

Specificity and Affinity of Lac Repressor for the Auxiliary Operators *O2* and *O3* Are Explained by the Structures of Their Protein–DNA Complexes

Julija Romanuka¹, Gert E. Folkers¹, Nikolaos Biris¹,
Evgeny Tishchenko², Hans Wienk¹, Alexandre M. J. J. Bonvin¹,
Robert Kaptein¹ and Rolf Boelens^{1*}

¹*Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands*

²*Department of Medical Genetics and Microbiology, Medical Sciences Building, The University of Toronto, Toronto, Ontario, Canada M5S 1A8*

Received 23 February 2009;
received in revised form
8 May 2009;
accepted 13 May 2009
Available online
18 May 2009

The structures of a dimeric mutant of the Lac repressor DNA-binding domain complexed with the auxiliary operators *O2* and *O3* have been determined using NMR spectroscopy and compared to the structures of the previously determined Lac–*O1* and Lac–nonoperator complexes. Structural analysis of the Lac–*O1* and Lac–*O2* complexes shows highly similar structures with very similar numbers of specific and nonspecific contacts, in agreement with similar affinities for these two operators. The left monomer of the Lac repressor in the Lac–*O3* complex retains most of these specific contacts. However, in the right half-site of the *O3* operator, there is a significant loss of protein–DNA contacts, explaining the low affinity of the Lac repressor for the *O3* operator. The binding mode in the right half-site resembles that of the nonspecific complex. In contrast to the Lac–nonoperator DNA complex where no hinge helices are formed, the stability of the hinge helices in the weak Lac–*O3* complex is the same as in the Lac–*O1* and Lac–*O2* complexes, as judged from the results of hydrogen/deuterium experiments.

© 2009 Elsevier Ltd. All rights reserved.

Edited by P. Wright

Keywords: lac headpiece; lac operators; NMR; protein–DNA recognition

Introduction

The *Escherichia coli* Lac repressor controls the expression of genes involved in lactose metabolism.^{1,2} Effective downregulation of these genes is achieved by the presence of multiple Lac repressor operator binding sites within the *lac* operon. The main operator *O1* overlaps with the *lac* promoter and is essential for the function of the *lac* operon. Furthermore, the two auxiliary operators *O2* and *O3*, located 401 bp downstream of *O1* and 92 bp upstream of *O1*,

respectively, contribute significantly to transcriptional repression. *O1* is indispensable, since mutation or deletion of *O1* leads to an almost complete loss of repression even in the presence of both auxiliary operators.^{3,4} Inactivation of either *O2* or *O3* results in a slight decrease in repression, while combined loss of both *O2* and *O3* leads to a significant (~70-fold) decrease in repression.⁴ This cooperativity can be well explained as the tetrameric protein functions as a dimer of dimers that can simultaneously bind to the *O1* operator and to either of the auxiliary *O2* and *O3* operators,^{5,6} creating one of two alternative DNA loops.^{7,8}

Mutational studies revealed that the Lac repressor has different affinities for various operators, resulting in distinct repression efficiency.⁹ The differences in the operator affinities of the Lac repressor can be ascribed to the variation in their sequences. *O1* and *O2* operators have similar base-pair compositions, while the *O3* sequence differs significantly (Fig. 1). Natural *lac* operators are pseudo-palindromic sequences, where symmetry is broken by variations in the sequence between the two half-sites and by

*Corresponding author. E-mail addresses:
r.boelens@uu.nl; r.boelens@chem.uu.nl.

Abbreviations used: HTH, helix–turn–helix; NOD, nonoperator DNA; H/D, hydrogen/deuterium; NOE, nuclear Overhauser enhancement; 2D, two-dimensional; NOESY, NOE spectroscopy; RDC, residual dipolar coupling; 3D, three-dimensional; TOCSY, total correlated spectroscopy; HSQC, heteronuclear single quantum coherence.

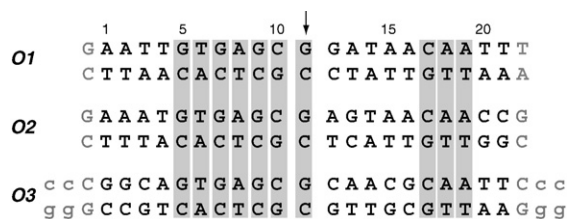


Fig. 1. Sequences of naturally occurring *lac* operators. The three natural operators *O1*, *O2*, and *O3* are aligned with the numbering referred to in the text and figures. Arrow denotes the central base pair. The bases conserved in all natural operators are highlighted. The two binding sites within each operator are referred to as the left (base pairs 1–10) and the right (base pairs 12–21). The extension of operator sequences is shown in gray.

insertion of the central G:C base pair (Fig. 1). The two half-sites differ significantly in their affinity for the Lac repressor when considered separately.¹⁰ Mutational analysis has suggested a greater overall contribution to binding by the left operator site because mutations in the left site sequence are more deleterious to repressor binding than those in the right site sequence.^{3,11}

The amino-terminal headpiece (residues 1–62) of the Lac repressor is responsible for operator recognition and binding. Residues 1–49 of the headpiece form a stable fold consisting of three helices. The first two helices comprise a canonical

helix–turn–helix (HTH) DNA-binding motif. The second (or recognition) helix of the HTH motif contains several residues that make contacts with DNA. The C-terminal residues Arg50–Gly58 of the headpiece form a so-called hinge region, which is unstructured in the absence of DNA. Initially, NMR structural studies were hindered by the unfavorable dynamics of the isolated monomeric Lac headpiece in complexes with natural operators probably due to its low affinity for them.¹² Inspired by the work of Falcon *et al.*,¹³ who introduced the mutation of Val by Cys at position 52 in the full-length Lac repressor, Kalodimos *et al.*¹⁴ prepared a covalently linked dimeric Lac headpiece (HP62V52C). This dimeric headpiece had an affinity for the natural operator in the picomolar range, comparable to that of the intact dimeric Lac repressor.¹⁴ Analysis of the structure of the HP62V52C–*O1* complex revealed a distinct binding mechanism of headpiece to the left and right halves of the operator.¹⁵ The global positioning of the dimer on the operator was dramatically asymmetric, resulting in a different pattern of specific contacts between the two half-sites demonstrating the intrinsic plasticity of the Lac headpiece. The structure of the left site of the complex was similar to that of the Lac headpiece bound to the fully symmetrical operator.¹⁶ In agreement with the crystal structure,¹⁷ hinge helices were folded and bound in the minor groove between base pairs 10 and 11, thereby introducing a kink in the DNA. This also confirmed a study by Spronk *et al.*, who showed

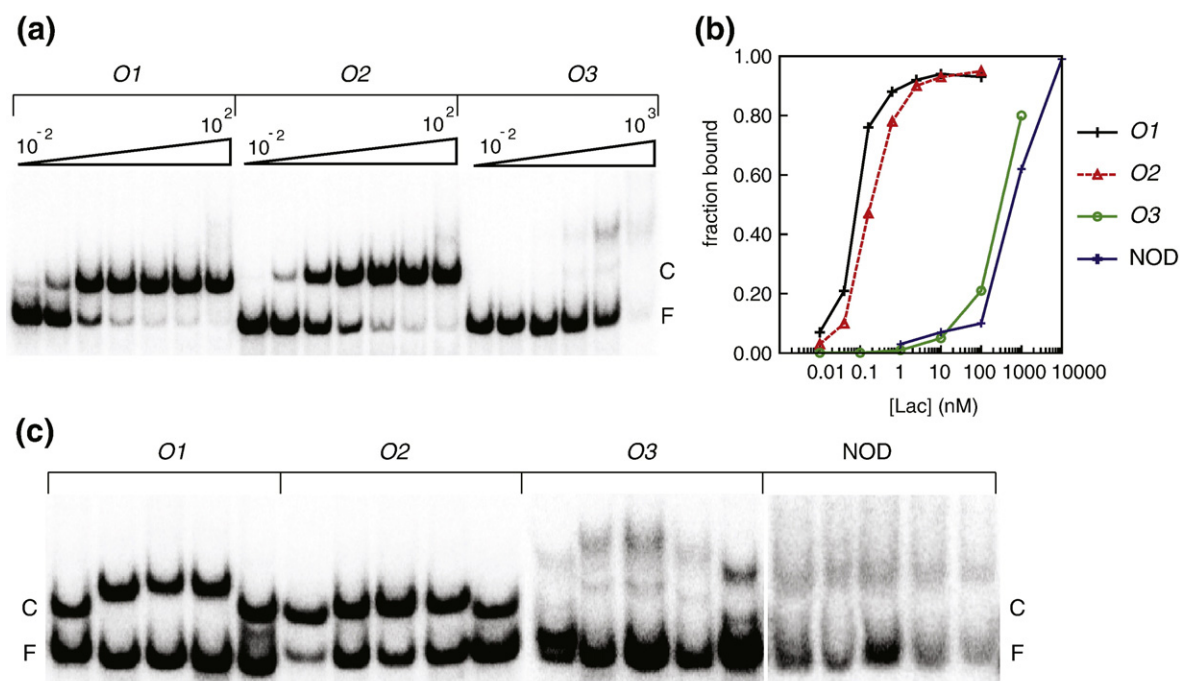


Fig. 2. DNA binding and bending experiments. (a and b) Representative DNA binding experiment using the indicated operators with a serial dilution of dimeric HP62V52C. The highest and lowest protein concentrations (nM) are indicated above each experiment. Different operator fragments are described in Fig. 1. (c) DNA bending experiment using the circular permutation method with the dimeric HP62V52C mutant. The free probe and the protein–DNA complex are indicated with F and C, respectively.

Table 1. Binding affinities and bend angles obtained with the dimeric Lac HP62 and various operators

	K_d^a (nM)	Bend ^b (°)	Bend ^c (°)
O1	0.05	25	33±8
O2	0.1	25	25±7
O3	100	40	14±9
NOD	200	<5	0

^a Binding affinities are presented as apparent K_d values (see Materials and Methods).

^b Bend angles were determined using the circular permutation method.

^c The average bend angles of the DNA operator in the structures of the ensemble as estimated with CURVES.

that the central base pair in natural *lac* operators acts as a spacer to create optimal spacing between two half-sites for the formation of the hinge helices and interaction with the DNA.^{12,16}

In an ongoing effort to understand the specificity and recognition of various operator sequences by the Lac repressor, we determined the NMR structures of the complexes of the dimeric Lac headpiece with its auxiliary operators O2 and O3. We compared these structures with that of the previously determined HP62V52C–O1¹⁵ and that of the Lac headpiece bound to a nonoperator DNA (NOD) fragment.¹⁸ Analysis of these complexes helps to understand how the Lac repressor recognizes its operators and explains significant differences in operator affinities.

Results and Discussion

Relative *in vitro* affinities and DNA bending

The binding affinities of the dimeric HP62V52C for natural *lac* operators and an artificial NOD fragment were compared using electrophoretic mobility shift assays (Fig. 2a and b; Table 1). The headpiece dimer binds the O1 and O2 operators with apparent equilibrium dissociation constants (K_d) of 0.05 and 0.1 nM, respectively. Thus, the dimeric Lac headpiece interacts with the O1 and O2 operators with comparable affinities. The O3 operator has a considerably lower K_d of 100 nM, surprisingly similar to the observed affinity for the NOD sequence (200 nM). On both operators, in agreement with the low affinity, dissociation of the protein–DNA complex during electrophoresis was observed, and on the O3 operator sequence, some additional complexes appeared to form, possibly due to nonspecific binding of HP62V52C (in surplus) to flanking NOD sequences.

To determine whether a headpiece can induce a bend on natural operators, we used the circular permutation method,¹⁹ essentially as performed before.¹⁶ The profile of the migration of the complexes with circularly permuted DNA fragments on a polyacrylamide gel showed that the dimeric headpiece bends both O1 and O2 operators

by an estimated angle of 25° (Fig. 2c; Table 1). The complexes formed on the O3 operator pose bend angles that are larger than but different from those of the other operator sequences and are close to experimental limits.

Stability of Lac–operator complexes from hydrogen/deuterium experiments

In order to establish the differences in the stabilities of the three Lac–operator complexes, we carried out hydrogen/deuterium (H/D) exchange experiments. A graph of protection factors (P) for all residues in the HP62V52C protein, defined as the ratio of the observed amide H/D exchange rates to the calculated rates of unprotected amino acids as explained in Materials and Methods, is shown in Fig. 3.

The protection factor values for the HTH motif domain (residues 1–49) correlate with its binding affinity for natural *lac* operators. HP62V52C protein binds to the O2 operator *in vitro* with a two-times-lower affinity compared to the O1 operator (Fig. 2a and b). This fits the slightly lower protection factors of the HP62V52C–O2 complex. The O3 operator has a very low affinity

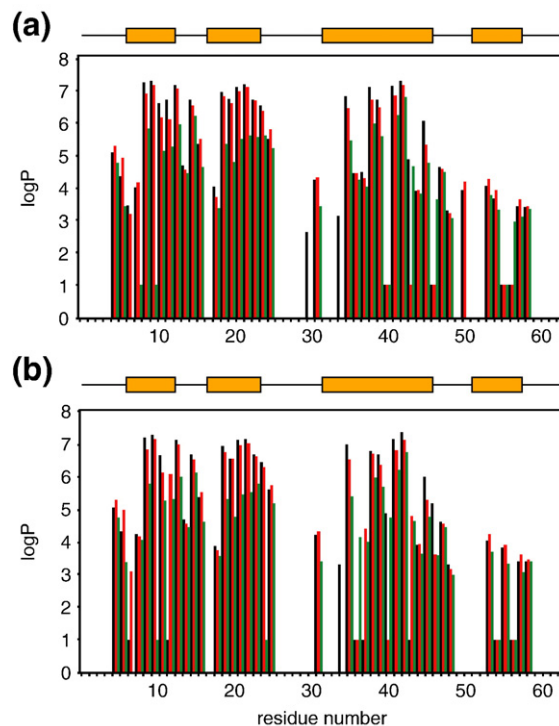


Fig. 3. Per-residue protection factors obtained from amide H/D exchange experiments for the Lac repressor headpiece bound to the natural operators O1, O2, and O3. (a) Protection factors for the left monomer for each complex. (b) Protection factors for the right monomer for each complex. Values for the O1, O2, and O3 complexes are shown in black, red, and green, respectively. A value of $\log P=1$ was arbitrarily given to residues not included in the analysis due to peak overlap. The elements of the secondary structure are shown above the graphs.

and also a significantly lower protection factor for residues in this region. The three *lac* operators are asymmetrical, and the two half-sites differ significantly in their affinities for the Lac repressor when considered separately.¹⁰ However, the protection factors of the Lac headpiece amides are the same for the left protein subunit and the right protein subunit in all three complexes (Fig. 3). This can be well explained by the cooperative binding of two Lac headpiece monomers to the operator and the considerable stabilization of the HTH motif due to DNA binding. H/D exchange would then only occur when the dimeric headpiece dissociates from the DNA and would next unfold.

In contrast, the protection factors for the hinge helices in the three complexes are very similar (Fig. 3). Thus, the hinge helices have the same stabilities in the three specific complexes even in the much weaker HP62V52C–O3 complex and thus

probably already occur while the Lac dimer is bound to DNA. In sharp contrast, the corresponding region in the nonspecific complex, where no hinge helices are formed, is poorly protected.¹⁸

Structure determination

The structures of a dimeric Lac HP62V52C in complexes with its auxiliary operators O2 and O3 were determined using NMR spectroscopy. The O2 and O3 operator sequences used in this study were 23 and 27 bp long, respectively (Fig. 1). For O2, we extended the naturally occurring 21-bp-long operator by one base at each terminus in order to prevent the DNA ends from fraying. For O3, we initially prepared also a 23-bp operator. However, it showed a very low affinity for the dimeric headpiece and abnormal nuclear Overhauser enhancement (NOE) connectivities in the two-dimensional (2D) NOE spectra of the free O3 operator. Therefore, 2 cyt bp

Table 2. Experimental restraints and structural statistics

	HP62–O1	HP62–O2	HP62–O3
rmsd (Å) with respect to mean (backbone/heavy) ^a			
HP62 (left subunit)	0.41±0.17/0.89±0.20	0.52±0.12/0.96±0.18	0.77±0.11/1.22±0.14
HP62 (right subunit)	0.37±0.09/0.88±0.13	0.59±0.12/0.95±0.12	0.92±0.20/1.37±0.15
HP62 (dimer)	0.63±0.14/1.03±0.16	0.79±0.13/1.11±0.14	1.10±0.17/1.50±0.15
Operator	0.83±0.22/0.76±0.20	0.96±0.24/0.85±0.22	1.61±0.37/1.45±0.36
HP62 operator	0.80±0.15/0.97±0.15	0.93±0.16/1.04±0.16	1.46±0.22/1.58±0.21
Number of experimental restraints			
Protein (dimer)			
Intraresidue NOEs	514	825	661
Sequential NOEs ($ i-j =1$)	419	513	329
Medium-range NOEs ($1 < i-j < 4$)	260	350	234
Long-range NOEs ($ i-j > 4$)	246	250	144
Interprotein	28	29	13
DNA NOEs	680	427	299
Protein–DNA NOEs	245	144	75
Total number of distance restraints	2392	2538	1755
Dihedral angle restraints	162	142	70
Dipolar coupling restraints (D_{HN})	84	91	72
rmsd from experimental restraints			
All distance restraints (Å)	0.06±0.002	0.03±0.002	0.03±0.002
Dihedrals angles (°)	0.70±0.48	0.88±0.64	0.46±0.18
D_{HN} (Hz)	0.18±0.013	0.13±0.012	0.39±0.013
Average Q -factor for dipolar coupling restraints ^b			
D_{HN}	0.18±0.018	0.13±0.012	0.32±0.01
Restraint violations ^c			
NOE distances with violations >0.3 Å	8±2.2	1.0±1.1	3.0±1.26
Dihedrals with violations >5°	1.4±1.9	1.8±2.1	0
D_{HN} with violations >0.5 Hz	1±0.5	0.6±0.6	11.5±5.47
rmsd from idealized geometry			
Bonds (Å)	0.005±0.0003	0.004±0.001	0.005±0.0004
Angles (°)	0.71±0.03	0.64±0.036	0.63±0.02
Improper (°)	0.7±0.08	0.65±0.13	0.55±0.03
Ramachandran analysis (%) ^d			
Most favored regions	89.4±1.9	87.3±1.3	75.1±2.4
Additionally allowed regions	8.8±1.9	11.3±1.5	20.5±2.9
Generously allowed regions	1.6±1	1.3±0.8	3.4±1.3
Disallowed regions	0.3±0.4	0.1±0.3	1±0.7
CNS energies			
E_{vdw} (kcal mol ⁻¹)	-962±37	-1081±27	-981±21
E_{elec} (kcal mol ⁻¹)	-5404±79	-5467±112	-5443±98

Statistics are based on 20 final structures for the HP62V52C–O1 and HP62V52C–O2 complexes and 10 final structures for the HP62V52C–O3 complex.

^a Residues 4–58 for the protein and base pairs 1–20 for the operator are taken into account.

^b Q -factors were calculated with PALES.

^c There were no NOE violations larger than 0.5 Å.

^d Residues 4–59 were taken into account.

were added on both sides to avoid base-pair mismatch during annealing and to prevent fraying ends.

NMR titrations of the headpiece with its cognate operators show that all NMR signals of the free and bound proteins are in slow exchange, corresponding to low offrates of the headpiece in specific complexes. The complexes show a double set of resonances for each headpiece monomer, since they are bound to different operator half-sites. The structures of the two HP62V52C–DNA complexes were solved by heteronuclear double-resonance and triple-resonance NMR spectroscopies using ^{15}N - and ^{13}C -labeled protein and unlabeled oligonucleotides. As described in Materials and Methods, structural information was derived from NOE spectroscopy (NOESY) spectra, Talos analysis, H/D exchange, and residual dipolar couplings (RDCs). The structure of the HP62V52C–O2 complex was calculated using 2771 experimental NMR restraints (Table 2). For the HP62V52C–O3 complex, a lower number (1897) of collected restraints were identified (Table 2). This can probably be attributed to a combination of resonance broadening effects asso-

ciated with residues contacting DNA, overall lower stability, and the larger size of this complex. This results in the significantly larger backbone rmsd values of the well-ordered segments for HP62V52C–O3 (1.46 Å) in comparison with those for the HP62V52C–O2 complex (0.93 Å). The HP62V52C–O3 complex also has the highest RDC *Q*-factor (HP62V52C–O1, 0.18 ± 0.01 ; HP62V52C–O2, 0.13 ± 0.01 ; HP62V52C–O3, 0.32 ± 0.01), which is a goodness-of-fit measure for RDCs.²⁰ The higher rmsd and *Q*-factor values could reflect conformational exchange on a timescale slower than nanoseconds and are possibly related to the heterogeneity observed in gel-shift assays for the O3 complex. Further structural statistics are shown in Table 2.

To avoid differences between the three complexes due to the structure determination protocol, we have refined the structure of HP62V52C bound to the main operator O1 with the same structure calculation protocol as for the O2 and O3 complexes and recalculated the structure of the O1 complex using the original restraints. The structural statistics for the HP62V52C–O1 complex are shown in Table 2. Overall, the current structure closely resembles

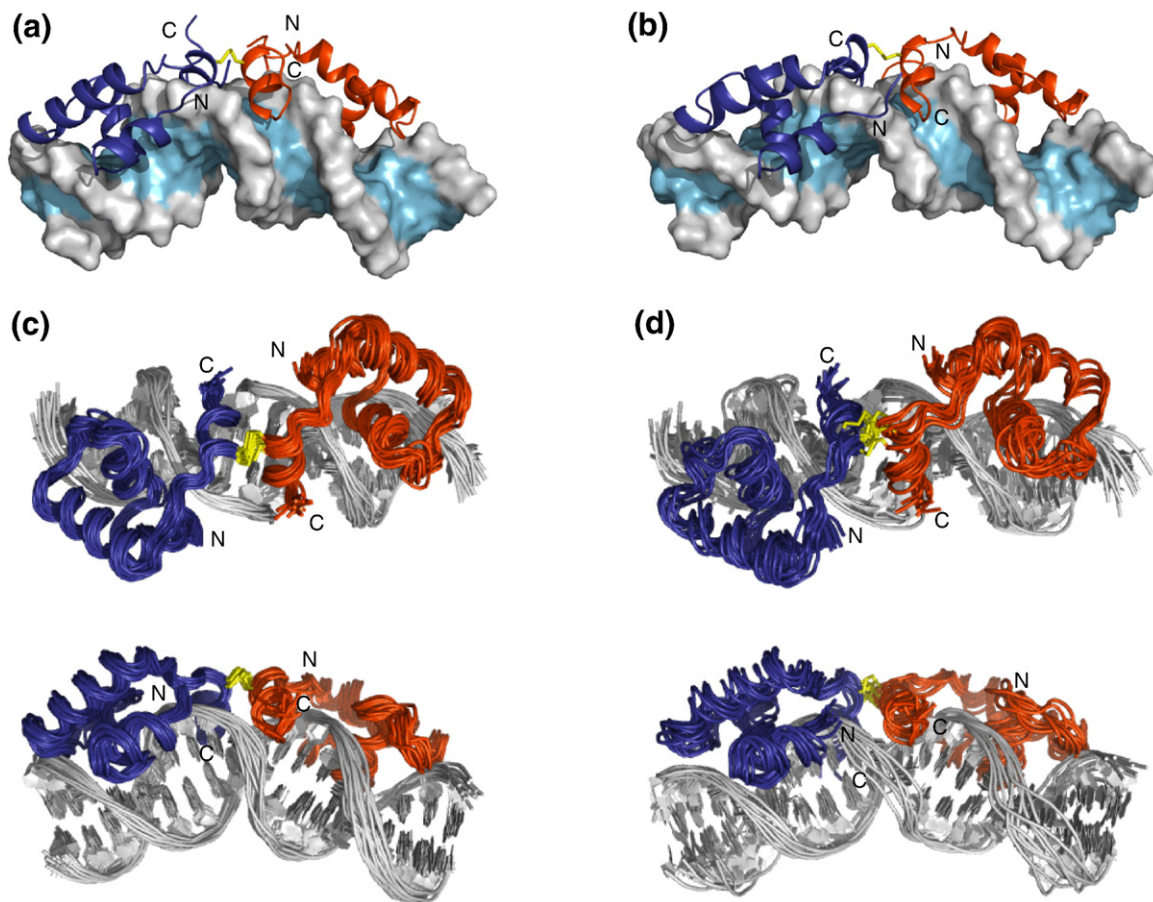


Fig. 4. Three-dimensional structures of the HP62V52C–O2 (a) and HP62V52C–O3 (b) complexes. The left and right Lac headpiece subunits are shown in dark blue and dark orange, respectively. The disulfide bond formed by cysteines at position 52 is shown in yellow. Ensemble of the 20 lowest-energy structures of the HP62V52C–O2 complex (c) and ensemble of the 10 lowest-energy structures of the HP62V52C–O3 complex (d) shown in two different views rotated by 90°. For clarity, only the backbone atoms of amino acids Val4–Lys59 and of the DNA heavy atoms of base pairs 2–20 are shown.

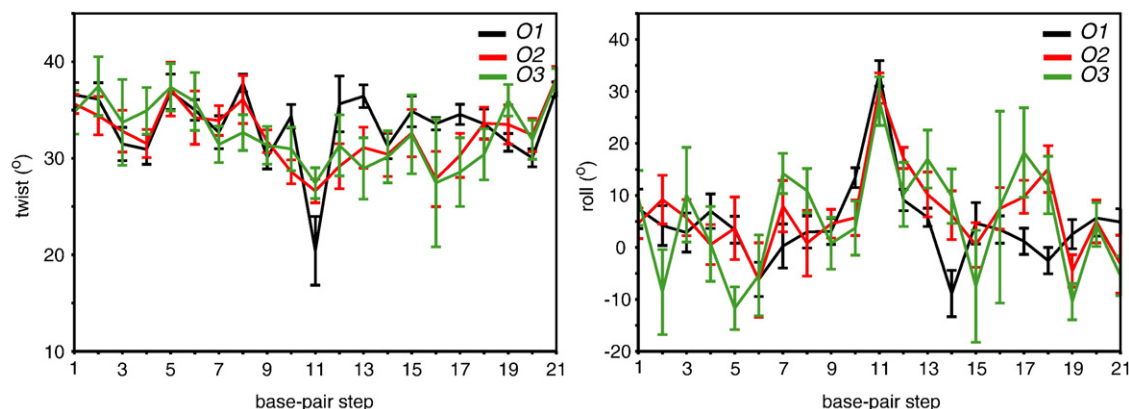


Fig. 5. Inter-base-pair DNA helical parameters of the *lac* operators *O1* (black), *O2* (red), and *O3* (green) in complex with dimeric HP62.

the previously reported structure of the HP62V52C–*O1* complex.

Structure description

The ensembles of the lowest-energy structures of the HP62V52C–*O2* and HP62V52C–*O3* complexes are presented in Fig. 4. As expected, the overall topology of newly determined complexes is identical with that of the HP62V52C in complex with the *O1* operator. The α -helical contents are the same in both the left protein subunit and the right protein subunit, and the recognition helix is aligned along the base-pairing edges of nucleotides in the major groove in all complexes. The folded hinge helices are placed in the minor groove.

Intercalation of the side chain of Leu56 between base pairs 10 and 11 introduces a kink in the DNA, which is similar for all three structures (Table 1). This kink is also reflected in the DNA helical parameters for various complexes, showing a large deformation at the central base pair (Fig. 5). The lower number of restraints for the HP62V52C–*O3* complex resulted in a less accurate definition of the operator (rmsd, 1.6 Å), precluding a detailed comparison of the DNA structures.

A good correlation is found between the binding strength and the accessible surface area that is buried upon the binding of the Lac headpiece to its natural operators. The largest buried surface area is found in the HP62V52C–*O1* complex (total, ~ 3800 Å²). The HP62V52C–*O2* complex has a

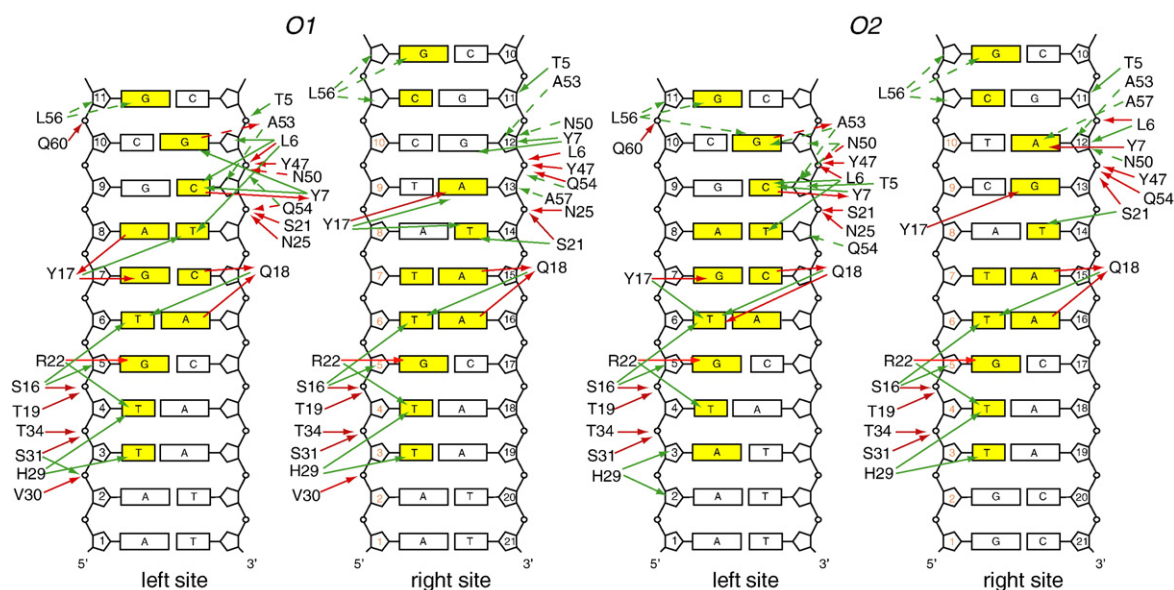


Fig. 6. Comparison of protein–DNA contacts in the HP62V52C–*O1* and HP62V52C–*O2* complexes. The bases that are recognized specifically are shown in yellow. Continuous and broken lines indicate interactions observed in the majority (50%) of the structures in the final ensembles in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively.

buried accessible surface of $\sim 3500 \text{ \AA}^2$, corresponding with a slightly lower affinity as compared to the *O1* complex. The HP62V52C-*O3* complex has the smallest accessible surface area ($\sim 3000 \text{ \AA}^2$) and the smallest binding affinity of all.

Comparison of complexes

O1 versus *O2*

Overlay of headpiece monomers in the *O1* and *O2* complexes shows a backbone rms difference of only 0.9 \AA between the *O1* half-site and the *O2* half-site. A full overlay of the two complexes shows a larger rmsd of $\sim 2.7 \text{ \AA}$, since there is a variation in the detailed DNA geometry and in the DNA bend between the two half-sites in the different conformers of the ensembles. A schematic summary of the consensus protein-DNA contacts in the ensembles of the HP62V52C-*O1* and HP62V52C-*O2* complexes is given in Fig. 6, and representative structures for these two complexes are shown in Fig. 7.

The left half-sites of the *O1* and *O2* operators only differ at single position 3 (cf. Fig. 1). As a result, the structures for the left half-sites are very similar, and the network of contacts is largely preserved. Hydrogen-bonding contacts of Leu6, Ser16, Thr19, Ser21, Asn25, Ser31, and Thr34 anchor Lac headpiece to the DNA backbone either with their side chains or with their backbone, or with both, whereas sequence-specific contacts are present between the side chains of Tyr17, Gln18, and Arg22 and their partner base. The sequences of the *O1* and *O2* left

half-sites differ at base pair 3 (T:A3 versus A:T3, respectively), and this results (for the *O2* complex) in a loss of hydrophobic contacts of His29 with the methyl groups of Thy3 and Thy4 and a loss of the hydrogen bond of the backbone of Val30 with the phosphate backbone.

In the right half-site of the *O1* and *O2* operators, 4 bp are different: base pairs 12, 13, 20, and 21 (Fig. 1). However, there are no specific contacts between the headpiece and base pairs 20 and 21. Thus, these bases do not contribute significantly to the recognition mechanism, in agreement with both *in vitro* and *in vivo* mutational studies.²¹ The differences in base pairs 12 and 13 result in a slightly different binding of the headpiece to the right half-sites in the *O1* and *O2* complexes. First, the anchoring contacts by Leu6 and Ser21 are shifted 1 bp towards the center of the operator in the HP62V52C-*O2* complex. This results in a positioning change of the recognition helix along the major groove (cf. Fig. 7) and a slightly different angle between the axis of the DNA double helix and the recognition helix: 74° and 96° for the HP62V52C-*O1* and HP62V52C-*O2* complexes, respectively. This also leads to a different conformation of the side chains of Tyr7 and Tyr17 contacting base pairs 12 and 13.

Tyr17, together with Gln18 and Arg22, is generally considered to be the most important residue for the specific recognition of the *lac* operator because it participates in sequence-specific protein-DNA interactions, which are conserved in most half-sites of the *O1* and *O2* operators. However, a comparison of the HP62V52C-*O1* and HP62V52C-*O2* structures

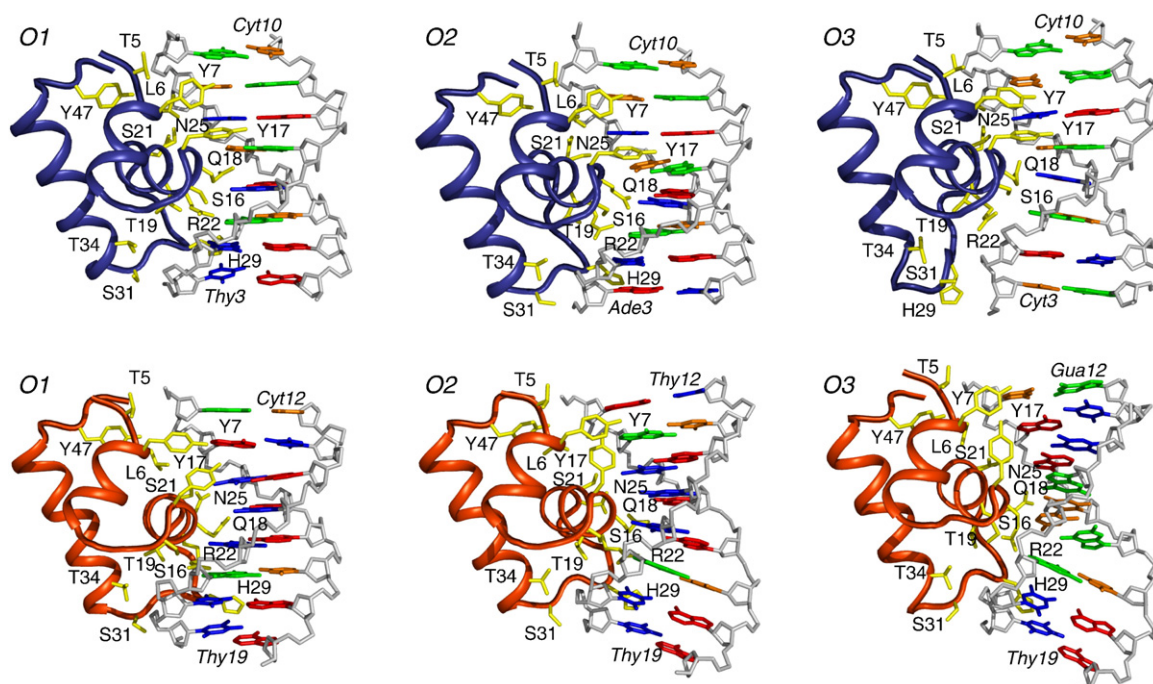


Fig. 7. Protein-DNA interactions observed in the HP62V52C-*O1*, HP62V52C-*O2*, and HP62V52C-*O3* complexes. The left monomer of each complex is shown in dark blue, whereas the right monomer is shown in dark orange. Residues involved in protein-DNA interactions are shown in yellow.

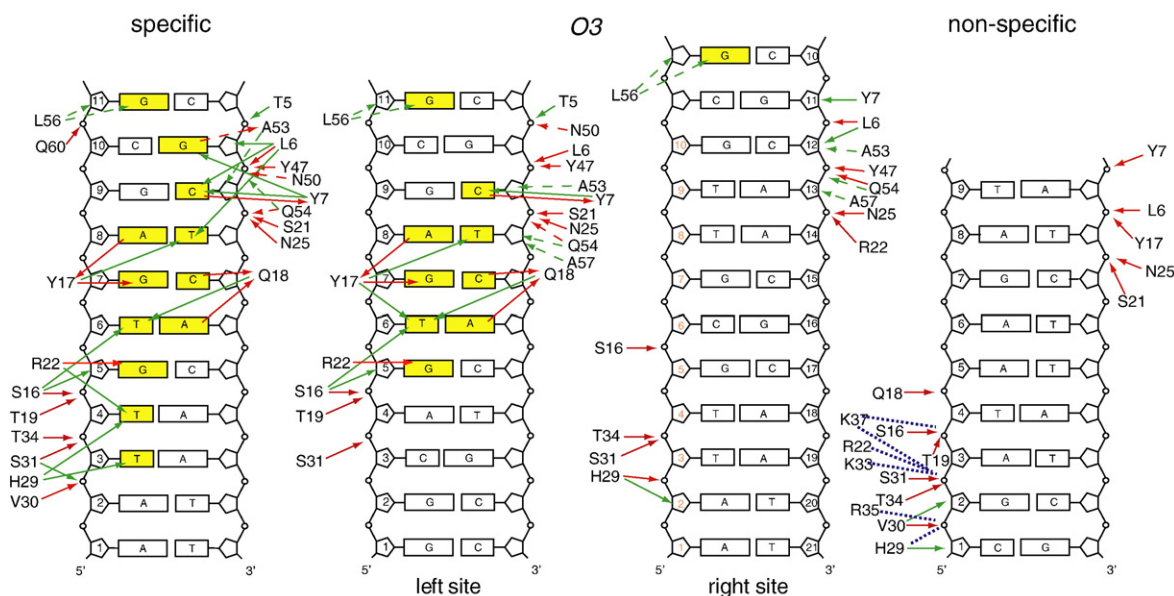


Fig. 8. Comparison of protein–DNA contacts in the HP62V52C–O3 and HP62V52C–NOD complexes. The bases that are recognized specifically are shown in yellow. Continuous and broken lines indicate interactions in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively.

shows that Tyr17 adjusts its hydrogen-bonding contacts at the two right half-sites. This emphasizes the intrinsic plasticity of the headpiece to optimally recognize the substituted base pairs in the two operators. Also, the DNA conformation of the O1 and O2 operators is slightly different at base pairs 12 and 13 and even so at conserved base pairs 16 and 17 (cf. Fig. 5). Thus, both protein and DNA can adapt to form an optimal interaction interface.

O1 versus O3

The overlay of the Lac headpiece monomer bound to the left half-site of the O1 and O3 operators results in a relatively large rmsd value (~ 1.5 Å) for protein backbone atoms as compared to the value of ~ 0.9 Å for the O1 and O2 complexes. However, the interactions between the key residues Tyr17, Gln18, and Arg22 in the major groove and the majority of anchoring contacts are preserved in the left Lac headpiece subunit (cf. Fig. 8). The most significant difference between the protein structures occurs in the loop from Val24 to Thr34. This difference is due to the substitutions of T:A3 and T:A4 in the O1 operator by C:G3 and A:T4 in the O3 operator. Due to these substitutions, not only sequence-specific contacts but also several anchoring contacts are absent in the HP62V52C–O3 complex. The importance of these interactions was also shown by mutational studies.²²

In contrast to the other half-sites, in the right O3 half-site, we observed no specific interactions for the side chains of Tyr17, Gln18, and Arg22. The network of contacts is also different from that observed in the structure of the dimeric HP62 bound to the NOD.¹⁸ Only for the side chain of Arg22 was a hydrogen

bond with DNA phosphate backbone identified (cf. Fig. 8). Most of the anchoring contacts, as well as the interactions between the hinge helix and the minor groove, are still present. This nonspecific binding mode of the dimeric Lac headpiece to the right half-site of the O3 operator is insufficient to account for the weak binding of the Lac repressor to this auxiliary operator. Experiments by Betz *et al.*³ showed that mutations introduced at positions 14 and 15 had only intermediate effects on repressor affinity, while substituting A:T4 with G:C in the left half-site led to a significant reduction in repressor affinity.²² Therefore, the lower affinity of the dimeric Lac headpiece for the O3 operator is due to a combined loss of contacts at both halves and a smaller buried surface area.

Plasticity and affinity

The dimeric Lac headpiece adapts its precise conformation so that it can optimally interact with the operator binding interface. This intrinsic plasticity is seen at different levels. At a global view, both protein monomers adopt alternative conformations in order to recognize the different halves of the natural operators and to compensate for the presence of the additional central G:C base pair. At the same time, the DNA adjusts its conformation with a similar bending in all three operator complexes. At a more local level, the protein side chains involved in the sequence-specific contacts adjust their precise orientations for optimal interaction. Thus, Lac headpiece operator recognition is not a simple read-out of favorable interactions with individual base pairs but rather results from a complex interplay between complementarity and conformational

adaptability, in agreement with the results of previous thermodynamic analyses.^{23,24}

The affinities of Lac headpiece for the three operators are very different. Still the DNA binding of hinge helices appears to be very similar in all three complexes. The contacts in the DNA minor groove are the same, and the folding stabilities as probed in H/D exchange are almost identical. This indicates that hinge helix formation alone cannot account for the high affinity of the Lac repressor. More likely, the cooperative folding of the hinge helices contributes to specificity by binding at the central GCG sequence and assists in positioning the HTH motif domain with respect to the operator. Formation of favorable contacts between residues in the HTH domain and DNA will then further increase the affinity of the Lac repressor for its natural operators.

A dual role for the O3 operator?

Lac repressor binds to both the O1 operator and the O2 operator with almost similar high affinities. This agrees well with the structures and binding affinities of the HP62V52C-O1 and HP62V52C-O2 complexes. However, when the Lac repressor tetramer binds simultaneously to the O1 and O2 operator sites, only weak cooperativity is observed.^{8,9} This has been described in the long distance and high level of flexibility of the DNA loop between the O1 site and the O2 site. In contrast, the distance between the O1 site and the O3 site is significantly shorter and, in fact, O3 is even optimally positioned in the DNA for tight cooperative enhancement for Lac tetramer binding.^{25,26} However, the affinity of the Lac repressor for O3 is relatively low, as explained by the structure and binding affinity of the HP62V52C-O3 complex, where we note that binding at the right half-site of O3 resembles that of a weak nonspecific DNA complex. Lac repressor binding to O3 can only weakly enhance *lac* operon repression.^{4,9} This then raises the question on the role of the O3 site. This operator is immediately followed by a CAP binding site. This permits simultaneous binding of both proteins to their binding site, resulting in the cooperativity of CAP and Lac repressor binding either due to direct interaction^{27,28} or by increased looping.^{8,29,30} The presence of a weak operator O3 next to the CAP binding site would result in a relatively weak CAP/Lac coupling, which does not strongly interfere with the direct control of the *lac* operon. In addition, we note that O3 is also located at the 5' end of the *LacI* gene, where the stop codon TGA is part of the left half of the O3 site. Thus, an additional role of O3 could also be an inverse one (i.e., not only as an auxiliary operator in conjunction with O1 but also as a self-inhibitor in Lac repressor production due to Lac repressor binding at O3). The strict conservation of both the presence and the positioning of the operators in all available genome sequences underscores importance, although it needs to be mentioned that the small number of *lac* operon sequences makes it hard to firmly establish this. Thus, the O3 operator would have a dual role:

repression of the *lac* operon and prevention of excessive transcription of the *LacI* gene. The presence of a weak operator would then permit effective *LacI* gene transcription until sufficient Lac repressor levels are obtained to bind the weaker operator and could contribute to the bistabilic behavior of the operon.^{31,32}

Materials and Methods

NMR sample preparation

Cloning, expression, and purification of the HP62V52C mutant have been described previously.¹⁴ Uniformly ¹⁵N- and ¹³C-labeled protein was grown in M9 minimal medium. The HPLC-purified natural *lac* operator O2 fragment was purchased from Carl Roth GmbH (Germany). The HPLC-purified O3 DNA fragment was purchased from Eurogentec. To form the double-stranded DNA, we mixed equimolar amounts of complementary strands, heated them to 95 °C for 5 min, and slow-cooled them to anneal over a period of several hours. The free HP62V52C sample contained 0.4 M KCl and 0.06 M potassium phosphate buffer (pH 5.8). For complex formation, protein was mixed with an equimolar amount of the operator, and buffer was exchanged with 0.02 M KCl and 0.01 M potassium phosphate buffer (pH 6.0), with the addition of 5% *d*₈-glycerol, 0.01% sodium azide, and 5% D₂O using the ultrafiltration device Amicon (Millipore). Complex formation was monitored with NMR spectroscopy. NMR samples contained 0.7 mM HP62V52C-DNA complex.

DNA binding and DNA bending assay

Gel-shift retardation and DNA bending assays were performed essentially as described.¹⁶ Binding experiments were performed with 140-bp DNA fragments containing the 21-bp *lac*-O1, *lac*-O2, *lac*-O3, or NOD. Comparable results were obtained with the operator sequences used for structural analysis under the sample conditions described above.

These fragments were the same as those used in the bending experiments. All the experiments were performed at 4 °C in 20 μl of reaction buffer containing 0.05 or 0.250 M KCl and 0.01 M Tris buffer (pH 8.1), with the addition of 1 mM ethylenediaminetetraacetic acid, 5% glycerol, and 0.1 mg ml⁻¹ bovine serum albumin.

NMR spectroscopy

All NMR experiments on the protein-DNA complexes were carried out at a temperature of 315 K on Bruker AVANCE spectrometers. Backbone assignments were based on the assignments available for the HP62-O1 operator complex. Assignments were further confirmed by analysis of three-dimensional (3D) HNCOC, HNCA, and HNCO experiments. Aliphatic side-chain assignments were derived from 3D H(C)(C)(CO)NH total correlated spectroscopy (TOCSY), (H)C(C)(CO)NH-TOCSY, and HCCH-TOCSY experiments. All abovementioned experiments were recorded on a Bruker AVANCE 600-MHz spectrometer equipped with a TXI probe with *z*-gradients. Aromatic ring resonances were assigned using 2D ¹H-¹H NOESY on a sample in D₂O. Intramolecular NOEs for the protein were identified from 2D ¹H-¹H

NOESY, 3D NOESY- ^{15}N heteronuclear single quantum coherence (HSQC), and 3D NOESY- ^{13}C HSQC experiments acquired at 750 MHz and using a mixing time of 80 ms. DNA resonances and protein-DNA NOE signals were assigned in 2D ^1H - ^1H NOESY and ^{15}N , ^{13}C double-filtered 2D NOE experiments acquired in H_2O and D_2O at 750 and 900 MHz, respectively. DNA assignments were obtained using conventional sequential assignment methodology for nucleic acids.³³ Residual $^1D_{\text{HN}}$ dipolar couplings were determined from the difference in $^1J_{\text{NH}}$ at two different fields—21.14 T (^1H frequency, 900.21 MHz) and 11.75 T (^1H frequency, 500.28 MHz)—due to the natural alignment of the DNA. The values for $^1J_{\text{NH}}$ coupling were extracted from a series of 30 J -modulated 2D ^{15}N HSQC spectra.³⁴ The intensities of the peaks in the HSQC spectra were measured at each dephasing period, and the data were fitted to the equation described in Tjandra *et al.*, giving $^1D_{\text{NH}}$ scalar couplings.³⁴ The data were fitted using the nonlinear least-squares Levenberg-Marquardt algorithm implemented in the program Gnuplot 3.7†. All spectra were processed using the NMRPipe software package³⁵ and analyzed with NMRView.³⁶

Structure calculation

The structure determination protocol was similar to the one used for the previously determined HP62V52C-O1 complex.¹⁵ First, the structure of the protein in the bound state was determined and refined with RDCs. Then this structure was docked onto the standard 23 bp B-form DNA sequence using HADDOCK2.0,^{37,38} with manually identified intermolecular NOEs used as restraints.

The chemical shifts of ^{15}N , ^1H , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^{13}\text{C}'$ were used as inputs for TALOS³⁹ to predict dihedral angles that were subsequently used as restraints in the calculations. NOE peak lists were generated automatically with NMRView³⁶ and edited manually to remove obvious artifacts. The peak lists, together with chemical shift tables, were used for automated NOESY assignment and structure calculation using ARIA1.3.^{40,41} A set of manually assigned NOEs was used to facilitate the automated assignment procedure. The automatically assigned NOEs were carefully checked and converted into distance restraints. This set of distance restraints, in combination with dihedral angle restraints, was then used to calculate the final ensemble of structures. Two hundred structures were generated, and the 50 lowest-energy structures were refined in water. The 20 lowest-energy structures were then further refined with $^1D_{\text{HN}}$ restraints. The natural alignment of DNA molecules by high magnetic fields was used to measure one-bond $^1D_{\text{HN}}$ dipolar couplings.⁴² For HP62V52C in complex with DNA, this resulted in maximum RDCs of 2.7 Hz (O1), 2.4 Hz (O2), and 2.7 Hz (O3). RDCs were included as harmonic restraints during the refinement steps of the calculations of the dimer alone and of the protein-DNA complex. The value of the RDC force constant ($100 \text{ kcal mol}^{-1} \text{ Hz}^{-1}$) was chosen such that the difference between the backcalculated values and the experimental values fell within the experimental error. The axial and rhombic components of the molecular alignment tensor D_a and R , respectively, were initially determined by PALES2.0⁴³ and further refined by calculating a few structures each time, where D_a and R were backcalculated

from structures in order to minimize the total energy of the protein-DNA structures and Q -factor.

Docking of the HP62-V52C dimer ensemble on natural *lac* operators was performed with HADDOCK2.0^{37,38} on the basis of intermolecular distance restraints. Intermolecular NOEs were converted into distance restraints using NMRView. They were classified into four distance ranges: 1.8–2.8, 2.8–3.4, 3.4–4.5, and 4.5–5.0 Å, corresponding to strong, medium, weak, and very weak NOEs, respectively. Observation of proton signals from the hydroxyl groups is an indication that these atoms are protected from exchange with solvent, suggesting that they participate in a hydrogen-bond interaction. Therefore, they were included during docking as hydrogen-bond restraints. Another set of hydrogen-bond restraints was identified from H/D exchange experiments, with acceptors for the hydrogen bond on the DNA side identified from preliminary structure analysis. Hydrogen bonds were introduced as ambiguous restraints with an upper bound of 2 Å (H-OP) or 3 Å (N-OP or O-OP) at later docking stages.

The starting B-form DNA structure was built with the 3D-DART server‡. Watson-Crick base pairing and the planarity of purine and pyrimidine rings were ensured by a set of artificial restraints. Backbone dihedral angle restraints for B-DNA were used in order to maintain the conformations. Two middle bases were defined as fully flexible. In total, 400 structures were generated, and the 200 lowest-energy structures were selected for refinement in explicit water. The final structures were selected based on energy and absence of restraint violations larger than 0.5 Å. The structures of the HP62V52C-O2 and HP62V52C-O3 complexes were analyzed for comparison with previously published HP62V52C-O1 structure. To avoid possible errors due to different structure calculation protocols, we recalculated the structure of the HP62V52C-O1 complex using the same structure calculation protocol and available restraints.

Protein-DNA interactions were analyzed with HBPLUS⁴⁴ and HADDOCK2.0^{45,46} programs. The final set of intermolecular hydrogen bonds is a consensus for outputs from both programs. DNA helical parameters and bend angles were analyzed with CURVES,⁴⁷ and the RDC quality factor (Q -factor)⁴⁸ was evaluated with PALES2.0.⁴⁹ The quality of the protein structures was analyzed using PROCHECK,⁵⁰ MOLMOL⁵¹ and PyMOL§ software were used for structure analysis and figure preparation.

H/D exchange

Amide H/D exchange experiments were carried out on samples of HP62V52C-O1, HP62V52C-O2, and HP62V52C-O3 complexes prepared as described above. The samples were lyophilized and afterwards dissolved in D_2O to start the exchange. The pH of the samples was adjusted to 6 (pD 6.4) by adding either NaOD or DCl. The experiments were performed at 315 K. Progress of the exchange process was followed by collecting a series of successive 2D ^{15}N HSQC spectra. The dead time was approximately 15 min. The spectra were acquired on a Bruker AVANCE 600 NMR spectrometer equipped with a cryoprobe. Another set of experiments was carried out at pH 9 to extract decays for the most slowly exchanging amides. The intensities of cross-peaks were obtained using

† www.gnuplot.info

‡ <http://haddock.chem.uu.nl/dna>

§ www.pymol.sourceforge.net

a standard integration routine in NMRView. The time-dependent decay of peak intensities was fitted to a three-parameter (corresponding to intensity at time 0, decay rate k_{ex} , and a residual term to take into account the residual nondeuterated water) single-exponential decay function using GnuPlot. Rate constants were then extracted from the fitting results and subsequently used to calculate residue-specific protection factors. Values for sequence-specific intrinsic rate constants (k_{ch}) were calculated using the spread sheet available from S. Englander's Web sitell. Protection factors were determined from the relation $P = k_{\text{ch}}/k_{\text{ex}}$, where P is the protection factor, and k_{ex} and k_{ch} are as defined above.

Accession codes

The atomic coordinates of the final structures and restraint lists have been deposited in the Protein Data Bank (refined HP62V52C-O1, accession code 2KEI; HP62V52C-O2, accession code 2KEJ; HP62V52C-O3, accession code 2KEK).

Acknowledgements

The authors are grateful to Dr. R. W. Wechselberger for assistance with NMR measurements. This work was financially supported by the Netherlands Foundation for Chemical Research (NWO-CW) and the European Commission (project 031220; Spine2 complexes).

References

- Lewis, M. (2005). The lac repressor. *C. R. Biol.* **328**, 521–548.
- Wilson, C. J., Zhan, H., Swint-Kruse, L. & Matthews, K. S. (2007). The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell. Mol. Life Sci.* **64**, 3–16.
- Betz, J. L., Sasmor, H. M., Buck, F., Insley, M. Y. & Caruthers, M. H. (1986). Base substitution mutants of the lac operator: *in vivo* and *in vitro* affinities for lac repressor. *Gene*, **50**, 123–132.
- Oehler, S., Eismann, E. R., Kramer, H. & Muller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *EMBO J.* **9**, 973–979.
- Gilbert, W. & Maxam, A. (1973). The nucleotide sequence of the lac operator. *Proc. Natl Acad. Sci. USA*, **70**, 3581–3584.
- Reznikoff, W. S., Winter, R. B. & Hurley, C. K. (1974). The location of the repressor binding sites in the lac operon. *Proc. Natl Acad. Sci. USA*, **71**, 2314–2318.
- Kramer, H., Amouyal, M., Nordheim, A. & Muller-Hill, B. (1988). DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J.* **7**, 547–556.
- Kramer, H., Niemoller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B. & Muller-Hill, B. (1987). Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.* **6**, 1481–1491.
- Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Muller-Hill, B. (1994). Quality and position of the three lac operators of *E. coli* define efficiency of repression. *EMBO J.* **13**, 3348–3355.
- Sasmor, H. M. & Betz, J. L. (1990). Symmetric lac operator derivatives: effects of half-operator sequence and spacing on repressor affinity. *Gene*, **89**, 1–6.
- Sadler, J. R., Sasmor, H. & Betz, J. L. (1983). A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl Acad. Sci. USA*, **80**, 6785–6789.
- Spronk, C. A., Bonvin, A. M., Radha, P. K., Melacini, G., Boelens, R. & Kaptein, R. (1999). The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure*, **7**, 1483–1492.
- 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.
- Kalodimos, C. G., Folkers, G. E., Boelens, R. & Kaptein, R. (2001). Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. *Proc. Natl Acad. Sci. USA*, **98**, 6039–6044.
- Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R. & Kaptein, R. (2002). Plasticity in protein–DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain. *EMBO J.* **21**, 2866–2876.
- Spronk, C. A., Folkers, G. E., Noordman, A. M., Wechselberger, R., van den Brink, N., Boelens, R. & Kaptein, R. (1999). Hinge-helix formation and DNA bending in various lac repressor–operator complexes. *EMBO J.* **18**, 6472–6480.
- Bell, C. E. & Lewis, M. (2001). Crystallographic analysis of Lac repressor bound to natural operator O1. *J. Mol. Biol.* **312**, 921–926.
- Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennegues, M., Boelens, R. & Kaptein, R. (2004). Structure and flexibility adaptation in non-specific and specific protein–DNA complexes. *Science*, **305**, 386–389.
- Wu, H. M. & Crothers, D. M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509–513.
- Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837.
- Caruthers, M. H. (1980). Deciphering the protein–DNA recognition code. *Acc. Chem. Res.* **13**, 155–160.
- Reznikoff, W. S. (1984). *The Microbe 1984: Part II. Prokaryotes and Eukaryotes* Cambridge University Press, Cambridge.
- Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E. *et al.* (1997). Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.* **267**, 1186–1206.
- Mossing, M. C. & Record, M. T., Jr (1985). Thermodynamic origins of specificity in the lac repressor–operator interaction. Adaptability in the recognition of mutant operator sites. *J. Mol. Biol.* **186**, 295–305.
- Muller, J., Barker, A., Oehler, S. & Muller-Hill, B. (1998). Dimeric lac repressors exhibit phase-dependent co-operativity. *J. Mol. Biol.* **284**, 851–857.

26. Muller, J., Oehler, S. & Muller-Hill, B. (1996). Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* **257**, 21–29.
27. Fried, M. G. & Daugherty, M. A. (2001). *In vitro* interaction of the *Escherichia coli* cyclic AMP receptor protein with the lactose repressor. *J. Biol. Chem.* **276**, 11226–11229.
28. Vossen, K. M., Stickle, D. F. & Fried, M. G. (1996). The mechanism of CAP–lac repressor binding cooperativity at the *E. coli* lactose promoter. *J. Mol. Biol.* **255**, 44–54.
29. Kuhlman, T., Zhang, Z., Saier, M. H., Jr & Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **104**, 6043–6048.
30. Narang, A. (2007). Effect of DNA looping on the induction kinetics of the lac operon. *J. Theor. Biol.* **247**, 695–712.
31. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B.I. & Van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, **427**, 737–740.
32. Santillan, M., Mackey, M. C. & Zeron, E. S. (2007). Origin of bistability in the lac operon. *Biophys. J.* **92**, 3830–3842.
33. Scheek, R. M., Zuiderweg, E. R., Klappe, K. J., van Boom, J. H., Kaptein, R., Ruterjans, H. & Beyreuther, K. (1983). Lac repressor headpiece binds specifically to half of the lac operator: a proton nuclear magnetic resonance study. *Biochemistry*, **22**, 228–235.
34. Tjandra, N., Grzesiek, S. & Bax, A. (1996). Magnetic field dependence of nitrogen–proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264–6272.
35. 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.
36. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352.
37. de Vries, S. J., van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V. *et al.* (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
38. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.
39. Wishart, D. S. & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363–392.
40. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
41. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, **19**, 315–316.
42. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997). Use of dipolar ^1H – ^{15}N and ^1H – ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4**, 732–738.
43. Zweckstetter, M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–690.
44. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
45. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. (2003). HADDOCK: protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.
46. de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V. *et al.* (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
47. Lavery, R. & Sklenar, H. J. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91.
48. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837.
49. Zweckstetter, M. & Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122**, 3791–3792.
50. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A. E. M., Nabuurs, S. B. *et al.* (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.
51. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.