biol.* **259**, 761-773.
9. Spronk, C.A.E.M., Slijper, M., van Boom, J.H., Kaptein, R. & Boelens, R. (1996). Formation of the hinge-helix in the lac repressor is induced upon binding to the *lac* operator. *Nat. Struct. Biol.* **3**, 916-919.
10. Zwieb, C., Kim, J. & Adhya, S. (1989). DNA bending by negative regulatory proteins: *Gal* and *Lac* repressors. *Genes Dev.* **3**, 606-611.
11. Slijper, M., *et al.*, & Kaptein, R. (1997). Backbone and sidechain dynamics of lac repressor headpiece (1–56) and its complex with DNA. *Biochemistry* **36**, 249-254.
12. Kleina, L.G. & Miller, J.H. (1990). Genetic studies of the lac repressor XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *J. Mol. Biol.* **212**, 295-318.
13. Markiewicz, P., Kleina, L.G., Cruz, C., Ehret, S. & Miller, J.H. (1994). Genetic studies of the lac repressor XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as 'spacers' which do not require a specific sequence. *J. Mol. Biol.* **240**, 421-433.
14. Zuiderweg, E.R.P., Boelens, R. & Kaptein, R. (1985). Stereospecific assignments of ^{1}H-NMR methyl lines and conformation of valyl residues in the lac repressor headpiece. *Biopolymers* **24**, 601-611.
15. Falcon, C.M., Swint-Kruse, L. & Matthews, K.S. (1997). Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. *J. Biol. Chem.* **272**, 26818-26821.
16. Ebright, R.H. (1986). Evidence for a contact between glutamine-18 of lac repressor and base pair 7 of *lac* operator. *Proc. Natl Acad. Sci. USA* **83**, 303-307.
17. Sartorius, J., Lehming, N., Kisters, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1989). Lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to *lac* operator variants with multiple exchanges. *EMBO J.* **8**, 1265-1270.
18. Sartorius, J., Lehming, N., Kisters-Woike, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1991). The roles of residues 5 and 9 of the recognition helix of lac repressor in *lac* operator binding. *J. Mol. Biol.* **218**, 313-321.
19. Otting, G. (1997). NMR studies of water bound to biological molecules. *Progr. Nucl. Magn. Reson. Spect.* **31**, 259-285.
20. Penin, F., *et al.*, & Deléage, G. (1997). Three dimensional structure of the DNA-binding domain of the fructose repressor from *Escherichia coli* by ^{1}H and ^{15}N NMR. *J. Mol. Biol.* **270**, 496-510.

21. Weickert, M.J. & Adhya, S. (1992). A family of bacterial regulators homologous to Gal and lac repressors. *J. Biol. Chem.* **267**, 15869-15874.
22. Nguyen, C.C. & Saier, M.H.J. (1995). Phylogenetic, structural and functional analysis of the LacI-GalR family of bacterial repressors. *FEBS Lett.* **377**, 98-102.
23. Arvidson, D.N., Lu, F., Faber, C., Zalkin, H. & Brennan, R.G. (1998). The structure of PurR mutant L54M shows an alternative route to DNA kinking. *Nat. Struct. Biol.* **5**, 436-441.
24. Pedersen, H. & Valentin-Hansen, P. (1997). Protein-induced fit: the CRP activator protein changes sequence-specific DNA recognition by the *Cyt*R repressor, a highly flexible LacI member. *EMBO J.* **16**, 2108-2118.
25. Spronk, C.A.E.M., *et al.*, & Kaptein, R. (1999). Hinge-helix formation and DNA-bending in various *lac* repressor–operator complexes. *EMBO J.,* in press.
26. Slijper, M. (1996). *NMR-studies of protein–DNA interactions in the* lac *operon*. PhD thesis, Utrecht University, The Netherlands.
27. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.
28. Cavanagh, J., Fairbrother, W.J., Palmer, A.G. & Skelton, N.J. (1996). Protein NMR Spectroscopy: Principles and Practice. Academic press, California, USA.
29. Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510-7516.
30. Slijper, M., Kaptein, R. & Boelens, R. (1996). Simultaneous ^{13}C and ^{15}N isotope-editing of biomolecular complexes. Application to a mutant lac repressor headpiece DNA complex. *J. Mag. Res. B* **111**, 199-203.
31. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. John Wiley and sons, New York.
32. Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R. & Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/*boxB* RNA complex. *J. Am. Chem. Soc.* **119**, 6711-6721.
33. Fletcher, C.M., Jones, D.N.M., Diamond, R. & Neuhaus, D. (1996). Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR* **8**, 292-310.
34. Kim, S-G., Lin, L-J. & Reid, B.R. (1992). Determination of nucleic acid backbone conformation by ^{1}H NMR. *Biochemistry* **31**, 3564-3574.
35. Brünger, A.T. (1992). X-PLOR. A system for X-ray Crystallography and NMR. Yale University Press, New Haven, CT.
36. Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins* **17**, 297-309.
37. MacKerell, Jr. A.D., *et al*., & Karplus, M. (1992). Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB J.* **6**:A143.
38. Jorgenson, W.L., Chandrasekhan, J., Madura, J., Impley, R.W. & Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935.
39. Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. (1977). Numerical integration of the cartesian equations of motions of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327-341.
40. Steinbach, P.J. & Brooks, B.R. (1994). New spherical-cutoff methods for long range forces in macromolecular simulation. *J. Comp. Chem.* **15**, 667-683.
41. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. & Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486.
42. Lu, X-J., El Hassan, M.A. & Hunter, C.A. (1997). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J. Mol. Biol.* **273**, 668-680.

Supplementary material

The solution structure of Lac repressor headpiece 62 complexed to a symmetrical *lac* operator

Christian AEM Spronk, Alexandre MJJ Bonvin, Plachikkat K Radha, Giuseppe Melacini, Rolf Boelens and Robert Kaptein

Structure December 1999, 7:1483–1492

Figure S1

Averaged helical parameters of *lac* operator in complex with HP62. Shown are only roll, twist and slide that deviate most at the center of the DNA and are indicative for the observed bend of 45° in the DNA helical axis upon DNA binding of Lac repressor.

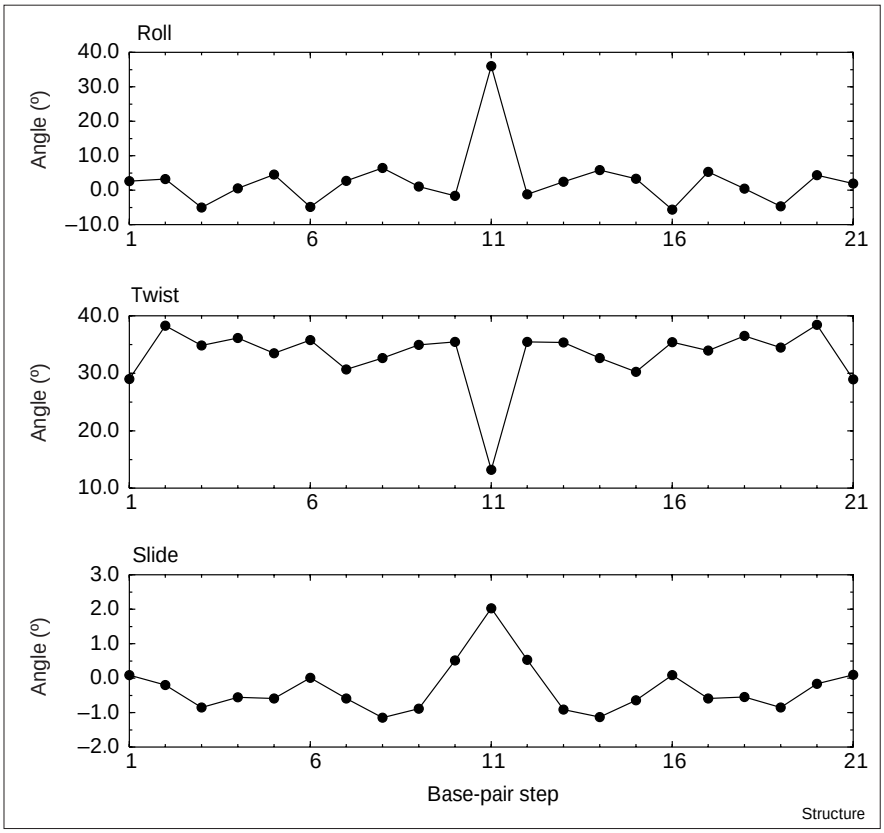


Table S1

Apolar interactions between HP62 monomers.

Residues	Contact type	Atoms	Occurrence (%)
Val52–Val52′	Backb–side	C′–Cγ1	27
Val52–Val52′	Backb–side	Cα–Cγ1	26
Val52–Val52′	Side–side	Cγ1–Cγ1	88
Val52–Gln55′	Side–side	Cγ1–Cβ	97
Val52–Leu56′	Backb–side	C′–Cδ1	67
Val52–Leu56′	Side–side	Cβ–Cδ1	74
Val52–Leu56′	Side–backb	Cγ1–Cα	60
Val52–Leu56′	Side–side	Cγ1–Cβ	94
Val52–Leu56′	Side–side	Cγ1–Cδ1	84
Ala53–Leu56′	Backb–side	Cα–Cδ1	99
Ala53–Leu56′	Backb–side	Cα–Cδ2	99
Ala53–Leu56′	Side–side	Cβ–Cδ1	92
Leu56–Leu56′	Side–side	Cβ–Cδ2	51
Leu56–Leu56′	Side–side	Cδ2–Cδ2	60

Apolar contacts were analyzed by counting the occurrence of C–C distances smaller than 4 Å in 77 coordinate sets obtained from the last 3 ps of the MD trajectories of 11 structures. Listed are only those interactions with an average occurrence in the dimer of more than 25%. Backb, backbone contact; side, sidechain contact.

Table S2

Apolar contacts between HP62 and the *lac* operator.

Residues	Contact type	Atoms	Occurrence (%)
Thr5–Gua10	Side–backb	Cγ2–C5′	46
Leu6–Gua10	Side–sugar	Cδ2–C2′	100
Leu6–Gua10	Side–backb	Cδ2–C3′	37
Leu6–Cyt9	Side–base	Cδ1–C6	88
Leu6–Cyt9	Side–base	Cδ2–C6	81
Leu6–Thy8	Side–base	Cδ1–C5m	62
Tyr7–Cyt9	Side–base	Cζ–C5	43
Tyr17–Thy6	Side–base	Cε2–C5m	26
Tyr17–Thy8	Side–base	Cδ1–C5m	26
Gln18–Cyt7	Side–base	Cδ–C5	29
His29–Ade2	Side–sugar	Cδ2–C2′	75
His29–Ade2	Side–sugar	Cε1–C2′	94
His29–Ade2	Side–backb	Cε1–C3′	81
His29–Ade2	Side–backb	Cε1–C5′	48
His29–Ade2	Side–sugar	Cγ–C2′	44
His29–Thy3	Side–base	Cδ2–C5m	61
Ala53–Gua10	Side–sugar	Cβ–C1′	73
Ala53–Gua10	Side–backb	Cβ–C3′	34
Ala53–Gua10	Side–backb	Cβ–C4′	88
Ala53–Cyt9	Side–backb	C′–C4′	95
Ala53–Cyt9	Side–backb	Cβ–C4′	100
Ala53–Cyt9	Side–backb	Cβ–C5′	82
Leu56–Gua11	Side–base	Cδ2–C2	85
Lys59–Thy8	Side–backb	Cγ–C5′	25
Lys59–Cyt7	Side–backb	Cβ–C5′	31

Apolar contacts were analyzed by counting the occurrence of C–C distances smaller than 4 Å in 77 coordinate sets obtained from the last 3 ps of the MD trajectories of 11 structures. Listed are only those interactions with an average occurrence in the dimer of more than 25%. Side, sidechain; backb, backbone.

Table S3

Hydrogen bonds between HP62 and *lac* operator*.

Residues (donor–acceptor)	Contact type (donor–acceptor)	Atoms (donor–proton–acceptor)	Occurrence (%)
Leu6–Cyt9	Backb–backb	N–HN–O2P	100
Cyt9–Tyr7	Base–side	N4–H41–Oζ	96
Tyr17–Gua7	Side–base	Oζ–Hζ–O6	48
Cyt7–Gln18	Base–side	N4–H41–Oε1	50
Ade6–Gln18	Base–side	N6–H61–Oε1	27
Ser21–Thy8	Side–backb	Oγ–Hγ–O2P	48
Asn25–Thy8	Side–backb	Nδ2–Hδ21–O1P	30
Asn25–Thy8	Side–backb	Nδ2–Hδ22–O1P	29
His29–Thy3	Side–backb	Nδ1–Hδ1–O2P	26
Thr34–Thy4	Side–backb	Oγ1–Hγ1–O1P	62
Tyr47–Cyt9	Side–backb	Oζ–Hζ–O2P	83
Tyr47–Cyt9	Side–backb	Oζ–Hζ–O5’	39
Asn50–Cyt9	Backb–backb	N–HN–O1P	94
Asn50–Cyt9	Side–backb	Nδ2–Hδ21–O1P	56
Gln54–Thy8	Side–backb	Nε2–Hε21–O1P	66
Gln60–Cyt7	Backb–backb	N–HN–O1P	27

Hydrogen bonds were analyzed in 77 coordinate sets obtained from the last 3 ps of the MD trajectories of 11 structures. Listed are only those interactions with an average occurrence in the dimer of more than 25%. Backb, backbone; side, sidechain.

Table S4

Protein–DNA distance constraints.

Residue	Atoms	Nucleotide	Atoms	Distance (Å)
Thr5	Hβ	Gua10	H3′	5.00
Thr5	Hγ2*	Gua10	H3′	2.80
Leu6	Hδ1*	Gua10	H3′	4.50
Leu6	Hδ1*	Cyt9	H3′	3.50
Leu6	Hδ1*	Cyt9	H5′ + H5″	4.50
Leu6	Hδ1*	Cyt9	H5	4.50
Leu6	Hδ1*	Cyt9	H6	3.50
Leu6	Hδ1*	Thy8	H6	4.50
Leu6	Hδ2*	Gua10	H1′	4.50
Leu6	Hδ2*	Gua10	H3′	4.50
Leu6	Hδ2*	Cyt9	H3′	4.50
Leu6	Hδ2*	Cyt9	H5′ + H5″	4.50
Leu6	Hδ2*	Cyt9	H5	3.50
Leu6	Hδ2*	Cyt9	H6	2.80
Tyr7	Hδ*	Gua10	H8	4.50
Tyr7	Hε*	Gua10	H3′	6.00
Tyr7	Hε*	Gua10	H8	3.50
Tyr7	Hε*	Cyt9	H5	4.50
Tyr7	Hε*	Cyt9	H6	4.50
Tyr17	Hε* + Hδ*	Cyt9	H5	4.50
Tyr17	Hδ*	Thy8	C5m	2.80
Tyr17	Hε*	Thy8	C5m	3.50
Gln18	Hε*	Cyt7	H5	3.50
Asn25[†]	Hδ*	Thy8	O*P	2.50
Asn25[†]	Nδ2	Thy8	O*P	3.50
His29	Hβ*	Thy3	H6	4.50
His29	Hδ2	Ade2	H3’	4.50
His29	Hδ2	Thy3	C5m	4.50
His29	Hδ2	Thy4	C5m	4.50
His29	Hε1	Ade2	H2″	4.50
His29	Hε1	Ade2	H2′	4.50
His29	Hε1	Ade2	H3′	4.50
His29	Hε1	Thy3	C5m	4.50
Tyr47[†]	Hζ	Cyt9	O*P	2.50
Tyr47[†]	Oζ	Cyt9	O*P	3.50
Tyr47	Hζ	Cyt9	H3′	4.50
Ala53	Hα	Gua10	H1′	4.50
Ala53	Hα	Cyt9	H1′	3.50
Ala53	Hβ*	Gua10	H1′	2.80
Ala53	Hβ*	Cyt9	H1′	3.50
Ala53	Hβ*	Cyt9	H4′	2.80
Ala53	Hβ*	Cyt9	H6	4.50
Gln54	Hγ*	Cyt9	H4′	4.50
Gln54	HN	Cyt9	H4′	4.50
Leu56[#]	Hδ1*	Cyt10	H1′	4.50
Leu56[#]	Hδ1*	Gua10	H1′	4.50
Leu56[#]	Hδ2*	Cyt10	H1′	4.50
Leu56[#]	Hδ2*	Gua10	H1′	4.50
Leu56[#]	Hδ2*	Gua10	H1	3.50
Leu56[#]	Hδ2*	Gua10	NH2	4.50
Leu56[#]	Hδ2*	Cyt9	H1′	4.50

Table S4 continued

Protein–DNA distance constraints.

Residue	Atoms	Nucleotide	Atoms	Distance (Å)
Ala57	Hα	Cyt10	H1′	3.50
Ala57	Hβ*	Gua9	H1	4.50
Ala57	Hβ*	Cyt10	H1′	4.50
Ala57	Hβ*	Gua10	H1′	4.50
Ala57	Hβ*	Gua10	H1	4.50
Ala57	Hβ*	Gua10	NH2	4.50
Ala57	Hβ*	Cyt9	H1′	2.80
Ala57	Hβ*	Cyt9	H4′	4.50
Ala57	Hβ*	Thy8	H1′	2.80
Ala57	Hβ*	Thy8	H6	5.00

Distances are upper bound limits used in the structure calculations. Asterisks are used as wildcards for groups of atoms. [#]Indicates the use ambiguous constraints where no distinction can be made between atoms of the two DNA strands. [†]Indicates the use of hydrogen bonding restraints.

Table S5

Observed protein–protein NOEs.

Residue	Atoms	Residue	Atoms (Å)
Val52	Hα	Val52′	Hγ1*
Val52	Hβ	Val52′	Hγ**
Val52	Hγ1*	Val52′	Hγ1*
Val52	Hγ2*	Val52′	Hγ2*
Val52	Hγ**	Gln55′	Hβ*
Val52	Hγ1*	Leu56′	Hδ1*
Val52	Hγ**	Leu56′	Hα
Ala53	Hβ*	Leu56′	Hδ1*

Only NOEs observed in the filter experiments are listed. Asterisks are used as wildcards for groups of atoms. Double asterisks are used where, because of overlap, no distinction could be made between the two prochiral methyl groups of Val52.