NMR Studies of the Allosteric Effectors of the *Lac* Operon

Jūlija Romaņuka

Cover artwork by Oleksiy Sokolov
Cover design by Luca Codutti

Het werk beschreven in dit proefschrift werd gefinancierd de Nederlandse Organisatie voor Wetenschappelijk Onderzoek, gebied Chemische Wetenschappen (NWO-CW).

ISBN 978-90-393-5024-9

NMR Studies of the Allosteric Effectors of the *Lac* Operon

NMR studies van de allosterische effectoren van het *lac* operon

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van der graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 14 april 2009 des middags te 12.45 uur

door

Jūlija Romaņuka

geboren op 28 maart 1981, te Voronez, Rusland

Promotoren: Prof. dr. R. Boelens
Prof. dr. R. Kaptein

Co-promotor: Dr. G. E. Folkers

To my mother
In loving memory of my farther

Contents

Chapter 1

General introduction

Escherichia coli (E.coli), like all organisms, can adapt their metabolism, and thus their enzymes to the carbon source to which they are exposed. When glucose is abundant in the growth medium, the bacteria utilize it exclusively, even if other sugars are present. Only when glucose is depleted will *E.coli* upregulate the expression of proteins that transport and metabolize other carbon sources, such as lactose. The cluster of structural genes that are necessary for lactose utilization, along with their associated promoters and operators, is called the *lac* operon.

In 1961 Jacques Monod and François Jacob presented their general model of gene control in bacteria based on the *lac* operon, which survives essentially unchanged in contemporary textbooks. Together with another prominent scientist, André Lwoff, they were awarded the Nobel Prize in physiology or medicine in 1965 "for their discoveries concerning genetic control of enzyme and virus synthesis".

The *lac* operon

The *lac* operon consists of three structural genes: *lacZ*, *lacY* and *lacA*, which encode three proteins involved in lactose metabolism: β-galactosidase, galactoside permease and thiogalactoside transacetylase, respectively.

Galactoside permease (LacY) is a transmembrane protein that pumps lactose into the cell (reviewed in Kaback, 2005). β-Galactosidase (LacZ) is an intracellular enzyme that cleaves the disaccharide lactose into glucose and galactose (reviewed in Matthews, 2005). These sugars are further metabolized through the action of enzymes encoded in other operons. Thiogalactoside transacetylase (LacA) (reviewed in Roderick, 2005) is an enzyme that transfers an acetyl group from coenzyme A (CoA)

to the hydroxyl group of β-galactosides. This facilitates removal of toxic galactosides from the cell. These three structural genes are transcribed together in a single messenger RNA (polycistronic mRNA).

Expression of these genes is under control of the regulatory gene (*lacI*) which is located upstream of the *lacZ* gene and is transcribed from its own weak promoter (p_i). This gene encodes a protein called Lactose repressor (Lac repressor) (reviewed in Lewis, 2005; Wilson *et al.*, 2007). Expression of the Lac repressor is constitutive, i.e the gene is transcribed continually.

There are three regulatory sites located between the structural genes and *lacI*. They are the cAMP receptor protein (CRP) binding site, the promoter, and the *lac* operator. Figure 1.1 summarizes the structure and regulation of the *lac* operon.

Regulation of *lac* operon

When both glucose and lactose are provided to *E.coli*, the cells preferentially metabolize glucose. Under these conditions the Lac repressor protein binds with high affinity to its specific operator DNA sequence. In turn, RNA polymerase (RNAP) binding, elongation and/or initiation have been reported to be compromised, since the operator overlaps with the promoter (reviewed in Borukhov & Lee, 2005). Despite this elaborate binding complex, repression is not absolute. Even in the repressed state, each cell has a few copies of β-galactosidase and galactoside permease, presumably synthesized on the rare occasions when the repressor transiently dissociates from the operator. This basal level of transcription is essential to operon regulation.

When lactose is available in the cellular environment, these few copies of galactoside permease and β-galactosidase proteins transport this sugar and convert a small fraction of lactose into 1,6-allolactose. Binding of 1,6-allolactose to the Lac repressor reduces its affinity for the operator sequences, thereby allowing transcription of the polycistronic *lacZYA* mRNA by RNAP. Whereas allolactose is the natural inducer of the Lac repressor (Jobe & Bourgeois, 1972a), there exist also non-natural inducers such as 1-isopropyl-β-D-thiogalactoside (IPTG) (Jacob & Monod, 1961b). Upon binding of IPTG, the affinity of the Lac repressor for DNA is lowered 1000-fold (Barkley *et al.*, 1975).

The release of the Lac repressor from the operator is essential but not sufficient for effective transcription of the *lac* operon. The activity of RNA polymerase is also enhanced by a general transcription factor called the cAMP receptor protein (CRP; also known as catabolite activator protein, CAP) (Reznikoff, 1992; Busby & Ebright, 1999). This is one of two mechanisms by which glucose controls the *lac* operon and is called catabolite repression. When glucose levels are depleted, adenylyl cyclase produces high levels of cyclic AMP (cAMP) from ATP. The substrate cAMP subsequently binds to CRP, induces a conformational change, resulting in a conformation competent for DNA binding. The CRP-cAMP complex binds a DNA sequence in the promoter region located just upstream of the *lac* promoter (Figure 1.1). This DNA binding event enhances the affinity of RNA polymerase for the promoter and thereby increases the expression of the *lac* structural genes.

The second mechanism by which glucose affects the lactose operon is through a phenomenon known as inducer exclusion (Figure 1.1). Inducer exclusion is mediated by the phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS).

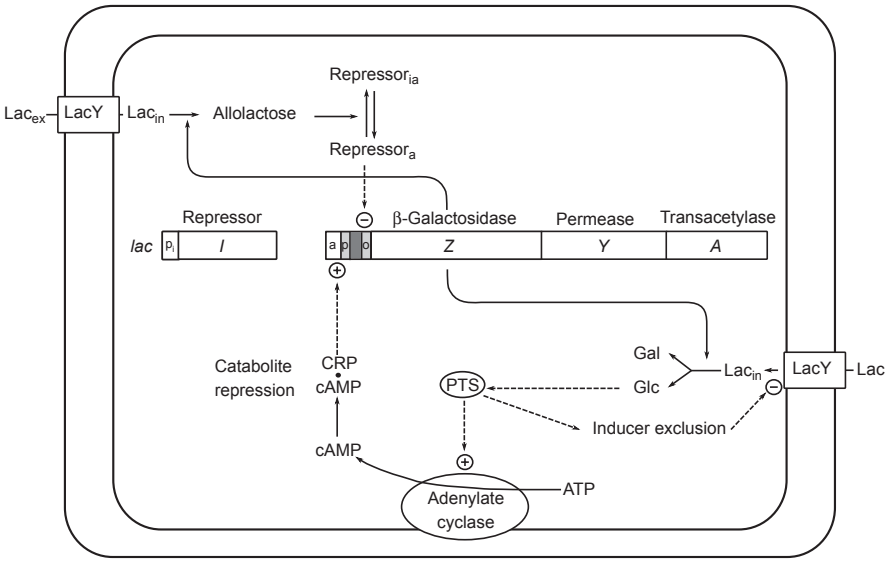


Figure 1.1: *Schematic representation of the lac operon and of lactose metabolism in E. coli. Symbols indicate: a, cAMP receptor protein binding sequence; p, promoter; o, operator; solid lines, metabolic reactions; dotted lines, and* ⊕ *or* ⊖ *positive or negative regulatory interactions; Repressor a or ia, active or inactive repressor. Adapted from Lengeler (1993). See main text for further details.*

Although the exact mechanism of inducer exclusion is not known, it is known that subunit IIA^{Glc} of the PTS interacts with the lac permease and prevents lactose transport in the presence of glucose (Osumi & Saier Jr, 1982; Postma *et al.*, 1996; Saier *et al.*, 1996).

The Lac repressor

Genetic analysis of the lactose metabolism system permitted the definition of the *I* gene, linked to but distinct from the *Z* and *Y* loci, and controlling the inducibility of the *lac* enzymes. I^- mutants synthesized maximal levels of β-galactosidase and transacetylase in the absence of inducer (reviewed in Jacob & Monod, 1961a). In 1959 PArdee, JAcob and MOnod performed so-called PaJaMo experiment, which was based on the phenomenon of bacterial conjugation (Pardee *et al.*, 1959). In the PaJaMo experiment I^+ and Z^+ genes were transferred into I^-Z^- cells. Even in the absence of inducer, β-galactosidase synthesis was already active, gradually falling to the basal rate after 1 hour. The maximal rate could be restored by the addition of IPTG. This experiment demonstrated the repressing activity, reflecting the requirement to build up a sufficient level of a cytoplasmic product in order to convert the i^- (constitutive) cells to the i^+ (inducible) cells. These observations indicated that this gene codes for a molecule which inhibits, or represses, the biosynthesis of the *lac*

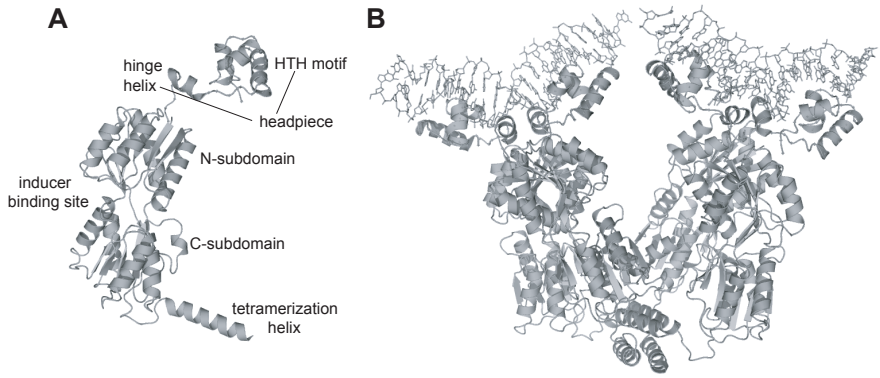


Figure 1.2: *Crystal structure of the Lac repressor bound to operator DNA (PDB code 1LBG) (Lewis* et al.*, 1996). (**A**) Structure of the monomer showing the individual structural units of the repressor. (**B**) The V-shaped tetrameric repressor bound to two perfectly symmetric ideal operator DNA.*

enzymes. J. Monod called the controlling substance a 'repressor' and raised the question about the nature of the controlling element. The answer was given by Gilbert & Müller-Hill (1966) when they isolated the Lac repressor protein using equilibrium dialysis. The experiment was based on the assumption that there should be an interaction between the repressor and inducer. The repressor bound to ^{14}C IPTG with a K_m of 1.3×10^{-6} M for wild type strains. High affinity permitted the detection, quantification and purification of the protein of 150,000-200,000 Da which is now called Lac repressor. Ever since the Lac repressor has been extensively studied by a variety of techniques and has received considerable interest in the fields of molecular and structural biology.

The structure of Lac repressor

It took almost 20 years after its first isolation before structural data became available for the Lac repressor. Kaptein *et al.* (1985) determined the structure of the isolated DNA-binding domain by NMR spectroscopy. Ten years later the core domain of the Lac repressor was crystallized and its structure was solved to 4.5 Å resolution (Friedman *et al.*, 1995). Shortly after, the full Lac repressor was crystallized both in complex with DNA and in complex with IPTG (Lewis *et al.*, 1996).

The Lac repressor consists of four identical monomers of 37.5 kDa each. Structurally it can be viewed as a dimer of dimers. The structure of the monomeric unit of the Lac repressor is shown in Figure 1.2. Each monomer consists of the following structural units: the amino-terminal headpiece (HP) (1-62) which is responsible for DNA operator recognition and binding, the core domain (63-329) which encompasses the inducer binding site and dimerization interface, and the carboxy-terminal tail (330-360) which is responsible for tetramer formation.

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 with a short β-hairpin turn containing a glycine residue in between as found in a large number of DNA binding proteins (reviewed in Aravind *et al.*, 2005). The second helix of the HTH motif is usually called the recognition helix since several residues make contacts to DNA. The HTH motif domain is connected to the core domain by a flexible linker often referred to as the hinge region (50-62).

The structure of the core domain belongs tothe very diverse periplasmic binding protein superfamily (Tam & Saier Jr, 1993). The core consists of two subdomains: the N-subdomain, containing residues 63-161 and 293-320, and the C-subdomain, with residues 162-289 and 321-329. The two subdomains have similar folds: six-stranded parallel beta sheet, surrounded by four alpha helices. The inducer binding site is located at the junction of the two subdomains.

The tetramerization of Lac repressor is mediated by the cooperative folding of four leucine rich sequences located at the C-terminal tail (residues 330-360) (Alberti *et al.*, 1991). These leucine heptad repeats have the ability to form four-helical bundles with an antiparallel arrangement of the helices (Alberti *et al.*, 1993; Lewis *et al.*, 1996).

Allostery and Lac repressor

Monod *et al.* (1963) proposed that the regulatory mechanism of the Lac repressor involved allostery. Allosteric proteins typically have at least two separate binding sites. In the case of Lac repressor one site of the protein is bound to the operator and blocks gene expression, which is released when a low molecular weight inducer molecule binds another site. This section gives a general overview of allostery as a concept.

Allostery

The term 'allostery' was introduced by J. Monod who called it 'the second secret of life' (second only to the genetic code) (Monod *et al.*, 1963; Monod, 1977). Allostery ('allo-steric = other-space') is defined as a regulatory mechanism when the binding of a ligand to one site affects the ligand binding properties of other sites on the same protein. If the first and second ligands are identical, the interaction is termed homotropic. When the ligands that bind to the different sites are different, the interaction is called heterotropic.

The first observation of allosteric behavior is attributed to Christian Bohr who reported in 1903 that the process of oxygen binding to hemoglobin (Hb) follows a sigmoidal curve or is cooperative (Bohr, 1903) as opposed to the normal hyperbolic curve associated with non-cooperative binding. After Adair demonstrated that Hb contains four oxygen binding hemes per molecule (Adair, 1925), it was well-understood that the binding of oxygen to the first heme must increase the oxygen-affinity of the second heme and of successive hemes either via direct contacts between the neighboring four heme groups or through conformational changes of the carrier protein (globin). In the 1960's this became clear when X-ray structures of the myoglobin and hemoglobin were determined. The crystal structure of the protein

revealed that the four heme groups of the Hb tetramer lie more than 20 Å away from each other and that a conformational change occurs in the protein when oxygen is coordinated to hemoglobin (Perutz *et al.*, 1960; Muirhead & Perutz, 1963; Perutz *et al.*, 1964). Around the same time experiments by Gerhart and Pardee demonstrated the existence of regulatory sites of enzymes and showed that these sites are structurally distinct from the catalytic sites (Gerhart & Pardee, 1961, 1962). Again, interactions between the sites could be explained by conformational changes throughout the protein induced by the binding of the regulatory molecule. The discovery of control of gene expression by repression (Jacob & Monod, 1961a) was the last piece of evidence that suggested that allosteric hemoglobin interactions may not be an isolated phenomenon but rather one example of general regulatory mechanisms. In this case the repressor was assumed to be an allosteric protein possessing two distinct sites, one of which binds operator DNA, the other the effector (Monod *et al.*, 1963).

Ever since, many models were developed to rationalize how the affinities of binding sites of proteins can be modulated. From the beginning there have been two opposing models for allostery: the Monod-Wyman-Changeux (MWC) or concerted model (Monod *et al.*, 1965) and the Koshland-Nemethy-Filmer (KNF) or sequential model (Koshland *et al.*, 1966). The MWC and the KNF models are limiting cases of a more general scheme for allosteric interactions (Hammes & Wu, 1971). This scheme allows the individual subunits to freely take on either of two conformational forms, regardless of the number of ligands that are bound (Figure 1.3(A)).

The MWC model can be described by a small number of parameters and can be summarized as follows:

- Allosteric proteins are symmetrical oligomers.
- There is only one binding site on each oligomer for each effector.
- The allosteric oligomers can exist in at least two states called R (relaxed) and T (tense). In the absence of ligand the two states are in equilibrium.
- Binding of ligand shifts the equilibrium towards the R conformation to which it binds with higher affinity.
- The signaling oligomers undergo reversible transitions between states through the change in quaternary organization, but the symmetry must be preserved. Thus, the transition is concerted.

The simplicity of mathematical description of the MWC model made it the most favored model for data analysis (reviewed in Changeux & Edelstein, 2005). As Figure 1.3(A) shows there are only two states, each with its own ligand binding dissociation constant. The most obvious limitation of the MWC model is its inability to account for negative cooperativity because a ligand can pull the conformational equilibrium only towards the form with a high affinity binding.

The KNF model avoids the assumption of symmetry but uses another simplifying feature. It assumes that the progress from T to the ligand bound R state is a sequential process, so that a series of intermediate conformations are obtained (diagonal in Figure 1.3). Each of the intermediate conformations can have its own unique values for the microscopic dissociation constants. As the protein switches from one state to

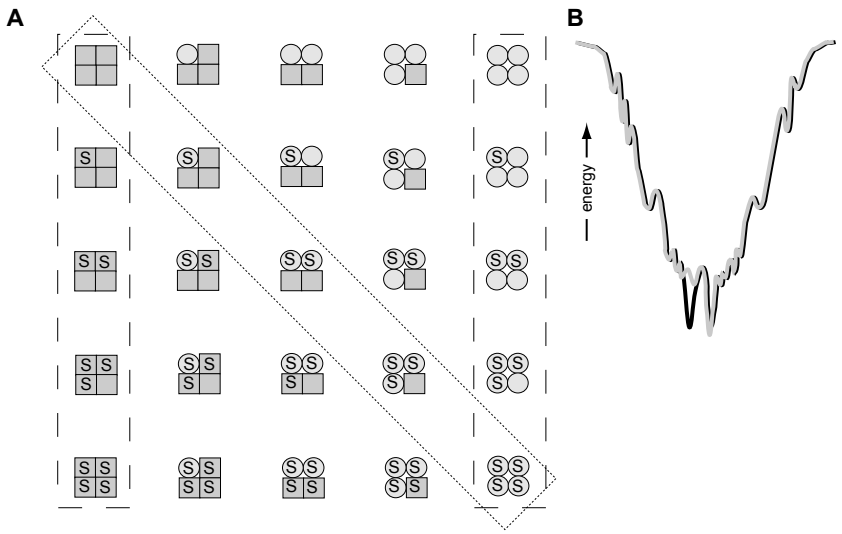


Figure 1.3: *Various models for allostery for tetrameric hemoglobin. (A) A general allosteric model for binding of substrate, S, to a four-subunit protein. The squares and circles are different conformations of the subunits. The MWC model involves only species enclosed by dashed rectangles, whereas the simplest sequential scheme involves the forms enclosed by the diagonal dotted rectangle. The free substrate and arrows between the states are omitted for the sake of clarity. Adapted from Hammes & Wu (1971) (B) Allosterically reshaping the energy landscape of a protein. Adapted from Swain & Gierasch (2006). Black represents the landscape in the absence of ligand and gray represents the ligand-bound state. Ligand binding to a native state with several energetically comparable states may stabilize one of those states at the expense of others, reducing conformational heterogeneity.*

another the successive dissociation constants can decrease or increase. This makes it possible to explain a wide variety of data. Systems with negative cooperativity cannot be described by the MWC model, but can be described in the framework of the KNF model (Stevens *et al.*, 2001; Koshland & Hamadani, 2002).

Since the introduction of classical models for long-range communication in proteins, the concept of allostery has evolved. The current view of allostery is based on the landscape paradigm (Weber, 1972; Kumar *et al.*, 2000; Swain & Gierasch, 2006) that was adopted from protein folding studies and is somehow similar to the MWC model. Under native conditions proteins have an energy landscape with many local minima corresponding to an ensemble of pre-existing structures with similar but discrete energy levels. Ligand binding leads to the selection of one or more conformational states from the large ensemble. Figure 1.3(B) depicts a sample scenario of the growing array of possible allosteric mechanisms based on the energy landscape reshaping theory. An allostery model based on the existence of multiple conformations in the protein's native state can also account for allostery without distinct conformational changes (Cooper & Dryden, 1984) because the frequencies of fluctuation also carry information. Therefore communication between distant sites could occur via

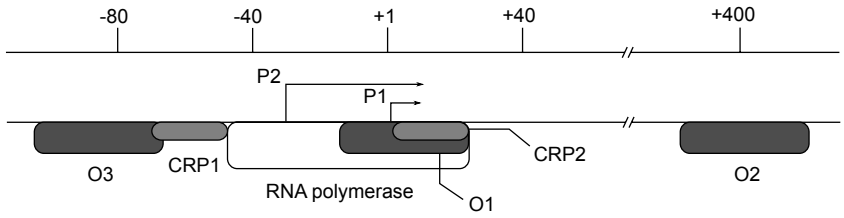


Figure 1.4: *Schematic representation of the regulatory DNA sequences of the lac operon. The cAMP receptor protein binding sites are designated CRP1 and CRP2. The two alternative transcription start sites P1 and P2 are indicated by arrows. Location of the three natural lac operators are denoted by O1, O2, and O3.*

changes in these dynamic frequencies (reviewed in Kern, 2003). In studies of dynamically driven allostery, NMR is very useful because it allows not only detailed characterization of the effects of binding on the structure but also on internal motions in various time-scales and on the populations in the ensemble of inter-converting states (for review see Markwick *et al.*, 2008). In contrast to the classical models for allostery that all assume a multimeric nature of the allosteric molecules, many monomeric proteins feature conformational plasticity that translates into allosteric behavior (Volkman *et al.*, 2001). Other extensions that have been introduced involve mutations, covalent modifications and changes in conditions, such as pH, as allosteric effectors (Johnson, 1992; Goodey & Benkovic, 2008).

An interesting application of allostery is the possibility of using allosteric sites as drug targets. Allosteric sites have a number of advantages over classic active-site ligands (Christopoulos, 2002). Allosteric sites are usually not conserved between different proteins, allowing for higher selectivity. Using allostery can prevent overdosage since when the allosteric sites are fully occupied, no further effect is observed. The difficulty however with such allosteric sites is that allosteric effectors are not generally easy to identify.

Binding of effectors to the Lac repressor

Lac repressor has two different effectors: operator DNA and inducer. Thus Lac repressor is a heterotropic system.

The *lac* operators

The primary *lac* operator, *O1*, is located 10 base pairs (bp) downstream of the starting site for the transcription from the *lac* promoter. When the *lac* promoter region was sequenced, two other DNA fragments resembling *O1* were discovered (Reznikoff *et al.*, 1974; Gilbert *et al.*, 1975). These were called *O2* and *O3*. *O2* is located 401 bp downstream and *O3* 92 bp upstream of *O1* (Figure 1.4). Since the *O1* operator and promoter sites overlap, it is clear that binding of the Lac repressor to the *O1* operator interferes with the binding of and transcription by RNAP. The function of the auxiliary operators could be to increase the efficiency of the repression through an

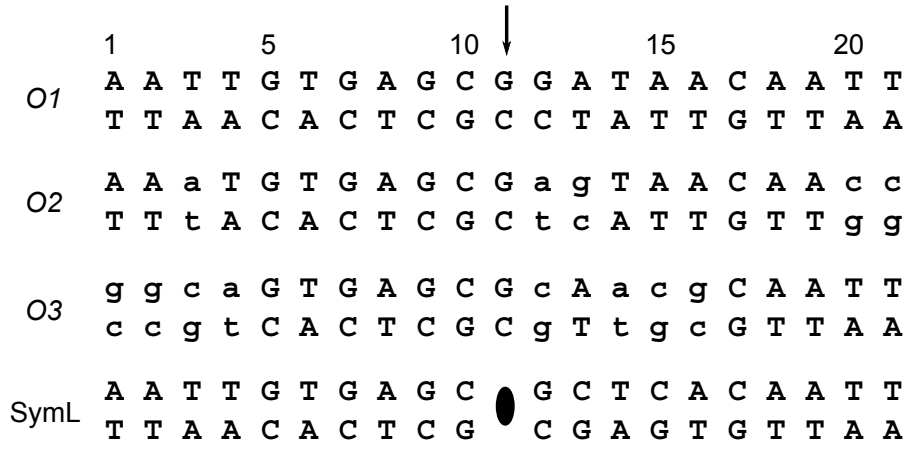


Figure 1.5: *Sequences of the natural lac operators and a symmetrical operator SymL. The central GC base pair is denoted by an arrow.*

increase in operator occupancy. Lac repressor, which is best described as a dimer of dimers, always binds with one dimer to the main operator *O1*, and the other dimer to one of the two secondary sites explaining the high affinity. As a result the intervening DNA is looped out (Krämer *et al.*, 1987, 1988). Additionally, when Lac repressor is bound to either *O2* or *O3* with one dimer, the local concentration of the second dimer at the *O1* site increases (Müller-Hill, 1998).

Natural *lac* operators are pseudo-palindromic (Figure 1.5). The two-fold symmetry is broken by a subtle variation in sequence between two half-sites and by insertion of a central G:C base pair that separates the two half-sites. First details about the repressor-operator interactions were elucidated from the analysis of constitutive mutations of the operators (O^c) (Miller & Reznikoff, 1978). The O^c operator constitutive mutation results in an increased basal level of expression of adjacent *lac* operon structural genes *in vivo* (Smith & Sadler, 1971) and a decreased affinity of the mutated operator DNA for the Lac repressor *in vitro* (Jobe *et al.*, 1974). Lac repressor interacts more strongly with the left half-site of the operator (Sasmor & Betz, 1990). Substitutions occurring in the left half-site (promoter-proximal) have a stronger effect on the affinity for repressor than when the mutations occurred in the right half. Also, no O^c mutations resulting in high constitutivity have been found on the right half of the operator (Miller & Reznikoff, 1978). Sadler *et al.* (1983) constructed an operator that is a perfect palindrome of the left half and lacks the central base pair (SymL) (Figure 1.5). The resulting synthetic operator binds 10 times stronger to the Lac repressor than wild type *O1*. Because of its high affinity for the Lac repressor and symmetry, the SymL operator has been used extensively in structural studies (Lewis *et al.*, 1996; Spronk *et al.*, 1999a; Bell & Lewis, 2000).

Specific vs. non-specific binding

The first structural details of repressor-operator interactions were elucidated from the NMR structure of the headpiece obtained by proteolysis (residues 1-51, HP51) bound to a 14 bp DNA fragment containing the left half of the *O1* site (Boelens *et al.*, 1987). When it was noted that headpiece HP56 (residues 1-56) would bind tighter and more specifically to a smaller 11 bp *O1* operator fragment (Lamerichs, 1989), detailed structural studies of HP56 bound to a *lac* half operator followed (Chuprina *et al.*, 1993). Lewis *et al.* (1996) showed that the hinge region in the Lac repressor is folded as a helix, as was similarly noted with the homologous Purine repressor (Schumacher *et al.*, 1994). However the X-ray structure of the Lac repressor-operator complex had a low resolution of 4.8 Å and could not provide details of the interactions between hinge helix and operator DNA. The solution structure of a longer headpiece HP62 construct (residues 1-62) bound to the symmetrical operator SymL, showed that folded hinge helices align in an antiparallel fashion and make both protein-protein and protein-DNA contacts in the minor groove (Spronk *et al.*, 1999a). Insertion of the hinge helices into the minor groove is accommodated by the opening of the minor groove and bending of operator by approximately 45° away from the protein, where intercalation of the sidechain of Leu56 in the central base-pair step is responsible for the bending of the operator. Extensive hydrophobic interactions between sidechains of residues Val52, Ala53 and Leu56 in the hinge helices create a dimerization interface within the headpiece that is crucial for a tight cooperative binding of the operator.

NMR structures of HP56 bound to the half-operator (Chuprina *et al.*, 1993) and HP62 bound to the SymL operator (Spronk *et al.*, 1999a) revealed that residues Tyr17 and Glu18 (positions 1 and 2 of the recognition helix) make hydrogen bonds to the edges of the bases in the major groove of the DNA. Together with Arg22 (position 5), which interacts favorably with a G:C bp, these residues are the main determinants of the specificity. This result was in full agreement with mutational studies that showed that it was possible to replace amino acids at positions 1 and 2 and the contacting bases to switch to the specificity of the Gal repressor (Lehming *et al.*, 1987). In addition, there are a number of anchoring hydrogen bonds to the DNA phosphate backbone (residues Leu6, Asn25, His29, Val30, Ser31 and Thr34) and hydrophobic contacts that stabilize binding (Figure 1.6).

To understand the DNA recognition mechanism of the Lac repressor several studies have been performed with the Lac repressor and the headpiece bound to the natural operators. However, since the crystal structure of the dimeric Lac repressor in complex with the main operator *O1* had a low resolution of 4 Å, the protein-DNA interactions in the major groove could not be described in high detail (Bell & Lewis, 2001). Similarly, unfavorable dynamics of HP62 in complex with the natural *O1*, *O2* and *O3* operators became an obstacle for NMR studies. Therefore Kalodimos *et al.* (2001) constructed a dimeric mutant of the HP62 by introducing a disulfide bond between partner hinges via the mutation of Val52 to Cys as inspired by a mutation in the full length Lac repressor that enhanced binding affinity (Falcon *et al.*, 1997). Also, the dimeric HP62V52C protein had five orders of magnitude higher affinity for the natural operator *O1* as compared with the monomeric form and gave high quality NMR spectra without exchange broadening. In **Chapter 2** NMR structures of the HP62V52C-*O2* and HP62V52C-*O3* complexes are described and compared with the

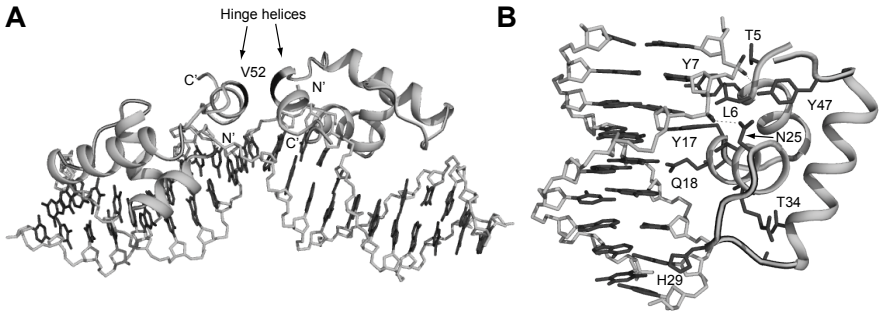


Figure 1.6: *Solution structure of the Lac headpiece HP62 bound to SymL operator (PDB code 1CJG). (A) The overall structure of the HP62-SymL complex. Position of Val52 is shown in dark grey. (B) Closer view of the interactions in the major groove. As an example, the hydrogen bond between the amide group of the side chain of Asn25 and the phosphate of Cyt9 is shown by dotted line.*

previously determined HP62V52C-*O1* complex structure (Kalodimos *et al.*, 2002) to gain further insight into the Lac repressor recognition mechanism of the natural operators.

The Lac repressor also binds to non-operator DNA. In fact, experiments with *E. coli* minicells demonstrated that *in vivo* 90 % of the repressor molecules are bound to the *E. coli* chromosome (Kao-Huang *et al.*, 1977). Lac repressor has a much lower affinity for the non-operator DNA than for the *lac* operators and this non-specific affinity is not changed by the binding of inducer to the repressor (Lin & Riggs, 1972; Riggs *et al.*, 1972; Butler *et al.*, 1977). However, the non-specific binding is highly dependent on salt concentration; for example, the association constant for the Lac repressor with non-operator DNA is decreased over 1000-fold by increasing the Na^+ concentration from 0.12 to 0.24 M (Butler *et al.*, 1977). This phenomenon was explained through analysis of the solution structure of the HP62V52C bound to the symmetrical non-operator (Kalodimos *et al.*, 2004) which revealed that most protein-DNA contacts in the non-specific complex are electrostatic, in line with earlier thermodynamic analysis (Mossing & Record, 1985).

It has been noted that Lac repressor binds its cognate DNA sequence with the association rate of 10^{10} $M^{-1}s^{-1}$, which is much higher than the maximal rate achievable by three-dimensional diffusion (10^8 $M^{-1}s^{-1}$) (Riggs *et al.*, 1970c,a; Berg *et al.*, 1981; Winter & Von Hippel, 1981; Winter *et al.*, 1981). Therefore it was proposed that the initial non-specific binding accelerates the rate that the repressor finds its operator, allowing it to 'slide' or 'hop' along the DNA before arriving to its target site (Winter *et al.*, 1981). Thus non-specific binding plays an important role in the cell.

Binding of inducers and anti-inducers

Different types of effectors bind to the Lac repressor at its ligand binding site. Based on their effect on the *lac* operon *in vivo*, they have been classified as inducers, anti-inducers and neutral ligands. Galactosides that function as anti-inducers inhibit in-

duction by inducers such as IPTG by competing for the binding site.

Barkley *et al.* (1975) quantified the interactions between the effectors and the Lac repressor in the absence and presence of the *lac* operator. Anti-inducers generally bind to repressor with lower affinity than inducers. The most potent anti-inducer is o-nitrophenyl-β-D-fucoside (ONPF) with a moderate affinity of approximately 10^3 M^{-1} (Jobe & Bourgeois, 1972b) whereas IPTG has a high affinity for the repressor ($\sim 10^6$ M^{-1}). The dissociation rate of the Lac repressor from the operator is 1000-fold faster in the presence of the gratuitous inducer IPTG. Anti-inducers on the other hand have a small stabilizing effect on the repressor-operator complex, at most five-fold in the case of lactose. ONPF decreases the dissociation rate only 3-fold. It has also been found that inducers bind with greater affinity to free repressor than to repressor-operator complex (Barkley *et al.*, 1975; O'Gorman *et al.*, 1980) whereas anti-inducers show a mixed behavior. Natural anti-inducers, glucose and lactose, satisfy the criterion of anti-inducing ligands, binding with greater affinity to repressor-operator complex than to the free repressor. ONPF exhibits anomalous behavior in the sense that it stabilizes the repressor-operator complex, yet seems to bind with greater affinity to the free repressor than to the repressor-operator complex. The non-natural inducers and anti-inducers are frequently used in the biochemical and biophysical experiments. ONPF is also used in structural studies because of its stabilizing effect on the Lac repressor-operator complex (Bell & Lewis, 2000; Daber *et al.*, 2007).

Analysis of the X-ray structures of the Lac repressor-IPTG and Lac repressor-ONPF complexes (Daber *et al.*, 2007) revealed that anti-inducers are competitive inhibitors of inducer binding because they both bind at the same binding site. Both ligands are anchored in the binding site in an identical fashion, but differ in the hydrogen bonding around the C6 position of the pyranose ring. As a fucoside does not have the O6 hydroxyl, it cannot form the water-mediated hydrogen bonding network that crosslinks the N- and C-terminal subdomains as is observed with the inducer. This hydrogen bonding bridge between two subdomains stabilizes a conformation that has a reduced affinity for the operator. From the comparison of structures of repressor-ONPF and repressor-operator-ONPF complexes the authors also concluded that ONPF itself is not responsible for stabilizing the repressed state, since the anti-inducer interacts with the repressor in a fashion independent of repressor conformation. This is in line with the current NMR study that showed that ONPF does not induce any conformation changes upon binding to the core domain of the Lac repressor (**Chapter 4**).

Allosteric mechanism

The allosteric mechanism of the Lac repressor has been analyzed by comparing structures of the Lac repressor tetramer in three crystal forms: apo, DNA bound, and IPTG bound. The conformation of the apo repressor was very similar to the inducer bound form with a rmsd value of only 0.4 Å. In contrast, the unliganded conformation and DNA bound conformation were quite different with a rmsd value of 1.9 Å (Lewis *et al.* (1996)). Therefore, the authors assumed that the Lac repressor can adopt two conformations: repressed and induced. The major difference between those two conformations was the orientation of the N-subdomains relative to each other. The structural change at the N-subdomains altered the dimerization interface. This in turn

caused the displacement of the C_α carbon of residue 62, which moved 3.5 Å away from its two-fold related mate in the IPTG bound conformation. This results in the disruption of the interactions of the hinge helices, reduction of the affinity of the repressor for DNA, and release of the HTH motif from the DNA.

Because the X-ray structures of the tetrameric Lac repressor bound to the DNA operator was determined at 4.8 Å resolution it could not provide details of the side chain interactions at the dimerization interface within the N-subdomain. Therefore Bell & Lewis (2000) constructed a dimeric Lac repressor and determined its structure in complex with a symmetrical DNA operator and the anti-inducer ONPF at a much higher resolution of 2.6 Å. Comparison of this structure with the IPTG bound structure provided a detailed view of how the interactions at the dimer interface are altered in switching from the operator-bound to the inducer-bound conformation (Figure 1.7). For example, in the induced state, a salt-bridge is formed between Lys84 of one monomer and Glu100 of the other monomer, but in the repressed conformations this salt-bridge is broken. Lys84 is wedged between two β-strands in the induced conformation, but in the repressed conformation this lysine contacts the carbonyl oxygen of Val94 of the same monomer and Val96 of the other monomer. Near the inducer binding site, an intersubunit salt-bridge is formed between His74 of one monomer and Asp278 of the other monomer in the IPTG-bound conformation, but is broken in the operator bound conformation. It has been proposed that these interactions at the dimerization interface of the N-subdomain play a crucial role in the allosteric mechanism (Lewis *et al.*, 1996; Bell & Lewis, 2000).

The clearer electron density of the headpiece permitted the authors to identify extensive interactions between the N-subdomain of one monomer and the headpiece of the other monomer. At this interface, the loop connecting the HTH domain and the hinge helix (residues 46-51) contacts residues 113-118 of the N-subdomain. Thus, the headpiece of one monomer is anchored to the N-subdomain of the other monomer via a number of interactions. Mutations of residues at this interface result in the I^- phenotype, i.e. the repressor that fails to bind *lac* operator sequences, suggesting that these interactions are important for stabilization of the operator bound conformation. Moreover, the integrity of these interactions is dependent on the orientation of the N-subdomain, which in turn is dependent on inducer binding. Based on these observations, Bell & Lewis (2000) proposed that the disruption of the hinge helix upon inducer binding might be due to an alteration of intersubunit interactions rather than the physical pulling of the hinge helices from the minor groove as suggested before (Lewis *et al.*, 1996). On the basis of Molecular Dynamics simulations Flynn *et al.* (2003) suggested how the allosteric signal might be transmitted from the core to the headpiece. However, there is thus far no experimental data to confirm this.

Genetic studies

Miller and co-workers used 14 different suppressors of nonsense mutations to generate more than 4000 single amino acid substitutions throughout the protein (Miller, 1984; Markiewicz *et al.*, 1994; Suckowa *et al.*, 1996). Phenotypic analysis of the mutants identified regions of the Lac repressor that were important for DNA binding and for induction.

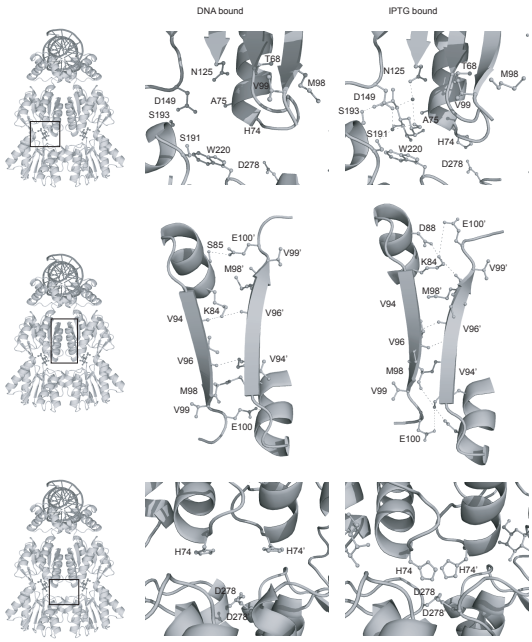


Figure 1.7: *Close-up views showing the intersubunit interactions in the Lac repressor (PDB code 1EFA). To the left of each panel, a dimeric repressor-operator-ONPF complex shows the locations of the relevant subunit interfaces (boxed region). The DNA is shown in a cartoon representation and the ONPF is shown in ball-and-stick. Hydrogen bonds are shown as dotted lines. Upper panel depicts interaction of IPTG with amino acids in the inducer binding pocket. The change of the interaction network at the monomer-monomer interface within the N-subdomain upon IPTG binding to the protein-DNA complex are shown the middle panel. Lower panel demonstrates formation of an ion pair involving the His74-Asp278 at the dimerization interface in the IPTG-bound (left) and how this ion pair is broken in the operator-bound (right) conformations.*

Since the headpiece is absolutely essential for DNA binding, it is not surprising that mutation of almost all residues in the HTH domain and the hinge helix alters the phenotype of the repressor to the I^- phenotype, i.e. a repressor that fails to bind DNA. Also mutations that compromise the integrity of the Lac repressor structure in the core result in a defective repressor molecule and I^- phenotype.

Mutations that result in the I^s phenotype, i.e. a protein that binds to the operator DNA with wild-type affinity but is incapable of induction, can be divided into two groups: (1) mutations at the inducer binding site that disrupt direct contacts to the inducer; and (2) mutations that affect a correct response to inducer. In particular, residues located at the dimerization interface of the N-subdomain exhibit a strong I^s phenotype.

Several mutants near the various interfaces with interesting properties have been characterized in high detail (Müller-Hartmann & Müller-Hill, 1996; Falcon *et al.*, 1997; Gerk *et al.*, 2000). For instance, the substitution of the apolar side chain of Lys84 increases the stability of the Lac repressor (Nichols & Matthews, 1997). Both the K84L and K84I substitutions have a ten-fold reduction in operator affinity, whereas the affinity of the K84A and K84M repressors remains similar to the wild-

type protein (Gerk *et al.*, 2000). Lys84 is located in a non-polar environment. Burial of this unpaired charge has a destabilizing effect on the wild-type protein. The energy that would be required to deprotonate and bury the two charged lysine residues of the dimer would be so great that the subunit interface would favor to undergo a conformational change where the charged side chain is exposed to the solvent. As a result the protein gains a conformational flexibility to function as a molecular switch.

Mutation at position 110 is thought to influence the equilibrium between the induced and repressed conformations of the Lac repressor (Müller-Hartmann & Müller-Hill, 1996). A repressor with an A110T substitution has a higher affinity for the inducer and a lower affinity for the *lac* operator, while the A110K mutant has the opposite phenotype. This observed shift closely matches the changes in equilibrium we observe by NMR when either operator or IPTG bind free Lac repressor (**Chapter 4**).

Outline of this thesis

The scope of this thesis is structural studies of the Lac repressor by NMR spectroscopy. In **Chapter 2** we study the complexes of Lac headpiece with its auxiliary operators *O2* and *O3*. Structural analysis and comparison with the previous HP62V52C-*O1* complex gives detailed insight into the operator recognition mechanism of the Lac repressor and our results can well explain the differences in affinity for different *lac* operators. In **Chapter 3** and **Chapter 4** we extended our NMR studies to a fully functional dimeric Lac repressor. In **Chapter 3** we describe the production of the thermostable dimeric K84M mutant of the Lac repressor that allowed us to record high quality NMR spectra suitable for a complete NMR resonances assignment. In **Chapter 4** we characterize all four functional states of the Lac repressor: free, inducer-bound, operator-bound and ternary complex through detailed analysis of their chemical shifts. Analysis of these complexes yielded profound new understanding of the allosteric mechanism of the Lac repressor, and supports the original MWC hypothesis.

Chapter 2

Specificity and affinity of Lac repressor for the auxiliary operators *O2* and *O3* is explained by the structures of their protein-DNA complexes.

Abstract

Here we present the solution structures of a dimeric mutant of the Lac repressor DNA binding domain complexed with the auxiliary operators *O2* and *O3*. To understand the Lac repressor recognition mechanism of various *lac* operators we compared the structures to the previously determined structures of the Lac DNA binding domain bound to the main operator *O1* (Kalodimos *et al.*, 2002) and to the non-operator DNA (Kalodimos *et al.*, 2004). Structural analysis of the *O1* and *O2* complexes show highly similar structures with very similar number of specific and non-specific contacts, which agrees with the similar affinities for the two operators. The DNA binding domain can adapt its side chains to compensate for small differences in the sequence of the two operators. The left monomer of the Lac repressor in the Lac-*O3* complex retains most of the specific contacts. However, dramatic alterations in the sequence of the right half-site of the *O3* operator lead to the loss of contacts between the protein and the DNA, explaining the low affinity of the Lac repressor for the *O3* operator. The binding mode in the right-half site resembles that found in the non-specific complex. In contrast to the Lac-non-operator DNA complex where hinge helices are not formed, in the weak Lac-*O3* complex the stability of the hinge helices is the same as in the Lac-*O1* and Lac-*O2* complexes as judged from the results of the hydrogen-deuterium exchange experiments.

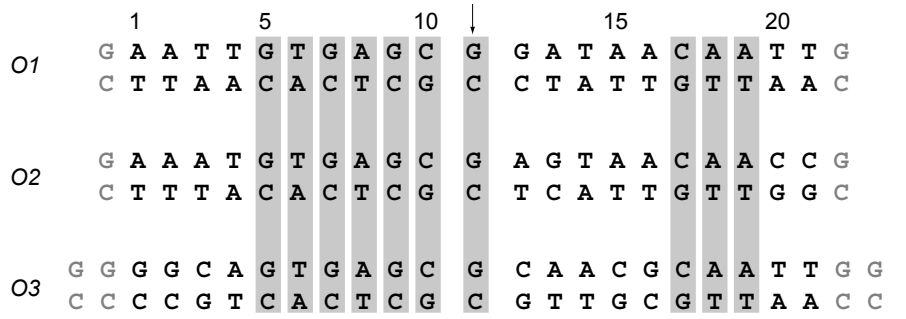


Figure 2.1: *Sequences of the naturally occurring lac operators. The three natural operators O1, O2 and O3 are aligned with the numbering referred to in the text and figures. The arrow denotes the central base pair. The bases conserved in all natural operators are highlighted. The two binding sites within each operator are referred to as the left (bp 1-10) and right (bp 12-21). Extension of the operators are shown in gray.*

Introduction

The Lac repressor is a regulatory protein that controls the expression of genes necessary for lactose metabolism in *Escherichia coli* (reviewed in Lewis, 2005; Wilson *et al.*, 2007). The effective downregulation of these genes is achieved by the presence of multiple Lac repressor binding sites within the *lac* operon. The tetrameric Lac repressor functions as a dimer of dimers, and simultaneously binds to the main operator *O1* and to either of the auxiliary operators *O2* and *O3* (Reznikoff *et al.*, 1974; Gilbert *et al.*, 1975, 1976), creating one of two alternative DNA loops (Krämer *et al.*, 1987, 1988). When the protein is bound to one of the two auxiliary sites, the other dimer is consequently in close proximity of the main operator and vice versa. This tethering of one site to another increases the local concentration of the protein near the promoter-proximal operator (Oehler *et al.*, 1994; Müller-Hill, 1998), thereby stabilizing the tetrameric complex on the promoter and increasing the repression efficiency.

The promoter-proximal operator *O1* is positioned 10 base pairs (bp) downstream of the promoter. The auxiliary operator *O2* is located 401 bp downstream of *O1* within the *lacZ* gene, while *O3* lies 92 bp upstream of *O1* within the *lacI* gene. Natural *lac* operators are pseudo-palindromic sequences, where the symmetry is broken by variations in the sequence between the two half-sites and by insertion of the central G:C base pair (Figure 2.1). The two binding sites differ significantly in their affinity for the Lac repressor when considered separately (Sasmor & Betz, 1990). 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 in the right site (Sadler *et al.*, 1983; Betz *et al.*, 1986).

Oehler *et al.* (1990) have shown that *in vivo* expression of β-galactosidase is repressed by a factor of 1300 in the presence of three active operators. Mutational studies revealed that the operator sequences possess distinct repression efficiency, reflected by the different affinity of the Lac repressor for the various operator sequences (Oehler *et al.*, 1994). For example a promoter containing only *O2* in the position of

O1 has only 10 % of the repressive capacity *in vivo* compared with *O1*, while *O3* can only marginally (0.3 %) repress transcription from the Lac promoter. The presence of a functional *O1* is essential as, even in the presence of both auxiliary operators, mutation/deletion of *O1* leads to an almost complete loss of repression (Betz *et al.*, 1986; Oehler *et al.*, 1990). While inactivation of either *O2* or *O3* results in a slight decrease of repression, the combined loss of both *O2* and *O3* leads to a significant ($\sim$70-fold) decrease of repression (Oehler *et al.*, 1990). Therefore the auxiliary operators contribute significantly to the transcriptional repression of the *lac* operon. The difference in the affinity of the Lac repressor for its operators can be ascribed to the variation in their sequences. *O1* and *O2* operators have similar base pair composition while the *O3* sequence differs significantly (Figure 2.1). Not only the sequence identity but also the position of the auxiliary operators is important. Müller *et al.* (1996) have shown that maxima of repression are found at inter-operator distances of 59.5, 70.5, 81.5, 92.5 and 115.5 bp. In the *lac* operon the weakest operator *O3* is located at one of the repression maximum corresponding to the distance of 92.5 bp.

Initial structural studies have used either a left site of the *O1* operator (Boelens *et al.*, 1987; Chuprina *et al.*, 1993) or a fully symmetric 'ideal' operator (SymL) (Lewis *et al.*, 1996; Spronk *et al.*, 1999a; Bell & Lewis, 2000) to learn about the recognition mechanism. The SymL operator which binds to the Lac repressor with the highest affinity is a palindrome of the left half-site of *O1* and lacks the central G:C base pair. The Lac repressor DNA binding domain or headpiece (HP) recognizes *lac* operators through a canonical helix-turn-helix (HTH) DNA-binding motif. The C-terminal residues Arg50-Gly58 of the headpiece form a so called hinge region which is unstructured in the absence of DNA. Solution structures of HP56 (residues 1-56) in complex with the left half-site of *O1* (Chuprina *et al.*, 1993) and of HP62 (residues 1-62) bound to the SymL operator (Spronk *et al.*, 1999a) provided details of the interactions between the residues of the recognition helix of the HTH motif and the bases in the major groove. By using a longer construct of the headpiece Spronk *et al.* (1999a) demonstrated that the hinge region undergoes a coil-to-helix transition upon binding to the SymL. The folded hinge helices align in an antiparallel fashion in the minor groove in the center of the operator and form extensive protein-protein and protein-DNA interactions. Because of these interactions the DNA structure shows an opening of the minor groove and a global bending of approximately 45°.

Spronk *et al.* (1999b) have shown that the central base pair in the natural *lac* operator acts as a spacer to create the optimal spacing between two half-sites for formation of the hinge helices and their interaction with the DNA because binding of the headpiece to the right half-site is shifted towards the center of the operator. The authors also showed that a coil-to-helix transition of the hinge region occurs upon complex formation between the headpiece and the natural operator *O1*. This mode of binding of the Lac repressor to the natural operator was supported by the X-ray structure of the dimeric Lac repressor bound to its main operator *O1* (Bell & Lewis, 2001). The analysis showed that the hinge region adopted an alpha helical conformation and was bound to the minor groove between the left half-site and the central G:C base pair. Due to the low resolution of the structure, the interactions between the repressor side chains and DNA bases could not be established. On the other hand NMR studies were hindered by unfavorable dynamics of the isolated monomeric Lac headpiece in complex with the *O1* operator probably due to the low affinity of the

Lac repressor DNA binding domain for the natural operators. To decrease the off-rate of the isolated HP for natural operators we have previously prepared a covalently linked dimeric Lac headpiece by the substitution of Val52 by Cys (HP62V52C). This mutation was previously reported in the full length protein to permit repression *in vivo* (Falcon *et al.*, 1997) while we found that this dimeric headpiece had comparable affinity for the natural operator as the intact dimeric Lac repressor (pM range) (Kalodimos *et al.*, 2001). High affinity of the Lac headpiece mutant for the *O1* operator permitted to solve the solution structure of the HP62V52C-*O1* complex to high resolution (Kalodimos *et al.*, 2002). This structure shed light on the distinct binding mechanism for HP to the left and right halves of the operator. In agreement with the crystal structure, the hinge helices are bound in the minor groove between bp 10 and 11, thereby introducing a kink in the DNA. The global positioning of the dimer on the operator was dramatically asymmetric, which resulted 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 the structure of the Lac headpiece bound to the SymL operator (Spronk *et al.*, 1999a).

In an ongoing effort to understand specificity and recognition of various operator sequences by the Lac repressor we determined the NMR structure of the dimeric Lac headpiece and its auxiliary operators *O2* and *O3*. We also compared structures of the HP62V52C-*O2* and the HP62V52C-*O3* complexes with the previously determined HP62V52C-*O1* structure (Kalodimos *et al.*, 2002) and the structure of the Lac HP bound to a non-operator DNA (NOD) fragment (Kalodimos *et al.*, 2004). The analysis of structures of the various Lac-operator complexes helps to understand the specific recognition of the natural operators by the Lac repressor and gives insight into the difference in the Lac headpiece-operator affinity from a structural point of view. The protein-DNA interactions are very similar between the HP62V52C-*O1* and HP62V52C-*O2* complexes as judged from the structural analysis. At the base pairs substitutions sites, the DNA contacting headpiece residues are able to adopt distinct conformations to compensate for these minor differences. The binding mode of the Lac headpiece to the weakest binder *O3* operator is a composition of the specific in the left-half site, similar to the *O1* and *O2* operators, and the non-specific binding modes. Despite this non-specific binding mode, the hinge helices are stable in the weakest HP62V52C-*O3* complex to the same extent as in other two specific complexes and to contrast to the HP62V52C-NOD complex where hinge region is unfolded (Kalodimos *et al.*, 2004).

Results

Relative *in vitro* affinities and DNA bending

The binding affinities of the dimeric HP62V52C for the natural *lac* operators and an artificial NOD fragment were compared using electrophoretic mobility shift assays (Figure 2.2(A, B); Table 2.1). The headpiece dimer binds the *O1* and *O2* operators with an apparent equilibrium dissociation constant (K_d) of 0.05 and 0.1 nM, respectively. Thus the dimeric Lac headpiece interacts with the *O1* and *O2* operators with comparable affinity. *O3* operator has a considerably lower K_d of 100 nM which

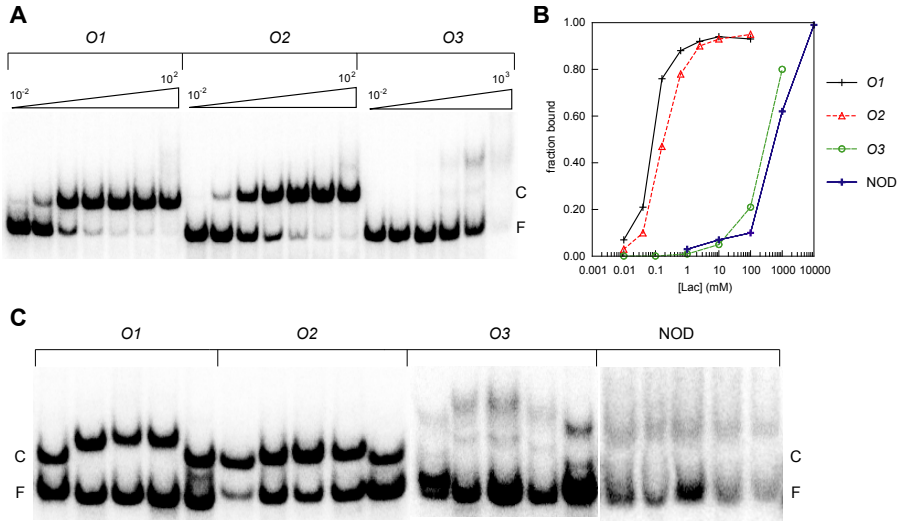


Figure 2.2: *DNA binding and bending experiments. (A, B) Representative DNA binding experiment using the indicated operators with a serial dilution of dimeric HP62V52C. The highest and lowest protein concentrations (nM) are indicated above each experiment. The different operator fragments are described in Figure 2.1. (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 2.1: *Binding affinities and bend angles obtained with the dimeric Lac HP62 and the various operators.*

	K_d[a] (nM)	Bend angle[b] (°)	Bend angle[c] (°)
HP62-*O1*	0.05	25	33±8
HP62-*O2*	0.1	25	25±7
HP62-*O3*	100	40	14 ±9
HP62-NOD	200	<5	0

[a]Binding affinities are presented as apparent K_d (see Materials and methods section)

[b]Bend angles were determined using the circular permutation method (Wu & Crothers, 1984).

[c]The average bend angles of the DNA operator in the structures of the ensemble as estimated with CURVES (Lavery & H., 1988).

is surprisingly similar to the observed affinity for the non-specific DNA sequence (200 nM). On both operators, in agreement with the low affinity, dissociation of the protein-DNA complex during electrophoresis was observed, and interestingly on the *O3* operator sequence at least two distinct complexes were formed.

To determine whether the headpiece can induce a bend on the natural operators, we used the circular permutation method (Wu & Crothers, 1984), essentially as per-

formed before (Spronk *et al.*, 1999b). The profile of the migration of the complexes with the 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° (Figure 2.2(C), Table 2.1). Remarkably the two distinct complexes formed on the *O3* operator posses larger but different bend angles than other operator sequences.

Hydrogen - deuterium experiments

We have carried out hydrogen-deuterium (H/D) exchange experiments in order to establish the difference in the stability of the three Lac-operator complexes. These experiments can give insight into the interaction mechanism by providing information about solvent accessibility. A graph of protection factors (P) for all residues in the HP62V52C protein defined as the ratio between the observed amide H/D exchange rates and calculated rates of unprotected amino acids as explained in the Materials and Methods section is shown in Figure 2.3.

The protection factor values in the HTH domain for the dimeric Lac headpiece directly reflect its affinity for the natural *lac* operators. The covalently linked Lac repressor headpiece binds to the *O2* operator *in vitro* with a two times lower affinity compared to the *O1* operator (Figure 2.2(A, B)). This is in agreement with the protection factors of HP62V52C-*O2* complex that are slightly lower. The *O3* operator is the weakest binder and also has the lowest protection factor values, with a maximal difference for the residues in the HTH motif. However the difference in the protection factors for the hinge helices is minimal for the three complexes, which agrees with the structural similarities for the hinge regions in all three structures (described below). In sharp contrast the corresponding region in the non-specific complex is poorly protected and hinge helices are not formed (Kalodimos *et al.*, 2004).

The natural *lac* operators are asymmetrical, and the two half sites differ significantly in their affinity for the Lac repressor when considered separately (Sasmor & Betz, 1990). However, the protection factors of the Lac headpiece amides are the same for the left and the right protein subunits in the three complexes (Figure 2.3(C, D, E)). The observation that the H/D exchange rates of the HTH domain correlate with the dissociation constant of the protein-operator complexes together with the found identity between the rates for the left and right half-sites for all operators suggests that exchange occurs via a free headpiece. In contrast the H/D exchange in the hinge helices is the same in the HP62V52C-*O1*, HP62V52C-*O2* complexes and even in the very weak HP62V52C-*O3* complex. This indicates that the exchange and thus the hinge helix unfolding already occurs while the protein is bound to DNA.

Structure determination

We have determined the solution structure of a dimeric Lac HP62V52C in complexes with its auxiliary operators *O2* and *O3*. The *O2* and *O3* operator sequences used in this study were 23 and 27 base pairs long, respectively (Figure 2.1). We extended the naturally occurring 21 base pairs long operators by one base at each terminus for *O2* in order to prevent double-stranded DNA ends from fraying. Initially we prepared a 23 base pairs double-stranded *O3* operator that showed very low affinity for the dimeric headpiece and abnormal NOE connectivity patterns in the 2D NOESY

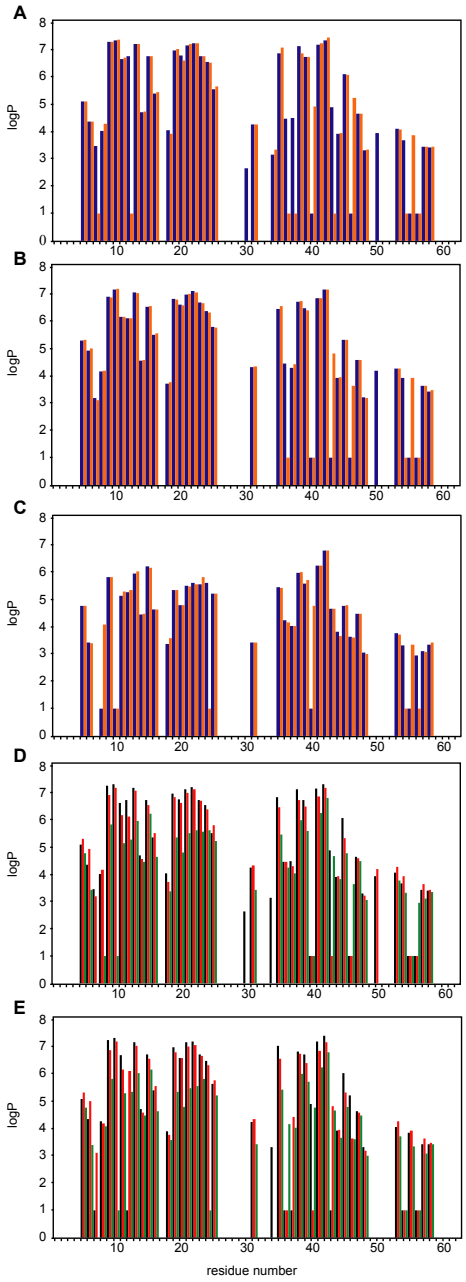


Figure 2.3: *Per residue protection factors obtained from amide H/D exchange experiments for the lac repressor headpiece bound to the natural operators O1 (A), O2 (B) and O3 (C). In (A, B, C) blue bars refer to the residues of the left subunit, whereas the orange ones to the right. (D) Protection factors for the left monomer for each complex. (E) 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 logP = 1 was arbitrarily given to residues not included in the analysis due to peak overlap. Full color figure on page 90.*

spectra of the free *O3* operator. Therefore three base pairs were added on both sides to avoid base pairs mismatch during annealing and also to prevent DNA ends fraying.

The structures of the two HP62V52C-DNA complexes were solved by heteronu-

Table 2.2: *Restraints statistics of HP62V52C-O1, HP62V52C-O2 and HP62V52C-O3 complexes*

	HP62-*O1*	HP62-*O2*	HP62-*O3*
Number of experimental restraints			
Protein (dimer)			
Intraresidue NOEs	514	825	661
Sequential ($\|i-j\|=1$)	419	513	329
Medium-range ($1<\|i-j\|\leq 4$)	260	350	234
Long-range ($\|i-j\|>4$)	246	250	144
Interprotein	28	29	13
DNA NOEs	680	427	299
Total number of distance restraints	2392	2538	1755
Dihedral angle restraints	162	142	70
Dipolar coupling restraints (D_{HN})	84	91	72
Total number of NMR restraints	2638	2771	1897

clear double and triple resonance NMR spectroscopy using ^{15}N- and ^{13}C-labeled protein and unlabeled oligonucleotides. The NMR titrations of the HP with its cognate operators show for all cross peaks that the free and bound protein are in the slow exchange regime, indicating low off-rates of the headpiece in the specific complexes. Also both complexes show asymmetry as judged by the presence of a double set of resonances due to the asymmetric nature of the operator sequences. Structural information was derived from NOESY spectra, dihedral angles, hydrogen-deuterium exchange and residual dipolar couplings (RDCs). One-bond $^{1}D_{HN}$ dipolar couplings could be measured for the protein due to the natural field dependent alignment of the DNA molecules caused by anisotropic magnetic susceptibility (Tjandra *et al.*, 1997). For HP62V52C in complex with DNA this results in maximum RDCs of 2.7 Hz (*O1*), 2.4 Hz (*O2*) and 2.7 Hz (*O3*).

The structure determination protocol was similar to the one used for the previously determined HP62V52C-*O1* complex (Kalodimos *et al.*, 2002). First the structure of the protein in the bound state was determined and refined with RDCs. Then this structure was docked onto the standard B-form DNA sequence using HADDOCK2.0 (Dominguez *et al.*, 2003; de Vries *et al.*, 2007) with the manually identified intermolecular NOEs as restraints. The structure of the HP62V52C-*O2* complex was calculated using 2771 experimental NMR restraints (Table 2.2). For the HP62V52C-*O3* complex a lower number of collected restraints were identified (1897) (Table 2.2). This can probably be attributed to a combination of resonance broadening effects associated with residues contacting DNA, overall lower stability and larger size of this complex. This results in the significantly larger backbone rmsd values for the well-ordered segments for HP62V52C-*O3* (1.46 Å) in comparison with 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 (Cornilescu *et al.*, 1998). The higher rmsd and Q-factor values could reflect conformational exchange

Table 2.3: *Structural statistics of HP62V52C-O1, HP62V52C-O2 and HP62V52C-O3 complexes*[a]

	HP62-*O1*	HP62-*O2*	HP62-*O3*
RMSD (Å) with respect to mean[b] *(backbone/heavy)*			
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
Operators	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
RMSD from experimental restraints			
All distance restraints (Å)	0.06± 0.002	0.03±0.002	0.03±0.002
Dihedral 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[c]			
D_{HN}	0.18±0.018	0.13±0.012	0.32±0.01
Restraints violations[d]			
Distances $>$ 0.3 Å	8±2.2	1.0±1.1	3.0±1.26
Dihedrals $>$ 5°	1.4±1.9	1.8±2.1	0
D_{HN} $>$ 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
Impropers (°)	0.7±0.08	0.65±0.13	0.55±0.03
Ramachandran analysis[e] (%)			
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)	-962±37	-1081±27	-981±21
E_{elec} (kcal/mol)	-5404±79	-5467±112	-5443±98

[a]Based on 20 final structures for HP62V52C-*O1* and HP62V52C-*O2* complexes and on 10 final structures of HP62V52C-*O3*.

[b] Residues 4-58 for the protein and bp 2-21 for the operator are taken into account.

[c] Q-factors were calculated with Pales (Zweckstetter & Bax, 2000).

[d]There were no NOE violations larger than 0.5 Å.

[e] Residues 4-59 were taken into account.

on a time scale slower than nanoseconds. Also the gel shift assays indicate the presence of significant structural heterogeneity for the *O3* complex, in contrast with the other complexes. Further structural statistics is shown in Table 2.3. To avoid differences in the structures of the three complexes due to the structure determination protocol we have refined the structure of HP62V52C bound to the main operator *O1* with RDCs in an identical manner and recalculated the structure of the complex us-

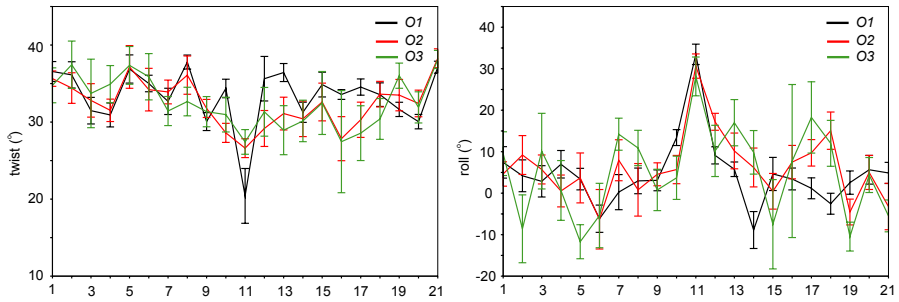


Figure 2.4: *Inter base pair DNA helical parameters of the lac operators O1 (black), O2 (red), and O3 (green) in complex with dimeric HP62. Full color figure on page 91.*

ing available restraints and the same structure calculation protocol as for the *O2* and *O3* operator complexes. The structural statistics for the HP62V52C-*O1* complex is given in Table 2.3. Overall the current structure resembles closely the structure of the previously reported HP62V52C-*O1* complex.

Structure description

The ensemble of the lowest energy structures of the HP62V52C-*O2* and HP62V52C-*O3* complexes are presented in Figure 2.5. As expected, the overall topology of newly determined complexes is identical to the HP62V52C in complex with the *O1* operator. The alpha-helical content is the same in both the left and right protein subunits and the recognition helix is aligned with its axis parallel to base-pairing edges of the 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 bp 10 and 11 introduces a kink in the DNA, which is distinct for all three structures (Table 2.1). This difference is reflected in the DNA helical parameters for the various complexes, showing large but distinct deformations around the central base pair (Figure 2.4). 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.

We found a good correlation between the binding strength and the accessible surface area that is buried upon binding of the Lac headpiece to its natural operators. The largest buried surface area is found in the HP62V52C-*O1* complex (total ∼3800 Å^2). When the dimeric Lac headpiece binds to the *O2* operator, ∼3500 Å^2 of the total solvent-accessible surface is buried which is very similar to the HP62V52C-*O1* complex. The HP62V52C-*O3* complex has the smallest accessible surface area (∼3000 Å^2) if compared to the buried surface area for other two complexes.

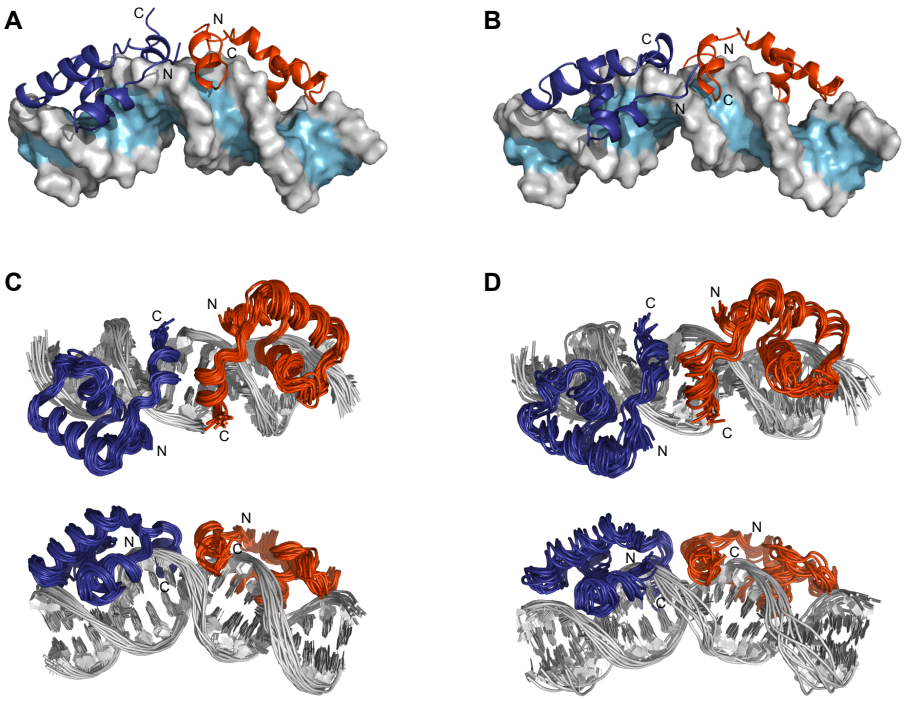


Figure 2.5: *Three-dimensional structures of the HP62V52C-O2 (A) and HP62V52C-O3 (B) complexes. The left and right Lac HP subunits are colored dark blue and dark orange, respectively. The ensemble of the 20 lowest energy structures of the HP62V52C-O2 complex (C) and the ensemble of the 10 lowest energy structures of the HP62V52C-O3 complex shown in two different views rotated by 90°. For clarity, only the backbone atoms of amino acids Val4-Lys59 and DNA heavy atoms of bp 2-20 are shown. Full color figure on page 91.*

Comparison of the complexes

O1 vs O2

The pairwise rms differences between the HP62V52C-*O1* and HP62V52C-*O2* complexes for residues 6 to 58 of the headpiece and bp 3 to 19 of the DNA are ∼2.7 Å for the protein and DNA backbone atoms together, ∼2.2 Å for the protein backbone and ∼2.5 Å for the DNA backbone atoms. A schematic summary of the contacts derived from the calculated structures of the HP62V52C-*O1* and HP62V52C-*O2* complexes is given in Figure 2.6. The structure of the left site in the two complexes is very similar, as a result the network of contacts is largely preserved (Figure 2.7). First, the Lac headpiece is anchored in the correct orientation through hydrogen bonding contacts to the sugar backbone through many residues located in the HTH domain. More specifically, Leu6, Ser16, Thr19, Ser21, Asn25, Ser31 and Thr34 hydrogen-bond to the DNA backbone either with their side chains or with their backbone, or with both. The sequence-specific contacts are present between the side chains of Tyr17, Gln18

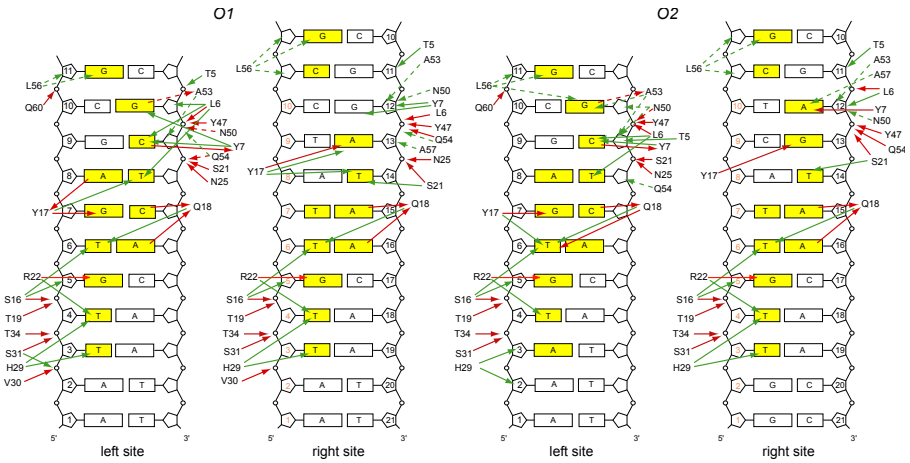


Figure 2.6: *Comparison of the protein-DNA contacts in the HP62V52C-O1 and HP62V52C-O2 complexes. The bases that are recognized specifically are colored in yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively. Full color figure on page 92.*

and Arg22 and their partner base (Figure 2.6).

The left half-sites of the *O1* and *O2* operators only differ at the single position 3 (Figure 2.1). In the HP62V52C-*O1* complex the side chain of His29 makes hydrophobic contacts with the methyl groups of Thy3 and Thy4. However the presence of the N7 atom of the Ade3 disrupts these interactions in the HP62V52C-*O2* complex. Also the pattern of hydrophobic interactions is changed for His29. It appears that in the HP62V52C-*O2* complex His29 makes hydrophobic contacts to the ribose rings of Ade2 and Ade3 while in the HP62V52C-*O1* complex only to the ribose ring of Thy3. Also the hydrogen bond between the backbone of Val30 and the phosphate group of Thy3 has not been observed in the HP62V52C-*O2* complex (Figure 2.6).

In the right half-site of the *O1* and *O2* operators four positions are different: 12, 13, 20 and 21 (Figure 2.1). However there are no specific contacts between the base pairs and the headpiece at either position 20 or 21. Thus these bases do not contribute significantly to the recognition mechanism in agreement with mutational studies both *in vitro* and *in vivo* (Caruthers, 1980). At position 12 the base pair G:C in *O1* is replaced by A:T in *O2* and at position 13 the base pair A:T in *O1* is switched to G:C in *O2* (Figure 2.1). These substitutions of the bases at position 12 and 13 of *O2* result in a slightly different binding of the dimeric Lac headpiece to the right half-site of the operator compared to the HP62V52C-*O1* complex. First, the anchoring contacts formed in the HP62V52C-*O2* complex by Leu6 and Ser21 are shifted one base pair towards the center of the operator, resulting in a positioning change of the recognition helix along the major groove (Figure 2.7). Also the angle between the axis of the DNA double helix and the recognition helix is slightly different: 74° and 96° for the HP62V52C-*O1* and HP62V52C-*O2* complexes, respectively. Substitution of

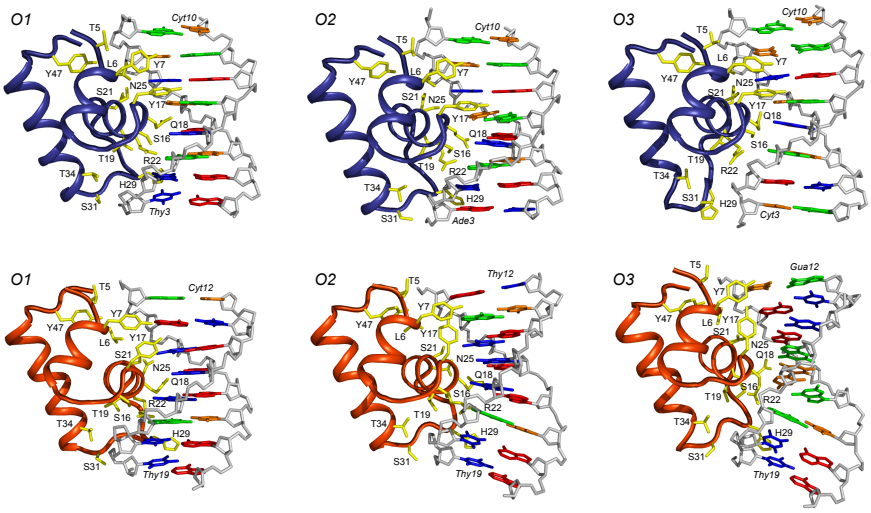


Figure 2.7: *Protein-DNA interactions observed in the HP62V52C-O1, HP62V52C-O2 and HP62V52C-O3 complexes. The left subunit of each complex is shown in dark blue, whereas the right subunit is shown in dark orange. Residues involved in protein-DNA interactions are shown in yellow. Full color figure on page 92.*

the bases at position 12 and 13 leads to a different conformation of the amino acids side chains contacting these base pairs. In the HP62V52C-*O1* complex there are no specific interactions with the G:C12. Although Tyr7 participates in hydrophobic interactions with the sugar group of Gua12. In *O1* A:T13 is recognized by Tyr17, the side chain of which hydrogen-bonds to the N7 atom of Ade13. Furthermore, the side chains of Tyr7 and Tyr17 show aromatic ring-stacking which assists in orienting the Tyr7 side chain towards DNA. In the HP62V52C-*O2* these ring-stacking interactions are preserved. However, the side chain of Tyr7 moves closer to the DNA so that its hydroxyl group can make a direct hydrogen bond to the N7 atom of Ade12. The substitution of Ade13 in the *O1* operator by Gua13 in *O2* still allows formation of the sequence-specific hydrogen bond between Tyr17 and the base N7 atom which is present in both adenine and guanine. However the Tyr17 aromatic ring is oriented such that it does not permit the hydrophobic interaction to Thy14 in the HP62V52C-*O2* complex. Tyr17 together with Gln18 and Arg22 are the most important residues for the specific recognition of the *lac* operator because they participate in the sequence-specific protein-DNA interactions which are conserved in both half-sites of the *O1* and *O2* operator. Due to the intrinsic plasticity of the headpiece, it is able to recognize specifically different base pairs in the left and right half-sites of *O1*. Current analysis of the HP62V52C-*O2* structure demonstrates once again the importance of the interactions between Tyr17 and DNA. This sequence-specific contact is preserved despite the base pair substitution in the HP62V52C-*O2* complex.

The DNA conformation of the *O1* and *O2* operators is not only different at the substitution site (positions 12 and 13) but also at the conserved positions 16 and 17

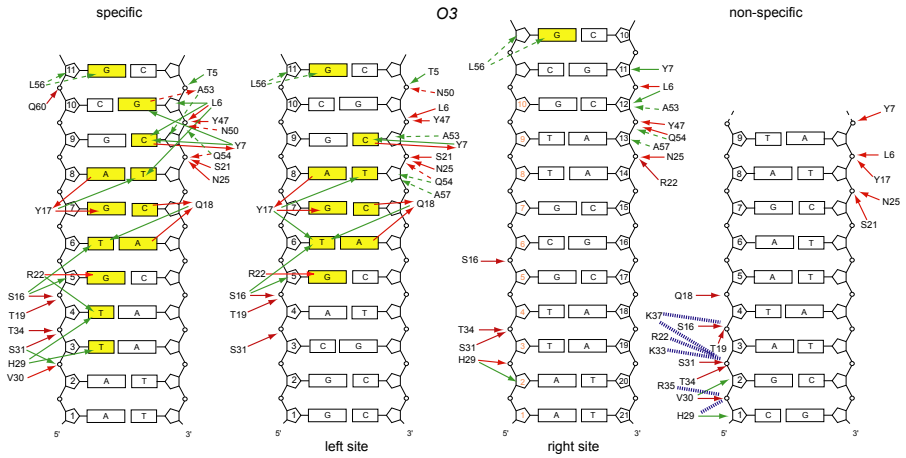


Figure 2.8: *Comparison of the protein-DNA contacts in the HP62V52C-O3 and HP62V52C-NOD complexes. The bases that are recognized specifically are colored in yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively. Dashed blue lines indicate electrostatic contacts. Full color figure on page 93.*

in the right half-site (Figure 2.4). Thus also DNA can adapt to form an optimal interaction interface. This further demonstrates the plasticity of both protein and DNA.

O1 vs O3

The overlay of the Lac headpiece monomer bound to the left half-site of the *O1* and to the *O3* operator results in a relatively large rmsd value (∼1.5 Å) for the protein backbone atoms as compared to the *O1* and *O2* complexes which is ∼0.9 Å. However the interactions between the key residues Tyr17, Gln18 and Arg22 in the major groove are preserved in the left Lac headpiece subunit. Also the majority of the anchoring contacts are present (Figure 2.8). 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*. Interestingly, due to these substitutions in the left half-site of the *O3* operator not only the sequence-specific contacts are absent in the HP62V52C-*O3* complex but also several anchoring contacts (Figure 2.8). The importance of these interactions was shown by mutational studies (Reznikoff, 1984).

The binding of the Lac headpiece to the right half-site of the *O3* operator is distinct from all the other complexes including the Lac-NOD complex (Kalodimos *et al.*, 2004). None of the above described interacting residues make sequence-specific contacts. However one of the key residues in the recognition mechanism of Lac repressor Arg22 now makes a hydrogen bond to the DNA phosphate backbone (Figure 2.7, 2.8). In the HP62V52C-*O3* complex all specific contacts are absent even for the conserved base pairs at positions 17, 18 and 19. The network of contacts is different from the contacts observed in the structure of the dimeric HP62 bound to the non-operator

DNA (Kalodimos *et al.*, 2004). Mutations introduced at position 14 and 15 had only intermediate effects on the repressor affinity (Betz *et al.*, 1986) while substitution of A:T4 by the G:C base pair lead to a significant reduction in the repressor affinity for the mutant operator (Reznikoff, 1984). Therefore the lower affinity of the *O3* operator for the Lac headpiece cannot be attributed to the loss of contacts only in the right half-site but is an effect of combined loss of contacts in both sites. The absence of base specific contacts in the right site of the HP62V52C-*O3* complex combined with partial loss of interactions in the left half-site explains the smaller buried surface area, and provide a good explanation for the lower affinity and dynamic behavior for the dimeric Lac headpiece bound to the *O3* operator.

Discussion

The binding of the hinge helices to the DNA minor groove is very similar in all three complexes regarding both the contacts to the DNA (Figure 2.6, 2.8) and stability (Figure 2.3) despite the differences in the binding affinity. This suggests that the hinge helix formation is not sufficient for an high affinity binding of the Lac repressor. Most likely cooperative folding of the hinge helices contributes to the specificity by binding at the central sequence and assists in positioning the HTH domain with respect to the operator. Formation of the favorable contacts between residues in the HTH domain and DNA will then further increase the affinity of the Lac repressor for its natural operators.

Our data indicate that the dimeric Lac headpiece can undergo conformational changes to optimize its operator binding interface. The Lac headpiece exhibit its intrinsic plasticity at different levels. At the global view, both protein subunits can adopt alternative conformations in order to recognize specifically asymmetrical natural operators and to compensate for the presence of the central G:C base pair. Also locally the side chains of the residues participating in the sequence specific contacts can change their conformation to retain interaction as in the HP62V52C-*O2* complex. Despite the difference of two base pairs in the *O2* operator right half-site compared to the strongest binder *O1* operator Lac headpiece can change its conformation to maintain the contacts. This is in line with the thermodynamics studies of the Lac repressor with variants of the *lac* operator containing single base pair substitutions. Analysis of various thermodynamical parameters suggested that Lac repressor does not rely on any simple read-out of favorable interactions with individual base pairs, but rather result from a complex interplay between complementarity and conformational adaptability (Mossing & Record, 1985; Frank *et al.*, 1997).

Lac headpiece can adapt its conformation to compensate for small changes in its cognate DNA sequence like left half-site versus right half-site or right half-site of the *O1* operator versus the right half-site of the *O2* operator. However if the sequence is significantly different like the overall sequence the *O3* operator, the protein loses the ability to contact DNA. The result is a significant decrease in the affinity for the operator. The *lac* system is remarkable in its evolution. It achieved high efficiency of repression by using multiple repressor binding sites. On the other hand it maintains a balance between the positive and negative control by placing a weak operator next to the catabolite activator protein binding site not to interfere with the positive control

of the *lac* operon (Oehler *et al.*, 1994).

Materials and methods

NMR sample preparation

Cloning, expression and purification of the HP62-V52C mutant was described previously (Kalodimos *et al.*, 2001). Uniformly ^{15}N and ^{15}N, ^{13}C-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, equimolar amounts of the complementary strands were mixed, heated to 95°C for 5 min and slow-cooled to anneal over a period of several hours. The free HP62V52C sample contained 0.4 M KCl, 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 to 0.02 M KCl, 0.01 M potassium phosphate buffer pH 6.0 with the addition of 5 % d8 -glycerol,
0.01 % sodium azide and 5 % D_2O using ultrafiltration device Amicon from Millipore. The complex formation was monitored with NMR spectroscopy. NMR samples contained 0.7 mM HP62-V52C-DNA complex.

DNA binding and DNA bending assay

The gel shift retardation and the DNA-bending assays were performed essentially as described (Spronk *et al.*, 1999b). Binding experiments were performed with 140 bp DNA fragments containing the 22 bp lac-*O1*, lac-*O2*, lac-*O3* or NOD sequences. Comparable results were obtained with the operator sequences used for structural analysis under the above described sample conditions.

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 M or 0.250 M KCl, 0.01 M Tris buffer pH 8.1 with the addition of 1 mM EDTA, 5 % glycerol and 0.1 mg/ml BSA.

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 3D HNCOCA, HNCA and HNCO experiments. Aliphatic sidechain assignments were derived from 3D H(C)(C)(CO)NH-TOCSY, (H)C(C)(CO)NH-TOCSY and HCCH-TOCSY experiments. All above mentioned experiments were recorded on Bruker AVANCE 600 MHz spectrometer equipped with TXI probe with z-gradients. Aromatic ring resonances were assigned using 2D ^{1}H-^{1}H NOESY on a sample in D_2O. Intra-molecular NOEs for the protein were identified from 2D ^{1}H-^{1}H NOESY, 3D NOESY-^{15}N-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 ^{1}H-^{1}H NOESY and

^{15}N,^{13}C double filtered 2D NOE experiments acquired in H_2O and D_2O at 750 MHz and 900 MHz. DNA assignments were obtained using conventional sequential assignment methodology for nucleic acids (Scheek *et al.*, 1983). Residual $^1D_{HN}$ dipolar couplings were determined from the difference in $^1J_{NH}$ at two different fields, 21.14 T (^{1}H frequency 900.21 MHz) and 11.75 T (^{1}H frequency 500.28 MHz) due to the natural alignment of the DNA. The values for the $^1D_{HN}$ coupling were extracted from a series of 30 J-modulated 2D ^{15}N-HSQC spectra (Tjandra *et al.*, 1996). 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.* (1996), giving the $^1D_{HN}$ scalar couplings. The data were fitted using the non-linear least-squares Levenberg-Marquardt algorithm implemented in the program Gnuplot 3.7 (www.gnuplot.info). All spectra were processed using the NMRPipe software package (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson & Blevins, 1994).

Structure calculation

The chemical shifts of the ^{15}N, ^{1}H, $^{13}C_\alpha$, $^{13}C_\beta$ and ^{13}C' were used as input for TALOS to predict dihedral angles which were subsequently used as restraints in the calculations. The NOE peak lists were generated automatically with NMRView (Johnson & Blevins, 1994) and edited manually to remove obvious artifacts. The peak lists together with the chemical shift tables were used for automated NOESY assignment and structure calculation using ARIA1.3 (Brunger *et al.*, 1998; Linge *et al.*, 2003a,b). 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 the dihedral angle restraints, was then used to calculate the final ensemble of structures. 200 structures were generated, and the lowest 50 energy structures were refined in water. The 20 lowest energy structures were then further refined with the $^1D_{HN}$ restraints. Residual dipolar couplings were included as harmonic restraints during the refinement steps of the calculation of the dimer alone and of the protein-DNA complex. The value of the RDC force constant,
100 kcalmol^{-1} Hz^{-2}, was chosen such that the difference between the back calculated and the experimental values fell within the experimental error. The axial and rhombic components of the molecular alignment tensor Da and R, respectively, were initially determined using Pales2.0 (Zweckstetter & Bax, 2000) and further refined calculating few structures each time where Da and R were back calculated from structures in order to minimize the total energy of the protein-DNA structures and Q-factor.

Docking of the HP62V52C dimer ensemble on the natural *lac* operators was performed with HADDOCK2.0 (Dominguez *et al.*, 2003; de Vries *et al.*, 2007) 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, 4.5-5.0 Å corresponding to strong, medium, weak and very weak NOEs, respectively. Observation of the 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 those signals were included during docking as hydrogen bonds restraints. Another set of hydrogen bond restraints was identified from hydrogen/deuterium exchange experiments with accep-

tors for the hydrogen bond on the DNA side identified from preliminary structure analysis. Hydrogen bonds were introduced as ambiguous restraints with upper bound of 2 Å (H-OP) or 3 Å (N-OP or O-OP) at the later docking stages.

The starting B-form DNA structure was built with 3D-DART server http://haddock.chem.uu.nl/dna/dna. 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 no restraint violations larger than 0.5 Å criteria. Structures of HP62-V52C-*O2* and HP62-V52C-*O3* complexes were analyzed for comparison with the previously published HP62-V52C-*O1* structure. To avoid possible errors due to different structure calculation protocols, the structure of HP62-V52C-*O1* complex was recalculated using the same structure calculation protocol and available restraints.

Protein-DNA interactions were analyzed with HBPLUS (McDonald & Thornton, 1994) and HADDOCK2.0 (Dominguez *et al.*, 2003; de Vries *et al.*, 2007) programs. The final set of intermolecular hydrogen bonds is a consensus for outputs from both programs. DNA helical parameters and the bend angle were analyzed with CURVES (Lavery & H., 1988), and the RDC quality factor (Q-factor) (Cornilescu *et al.*, 1998) was evaluated with PALES (Zweckstetter & Bax, 2000). The quality of the protein structures was analyzed using PROCHECK (Nederveen *et al.*, 2005). Molmol (Koradi *et al.*, 1996) and Pymol (www.pymol.sourceforge.net) software was used for structure analysis and figure preparation.

Hydrogen-Deuterium Exchange

Amide hydrogen-deuterium exchange experiments were carried out on samples of HP62-V52C-*O1*, HP62-V52C-*O2* and HP62-V52C-*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 the 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 a standard integration routine in NMRView. The time-dependent decay of peak intensities was fitted to a three-parameter (corresponding to the intensity at time zero, the decay rate 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 site at http://hx2.med.upenn.edu/download.html. Protection factors were determined from the relation $P = k_{ch}/k_{ex}$, where P is the protection factor, and k_{ex} and k_{ch} are as defined above.

Chapter 3

Expression of the isotopically labeled Lac repressor dimer: Tool box for production of proteins for NMR spectroscopy

Abstract

In bacteria gene expression is controlled by repressors, where binding of metabolic regulators to the repressors leads to allosteric changes in these proteins. X-ray crystallography combined with biochemical experiments provided an important insight on allosteric regulation by the Lac repressor. However, several structural aspects of its allosteric mechanism remain elusive that could possibly be resolved by NMR spectroscopy. The size of the Lac repressor makes it a challenging target for NMR studies putting constraints on sample preparation to overcome problems associated with NMR studies of larger proteins. We optimized protocols for production of deuterated functionally active thermostable dimeric Lac repressor and its core domain mutants. The Lac repressor core domain has never been obtained as a recombinant protein possibly due to the observed toxicity to the host cells. We overcame the core domain induced toxicity by co-expression of this domain with the full length Lac repressor combined with stringent control of culture conditions. Sensitivity of NMR measurements is dramatically affected by buffer conditions therefore a thermofluor buffer optimization screen was performed. Optimal buffer conditions have been defined that permit high-resolution NMR studies of the Lac repressor. The sample preparation strategy to overcome toxicity and to optimize protein sample conditions provides a broad range of universally applicable techniques for production of larger proteins for biochemical and biophysical studies.

Introduction

Expression of the genes necessary for lactose metabolism in *Escherichia coli* (*E. coli*) is under the control of Lactose (Lac) repressor. Under normal growth conditions Lac repressor binds to operator DNA and inhibits transcription. When the primary carbon source (glucose) has been depleted and lactose is available allolactose (a derivative of lactose) binds to the repressor. It induces a conformational change that releases the protein from the operator DNA. Lac repressor is a tetrameric protein comprised of monomers of 37.5 kDa, each consisting of (i) the helix-turn-helix DNA binding domain (residues 1-49), (ii) a hinge region (50-62), (iii) an inducer binding domain or core domain, (62-333) and (iv) a C-terminal tetramerization domain (334-360). The inducer binds in a cleft between the two subdomains of the core domain (reviewed in Lewis, 2005; Wilson *et al.*, 2007).

Lac repressor was studied in detail by X-ray crystallography; however the lack of electron density for the DNA binding domain in the absence of DNA, presumably due to its high intrinsic mobility (Friedman *et al.*, 1995; Lewis *et al.*, 1996; Daber *et al.*, 2007), gave an incomplete insight into allosteric regulation by the repressor. To better understand the molecular mechanism underlying the transitions occurring in the presence of the effectors NMR could be useful. While NMR studies are not hindered by the mobility of the DNA binding domain the size prevented detailed structural analysis of a functional Lac repressor. Therefore initial NMR studies were limited to the structure of the isolated DNA binding domain and its complexes with various DNA operators (Chuprina *et al.*, 1993; Spronk *et al.*, 1999a; Kalodimos *et al.*, 2002, 2004). Recent technical improvements increased the size of molecules amenable to NMR studies making the Lac repressor dimer a suitable but still challenging target for NMR spectroscopy. To be able to study this protein by NMR extensive optimization of parameters influencing protein stability and spectral sensitivity is required.

NMR studies of large proteins are difficult because of (i) overcrowding of the NMR spectra and (ii) poor sensitivity. In the case of multimeric proteins such as Lac repressor, the spectral overlap problem can be tackled through a 'divide and conquer strategy' in combination with 3D TROSY NMR methods where assignments of the individual domains are translated to the intact protein (Sprangers *et al.*, 2007). This requires expression of the stable core domain of the Lac repressor. The common though not so straightforward solutions to the problem of sensitivity and resolution include protein deuteration which reduces the NMR relaxation rates of the nuclei of interest and enhances peak intensities (Sattler & Fesik, 1996; Gardner & Kay, 1998). Moreover recording NMR spectra at elevated temperatures also increases sensitivity (Hua *et al.*, 2001; McElroy *et al.*, 2002; Boomershine *et al.*, 2003). We therefore used a previously reported thermostable yet fully functional mutant of the Lac repressor K84M (Gerk *et al.*, 2000).

Finally protein homogeneity, stability and solubility can dramatically influence the quality of the acquired spectra. Optimization can be achieved by selecting a suitable buffer composition based on the quality of the NMR spectra through trial and error. This requires large amounts of isotopically enriched protein and can be expensive and time consuming, especially for large ^{2}H,^{13}C,^{15}N-labeled proteins. Recently a highly efficient thermofluor method was developed for high-throughput buffer condition screening for X-ray crystallography (Ericsson *et al.*, 2006). We used this method

for an effective selection of suitable buffers for NMR sample preparation and applied it to the Lac repressor.

We here describe a detailed protocol for production of a stable, isotopically labeled dimeric Lac repressor and the isolated inducer binding core domain permitting us to study allostery of the Lac repressor using NMR spectroscopy. Our solution to the toxicity problem of the core domain and overall strategy for producing large proteins for NMR spectroscopy including thermostable mutant and efficient buffer condition screen may be of general interest to researchers meeting similar challenges.

Experimental Procedures

Construction of expression plasmids

In the first cloning step the fragment carrying Ligation Independent Cloning (LIC) extensions and a hexahistidine tag were excised from the pLICHIS plasmid using the restriction enzymes EcoRI and XbaI. This fragment was then inserted into a pET-3a vector digested with the same enzymes to produce the expression plasmid pET-LICHIS, that lacks the *lac* operator in the T7 promoter and the *LacI* gene.

The gene encoding for the dimeric Lac repressor (residues 1-333) was amplified from the *LacI* gene present on the pET15 backbone by PCR. The amplified fragment was cloned into the pET-LICHIS vector according to the EFC protocol (de Jong *et al.*, 2006). Using this pET-LICHISΔOΔlacI-lac333 construct as a template, a Lac333K84M mutant was generated using the previously described double mutant PCR method (de Jong *et al.*, 2006).

The plasmid pET-LICHISΔOΔlacI-lac333K84M was used as a template to construct the pET-LICHISΔOΔlacI-lac60-333K84M expression plasmid encoding for the core domain of the Lac repressor (residues 60-333). Because expression of the core domain appeared only possible in presence of functional Lac repressor, we introduced the *LacI* gene into the expression vector. Therefore the BglII-NruI fragment encoding the *LacI* gene was excised from pET15b (Novagen) and cloned into the corresponding sites of the core domain expression vector to produce the vector pET-LICHISΔO-lac60-333K84M. Cloning was confirmed by DNA sequencing.

^{2}H,^{13}C,^{15}N labeling

Lac60-333K84M was expressed in *E.coli* strain Rosetta (DE3) pLysS and Lac333K84M in *E.coli* BL21 (DE3). The transformed cells were grown overnight at 37°C on a LB agar plate supplemented with appropriate antibiotics (Rosetta (DE3) pLysS/pET-LICHISΔO-lac60-333K84M: 50 μg/ml ampicillin and 35 μg/ml chloramphenicol; BL21 (DE3)/pET-LICHISΔOΔlacI-lac333K84M: 50 μg/ml ampicilin) for selection. A 5-ml LB starter culture was inoculated with a freshly transformed colony of cells and grown until the culture became visibly turbid (OD $\sim$ 0.2). The cells were centrifuged for 10 minutes at 1200g at room temperature and resuspended in 20 ml M9/H_2O medium, prepared with unlabeled glucose and $^{14}NH_4Cl$. At OD_{600} of approximately 0.2 the cells were centrifuged and resuspended in 100 ml M9/D_2O medium supplemented with 2 g/L [U-^{2}H, ^{13}C]-glucose and 0.5 g/L $^{15}NH_4Cl$ as sole

carbon and nitrogen sources, respectively. When the OD reached approximately 0.2, the culture was diluted to 200 ml and cultured again to OD_{600} of 0.2 and subsequently transferred to 1 L medium. Overexpression of the Lac60-333K84M was obtained by the addition of 0.5 mM IPTG at OD_{600} of 0.6, growth was continued until the cultures reached the stationary growth phase yielding a final OD_{600} of 1.5. Overexpression of the Lac333K84M was obtained without addition of IPTG; cultures were grown until the stationary phase, yielding an OD_{600} of 1.5. Cells were harvested by centrifugation at 6000g for 20 min at 4°C. The pellets were resuspended in 20 ml lysis buffer (50 mM potassium phosphate buffer pH 8.0, 300 mM NaCl, 20 mM imidazole, freshly added 0.2 % Triton, 1 mM PMSF, 0.1 μl/ml protease inhibitor cocktail, 0.5 mg/ml lysozyme) and frozen at -80°C. Generally, the pellet from 500 ml culture of OD_{600} = 1.5 was resuspended in 10 ml of lysis buffer.

^{2}H,^{15}N labeling

The [^{15}N]- and [^{2}H,^{15}N]-labeled proteins were expressed using an auto-induction medium (Studier, 2005; Tyler *et al.*, 2005). The low amount of lactose present in the auto-induction medium combined with the low affinity of (allo)lactose for the Lac repressor leads to the formation of the unliganded Lac repressor proteins after purification. Transformation and preculturing were carried out using the same procedures as for [^{2}H,^{13}C,^{15}N]-labeled proteins. A 100-ml starter M9/D_2O culture (unlabeled glucose, $^{14}NH_4Cl$) was centrifuged at 1200g for 15 min at room temperature to collect the cells. The pelleted cells were transferred to 1 L auto-induction/D_2O medium supplemented with 1 g/L $^{15}NH_4Cl$ as a sole nitrogen source and 5 g/L [U-^{2}H]-glycerol. The final cell density was 4.5-5 at OD_{600} for both proteins. The harvested cells (500 ml culture) were resuspended in 25 ml lysis buffer and frozen at -80°C.

Purification

The hinge region is highly susceptible to proteolytic cleavage. Therefore special care must be taken during purification to avoid degradation of the protein. Sonication was performed on ice using a short (10 sec) pulse, while the pre-cooled buffers contained whenever possible freshly added PMSF and/or EDTA. The samples were kept on ice during loading and the collected fractions were transferred to ice immediately after collection. All purification steps were performed consecutively, as interruption of the purification procedure at any stage had an adverse effect on long term sample stability.

After sonication, the crude lysate was cleared by centrifugation at 15,000g for 45 min at 4°C. The supernatant was loaded onto a 1.7 ml Nickel POROS metal chelating column (PerSeptive Biosystems), the column was washed with 30 column volumes of binding buffer (50 mM potassium phosphate buffer pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM PMSF, 1 mM β-mercaptoethanol) and protein was eluted using an isocratic elution in binding buffer containing 500 mM imidazole. Protein containing fractions were buffer exchanged using a Sephadex G25 column to 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM PMSF, and 1 mM DTT. Subsequently NaCl was diluted to 35 mM and protein solution was loaded on a Poros HQ anion exchange column (PerSeptive Biosystems). Elution was employed with a NaCl gradient from 15 to 450

mM, both proteins eluted at approximately 60 mM NaCl. A gel filtration step using a Superdex G75 column (GE Healthcare) in 50 mM potassium phosphate buffer pH 7.5, 400 mM KCl, 5 % (v/v) glycerol, 1 mM EDTA, 1 mM PMSF, 1 mM DTT and 0.01 % NaN_3, was required only for the Lac333K84M protein.

Removal of bound IPTG molecules and exchange of labile ^2H atoms for ^1H in the [^2H,^{13}C,^{15}N]-labeled Lac60-333K84M protein was performed by incubation for 3 hours at 60°C after dilution of the sample with 50 mM potassium phosphate buffer pH 9 to a concentration of 44 μg/ml. Subsequent buffer exchange to a suitable buffer and concentration was performed by ultrafiltration using a 5 kDa cutoff Centricon (Amicon).

DNA operator preparation

The HPLC purified palindromic 22bp (5'-GAATTGTGAGCGCTCACAATTC-3') *lac* SymL operator DNA fragment was purchased from Eurogentec. DNA fragment was dissolved in the 50 mM potassium phosphate buffer pH 7.5 containing 250 mM KCl. To form double-stranded DNA, the solution was heated to 95°C for 5 min and slow-cooled to anneal over a period of several hours. Then operator DNA solution was dialyzed against water and lyophilized.

Buffer condition optimization

Optimization of buffer conditions was performed using the MyiQ real-time PCR system (BioRad) essentially as described before (Ericsson *et al.*, 2006). 2.5 μl of protein solution ($\sim$0.5 mM) was diluted with 17.5 μl of buffer, and 5 μl of 60 times water diluted SYPRO Orange (Sigma Aldrich). The signal intensity of the fluorescent dye was monitored from 5°C to 99°C with increments of 0.5°C per step ($\sim$30 seconds per step).

DNA binding assays

The gel retardation assays were performed essentially as described before using the same buffers (Spronk *et al.*, 1999b). The experiments were performed at 4°C, although identical results were obtained at 20°C or 37°C.

NMR spectroscopy

NMR samples of Lac60-333K84M contained $\sim$0.6 mM of [^2H,^{13}C,^{15}N]-labeled protein in 50 mM potassium phosphate buffer pH 6.5, 1 mM DTT, 400 mM [U-^2H]-glycine, 5 % D_2O, 0.01 % NaN_3. TROSY spectrum of [^2H,^{13}C,^{15}N]-labeled Lac60-333K84M was acquired at 318 K on a Bruker Avance 900 MHz NMR spectrometer equipped with a standard TXI probe.

The NMR sample of Lac333K84M bound to a symmetrical DNA operator contained $\sim$0.1 mM of [^2H,^{13}C,^{15}N]-labeled protein in 20 mM KPO_4 pH 6.5, 1 mM DTT, 10 mM KCl, 5 % [U-^2H]-glycerol, 5 % D_2O, 0.01 % NaN_3, trace amount of EDTA-free protein inhibitor cocktail (Roche). A TROSY spectrum of [^2H,^{13}C,^{15}N]-labeled Lac333K84M in complex with the symmetrical DNA operator was acquired

at 315 K on a Bruker Avance 900 MHz NMR spectrometer equipped with a TCI cryoprobe.

Results and discussion

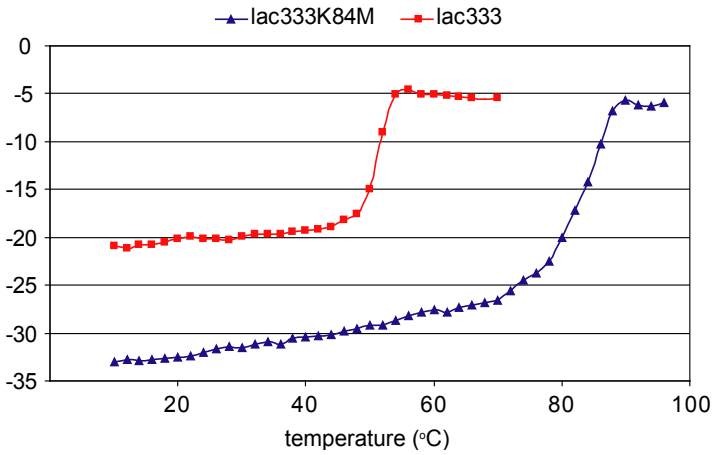


Figure 3.1: *CD spectra of the dimeric Lac repressor wild type (□) and K84M mutant (△). CD spectra were measured at 222 nm on Jasco J-810 instrument with increasing temperature. The buffer used was 80 mM potassium phosphate, pH 7.5, with 0.3 mM DTT.*

When studying large proteins by NMR spectroscopy, it is generally advantageous to measure at higher temperatures to reduce the overall correlation time τ_c, leading to a decrease in the linewidth of individual resonances, giving higher signal intensity and less overlap in NMR spectra. To study allostery of the Lac repressor by NMR we have constructed a thermostable Lac repressor mutant only able to form dimers by deleting the C-terminal residues 334-360 (Chen & Matthews, 1992) and introducing a lysine to methionine mutation at position 84. Gerk *et al.* (2000) reported that this mutation leads to resistance to heat denaturation up to 86°C for the tetrameric construct *in vitro* while maintaining functionality *in vivo*. We confirm the increased thermostability for this Lac repressor mutant using CD spectroscopy (Figure 3.1). The remarkable increase in the thermostability of the protein is thought to be the result of strengthening of the dimerization interface formed by the core domain of the protein (Nichols & Matthews, 1997). As a consequence only the core domain is truly heat resistant while the less stable headpiece is unaffected by this mutation resulting in a two-stage melting transition for the dimer (Figure 3.1). There is a clear transition at 85°C corresponding to the unfolding of the core domain. The unfolding of the headpiece is more gradual that starts at approximately 25°C and continues to 55°C. The increased stability of the core however will permit us to obtain the resonance assignments of the core domain that can subsequently be translated to the complete dimeric Lac repressor.

Expression and purification of the Lac repressor dimer

Previous experience with the Lac headpiece overexpression revealed that the presence of an operator in the T7 promoter, required to prevent protein expression in the absence of inducer, interferes with the Lac repressor overexpression (unpublished results; Slijper, 1996). Similar problems can be anticipated for the Lac repressor dimer (Lac333K84M) overexpression. Furthermore overexpression protocols classically use IPTG for overexpression, this would lead to the formation of a ligand-bound repressor. Finally the presence of the *LacI* gene in the vector backbone of the generally used vectors for recombinant protein production could possibly result in the formation of a heterodimeric Lac repressor sample composed of wildtype LacI and the Lac repressor dimer. To overcome these problems we constructed an N-terminal his-tag containing ligation independent cloning expression vector that lacks the *LacI* gene and the Lac repressor binding site in the T7 promoter sequence. The dimeric Lac333K84M protein could be expressed and purified effectively using this vector, and importantly significant protein expression was also obtained in the absence of inducer, probably as a result of the leaky T7 promoter. This enabled us to study the repressor both in the apo and holo forms with various inducers and anti-inducers.

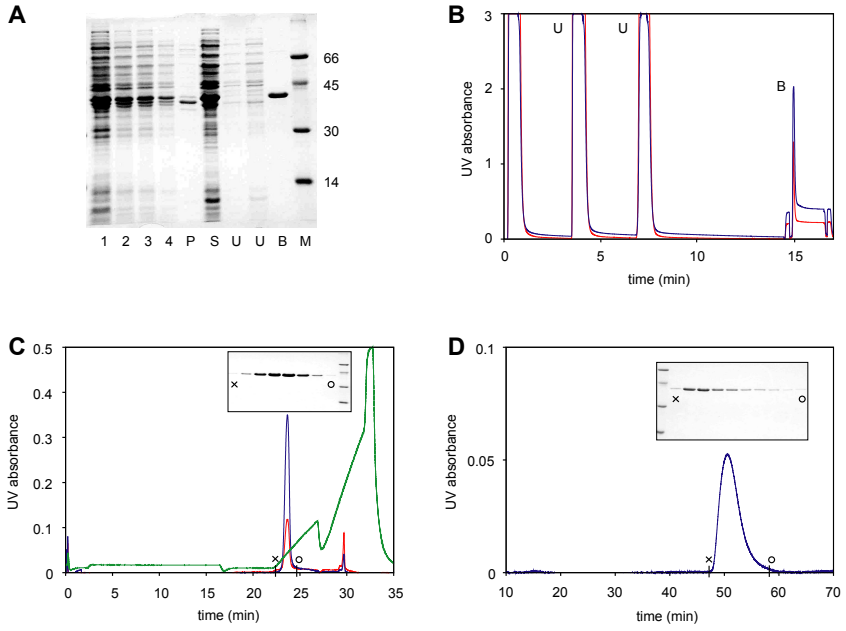


Figure 3.2: *Expression and purification of Lac333K84M expressed in E. coli BL21 (DE3) grown on M9 minimal medium and analyzed by 12.5 % SDS-PAGE. SDS-PAGE results of the purification for the eluted peak fractions are shown above the elution profiles. (A) Lanes are marked as: 1, 2, 3, 4, samples taken during growth; P and S, insoluble and soluble fractions after lysis, respectively; U, unbound to metal chelate column; B bound fraction; M, molecular weight standard. (B) Metal chelate column elution profile. (C) Anion exchange chromatography elution profile. (D) Gel filtration elution profile.*

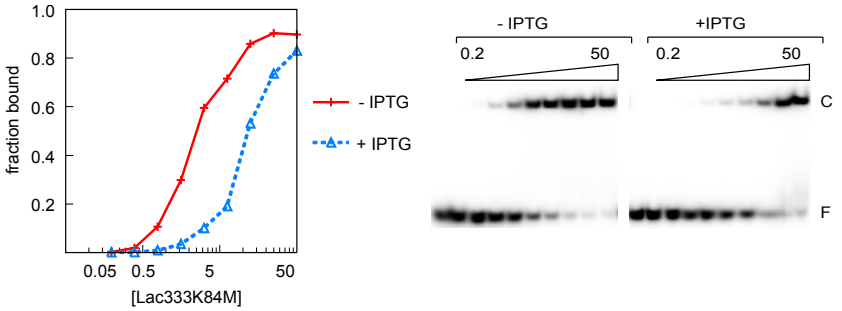


Figure 3.3: *DNA binding experiment using SymL operator in the absence and presence of IPTG for the Lac repressor K84M mutant. The highest and lowest protein concentration (nM) are indicated above each experiment. The free probe and the protein-DNA complex are indicated with F and C, respectively.*

Culturing BL21 (DE3) cells, transformed with the Lac333K84M expression plasmid until the stationary phase, resulted in production of large amounts of the Lac repressor dimer in the absence of inducer (Figure 3.2(A)). The soluble his-tagged Lac repressor dimer, was purified using successive metal chelate affinity (Figure 3.2(B)), anion exchange (Q-Sepharose) (Figure 3.2(C)), and gel filtration (Superdex G75) chromatography (Figure 3.2(D)). Although the protein was essentially pure after the his-tag purification as judged by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 3.2(A)), subsequent purification was needed to prevent cleavage within the hinge region (data not shown), resulting in greater than 99 % pure protein.

To confirm that the protein was functional we tested the *in vitro* affinities of the Lac333K84M for an idealized operator (Sadler *et al.*, 1983) in the absence and presence of inducer (Figure 3.3) using electrophoretic mobility shift assays. In the absence of inducer, the apparent dissociation constant (K_d) for the dimeric Lac333K84M-operator interaction is 2.9$\pm$1.0 nM, while in the presence of saturating concentrations of inducer, the dissociation constant increased to 10.3$\pm$3.1 nM. Similar results were obtained using either the low or high salt buffers used during NMR measurements.

Expression and purification of the core domain

To express the core domain (Lac60-333K84M) we initially cloned the gene encoding this region in the pET3c derived expression construct (pET-LICHISΔOΔlacI-lac60-333K84M). Transformation of this plasmid in various bacterial strains including BL21 (DE3), Rosetta (DE3) and Rosetta (DE3) pLysS failed to produce any transformants. Only in Rosetta Blue (DE3) significantly smaller colonies were obtained than with other plasmids (Figure 4.16(A)). The Rosetta Blue (DE3) strains possess a mutation in the LacI promotor (I^q) leading to a 10-fold increase in the Lac repressor expression (Calos, 1978). To confirm that the absence of a functional Lac repressor and not the genetic difference between these various strains causes the growth inhibition, we performed transformation in T7 express and T7-Express-I^q. Only in the latter strain colonies were obtained, while transformation with the plasmid expressing

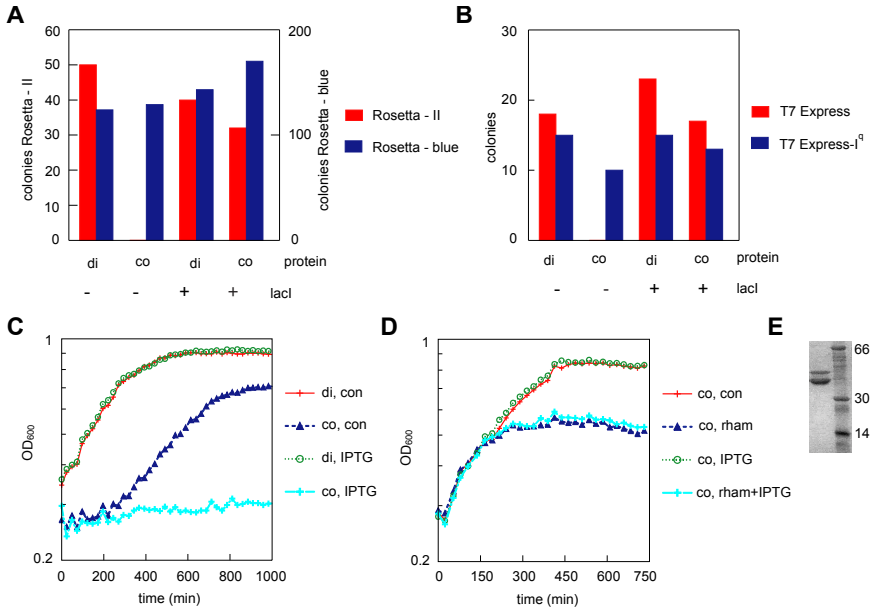


Figure 3.4: *(A) Number of colonies obtained after transformation of pET-LICHIS expression constructs, that lack (+) or contain (-) the full length LacI gene in the vector, for either the dimeric K84M Lac repressor mutant (di) or the core domain (co) in E. coli Rosetta-II (DE3) pLysS or Rosetta-blue. (B) As in A but using the E.coli strains T7-Express or T7-Express-I^q. (C) Growth of Rosetta-II (DE3)-pLysS cultures transformed with the LacI containing pET-LICHIS expression plasmids for the dimeric Lac repressor or the core. OD_{600} is measured at regular intervals, if OD_{600} is above 0.35, IPTG (final concentration 1 mM) or medium (con) is added to the cultures. Note that the OD_{600} is measured in microplates where the path length is not 1 cm. The here obtained plate reader values can be converted by the following empirical relation: OD_{600} (1 cm) = -0.65+4.5×OD_{600}(plate reader). (D) As in C but using E.coli KRX, an E.coli K12 strain containing the T7 RNA polymerase gene under control of the rhamnose repressor. Induction is performed by the addition of 0.1 % rhamnose and or 1 mM IPTG. (E) Co-purification of the LacI gene product with Lac60-333K84M. Full color figure on page 94.*

Lac333K84M was successful in both strains (Figure 4.16(B)). Affinity purification of the his-tagged protein from the cleared lysate of Rosetta Blue (DE3) cells and subsequent SDS-PAGE revealed next to the expected 30 kDa Lac repressor core domain, an additional band of 38-40 kDa, being the predicted molecular weight of the *LacI* gene product (Figure 4.16(E)). Probably the LacI protein was co-purified with the his-tagged core domain protein by forming a stable heterodimer. This further suggests that by a yet unknown mechanism the presence of an overexpression plasmid for the Lac repressor core domain can only occur in strains that have sufficiently high LacI expression. We argue that the core domain homodimer expression during transformation or subsequent growth process (even in the absence of inducer) interferes with normal function of the Lac repressor and thereby inhibits bacterial growth by forming non-functional heterodimers. Rosetta Blue (DE3) produces enough en-

dogenous LacI protein to overcome the inhibitory effect of the overexpression of the core domain. To test this hypothesis we cloned the wildtype *LacI* gene in the pET3c derived Lac repressor core expression plasmid. This plasmid, pET-LICHISΔO-lac60-333K84M could be transformed into all strains that were unable to form colonies in the absence of LacI overexpression (Figure 4.16(A, B)). From these experiments we conclude that LacI overexpression is required to support maintenance of the core domain expression plasmid most probably by preventing premature expression of the toxic core domain. To directly test toxicity of the core domain we next performed a OD_{600} monitored growth experiment in Rosetta-II in the absence or presence of IPTG for cultures containing overexpression plasmids for either the dimeric Lac repressor or the core domain. While growth of the dimeric Lac repressor was not inhibited by the addition of IPTG growth of the core domain protein was blocked almost immediately following IPTG addition in both minimal medium and LB medium. The growth inhibitory effect of the Lac60-333K84M protein can also be observed during uninduced growth of these bacterial cultures as shown by the lower final OD_{600} obtained for this strain compared with the dimeric Lac repressor overexpression strain (Figure 4.16(C)). The core overexpression-dependent inhibitory effect is independent of the genetic background as the *E.coli* K12 derivative KRX shows similar inhibitory response. Experiments in this strain further show that the Lac repressor dependent T7 polymerase expression is not causing growth inhibitory effect as induction of T7 RNA polymerase expression by rhamnose is sufficient to inhibit the growth as in the DE3 containing expression strains. In agreement, addition of both rhamnose and IPTG did not lead to a further increase in toxicity (Figure 4.16(D)).

Importantly if during preculturing the bacteria are allowed to grow until OD_{600} values higher than 0.6 during the subsequent induction phase *no* Lac repressor core domain protein is found. It is widely accepted that when a bacterial culture reaches the stationary phase or faces metabolic stress, an increase in cAMP levels is obtained, suggesting that IPTG independent leaky expression is caused by cAMP accumulation. This argues that when cultures approach the stationary phase expression of the core protein is initiated leading to severe growth inhibition (Figure 4.16(C)). When such a stressed culture is allowed to adapt to these conditions by means of genetic selection a mutant is obtained that supports growth but fails to produce protein. The observed growth inhibitory effect can be avoided by transferring precultures at an OD_{600} of approximately 0.2 with limited dilution (5-fold) keeping the bacterial cultures in the exponential growth phase throughout all stages of culturing. Using these tightly controlled culturing schemes Rosetta (DE3) pLysS strain permitted IPTG induced overexpression of the Lac repressor core domain to appreciable levels.

Purification of the his-tagged protein using metal chelate affinity and anion exchange chromatography yielded a highly pure and stable protein (Figure 3.5). Despite the problems associated with the growth inhibition by overexpression of the Lac repressor core domain we succeeded in the purification of 4.4 mg of essentially pure and highly stable protein per liter minimal media.

Back exchange of deuterated amide resonances and IPTG removal

After purification of the [^{2}H,^{13}C, ^{15}N]-labeled core domain protein overexpressed by IPTG induction we compared its fingerprint ^{15}N-TROSY spectrum with the protein

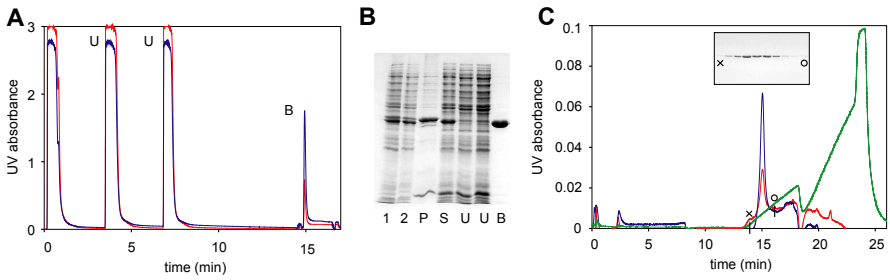


Figure 3.5: *Expression and purification of Lac60-333K84M expressed in E. coli Rosetta-II (DE3) pLysS grown on M9 minimal medium and analyzed by 12.5 % SDS-PAGE. (A) Metal chelate column elution profile. (B) Lanes are marked as: 1 and 2, samples taken during growth; P and S, insoluble and soluble fractions after lysis, respectively; U, unbound to metal chelate column; B bound fraction. (C) Anion exchange chromatography elution profile. SDS-PAGE results of the purification for the eluted peak fractions are shown above the elution profile.*

obtained using auto-induction medium that is devoid of IPTG (see Material and Methods section). As judged by NMR analysis approximately 50 % of the protein was in the IPTG bound conformation (data not shown). To be able to study the repressor under both liganded and unliganded condition the IPTG should be removed from the core protein. Furthermore to allow HN-correlated NMR spectroscopy of deuterated proteins the backbone ^{2}H amide atoms must be exchanged back to ^{1}H. This is usually done by chemical unfolding, followed by refolding the protein in the presence of H_2O. Neither chemical nor thermal unfolding-refolding back exchange procedure could be used for the Lac repressor dimer and its core domain as the unfolding is not fully reversible for both proteins (Schnarr & Maurizot, 1981). We therefore performed the amide exchange reaction by performing only a prolonged incubation in H_2O at elevated temperature (60°C). This procedure yielded the core domain protein free of IPTG with 89 % of the backbone amide atoms exchanged to protons. For the intact Lac repressor 93 % of the backbone amides were exchanged to protons during purification and attempts to further increase this percentage by thermal or chemical unfolding were not successful.

Buffer Optimization

It is well known that pH, ion type and concentration as well as the presence of additives contribute to the stability of proteins (Roberts, 1993). In addition resolution of NMR experiments is influenced by both pH and salt concentration. To optimize measurement conditions and long term protein stability, a buffer screen is performed using the recently developed thermofluor analysis (Ericsson *et al.*, 2006). We first determined the optimal pH by testing 5 different buffers (50 mM), with a pH range from 4.5 to 9.0 at three different NaCl concentrations. Figure 4.17(A) shows dependence of the T_m for Lac60-333K84M on the pH. Significant destabilization is observed for buffers with a pH below 6.0, as the T_m decreases more than 30°C using these buffers. We have chosen to work at pH of 6.5, because the protein is stable at this pH and it

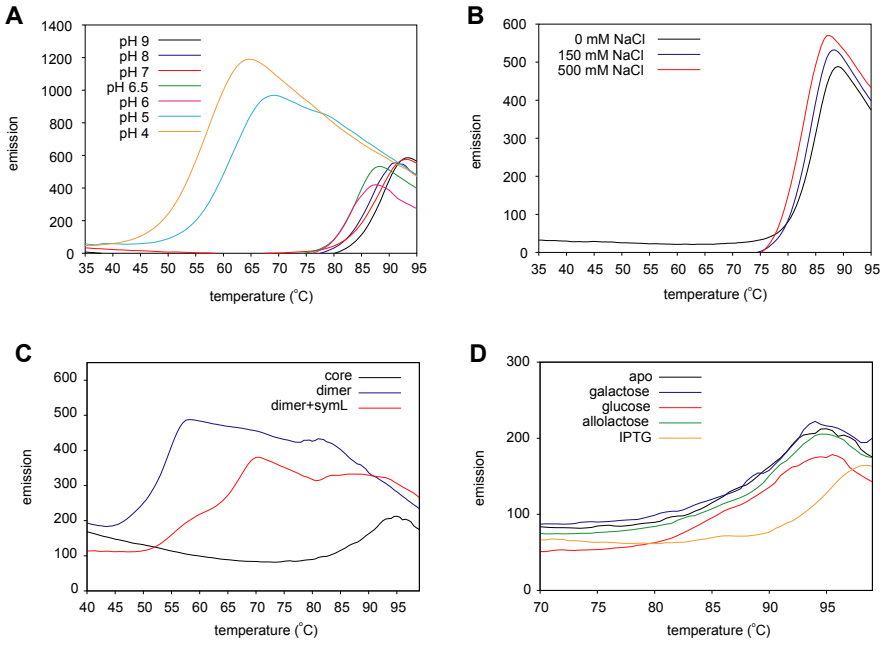


Figure 3.6: *Melting experiments. (A) Lac60-333K84M T_m dependence on the buffer pH. (B) Lac60-333K84M T_m dependence on the salt concentration. (C) Comparison of the melting temperature for the apo Lac333K84M, Lac60-333K84M and Lac333K84M bound to DNA. (D) Influence of different ligands on the stability of Lac60-333K4M. Full color figure on page 95.*

is generally beneficial for NMR spectroscopy to work at low pH due to the slower exchange of the labile protons with water. The thermal stability is only marginally influenced by the ionic strength (Figure 4.17(B)) therefore NaCl was not added to NMR buffer. Addition of glycerol or glycine (data not shown) increases the stability even further. Therefore we added 400 mM of [U-^{2}H]-glycine to the NMR sample of core domain. The T_m of Lac60-333K84M under chosen condition was 90°C and protein can be characterized as very stable. Importantly conditions used in the thermofluor analysis were simultaneously tested in stability tests by incubation for 3 days at 37°C showing a good correlation between the T_m and the long term stability judged by SDS-PAGE analysis of the soluble fraction. NMR measurements on a subset of conditions further support the hypothesis that thermofluor analysis could be used to identify buffers suitable for NMR analysis (data not shown).

Thermal denaturation experiments with the Lac333K84M show three unfolding transitions separated by a plateau (Figure 4.17(C)). We interpret this as initial unfolding of some excess of the free headpiece at a melting temperature of 47°C. At 69°C the headpiece bound to the operator is completely unfolded, the third transition with a T_m of 90°C corresponds to the melting temperature of the isolated core domain. The final buffer for the NMR studies of the Lac repressor dimer was an optimal buffer

for the core domain with addition of 400 mM NaCl because of the salt stabilizing effect of the headpiece. Due to the stabilizing effect of the operator DNA on the Lac333K84M however, the protein-DNA complex can be studied at lower salt concentration (20 mM KCl). The stabilizing effect of the DNA binding is reflected in the shift of the T_m of the protein-DNA complex to 69°C (Figure 4.17(C)).

We have also studied the effect of different inducers on the T_m of the core domain (Figure 4.17(D)). The most potent inducer of the Lac repressor, IPTG, has a stabilizing effect on the Lac repressor as judged from the increase in the melting temperature (from approximately 92°C to 97°C).

NMR spectroscopy

Knowing that signal-to-noise ratio of an NMR spectrum increases with the protein concentration, it is important to maximize the protein concentration. The Lac repressor dimer has the tendency to form soluble aggregates at higher protein concentrations as judged by ^{15}N-TROSY measurements. We routinely obtain 0.1 mM samples for the Lac repressor dimer and 0.6 mM samples for the core domain alone under the optimized conditions. Figure 3.7(A) shows a TROSY spectrum of the [^{2}H,^{13}C,^{15}N]-labeled Lac60-333K84M, the well dispersed peaks are indicative for a properly folded protein.

The TROSY spectrum of Lac333K84M bound to the SymL operator (Figure 3.7(B)) shows that addition of an idealized operator leads to significant spectral changes indicating that the Lac repressor protein is fully functional. Most importantly, after 2 weeks NMR experiments sample quality, as judged by SDS-PAGE (Figure 3.7(C)) and NMR spectroscopy was unchanged, permitting detailed structural analysis.

Conclusions

We have successfully produced deuterated, stable samples of the Lac repressor dimer and its core domain that can be used for extensive NMR studies to answer the question of allostery of the Lac repressor (Chapter 4). Optimization of conditions most suitable for NMR measurements and long term stability as needed for lengthy NMR recordings was effectively performed using a high-throughput thermofluor method. The proteins were fully functional and were able to bind ligands and DNA.

Expression of the core domain of the Lac repressor and of other members of Lac family (Tungtur *et al.*, 2007) appears to be toxic to the host cell. We found a generally applicable procedure to overcome the overexpression induced toxicity. Next to the well established requirement for high level of LacI expression, we found that tight control of the culture conditions can prevent premature protein production by leakiness of the T7 promoter. Keeping the cultures in the exponential growth phase at all stages during preculturing was essential to obtain overexpression of the toxic Lac repressor core domain. This culture regime probably effectively prevents cAMP accumulation, which is known to contribute to Lac repressor controlled gene regulation *in vivo* (Inada *et al.*, 1996) and also to heterologous gene expression. These tightly controlled pre-culture conditions can be useful to boost the overexpression of poorly expressed and possibly toxic proteins in *E.coli*.

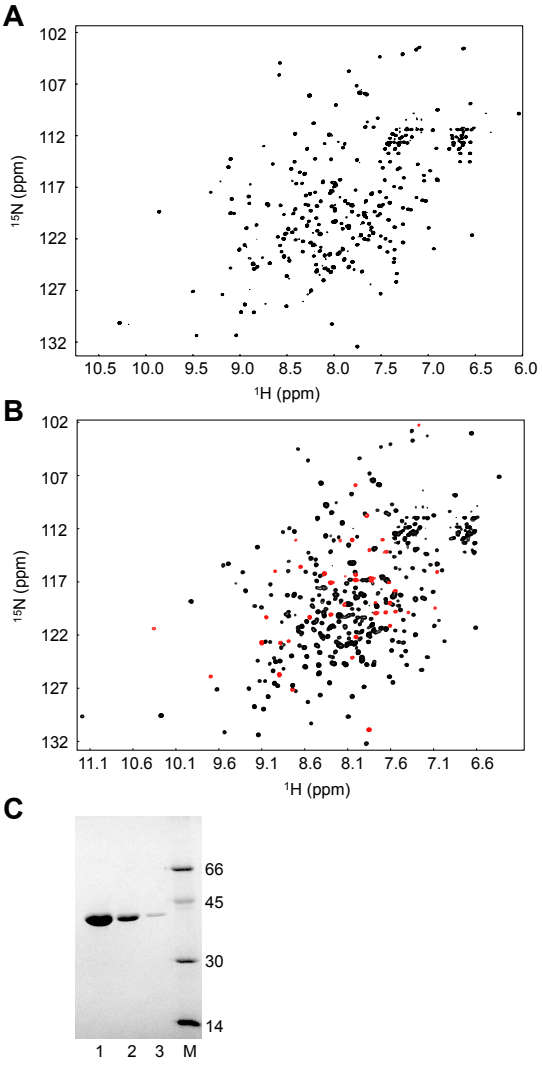


Figure 3.7: *(A) TROSY spectrum of [^{2}H,^{13}C,^{15}N]-Lac60-333K84M recorded at 318K. (B) TROSY spectrum of [^{2}H,^{13}C,^{15}N]-Lac333K84M-DNA complex recorded at 315 K. NMR resonances of the backbone amides of the headpiece are shown in red. (C) Long term stability of Lac333K84M analyzed by 12.5 % SDS-PAGE. Lanes 1, 2 and 3 correspond to different protein concentrations; M, molecular weight standard.*

Chapter 4

On the nature of allosteric transitions of the Lac repressor

Abstract

High-resolution NMR spectroscopy enables us to characterize the allosteric transitions between various functional states of the Lac repressor for which structures are known from X-ray crystallography. In the absence of ligands, Lac repressor exists in dynamic equilibrium between DNA bound and inducer bound conformations. Binding of either effector shifts this equilibrium towards one of the states. Our analysis of the ternary complex between the repressor, operator DNA and inducer shows how inducer binding results in the disruption of the contacts between the inducer binding domain and DNA binding domains, destabilization of the dimerization interface within the DNA binding domain, and how this leads to release of the Lac repressor from the operator DNA.

Introduction

Cellular signaling and metabolic response are mostly under control of allosteric regulation. Originally two models have been proposed to explain the behavior of allosteric proteins: the Monod-Wyman-Changeux (MWC) model (Monod *et al.*, 1965) and the Koshland-Nemethy-Filmer (KNF) model (Koshland *et al.*, 1966). The MWC model or concerted model postulates that in the absence of ligands the protein is in equilibrium between two reversibly accessible states called for historical reasons T (tense) and R (relaxed). Binding of the ligand shifts the equilibrium towards the conformation to which it binds with the higher affinity. The protomers of the oligomeric allosteric molecule undergo this transition simultaneously or in a concerted manner. The KNF or sequential model supposes that the conformational and affinity effects occur sequentially as a consequence of binding each ligand. Another difference between the two classical models is that in the KNF model the R conformation is induced only upon ligand binding via an induced-fit mechanism, which is in contrast

to a pre-existing equilibrium between R and T states imposed by the MWC model. An early example of an allosteric regulatory protein is the Lactose (Lac) repressor, which controls the *lac* operon. Even though Lac repressor is a very well studied system experiments to distinguish between the MWC and the KNF models have remained inconclusive (O'Gorman *et al.*, 1980).

Lac repressor binds to specific operator sequences within *Escherichia coli* (*E. coli*) DNA to repress transcription of genes required for lactose metabolism. When the inducer allolactose (Jobe & Bourgeois, 1972a) or the non-natural inducer isopropyl β-D-thiogalactoside (IPTG) (Jacob & Monod, 1961b) binds to the Lac repressor, the protein changes its conformation to one with a much lower affinity for the *lac* operator. This lower affinity results in the start of transcription of the structural *lac* genes.

Extensive biochemical and structural studies gave considerable insight in the architecture and different conformations of the tetrameric Lac repressor (reviewed in Lewis, 2005; Wilson *et al.*, 2007). Lac repressor is best described as a dimer of dimers with two DNA binding sites and four inducer binding sites. Each monomer consists of the amino-terminal headpiece (HP, residues 1-62) which is responsible for the operator recognition; the core or inducer binding domain (63-333); and the carboxy-terminal tetramerization helix (334-360). The inducer binding site is located in the cleft between N- and C-terminal subdomains of the core. The functional DNA binding unit of the Lac repressor is the dimer which is held together via an extensive dimerization interface on the core domain. The amino-terminal part of the headpiece (residues 1-49) comprises a canonical helix-turn-helix (HTH) DNA binding motif (Kaptein *et al.*, 1985) which is linked to the core by a linker or hinge region (50-62). When the Lac repressor or isolated dimeric HP binds with its HTH motif to the major groove of operator DNA, the hinge region folds into an alpha helix. The hinge helix protrudes into the minor groove and introduces a kink in the operator since Leu56 intercalates between the central G:C base pair step of the operator (Lewis *et al.*, 1996; Spronk *et al.*, 1999a,b; Bell & Lewis, 2000). A large number of the protein-DNA interactions are responsible for the tight complex of the Lac repressor with its operator (Spronk *et al.*, 1999a; Kalodimos *et al.*, 2002). Upon binding of the Lac repressor to the operator, extensive interactions between the HP of one monomer and the core domain of the other monomer are formed that further stabilize the DNA-bound conformation (Bell & Lewis, 2000). IPTG binding to the Lac repressor core causes a reorientation of their N-subdomains which alters both the intramolecular interactions between the N- and C-subdomains of the monomer and the interactions between the two N-subdomains in the dimer (Lewis *et al.*, 1996; Friedman *et al.*, 1995; Daber *et al.*, 2007). However, there is no consensus on how this inducer signal is transferred from the core to the HP to cause a decrease in DNA binding affinity. It has been argued that physical pulling of the hinge helix from the minor groove of the operator as a result of the reorientation of the N-subdomains reduces the affinity of the repressor for the operator (Lewis *et al.*, 1996). However, observation of the interdomain contacts between HP and core in the higher resolution structure (Bell & Lewis, 2000) together with the results of genetic studies (Suckowa *et al.*, 1996) addressed the importance of these interactions in the allosteric mechanism of the Lac repressor. Bell & Lewis (2000) proposed that the release of the hinge helix from the minor groove of the operator may be the result of an alteration of intersubunit interactions rather than

physical displacement. X-ray structures of free and IPTG bound repressor could not resolve this, since in the absence of DNA both systems lack electron density for the headpiece. In principle, structural analysis of the ternary complex between repressor, operator and inducer could establish the contribution of the interdomain contacts and unfolding of the hinge helix to the overall allosteric mechanism of the Lac repressor. The existence of the ternary complex was established by kinetic experiments (Riggs *et al.*, 1970b), but so far no structural information has been available for it.

There also appears to be a discrepancy between results from the various studies on the free form of the Lac repressor. Optical and chemical modification studies have demonstrated substantial conformational changes that occur in the core region upon IPTG binding (Laiken *et al.*, 1972; O'Gorman & Matthews, 1977; Whitson *et al.*, 1984; Daly & Matthews, 1986). In contrast, the crystal structure of the apo Lac repressor tetramer is very similar to the structure of the protein bound to IPTG (with a rmsd of only 0.4 Å), and is different from the DNA-bound state (Lewis *et al.*, 1996). Also a more recent crystal structure of the dimeric IPTG bound repressor shows no significant conformational differences with the free state (Daber *et al.*, 2007).

For the allosteric model of the Lac repressor there are at least four relevant states: free, inducer- and operator-bound and ternary complex. For two of them (inducer- and operator-bound) detailed structural information is available via crystallography. The other two however have resisted detailed structural analysis. Structural data is available for the free core of Lac repressor where the DNA binding domain was removed, and for the key-regulatory ternary complex detailed structural information is fully lacking. In the present study we found that the two lacking allosteric states are intrinsically dynamic and possibly for that reason have resisted prior analysis.

Recent advances in methodology and new labeling techniques have considerably extended the size limit of Nuclear Magnetic Resonance (NMR) spectroscopy (Pervushin *et al.*, 1997; Fiaux *et al.*, 2002; Tugarinov *et al.*, 2005), which now allows a complete analysis of the dimeric Lac repressor (76 kDa) in various allosteric states under identical conditions. For this study a fully functional dimeric Lac repressor mutant containing a stabilizing K84M mutation (Gerk *et al.*, 2000) and lacking the tetramerization domain was produced. NMR chemical shift changes between various functional states of the intact Lac repressor dimer (apo, protein-inducer complex, protein-DNA complex, and protein-DNA-inducer complex) could be monitored and could be mapped on available X-ray structures. The NMR results showed that spectral properties of the core of the Lac repressor can be described by only two structural states. Whereas in the free repressor these two states are in equilibrium, binding to the operator or inducer shifts the equilibrium to one or other of these two states, maintaining the symmetry of the dimer. Therefore, our NMR results give strong support for the MWC model for the allosteric function of Lac repressor.

Results

Apo Lac repressor

^{15}N-TROSY NMR spectra of the isolated core and HP domains were compared with that of the intact protein (Figure 4.1(A, B)). The NMR spectra of the core domain

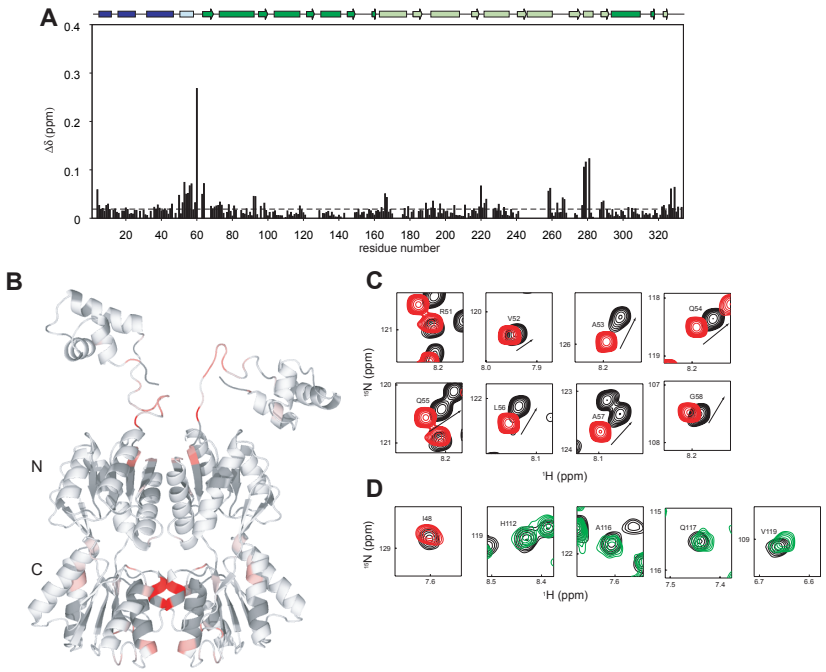


Figure 4.1: *NMR analysis of the apo form of Lac repressor dimer. (A) Histogram of combined chemical shift changes ($\Delta\delta$) between the isolated domains (HP (residues 1-62) and core domain (residues 60-333)) and the intact Lac repressor dimer. Secondary structure elements of the different modules of the Lac repressor are shown at the top of the histogram: headpiece is shown in dark blue, hinge - light blue, the N-subdomain (residues 62-161 and 293-320) is shown in dark green, C-subdomain (residues 162-289 and 321-329) in light green. (B) $\Delta\delta$ values that are larger than average are mapped onto the model of the free Lac repressor dimer as a gradient of white-red. Residues that were not assigned either due to overlap or incomplete backbone exchange are shown in darker shade of gray. (C, D) Comparison of ^{15}N-TROSY spectra of isolated headpiece (red), isolated core (green) and intact Lac repressor dimer (black). Arrows point in the direction of alpha helix formation. Full color figure on page 96.*

without DNA binding domain and that of the intact dimeric Lac repressor show that the proteins are well structured. Absence of significant chemical shift differences of backbone amides for residues Thr5-Ile48 in the two spectra reveals that the structure of the HTH domain is unaffected by the presence of the core. In addition, core residues His112-Ser120 (Figure 4.1(D)) that are known to contact HP in the presence of DNA (Bell & Lewis, 2000), are also unaffected. The lack of significant NMR chemical shift changes for these residues implies that there are no substantial direct interactions between the N-subdomain of the protein core and the headpiece in the apo form of the Lac repressor dimer.

In contrast, in the hinge region the amides of residues Asn50, Val52, Ala53, Gln54, Gln55, Leu56 and Ala57 have distinct chemical shift changes in the intact

Lac repressor with respect to the isolated Lac headpiece. Analysis of the direction of the shifts indicates that possibly a small fraction ($\sim 5\%$) of the unstructured ensemble of the hinge region of the HP is already in helical conformation even without being bound to DNA.

Analysis of the ^{15}N NMR relaxation data of the isolated monomeric headpiece (residues 1-56) (Slijper *et al.*, 1997) and the headpiece in the intact Lac repressor dimer support this picture (Supplementary text S1, Figure 4.8). The overall rotational tumbling for the HTH domain connected to the core is much faster than expected for a system in which the complete repressor forms a single rigid dimer with the core (τ_c = 10.9 ns *vs.* about 79 ns). However the tumbling time is slower than of a free monomeric headpiece (10.9 ns *vs.* 5.65 ns). This also indicates that the hinge region becomes partially ordered when the core is attached to the headpiece.

Remarkably, also chemical shift differences between the isolated core domain and the intact dimer are seen in the dimerization interface within the C-subdomain. These differences indicate that long-range (over 40 Å) interactions exist between the C-subdomain and the N-terminal end of the N-subdomain where the headpiece is attached. This remote interaction is possibly linked to subtle conformational differences and reflects the internal link between remote parts of the Lac repressor.

Repressor-operator complex

To monitor structural changes associated with specific DNA binding, we compared the NMR chemical shifts of the Lac repressor dimer free and bound to a 22 base-pair symmetrical operator (SymL). As expected, the largest shifts are observed in the HP region (Figure 4.2(A, B)). The ^{15}N and ^{1}H chemical shifts of the hinge region residues indicate that a hinge helix is formed similarly as has been shown for the isolated HP in complex with SymL (Spronk *et al.*, 1999a). Aside from these direct DNA-protein contacts, substantial chemical shift changes were observed for residues Val111-Gly121 in the core domain that correlate well with the interdomain contacts between the Lac dimer core and the headpiece seen in the crystal structure of the repressor-operator complex (Bell & Lewis, 2000).

A closer look shows that the chemical shift changes induced by operator binding to the headpiece propagate all the way to the inducer binding site (residues Ala75, Asp274) through several residues located at the dimerization interface that were implicated in allosteric signaling before, i.e. Asp88, Ser93, Val96 and Met98 (Bell & Lewis, 2000). Even residues at the junction of the N- and C-subdomains (residues His74, Glu277, Gln291 and Asp292) near the inducer binding site are affected by operator binding in agreement with a distinct conformation at the inducer binding site with a possibly altered affinity. Indeed we noted substantial differences in affinity in a gel shift assay that tested IPTG binding in absence or presence of operator (Figure 4.2(C)). A 25-fold higher concentration of IPTG is required to obtain half maximal inhibition of DNA binding when IPTG is added to the pre-formed repressor-operator complex compared to the concentration required to obtain the same inhibition when added to the apo protein. This indicates that DNA binding to HP lowers the affinity of the core for IPTG. These results are fully in line with previous fluorescence spectroscopy observations for the full length Lac repressor (Dunaway *et al.*, 1980).

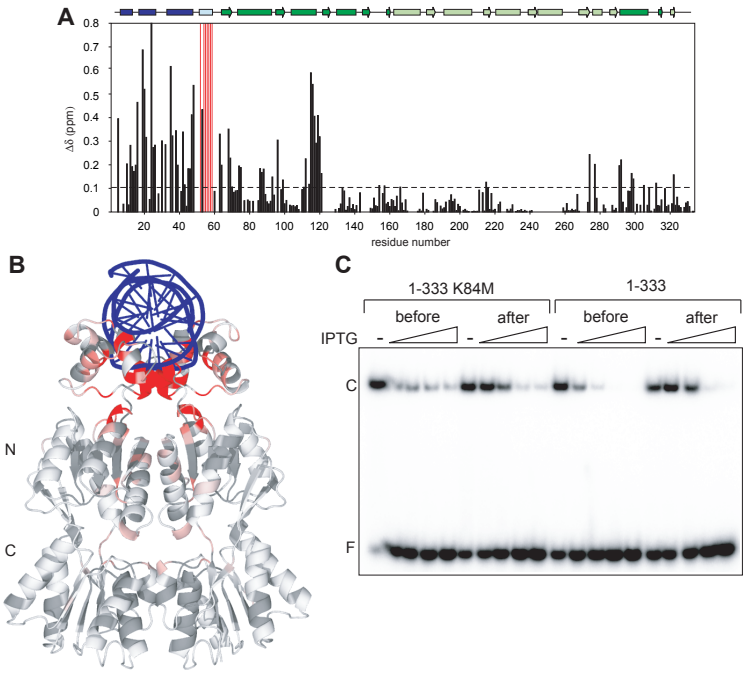


Figure 4.2: *DNA binding to the Lac repressor dimer. (A) The bar diagram shows the differences in chemical shift for backbone resonances of the free and the protein bound to the idealized symmetric operator (SymL). The hinge region is show in red to emphasize formation of the hinge helix. (B) Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to symmetrical operator and ONPF (1EFA) as a gradient of white-red. (C) Electrophoretic mobility shift assay of Lac repressor binding to the SymL operator, performed in the presence of 0.2 nM wild type Lac dimer or 3.1 nM K84M in the presence or absence of 1, 10, 100, 1000 µM IPTG, added either prior to or after protein-DNA complex formation. The free probe and the protein-DNA complex are indicated with F and C respectively. Full color figure on page 97.*

Inducer and anti-inducer complexes

Addition of IPTG to the apo form of the Lac repressor core domain results in major spectral changes that propagate from the inducer binding site to the outer surface of the N-subdomain (Figure 4.3(A, C)) that involves also changes for residues Val111-Gly121 that are the anchor points for HP binding in the protein-DNA complex. Many of the observed chemical shift differences, in particular those at the interface between the N-subdomains, are opposite to those that occur upon operator binding (Figure 4.4) and as we explain below are due to a shift in conformational equilibrium. Another issue is whatever upon IPTG binding the symmetry of the dimeric core domain would be maintained. The addition of excess IPTG results in a tight complex with a separate set of resonances for the bound conformation, and which converts only slowly

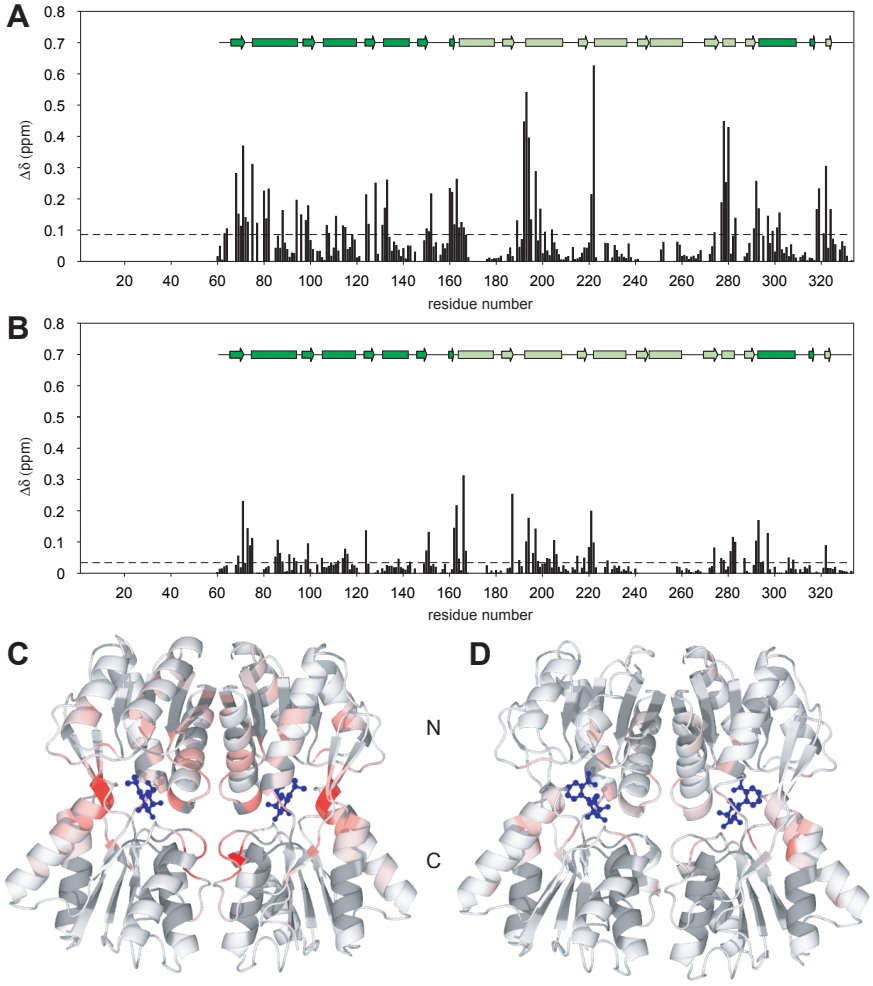


Figure 4.3: *Lac repressor core dimer binding to inducer IPTG or anti-inducer ONPF. (A, C) The bar diagram shows differences in chemical shift between backbone resonances of the free and IPTG bound Lac repressor core domain. Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to IPTG (2P9H) as a gradient of white-red. The ligands are shown in blue ball-stick representation. (B, D) The bar diagram shows the differences in chemical shift for backbone resonances of the free and ONPF bound Lac repressor core domain. Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to ONPF (2PAF) as a gradient of white-red. (E) Comparison of* ^{15}*N-TROSY spectra of the free core domain (black), the Lac repressor dimer bound to operator DNA (red), the core domain bound to IPTG (blue) and the core domain bound to ONPF (green). Full color figure on page 98.*

on the NMR chemical shift time scale with the free protein. When the Lac repressor

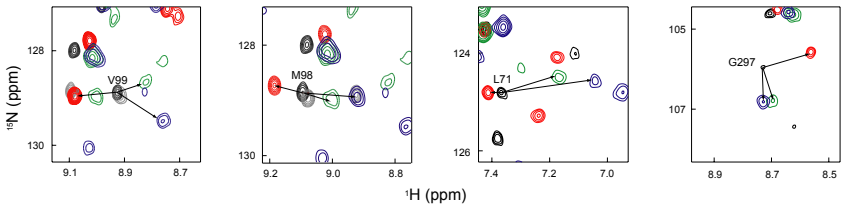


Figure 4.4: *Comparison of ^{15}N-TROSY spectra of the free core domain (black), the Lac repressor dimer bound to operator DNA (red), the core domain bound to IPTG (blue) and the core domain bound to ONPF (green). Full color figure on page 100.*

core is only partially saturated with IPTG, affected residues show only two sets of resonances corresponding to the free and bound conformations of the protein. The absence of any additional states even at the dimer interface indicates that IPTG binds to the Lac repressor monomer in a highly cooperative fashion, preserving symmetry of the dimer.

Anti-inducers, such as orthonitrophenyl-β-D-fucoside (ONPF) that inhibit induction, compete with IPTG in binding to the inducer binding site (Riggs *et al.*, 1970b). NMR titrations of the free Lac repressor core domain with ONPF show that also ONPF binding is in the slow exchange regime as binding of IPTG. In contrast to IPTG, addition of the anti-inducer ONPF only shows chemical shift changes in the inducer binding pocket, but fails to propagate the signal towards the core-HP interface (Figure 4.3(B, D)). The Lac repressor-ONPF complex has increased affinity for DNA relative to the Lac repressor alone (Riggs *et al.*, 1970b), which may suggest that Lac repressor-ONPF adopts an DNA bound structure. The direction of the NMR shifts does not indicate that ONPF stabilize the repressor-operator complex by inducing the DNA-bound conformation in the Lac repressor (see below) (Figure 4.4). Thus ONPF appears only to compete with inducer binding but is not itself responsible for stabilizing the repressed state.

Ternary complex of repressor, operator and inducer

The formation of a ternary complex by addition of IPTG to the dimeric Lac repressor-SymL complex is associated with large chemical shift changes indicating a significant structural perturbation (Figure 4.5(A, B)). Not surprisingly, the largest changes occur around the inducer binding site. The NMR shifts at the inducer binding site in the ternary complex are similar to those of the Lac core-IPTG complex. However, more importantly, many large spectral changes occur at sites distant from the inducer binding pocket. IPTG binding causes widespread changes involving residues at the dimerization interface within the N-subdomain (residues Asp88, Val94, Val96). Another strongly affected region is the dimerization interface within the C-subdomain (residues Ser221, Ala222, Asp278, and Ser280). These changes are all slow exchange processes, which is in line with high affinity and slow off-rates that have been reported for the inducer binding to the Lac repressor (Barkley *et al.*, 1975).

Surprisingly there are no substantial changes in chemical shift in the HTH domain

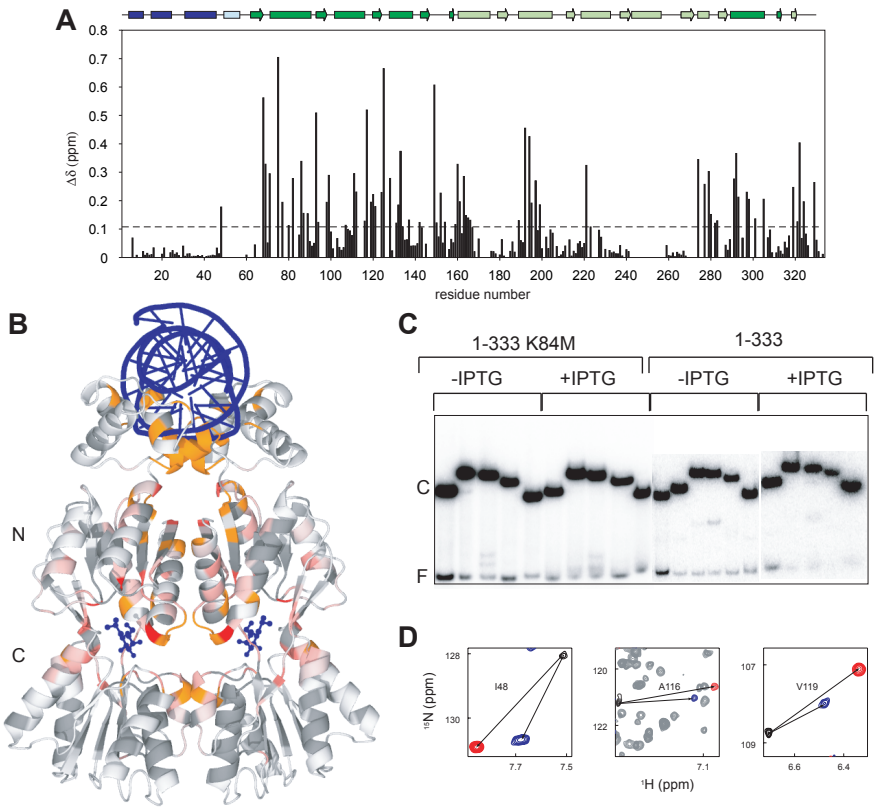


Figure 4.5: *Ternary complex of the Lac repressor dimer. (A) Chemical shift changes between repressor-SymL and repressor-SymL-IPTG complex. (B) Chemical shift changes plotted on a model for the ternary complex as a gradient of white-red. Residues that were affected but could not be assigned either due to the line broadening or large signal shift are shown in orange. (C) Electrophoretic mobility shift assay using different circularly permuted SymL lac operator fragments, showing bending of the operator DNA in the ternary complex. The free probe and the protein-DNA complex are indicated with F and C respectively. (D) Comparison of ^{15}N-TROSY spectra of the free Lac repressor dimer (black), Lac repressor dimer bound to DNA (red) and the ternary complex (blue). Full color figure on page 99.*

upon IPTG addition to the protein-DNA complex. The absence of IPTG-induced chemical shift changes for all residues between Thr5 and Tyr47 indicates that the HTH domain remains bound to operator DNA similar to the Lac repressor-DNA complex. This is consistent with the high affinity of the Lac repressor to the operator in the presence of IPTG ($K_d \sim 1\ \mu M$) (Barkley *et al.*, 1975). DNA bending studies revealed that the IPTG bound repressor bends the operator similarly as in the protein-DNA complex (Figure 4.5(C)), arguing that the hinge helix is (to a large extend) intact. However NMR signals of backbone amides of residues in the hinge region disappeared possibly due to line broadening. Since there were no signals found

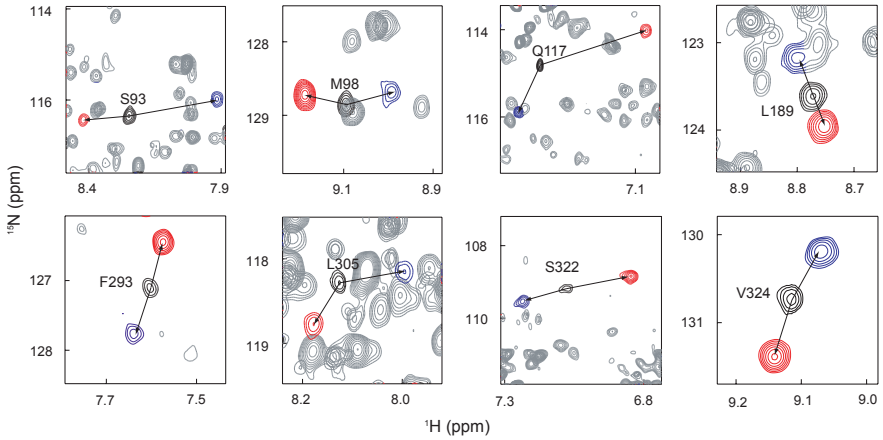


Figure 4.6: *Comparison of three conformations of the Lac repressor dimer. Overlay of ^{15}N-TROSY spectra of the free Lac repressor dimer (black), Lac repressor dimer bound to DNA (red) and the ternary complex (blue). Full color figure on page 100.*

at frequencies corresponding to an unstructured hinge region or at other frequencies. Therefore this behavior likely reflects intrinsic dynamics on the millisecond time scale rather than full helix unfolding.

Residues that are known to be involved in the core-HP interactions, i.e. Ile48, Val111-Gly121 undergo significant chemical shift changes upon IPTG addition, showing distinct chemical shift values for all forms of the Lac repressor (Figure 4.5(D)). This clearly indicates that the interface between the headpiece and core domain is changed in response to IPTG binding.

Finally, we compared all chemical shifts of the free Lac repressor, Lac repressor bound to the operator and the Lac repressor in the ternary complex for all residues. A striking pattern of change was observed in the amide ^{1}H and ^{15}N chemical shifts of many residues in the N-subdomain that were previously implicated in allostery: it appears in many cases that the resonance for the free protein occurs on a straight line connecting those of the repressor-operator complex and the ternary complex (Figure 4.6). The only exceptions are residues at the core-HP interface (Figure 4.5(D)) and residues directly contacting either operator or IPTG. This linear pattern is a clear signature of an allosteric protein that is in fast conformational equilibrium between two states, with the chemical shifts of the apo form representing a population-weighted average of the two bound states.

Discussion

In this study all functional states of the Lac repressor were characterized using NMR chemical shift mapping. By monitoring chemical shift changes upon inducer binding to either core domain or to the repressor-operator complex, residues that are criti-

cal for inducer signal transfer could be identified (Supplementary material text S2, Figure 4.9). The same set of residues was implicated before from genetic studies (Suckowa *et al.*, 1996; Pace *et al.*, 1997) and analysis of the crystal structures (Lewis *et al.*, 1996; Bell & Lewis, 2000). Therefore, the NMR chemical shift method is a sensitive, independent and reliable tool to identify conformational changes that occur during effector binding. Current NMR data resolves the existing discrepancy about

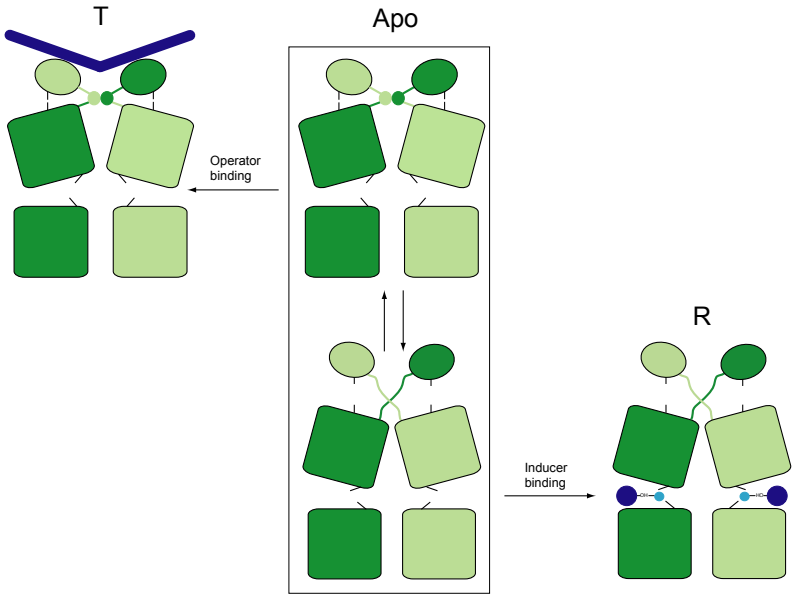


Figure 4.7: *The proposed model for allosteric regulation of the Lac repressor.*

the nature of the ligand-free form of the Lac repressor. The free Lac repressor is a highly dynamic molecule, that probably exists in a conformational equilibrium. This is not only the case for the headpiece where a flexible hinge has some alpha-helical propensity but in particular for the N-subdomain of the inducer binding domain. T-jump experiments showed that a rapid conformational transition between two states takes place in the Lac repressor in the absence of any inducers (Wu *et al.*, 1976). The amplitude of the fast relaxation effect decreased with increasing IPTG concentration, showing the presence of a dynamic equilibrium that can be shifted by IPTG. Current NMR data indicate that this equilibrium in the free Lac repressor dimer is between DNA- and inducer-bound conformations. In line with the classical MWC model, the linear chemical shift pattern in the N-subdomain of the dimeric Lac repressor demonstrates that binding of either allosteric effector shifts this pre-existing dynamic equilibrium towards one of the extreme states (Figure 4.7). This shows that the N-subdomain functions as a regulatory switch, very much according to the MWC model. Both upon binding of DNA and IPTG the transition is concerted for two protomers of the dimer without significantly populated intermediate states as judged from absence of extra resonances for intermediate conformations. This conservation

of symmetry in the allosteric transitions agrees well with the initial postulates of the MWC model.

Analysis of the ternary complex permits us to suggest a model for the transition from an operator-bound state to a state of non-specific DNA binding triggered by inducer binding. It is shown that the observed allosteric changes caused by the inducer involve the loss of contacts between the core domain and the headpiece. At this point the two headpieces remain bound to the operator with partially structured hinge helices. Taraban *et al.* (2008) recently characterized a ternary complex of Lac repressor-operator-inducer using a different protein construct by small-angle x-ray scattering (SAXS). Their measurements also indicate that under these conditions there is no complete unfolding of the hinge helix. However, the NMR data additionally show that the stability of the hinge helices is compromised due to the internal dynamics. The result is a reduced affinity of Lac repressor for the operator. *In vivo* the large excess of the non-specific low-affinity sites will effectively compete with the *lac* operator for repressor causing the Lac repressor to release from the operator and to bind non-specifically to DNA. The result is then further unfolding of the hinge helices (Kalodimos *et al.*, 2004).

In summary, at least three processes in the allosteric mechanism can be distinguished in the Lac repressor: the homotropic cooperative binding of IPTG, the heterotropic coupling of the *lac* operator and IPTG binding and the cooperative folding of the hinge helices. To conclude, the general allosteric mechanism of the Lac repressor can be described in the framework of the original MWC model.

Materials and Methods

Sample preparation

Plasmids expressing the Lac repressor dimer (residues 1-333, Lac333K84M) or core domain (residues 60-333, Lac60-333K84M), both including a N-terminal hexahistidine tag, were created using the LIC protocol (de Jong *et al.*, 2006). The core domain was expressed in *E.coli* strain Rosetta (DE3) pLysS, the dimer in *E.coli* BL21(DE3). For expression of [^{2}H,^{13}C,^{15}N]-labeled proteins, cells were grown in M9 minimal medium containing 2 g/l [^{2}H,^{13}C]-glucose and 0.5 g/l $^{15}NH_4Cl$ in 99.9 % D_2O/0.1 % H_2O. Proteins were purified using nickel affinity and anion exchange HQ chromatography. The Lac333K84M dimer was further purified by Superdex G75 gel filtration column.

IPTG-free Lac333K84M was obtained by growing host cells without induction, whereas expression of [^{2}H, ^{13}C,^{15}N]-labeled Lac60-333K84M was induced by 1 mM IPTG. In order to remove any bound IPTG molecules and back-exchange labile ^{2}H atoms for ^{1}H the protein was diluted to 44 μg/ml with 50 mM potassium phosphate buffer pH 9 and incubated for 3 hours at 60°C. Protein solutions were finally concentrated by ultrafiltration using a 5 kDa cutoff Centricon (Amicon) and simultaneously buffer was exchanged to an NMR suitable buffer. TROSY spectra of [^{2}H,^{13}C,^{15}N]-labeled Lac60-333K84M and [^{2}H,^{15}N]-labeled Lac60-333K84M grown in auto-induction media were compared to ensure that both were free of IPTG.

HPLC purified palindromic 22bp (5'-GAATTGTGAGCGCTCACAATTC-3') *lac*

operator (SymL) (Eurogentec) was dissolved in buffer containing 50 mM potassium phosphate buffer pH 7.5 and 250 mM KCl and annealed by heating the solution to 95°C for 5 min and slow-cooling over a period of several hours. Then operator DNA solution was dialyzed against water and lyophilized.

DNA binding and DNA bending assays

Gel retardation and the DNA bending assays were performed essentially as described in Spronk *et al.* (1999b).

NMR spectroscopy

The NMR sample used for resonance assignments contained ∼0.6 mM of the [^{2}H,^{13}C,^{15}N]-labeled Lac60-333K84M in 50 mM potassium phosphate buffer pH 6.5, 1 mM DTT, 400 mM [U-^{2}H]-glycine, 5 % D_2O and 0.01 % NaN_3. The NMR sample to study the protein-DNA complex contained 0.1 mM of [^{2}H,^{13}C,^{15}N]-labeled Lac333K84M mixed with an equimolar amounts of SymL in 20 mM potassium phosphate buffer pH 6.5, 1 mM DTT, 10 mM KCl, 5 % [U-^{2}H]-glycerol, 5 % D_2O, 0.01 % NaN_3 and trace amount of EDTA-free protein inhibitor cocktail (Roche). To form the ternary complex, 50 times excess of IPTG was added to the protein-DNA complex. The limited stability of Lac333K84M in the apo form precluded recording spectra necessary for backbone assignment. Thus first the backbone of the HP-less Lac60-333K84M was assigned and these assignments were subsequently transferred to Lac333K84M. Assignments of the backbone amide protons of the headpiece were available from previous studies (Spronk *et al.*, 1999a).

All NMR spectra were acquired on a Bruker AVANCE 900 MHz spectrometer equipped with a TCI HCN cryo probe. For the backbone assignment of the [^{2}H,^{13}C,^{15}N]-labeled Lac60-333K84M 3D TROSY-HNCA, TROSY-HNCOCA, TROSY-HNCO, TROSY-HNCACO, TROSY-HNCACB and a 3D NOESY-^{15}N-HSQC were recorded at 318 K. Substantial chemical shift differences between apo and IPTG-bound Lac60-333K84M required *de novo* resonance assignment of the IPTG-bound form of the Lac60-333K84M. Assignments of the Lac repressor dimer in complex with operator DNA were confirmed by 3D TROSY-HNCA. NMR data were processed using NMRPipe (Delaglio *et al.*, 1995) and analyzed with NMRView (Johnson & Blevins, 1994). The combined chemical shift change (in ppm) of a particular residue upon ligand binding was calculated as $\Delta\delta = [\Delta\delta_{HN}^2 + (\Delta\delta_N/R_{scale})^2]^{1/2}$, where the chemical shift scaling factor, R_{scale} = 6.5 (Mulder *et al.*, 1999). Data were mapped to existing crystal structures of the Lac repressor using Pymol (www.pymol.sourceforge.net).

Relaxation analysis

The relaxation parameters T_1, T_2, and ^{15}N heteronuclear Overhauser enhancements (hetNOE) were determined for the backbone amides of the DNA-binding domain in Lac328 construct on a Bruker 600 MHz spectrometer. Relaxation delays were 0 (2), 10, 20, 30, 40, 100, 200, 300, 400, 500 (2), 750, 1000 and 1300 ms for T_1 data and 4 (2), 8, 12, 16 (2), 20, 24, 28, 32, 36, 40, 52, 56, 72, 80 ms for T_2 data. Heteronuclear

NOE values were determined using signal intensity ratios from spectra with (I_{sat}) and without (I_{ref}) saturation of the amide protons using 120° pulses. The uncertainty of signal intensities was estimated from the baseline noise levels in the spectra. Typically, only peaks with signal-to-noise ratios above 5 were considered to be useful for the following analyses. Longitudinal and transverse relaxation rates were obtained by fitting a two-parameter exponential to the experimental intensity decay as a function of relaxation delay: $I(t)=I_0\exp(-Rt)$; fits were performed using CurveFit (Palmer, A. G.; Columbia University, New York, USA; unpublished). Uncertainties in the rate constants were derived from a Jackknife procedure. Relaxation data HP56 which were used for comparison, were published previously (Slijper *et al.*, 1997).

Supplementary text and figures

Text S1: Dynamic properties of the DNA-binding domain in the Lac repressor dimer

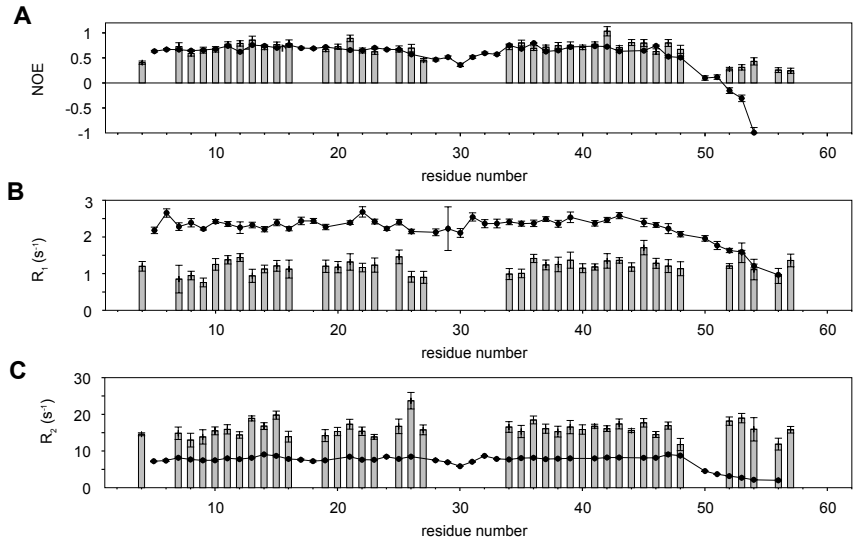


Figure 4.8: *Relaxation parameters for the headpiece residues of the unliganded Lac repressor as a part of Lac328 (gray bars) vs. isolated headpiece (residues 1-56) (filled triangles) and isolated HP62-V52C (filled circles). (A) Heteronuclear NOE. (B) Longitudinal relaxation rate R_1. (C) Transverse relaxation rate R_2.*

To study the motional behavior of the DNA-binding domain in the context of the dimeric Lac repressor we determined the relaxation parameters (R_1, R_2 and heteronuclear NOE) for the free Lac repressor dimer and compared them to the corresponding values for the isolated Lac DNA-binding domain HP56 (Figure 4.8).

Heteronuclear NOE values as well as R2-relaxation rates for residues 1-48 indicate that motional frequencies in this region are similar regardless whether the head-

piece is isolated or attached to the Lac repressor core (Figure 4.8). This supports the idea of the headpiece representing an independent structural unit.

The hinge region, however, behaves differently since R_2-relaxation rates of the intact Lac repressor dimer and isolated Lac headpiece are distinct. A continuous decrease in R_2 is observed for free HP56, reflecting the flexibility toward the C-terminus of the Lac repressor DNA binding domain. In contrast, R_2-levels do not change with respect to the remainder of the DNA-binding domain in the intact Lac repressor dimer, clearly showing that this region is not flexible when it is connected to the much larger core domain of the repressor and may even be partially folded. Smaller heteronuclear NOE values for the hinge region in both proteins show that the hinge helices are largely unfolded in the apo form.

Text S6: Allosteric transition of the Lac repressor

Chemical shift mapping permitted us to identify key events in the allosteric mechanism of Lac repressor whereas crystal structures provide information only on the endpoints of the transition. By comparing spectra of free, anti-inducer, and inducer bound Lac repressor we identified peaks that only shift by the latter addition, thereby elucidating residues that contribute to allosteric signaling. The trigger site involves residues Ser193, Asp149, Asn125 and Thr68 (Figure 4.9(A)). Then signal is transferred to the dimerization interface (Figure 4.9(B)) and interface between the N- and C- subdomains of the core domain (Figure 4.9(C)). This was reconciled by the mutational data (Suckowa *et al.*, 1996; Pace *et al.*, 1997). Substitutions at dimerization interface result in the I^s phenotype, the repressor which is incapable of transmitting allosteric signal. If protein carries mutations at interface between the N- and C- subdomains it becomes non-functional (I^- phenotype) due to the abolishment of functional dimer formation. Thus a good correlation is found between conformational transitions (monitored by chemical shift changes) and the reported effect of the various substitutions at these positions. The allosteric switch is the interface between the core domain and the headpiece involving contacts between Ile48 and Arg118 (Figure 4.9(D)). Mutation of these residues result in an I^- phenotype (Suckowa *et al.*, 1996), arguing that these residues are crucial for DNA binding underscoring the importance for the interdomain contacts for both binding and allosteric signaling. As a result of the loss of contacts between the core of one protomer and HP of the other protomer, the off-rate is enhanced leading to the lower affinity. The dramatic decrease in affinity for the operator by insertion of a glycine linker between the hinge helix and the core domain, underscores the importance of these contacts (Falcon & Matthews, 1999). Residues at the interdomain interface are not conserved between members of the Lac family members and might be important for differentiating their functions, in agreement using repressor mutants with a Pur inducer binding domain, (Tungtur *et al.*, 2007) show that position 48 is one of the key residues to determine allosteric response by the Pur co-repressor.

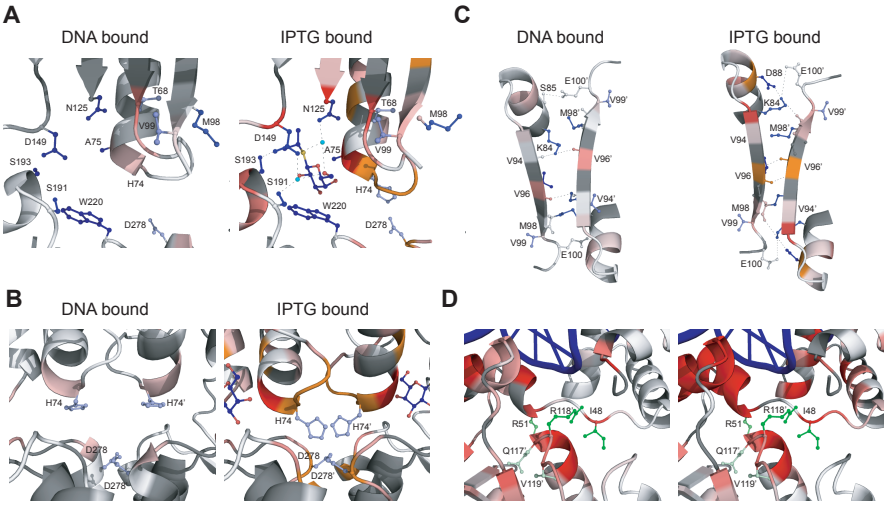


Figure 4.9: *Comparison of the crystallography, mutational and chemical shift data. The chemical shift changes of the backbone amides are plotted on the corresponding structures as a gradient of white-red. The following structures were used: Lac repressor bound to the symmetrical operator (1EFA) and Lac repressor core domain bound to IPTG (2P9H). Side chains of the important residues are shown. The strength of the I^s and I^- phenotypes is shown as gradient of blue and green, respectively. Hydrogen bonds are shown as dashed lines. (A) Chemical shift changes in the inducer binding site: upon DNA binding to the unliganded Lac repressor dimer (left) and upon binding of IPTG to the protein-DNA complex (right). (B) Comparison of interactions near the inducer binding site in the structures of the repressor bound to DNA (left) and IPTG (right). (C) Structural and chemical shift changes at the dimerization interface. (D) Interactions at the interface between the DNA binding domain and N-subdomain of the core. Left panel shows chemical shift differences between the unliganded and DNA bound Lac repressor dimer. Right panel shows chemical shift differences between the isolated headpiece bound to operator and core versus the intact protein bound to DNA. It is clear that the helix-turn-helix motif binds similarly to the isolated headpiece and in the intact dimer. The loop connecting the helical bundle to the hinge and the hinge helix are affected by the presence of the core showing formation of the interdomain contacts. Full color figure on page 101.*

Bibliography

Adair, G.S. (1925). A critical study of the direct method of measuring the osmotic pressure of haemoglobin. *Proc. Roy. Soc. (London) Ser. A*, **108 A**, 627–637.

Alberti, S., Oehler, S., von Wilcken-Bergmann, B., Krämer, H. & Müller-Hill, B. (1991). Dimer-to-tetramer assembly of lac repressor involves a leucine heptad repeat. *New Biol.*, **3**, 57–62.

Alberti, S., Oehler, S., von Wilcken-Bergmann, B. & Müller-Hill, B. (1993). Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. *EMBO J.*, **2**, 3227–3236.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. & Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.*, **29**, 231–262.

Barkley, M.D., Riggs, A.D., Jobe, A. & Bourgeois, S. (1975). Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry*, **14**, 1700–1712.

Bell, C.E. & Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nature*, **7**, 209–214.

Bell, C.E. & Lewis, M. (2001). Crystallographic analysis of Lac repressor bound to natural operator O1. *J. Mol. Biol.*, **312**, 921–926.

Berg, O.G., Winter, R.B. & Von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.

Betz, J.L., Sasmor, H.M., 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.

Boelens, R., Scheek, R.M., van Boom, J.H. & Kaptein, R. (1987). Complex of lac repressor headpiece with a 14 base-pair lac operator fragment studied by two-dimensional nuclear magnetic resonance. *J. Mol. Biol.*, **193**, 213 – 216.

Bohr, C. (1903). Theoretische Behandlung det quantitativen Verhaltnis bei der Sauerstoffaufnahme des Hamoglobin. *Zentr. Physiol.*, **17**, 682–689.

Boomershine, W.P., McElroy, C.A., Tsai, H.Y., Wilson, R.C., Gopalan, V. & Foster, M.P. (2003). Structure of Mth11/Mth Rpp29, an essential protein subunit of archaeal and eukaryotic RNase P. *Proc. Natl. Acad. Sci. USA*, **100**, 15398–15403.

Borukhov, S. & Lee, J. (2005). RNA polymerase structure and function at lac operon. *C. R. Biol.*, **328**, 576–587.

Brunger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J.S., Kuszewski, J., Nilges, N., Pannu, N., Read, R., Rice, L., Simonson, T. & Warren, G. (1998). Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Cryst.*, **D54**, 905–921.

Busby, S. & Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.*, **293**, 199–213.

Butler, A.P., Revzin, A. & Von Hippel, P.H. (1977). Molecular parameters characterizing the interaction of escherichia coli lac repressor with non-operator dna and inducer. *Biochemistry*, **16**, 4757–4768.

Calos, M. (1978). The DNA sequence for a low-level promoter of the lactose repressor gene and an 'up' promoter mutation. *Nature*, **274**, 762 – 765.

Caruthers, M. (1980). Deciphering the protein-DNA recognition code. *Acc. Chem. Res.*, **13**, 155–160.

Changeux, J.P. & Edelstein, S.J. (2005). Allosteric Mechanisms of Signal Transduction. *Science*, **308**, 1424–1428.

Chen, J. & Matthews, K.S. (1992). Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation. *J. Biol. Chem.*, **267**, 13843–13850.

Christopoulos, A. (2002). Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discov.*, **1**, 198–210.

Chuprina, V.P., Rullmann, 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.

Cooper, A. & Dryden, D.T.F. (1984). Allostery without conformational change - a plausible model. *Eur. Biophys. J.*, **11**, 103–109.

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.

Daber, R., Stayrook, S., Rosenberg, A. & Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. *J. Mol. Biol.*, **370**, 609–619.

Daly, T.J. & Matthews, K.S. (1986). Allosteric regulation of inducer and operator binding to the lactose repressor. *Biochemistry*, **25**, 5479–5484.

de Jong, R.N., Daniels, M.A., Kaptein, R. & Folkers, G.E. (2006). Enzyme free cloning for high throughput gene cloning and expression. *J. Struct. Funct. Genomics*, **7**, 109–118.

de Vries, S.J., van Dijk, A.D.J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. & Bonvin, A.M.J.J. (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimentional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

Dominguez, C., Boelens, R. & Bonvin, A.M.J.J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.

Dunaway, M., Olson, J., Rosenberg, J., Kallai, O., Dickerson, R. & Matthews, K. (1980). Kinetic studies of inducer binding to lac repressor-operator complex. *J. Biol. Chem.*, **255**, 10115–10119.

Ericsson, U.B., Hallberg, M.B., DeTitta, G.T., Dekker, N. & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Chem.*, **257**, 289–298.

Falcon, C.M. & Matthews, K.S. (1999). Glycine insertion in the hinge region of lactose repressor protein alters DNA binding. *J. Biol. Chem.*, **274**, 30849–30857.

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.

Fiaux, J., Bertelsen, E.B., Horwich, A.L. & Wuthrich, K. (2002). NMR analysis of a 900k groel-groes complex. *Nature*, **418**, 207–210.

Flynn, T.C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K.S. & Ma, J. (2003). Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Sci.*, **12**, 2523–2541.

Frank, D.E., Saecker, R.M., Bond, J.P., Capp, M.W., Tsodikov, O.V., Melcher, S.E., Levandoski, M.M. & Record, M.T. (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.

Friedman, A.M., O., F.T. & A., S.T. (1995). Crystal structure of the lac repressor core tetramer and its implications for DNA looping. *Science*, **268**, 1721–1727.

Gardner, K.H. & Kay, L.E. (1998). The use of ^{2}H, ^{13}C, ^{15}N multidimentional NMR to study the structure and dynamics of ptoteins. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Gerhart, J.C. & Pardee, A.B. (1961). Separation of feedback inhibition from activity of aspartate transcarbamylase (ATCase). *Fed. Proc.*, **20**, 224.

Gerhart, J.C. & Pardee, A.B. (1962). The enzymology of control by feedback inhibition. *J. Biol. Chem.*, **237**, 891–896.

Gerk, L.P., Leven, O. & Müller-Hill, B. (2000). Strengthening the dimerisation interface of Lac repressor increases its thermostability by 40 deg. C. *J. Mol. Biol.*, **299**, 805–812.

Gilbert, W. & Müller-Hill, B. (1966). Isolation of the lac respressor. *Proc. Natl. Acad. Sci. USA*, **56**, 1891–1898.

Gilbert, W., Gralla, J., Majors, J. & Maxam, A. (1975). Lactose operator sequences and the action of *lac* repressor. In H. Sund & G. Blauer, eds., *Symposium on Protein-Ligand Interaction*, 193–206, Berlin: Walter de Gruyter.

Gilbert, W., Majors, J. & Maxam, A. (1976). In *Dahlem Workshop on Chromosomes*, 167–178, Abakon Verlagsgesellschaft, Berlin.

Goodey, N.M. & Benkovic, S.J. (2008). Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.*, **4**, 474–482.

Hammes, G.G. & Wu, C.W. (1971). Regulation of enzyme activity. *Science*, **172**, 1205–1211.

Hua, Q.X., Dementieva, I.S., Walsh, M.A., Hallenga, K., Weiss, M.A. & Joachimiak, A. (2001). A thermophilic mini-chaperonin contains a conserved polypeptide-binding surface: combined crystallographic and NMR studies of the GroEL apical domain with implications for substrate interactions. *J. Mol. Biol.*, **306**, 513 – 525.

Inada, T., Kimato, K. & Aiba, H. (1996). Mechanism responsible for glucose-lactose diauxie *Escherichia coli* challenge to the cAMP model. *Genes to Cells*, **1**, 293–301.

Jacob, F. & Monod, J. (1961a). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol*, **3**, 318–356.

Jacob, F. & Monod, J. (1961b). On the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol.*, **26**, 193–211.

Jobe, A. & Bourgeois, S. (1972a). Lac repressor-operator interaction: VI. The natural inducer of the lac operon. *J. Mol. Biol*, **69**, 397–404.

Jobe, A. & Bourgeois, S. (1972b). Lac repressor-operator interaction: VIII. Lactose in an anti-inducer of the *lac* operon. *J. Mol. Biol*, **75**, 303–313.

Jobe, A., Sadler, J.E. & Bourgeois, S. (1974). Lac repressor-operator interaction: IX. The binding of lac repressor to operators containing O^c mutations. *J. Mol. Biol.*, **85**, 231–248.

Johnson, B.A. & Blevins, R.A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

Johnson, L. (1992). Glycogen phosphorylase: control by phosphorylation and allosteric effectors. *FASEB J.*, **6**, 2274–2282.

Kaback, H.R. (2005). Structure and mechanism of the lactose permease. *C. R. Biol.*, **328**, 557–567.

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.J.J., Salinas, K.R., 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.

Kalodimos, C.G., Biris, N., Bonvin, A.M.J.J., Levandoski, M.M., Guennuegues, M., Boelens, R. & Kaptein, R. (2004). Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*, **305**, 386–389.

Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W. & Von Hippel, P.H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proc. Natl. Acad. Sci. USA*, **74**, 4228–4232.

Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. & van Gunsteren, W.F. (1985). A protein structure from nuclear magnetic resonance data: lac repressor headpiece. *J. Mol. Biol.*, **182**, 179–182.

Kern, E.R.P., Dorothee and. Zuiderweg (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, **13**, 748–757.

Koradi, R., Billeter, M. & Wúthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

Koshland, D.E., Nemethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **5**, 365–385.

Koshland, J., Daniel E. & Hamadani, K. (2002). Proteomics and models for enzyme cooperativity. *J. Biol. Chem.*, **277**, 46841–46844.

Krämer, H., Niemöller, M., Amouyal, M., Revet, B., v Wilcken-Bergmann, B. & Müller-Hill, B. (1987). Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.*, **6**, 14811491.

Krämer, H., Amouyal, M., Nordheim, A. & Müller-Hill, B. (1988). DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J.*, **7**, 547–56.

Kumar, S., Ma, B., Tsai, C., Sinha, N. & Nussinov, R. (2000). Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.*, **9**, 10–19.

Laiken, S.L., Gross, C.A. & von Hippel, P.H. (1972). Equilibrium and kinetic studies of *Escherichia coli lac* repressor-iducer interactions. *J. Mol. Biol*, **66**, 143–155.

Lamerichs, R.M.J.N. (1989). *2D NMR studies of biomolecules: protein structures and protein-DNA interactions*. Ph.D. thesis, Bijvoet Center for Biomolecular Research, Utrecht.

Lavery, R. & H., S. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.

Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., v Wilcken-Bergmann, B. & Müller-Hill, B. (1987). The interaction of the recognition helix of lac repressor with lac operator. *EMBO J.*, **6**, 31453153.

Lengeler, J.W. (1993). Carbohydrate transport in bacteria under environmental conditions, a black box? *Antonie van Leeuwenhoek*, **63**, 275–288.

Lewis, M. (2005). The lac repressor. *C. R. Biol.*, **328**, 521–548.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247–1254.

Lin, S.y. & Riggs, A.D. (1972). Lac repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. *J. Mol. Biol.*, **72**, 671–690.

Linge, J., Habeck, M., Rieping, W. & Nilges, M. (2003a). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, **19**, 315–316.

Linge, J., Williams, M., Spronk, C., Bonvin, A. & Nilges, M. (2003b). Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.

Markiewicz, P., Kleina, L., Cruz, C., Ehret, S. & Miller, J. (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.

Markwick, P.R.L., Malliavin, T. & Nilges, M. (2008). Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol*, **4**, e1000168.

Matthews, B.W. (2005). The structure of E. coli [beta]-galactosidase. *C. R. Biol.*, **328**, 549–556.

McDonald, I.K. & Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777 – 793.

McElroy, C., Manfredo, A., Wendt, A., Gollnick, P. & Foster, M. (2002). TROSY-NMR studies of the 91 kDa TRAP protein reveal allosteric control of a gene regulatory protein by ligand-altered flexibility. *J. Mol. Biol.*, **323**, 463 – 473.

Miller, J.H. (1984). Genetic studies of the lac repressor: XII. Amino acid replacements in the DNA binding domain of the Escherichia coli lac repressor. *J. Mol. Biol.*, **180**, 205–212.

Miller, J.H. & Reznikoff, W.S., eds. (1978). *The Operon*. Cold Spring Harbor Laboratory.

Monod, J. (1977). *Chance and Necessity: Essay on the Natural Philosphy of Modern Biology*. Penguin Books.

Monod, J., Changeux, J.P. & Jacob, F. (1963). Allosteric proteins and cellular control systems. *J. Mol. Biol*, **6**, 306–329.

Monod, J., Wyman, J. & Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol*, **12**, 88–118.

Mossing, M.C. & Record, M.T. (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.

Muirhead, H. & Perutz, M.F. (1963). Structure of haemoglobin. A three-dimensional Fourier synthesis of reduced human haemoglobin at 5.5 Å resolution. *Nature*, **199**, 633–639.

Mulder, F., Schipper, D., Bott, R. & Boelens, R. (1999). Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.*, **292**, 111–123.

Müller, J., Oehler, S. & Müller-Hill, B. (1996). Repression of *lac* promotor as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.*, **257**, 21–29.

Müller-Hartmann, H. & Müller-Hill, B. (1996). The side-chain of the amino acid residue in position 110 of the lac repressor influences its allosteric equilibrium. *J. Mol. Biol.*, **257**, 473–478.

Müller-Hill, B. (1998). The function of auxiliary operators. *Mol. Microbiol.*, **29**, 13–18.

Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J., Nilges, M., Ulrich, E., Kaptein, R. & Bonvin, A. (2005). RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins*, **59**, 662–672.

Nichols, J.C. & Matthews, K.S. (1997). Combinatorial mutations of lac repressor. Stability of monomer-monomer interfaceis increased by apolar substitution at position 84. *J. Biol. Chem.*, **272**, 18550–18557.

Oehler, S., Eismann, E.R., Krämer, H. & Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *EMBO J.*, **9**, 973–979.

Oehler, S., Amouyal, M., Kolkhof, P., Wilcken-Bergmann, B. & Müller-Hill, B. (1994). Quality and position of the three *lac* operators of *E.coli* define efficiency of repression. *EMBO J.*, **13**, 3348–3355.

O'Gorman, R., Rosenberg, J., Kallai, O., Dickerson, R., Itakura, K., Riggs, A. & Matthews, K. (1980). Equilibrium binding of inducer to lac repressor-operator DNA complex. *J. Biol. Chem.*, **255**, 10107–10114.

O'Gorman, R.B. & Matthews, K.S. (1977). Fluorescence and ultraviolet spectral studies of Lac repressor modified with N-bromosuccinimide. *J. Biol. Chem.*, **252**, 3572–3577.

Osumi, T. & Saier Jr, M.H. (1982). Regulation of lactose permease activity by the phosphoenolpyruvate:sugar phosphotransferase system: evidence for direct binding of the glucose-specific enzyme III to the lactose permease. *Proc. Natl. Acad. Sci. USA*, **79**, 14571461.

Pace, H.C., Kercher, M.A., Lu, P., Markiewicz, P., Miller, J.H., Chang, G. & Lewis, M. (1997). Lac repressor genetic map in real space. *Trends Biochem. Sci*, **22**, 334 – 339.

Pardee, A.B., Jacob, F. & Monod, J. (1959). The genetic control and cytoplasmic expression of inducibility in the synthesis of B-galactosidase by E. coli. *J. Mol. Biol*, **1**, 165–178.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. & North, A.C.T. (1960). Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature*, **185**, 416–422.

Perutz, M.F., Bolton, W., Diamond, R., Muirhead, H. & Watson, H.C. (1964). Structure of haemoglobin. An X-ray examination of reduced horse haemoglobin. *Nature*, **203**, 687–690.

Pervushin, K., Riek, R., Wider, G. & Wthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipoledipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules insolution. *Proc. Natl. Acad. Sci. USA*, **94**, 12366–12371.

Postma, P.W., Lengeler, J.W. & Jacobson, G.R. (1996). Phosphoenolpyruvate-carbohydrate phosphotransferase systems. In F.C. Neidhardt, ed., *Escherichia coli and Salmonella. Cellular and Molecular Biology*, 1149–1174, ASM Press: Washington, DC.

Reznikoff, W. (1984). *The microbe 1984, Part II. Prokaryotes and Eukaryotes*. Cambridge University Press.

Reznikoff, W.S. (1992). Catabolite gene activator protein activation of lac transcription. *J. Bacteriol.*, **174**, 655–658.

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.

Riggs, A.D., Bourgeois, S. & Cohn, M. (1970a). Lac repressor-operator interaction: III. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.

Riggs, A.D., Newby, R.F. & Bourgeois, S. (1970b). Lac repressor-operator interaction: II. Effect of galactosides and other ligands. *J. Mol. Biol*, **51**, 303–314.

Riggs, A.D., Suzuki, H. & Bourgeois, S. (1970c). Lac repressor-operator interaction: I. Equilibrium studies. *J. Mol. Biol.*, **48**, 67–83.

Riggs, A.D., Lin, S. & Wells, R.D. (1972). Lac repressor binding to synthetic DNAs of defined nucleotide sequence. *Proc. Natl. Acad. Sci. USA*, **69**, 761–764.

Roberts, G.C.K., ed. (1993). *NMR of Macromolecules: A practical Approach*. Oxford University Press.

Roderick, S.L. (2005). The lac operon galactoside acetyltransferase. *C. R. Biolog.*, **328**, 568–575.

Sadler, J.R., Sasmor, H. & Betz, J.L. (1983). A perfectly symmetric *lac* operator binds to the *lac* repressor very tightly. *Proc. Natl. Acad. Sci. USA*, **80**, 6785–6789.

Saier, M.H., Ramseier, T.M. & Reizer, J. (1996). Regulation of carbon utilization. In F.C. Neidhardt, ed., *Escherichia coli and Salmonella. Cellular and Molecular Biology*, 1325–1343, ASM Press: Washington, DC.

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.

Sattler, M. & Fesik, S.W. (1996). Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure*, **4**, 1245 – 1249.

Scheek, R., Russo, N., Boelens, R. & Kaptein, R. (1983). Sequential resonance assignment in DNA ^{1}H NMR spectra by two-dimensional NOE spectroscopy. *J. Am. Chem. Soc.*, **105**, 2914–2916.

Schnarr, M. & Maurizot, J.C. (1981). Unfolding of lac repressor and its proteolytic fragments by urea: headpieces stabilize the core within lac repressor. *Biochemistry*, **20**, 6164–6169.

Schumacher, M., Choi, K., Zalkin, H. & Brennan, R. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, **266**, 763–770.

Slijper, M. (1996). *NMR studies of protein-DNA interactions in the lac operon*. Ph.D. thesis, Bijvoet Center for Biomolecular Research, Utrecht.

Slijper, M., Boelens, R., Davis, A.L., Konings, R.N.H., van der Marel, G.A., van Boom, J.H. & Kaptein, R. (1997). Backbone and side chain dynamics of lac repressor headpiece (1-56) and its complex with dna. *Biochemistry*, **36**, 249–254.

Smith, T.F. & Sadler, J.R. (1971). The nature of lactose operator constitutive mutations. *J. Mol. Biol.*, **59**, 273 – 305.

Sprangers, R., Velyvis, A. & Kay, L.E. (2007). Solution NMR of supramolecular complexes: providing new insights into function. *Nat. Meth.*, **4**, 697–703.

Spronk, C.A., Bonvin, A.M., Radha, P.K., Melacini, G., Boelens, R. & Kaptein, R. (1999a). The solution structure of lac repressor headpiece 62 complexed to a symmetrical *lac* operator. *Structure*, **7**, 1483 – 1492.

Spronk, C.A., Folkers, G.E., Noordman, A., Wechselberger, R., van den Brink, N., Boelens, R. & Kaptein, R. (1999b). Hinge-helix formation and DNA bending in various lac repressor-operator complexes. *EMBO J.*, **18**, 6472–6480.

Stevens, S.Y., Sanker, S., Kent, C. & Zuiderweg, E.R. (2001). Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. *Nat. Struct. Mol. Biol.*, **8**, 947–952.

Studier, F. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif*, **41**, 207 – 234.

Suckowa, J., Markiewiczb, P., Kleinac, L.G., Miller, J.H., Kisters-Woikea, B. & Müller-Hill, B. (1996). Genetic studies of the lac repressor XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J. Mol. Biol.*, **261**, 509–523.

Swain, J.F. & Gierasch, L.M. (2006). The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.*, **16**, 102–108.

Tam, R. & Saier Jr, M.H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.*, **57**, 320–346.

Taraban, M., Zhan, H., Whitten, A.E., Langley, D.B., Matthews, K.S., Swint-Kruse, L. & Trewhella, J. (2008). Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J. Mol. Biol.*, **376**, 466–481.

Tjandra, N., Grzesiek, S. & Bax, A. (1996). Magnetic field dependence of nitrogen-proton *J* splittings in ^{15}N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc*, **118**, 6264–6272.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. & Bax, A. (1997). Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Mol. Biol.*, **4**, 732–738.

Tugarinov, V., Choy, W.Y., Orekhov, V.Y. & Kay, L.E. (2005). Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. USA*, **102**, 622–627.

Tungtur, S., Egan, S.M. & Swint-Kruse, L. (2007). Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins*, **68**, 375–388.

Tyler, R.C., Sreenath, H.K., Singh, S., Aceti, D.J., Bingman, C.A., Markley, J.L. & Fox, B.G. (2005). Auto-induction medium for the production of [*U*-^{15}N]- and [*U*-^{13}C, *U*-^{15}N]-labeled protein for NMR screening and structure determination. *Protein Expr. Purif.*, **40**, 268–278.

Volkman, B.F., Lipson, D., Wemmer, D.E. & Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science*, **291**, 2429–2433.

Weber, G. (1972). Ligand binding and internal equilibriums in proteins. *Biochemistry*, **11**, 864–878.

Whitson, P.A., Burgum, A.A. & Matthews, K.S. (1984). Trinitrobenzenesulfonate modification of the lysine residues in lactose repressor protein. *Biochemistry*, **23**, 6046–6052.

Wilson, C.J., Zhan, H., Swint-Kruse, L. & Matthews, K.S. (2007). The lactose repressor system: paradigm for regulation, allosteric behavior and protein folding. *Cell. Mol. Life. Sci*, **64**, 3–16.

Winter, R.B. & Von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli lac repressor-operator interaction: equilibrium measurements. *Biochemistry*, **20**, 6948–6960.

Winter, R.B., Berg, O.G. & Von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*, **20**, 6961–6977.

Wu, F.Y.H., Bandyopadhyay, P. & Wu, C.W. (1976). Conformational transitions of the lac repressor from Escherichia coli. *J. Mol. Biol.*, **100**, 459 – 472.

Wu, H.M. & Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509–513.

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.

Summary

Bacteria, like all other organisms can respond to changing conditions in their environment such as nutrients. When glucose is plentiful, *Escherichia coli* utilizes it exclusively as its food source, even when other sugars are present. Only when the glucose supplies become exhausted will the bacterium upregulate expression of proteins that transport and metabolize other carbon sources, such as lactose. Genes that code for these proteins are grouped in the stretch of DNA called the *lac* operon. In order to regulate expression of the genes, the operon has a master switch: the Lac repressor protein. The repressor associates with a DNA regulatory element called the operator and this blocks synthesis of the genes. If the repressor is to function as a switch, it must be inducible; the switch must be able to turn on or turn off in response to a given chemical signal. To accomplish this the repressor binds an inducer, a metabolite that monitors the metabolic state. The inducer is a chemical signal that modulates affinity of the repressor for the operator and turns the switch on or off. In the presence of the inducer the repressor dissociates from the operator, which permits the expression of the genes. Binding of inducer changes the structure of the repressor such that different conformations of the repressor have different binding affinities for the operator DNA. This regulatory mechanism, when a ligand associates with a protein and alters its ability to perform a given function through changing its conformation, is called allostery. Therefore Lac repressor is allosterically regulated to react to changes in its environment. The major goal of this thesis is to study complexes between the Lac repressor and its ligands to better understand how the Lac repressor recognizes its cognate DNA sequences and how the allosteric mechanism of the repressor functions.

A general overview of the regulation of the *lac* operon and its components as well as introduction to the concept of allostery and various models for allostery is provided in **Chapter 1**. The Lac repressor consists of four subunits that are arranged as a dimer of dimer and are held together by a tetramerization domain. Each subunit has a so-called headpiece domain that uses a helix-turn-helix (HTH) motif for DNA binding. The dimeric core domain is responsible for the inducer binding and can adopt different conformations.

In order to understand how Lac repressor recognizes its operators, we determined the structures of various Lac headpiece-operator complexes. **Chapter 2** describes the NMR structures of a dimeric mutant of the Lac repressor DNA binding domain (DBD), or headpiece, bound to the natural *lac* operators. In addition to the main operator *O1*, the *lac* operon of *E.coli* has two auxiliary operators *O2* and *O3*. Due to the tetrameric nature of the Lac repressor it can simultaneously bind to the main, promotor-proximal operator *O1* with one dimer and to either of the auxiliary opera-

tors with the second dimer creating one of two alternative DNA loops. Natural *lac* operators are partially palindromic consisting of two half-sites where each half-site binds one Lac DBD monomer. For these studies we used the dimeric headpiece mutant V52C which has a covalent disulfide linkage between the two monomers and therefore a very high DNA affinity. The structure of the Lac-*O2* complex showed that the Lac DBD is able to change its conformation to compensate for the small differences in the sequence of the *O2* operator (compared to the *O1* operator), retaining high binding affinity. The number of contacts between the Lac DBD monomer and the left half-site of the *O3* operator is comparable to other complexes. Therefore this part of the binding mode is largely specific like in other complexes. On the other hand significant alterations in the sequence of the right half-site of the *O3* operator precluded formation of interactions. Therefore the dramatic decrease in the affinity of the Lac repressor for the *O3* operator can be explained by a smaller number of contacts between the DNA and the HTH motif. In fact, the interactions in the right subunit of the Lac-*O3* complex resemble that of a non-operator DNA complex. Interestingly, the stability of the C-terminal helix of the Lac repressor DBD, or hinge helix, which is formed only upon binding to the specific DNA sequences, is the same in all three complexes, as judged from the results of the hydrogen-deuterium exchange experiments, despite the difference in affinity. We argue that by binding to the central part of the operator hinge helices help in positioning the HTH motif domain in the correct position.

With the aim to understand how the inducer would affect DNA binding and hinge helix formation and how the inducer signal would be transmitted to the headpiece, we designed a dimeric Lac repressor construct. In **Chapter 3** and **Chapter 4** the structural studies are extended to the 76 kDa two-domain construct of the dimeric Lac repressor which encompasses the DNA binding domain and an inducer binding domain (core domain). The C-terminal tetramerization region was removed to permit only formation of dimers which reduces the molecular size and makes the repressor fully symmetric resulting in more tractable NMR spectra. **Chapter 3** describes several strategies that might be of general interest to overcome difficulties associated with studies of large molecular weight proteins by NMR spectroscopy. We produced a thermostable, isotopically labeled (^{2}H, ^{15}N, ^{13}C) Lac repressor dimer and its inducer binding domain with the aim to elucidate the allosteric mechanism of the Lac repressor. The availability of the thermostable, deuterated protein samples yielded high quality NMR spectra and a near-complete resonances assignment. **Chapter 3** also describes the optimization protocol which might be of general use for obtaining the best buffer conditions for a stable protein and to further improve the sensitivity of the NMR spectra.

Chapter 4 presents the analysis of various regulatory states of the Lac repressor under identical conditions using the chemical shift perturbation method. For its regulatory mechanism Lac repressor can exist in at least four different states: the free Lac repressor, the repressor bound to the DNA operator, the repressor bound to the inducer, and the ternary complex when both DNA and inducer are bound. By comparing the chemical shifts of the protein in the various states we are able to outline a path for the allosteric signal from the inducer binding site to the DNA binding domain which are 40 Å away from each other. The inducer binding signal is transferred to the DBD through the large conformational changes in the N-terminal subdomain of the

inducer binding domain. This part of the repressor acts as the main allosteric switch that can lock the DBD domain in either a tight or weak DNA bound conformation through formation of interdomain contacts.

Comparison of the chemical shifts of the free repressor, the repressor bound to the operator and the repressor bound to the inducer showed that the inducer binding domain is in fast dynamic equilibrium between two different conformational states. Binding of either operator or inducer simply shifts this equilibrium towards the conformation to which the Lac repressor binds with strong or weak affinity. This agrees well with one of the initial postulates of the Monod-Changeux-Wyman model for allostery which was introduced in **Chapter 1**. Further evidence that this model describes the allosteric mechanism of the Lac repressor best comes from the observation of the highly cooperative binding of the inducer to the core domain dimer without intermediate states.

We have currently based most of our interpretations of the allosteric model of the Lac repressor on NMR chemical shifts. Whereas these shifts are very sensitive markers for changes in the atomic environment, they do not provide direct structural information. The orientation of the various domains of the Lac repressor can be further monitored by Residual Dipolar Couplings (RDCs) whereas details of the DNA binding of the intact dimer could be investigated by incorporating paramagnetic labels either on the DNA or the protein site. The protein constructs that were prepared in **Chapter 3** may be very suitable for such studies in the future allowing precise structural and dynamical analysis of the allosteric mechanism. Thus there are highly promising possibilities for further investigations on the genetic switch of the *lac* operon.

Samenvatting

Zoals alle andere organismen kunnen ook bacteriën reageren op veranderende condities in hun omgeving, zoals voedingsstoffen. Wanneer glucose volop aanwezig is, gebruikt *Escherichia coli* dit als exclusieve voedselbron, dus zelfs wanneer andere suikers aanwezig zijn. Slechts wanneer de glucosevoorraden uitgeput raken verhoogt de bacterie de expressie van eiwitten die andere koolstofbronnen, transporteren en metaboliseren, zoals bijvoorbeeld lactose. Genen die voor deze eiwitten coderen, zijn gegroepeerd in het zogenaamde *lac* operon. Om expressie van deze genen te reguleren heeft dit operon een hoofdschakelaar: het Lac repressor eiwit. De repressor bindt aan een regulatie-element, de DNA-operator, en dit blokkeert synthese van de genen. Als de repressor als een schakelaar moet functioneren, moet het induceerbaar zijn; de schakelaar moet omgezet kunnen worden als respons op een gegeven chemisch signaal. In dit model bindt de repressor tevens een inducer, een metaboliet dat de metabolische toestand aangeeft. De inducer is een chemisch signaal dat de affiniteit van de repressor voor de operator moduleert of de schakelaar omzet. In aanwezigheid van de inducer dissocieert de repressor van de operator, hetgeen de expressie van de genen toestaat. Binding van de inducer verandert de structuur van de repressor zodanig dat verschillende conformaties van de repressor verschillende bindingsaffiniteiten voor het operator-DNA hebben. Dit regulerende mechanisme heet allosterie, en dus ook de Lac repressor wordt allosterisch gereguleerd door veranderingen in zijn omgeving. Het voornaamste doel van dit proefschrift is om complexen tussen Lac repressor en zijn liganden te bestuderen om beter te begrijpen hoe de Lac repressor zijn DNA herkent, en hoe het allosterische mechanisme van de repressor functioneert.

Een algemeen overzicht van de regulatie van het *lac* operon en zijn componenten, evenals een introductie van het concept van allosterie en verschillende modellen voor allosterie, is gegeven in **Hoofdstuk 1**. Er wordt beschreven dat de Lac repressor uit vier onderdelen bestaat die als een dimeer van dimeren georganiseerd zijn en bij elkaar gehouden worden door een tetramerisatie-domein. Elk onderdeel heeft een zogenaamd ”headpiece”-domein dat een helix-turn-helix (HTH) motief gebruikt om DNA te binden. Het dimere ”core”-domein is verantwoordelijk voor inducer-binding en kan verschillende conformaties aannemen.

Om te begrijpen hoe de Lac repressor zijn operatoren bindt, hebben we de structuren van verschillende Lac headpiece-operator complexen bepaald. **Hoofdstuk 2** beschrijft de NMR structuren van een dimere mutant van het DNA-bindende domein (DBD), oftewel de headpiece, van de Lac repressor gebonden aan de natuurlijke lac operatoren. Behalve de voornaamste operator, genaamd *O1*, heeft het *lac* operon van *E.coli* nog twee andere operatoren, namelijk *O2* en *O3*. Als gevolg van de tetramere

aard van de Lac repressor kan het tegelijkertijd binden met één dimeer aan de voornaamste operator *O1*, die bij de promoter ligt, met de tweede dimeer aan één van de andere operatoren, om zo één van twee verschillende DNA-lussen te creëren. Natuurlijke *lac* operatoren zijn gedeeltelijk palindroom, bestaande uit twee helften, waarbij elke helft één Lac DBD monomeer bindt. Voor deze studies gebruikten we de dimere headpiece mutant V52C die een covalente disulfide-binding heeft tussen de twee monomeren en derhalve een zeer hoge DNA-affiniteit. De structuur van het Lac-*O2* complex toonde dat het Lac DBD in staat is zijn conformatie te veranderen om te compenseren voor de kleine verschillen in de sequentie van de *O2*-operator ten opzichte van de *O1* operator, waardoor een hoge bindingsaffiniteit gewaarborgd blijft. Het aantal contacten tussen de Lac DBD monomeer en de linkerhelft van de *O3*-operator is vergelijkbaar met andere complexen. Daarom is dit deel van de bindingswijze grotendeels specifiek. Anderzijds verhinderen significante verschillen in de sequentie van de rechterhelft van de *O3*-operator de vorming van interacties. Daardoor kan de dramatische afname in de affiniteit van de Lac repressor voor de *O3* operator uitgelegd worden door een kleiner aantal contacten tussen het DNA en het HTH-motief. In feite lijken de interacties in het rechterdeel van het Lac-*O3* complex op die van een non-operator DNA complex. Het is interessant dat de stabiliteit van de C-terminale helix of "hinge"-helix van het Lac repressor DBD, die slechts gevormd wordt na binding aan de specifieke DNA sequenties, hetzelfde is in alle drie complexen ondanks het verschil in affiniteit, zoals afgeleid kon worden uit de resultaten van de waterstof-deuterium experimenten. We beargumenteren dat de hinge-helices helpen in de plaatsing van het HTH-motief in de correcte positie, door hun binding aan het centrale deel van de operator.

We hebben vervolgens een dimeer Lac repressor construct ontworpen, met als doel te onderzoeken hoe de inducer DNA-binding en hinge-helix vorming beïnvloedt en hoe het inducer-signaal aan het headpiece doorgegeven zou worden hebben. In **Hoofdstuk 3** en **Hoofdstuk 4** worden de structurele studies uitgebreid met het 74 kDa twee-domein construct van de dimere Lac repressor, dat het DNA-bindende domein en een inducer-bindend domein ("core"-domein) omvat. Het C-terminale tetramerisatie gebied is weggelaten om alleen vorming van dimeren toe te staan, wat de moleculaire grootte reduceert en de repressor volledig symmetrisch maakt. Dit resulteert in minder gecompliceerde NMR spectra. *Hoofdstuk 3* beschrijft verschillende strategieën die van algemeen belang kunnen zijn om moeilijkheden te overkomen met betrekking tot het bestuderen van eiwitten met een groot molecuulgewicht met behulp van NMR spectroscopie. Wij produceerden een thermostabiel, isotoop-gelabeld (^{1}H, ^{15}N, ^{13}C) Lac repressor dimeer en zijn inducer-bindingsdomein met als doel het allosterische mechanisme van de Lac repressor op te helderen. De beschikbaarheid van de thermostabiele, gedeutereerde eiwitten leidde tot hoge kwaliteit van de NMR-spectra en bijna complete resonantie-toekenning. *Hoofdstuk 3* beschrijft verder het optimalisatieprotocol dat van algemeen nut zou kunnen zijn om de beste buffercondities te verkrijgen voor een stabiel eiwit en om de gevoeligheid van de NMR-spectra verder te verbeteren.

Hoofdstuk 4 toont de analyse van verschillende regulatietoestanden van de Lac repressor bij identieke condities, waarbij we gebruik maken van de "chemical shift" verstoring-methode. Voor zijn regulatiemechanisme kan de Lac repressor in ten minste vier verschillende toestanden voorkomen: de vrije Lac repressor, de repres-

sor gebonden aan de DNA-operator, de repressor gebonden aan de inducer, en het drievoudige complex wanneer zowel DNA als inducer gebonden zijn. Door de chemical shifts van het eiwit in verschillende toestanden te vergelijken, waren we in staat om een route voor het allosterische signaal te schetsen vanaf de inducer-bindingslocatie naar het DNA-bindende domein, welke 40 Å uit elkaar liggen. Het inducer-bindingssignaal wordt naar het DBD overgebracht door grote conformationele veranderingen in het N-terminale deel van het inducer-bindende domein. Dit deel van de repressor treedt op als de belangrijkste allosterische schakelaar die het DBD-domein door de vorming van interdomein contacten ofwel in een stevig DNA-gebonden of in een zwak DNA-gebonden conformatie fixeert.

Vergelijking van de chemical shifts van de vrije repressor, de repressor gebonden aan de operator en de repressor gebonden aan de inducer liet zien dat het inducer-bindende domein in een snel dynamisch evenwicht verkeert tussen twee conformationele toestanden. Binding van ofwel de operator of de inducer verschuift dit evenwicht eenvoudig naar de conformatie waar de Lac repressor met sterke danwel zwakke affiniteit bindt. Dit komt goed overeen met één van de oorspronkelijke postulaten van het Monod-Changeux-Wyman model voor allosterie dat in **Hoofdstuk 1** geïntroduceerd was. Aanvullend bewijs dat dit model het allosterische mechanisme van de Lac repressor het best beschrijft komt van de observatie van de zeer coöperatieve binding van de inducer aan het core-domein zonder tussenliggende toestanden.

Tot nu toe hebben we de meeste van onze interpretaties voor het allosterische model van de Lac repressor gebaseerd op NMR chemical shifts. Hoewel deze shifts erg gevoelige markers zijn om veranderingen in de atomaire omgeving te volgen, leveren ze niet eenvoudig direct structuurinformatie. De oriëntatie van de verschillende domeinen van de Lac repressor kan verder worden bekeken met behulp van "Residual Dipolar Couplings" (RDCs), terwijl details van DNA-binding van het intacte dimeer onderzocht zou kunnen worden door paramagnetische labels in te brengen in het DNA of in het eiwit. De eiwitconstructen die in **Hoofdstuk 3** werden vervaardigd kunnen in de toekomst erg bruikbaar zijn voor zulke studies en maken een precieze structurele en dynamische analyse van het allosterische mechanisme mogelijk. Er zijn dus veelbelovende mogelijkheden voor verder onderzoek aan de genetische schakelaar van het *lac* operon.

Acknowledgement

I would like to express my deep and sincere gratitude to my promotor Rolf Boelens. Your wide knowledge, encouraging and supervision have provided a good basis for the present thesis. I wish to express my warm thanks to Gert who introduced me to the great world of the wet lab. You helped me a lot with your guidance, valuable advice and extensive discussions. I'm also deeply indebted to Rob Kaptein whose comments and direction were invaluable.

I would like to thank Roberto, Manuel, Karine, Hans, Heidi who were directly involved in the projects I worked on during my PhD. Thank you for good collaboration. Johan, thank you for your patience and help with my endless computer troubles. My deep gratitude goes to my two long-term office mates: Jeff and Klaartje. Computer guys, Marc, Sjoerd and Mikael, thank you for the scripts and useful discussions. Alex, thank you for your valuable help in docking protocols, cluster time and interest. Dev and Anding, we shared some tough moments. Good luck!

I thank all my former and present colleagues from the NMR department who created a stimulating scientific and friendly environment in the lab.

I also thank Vladislav Orekhov and Martin Billeter who opened the NMR field to me.

I deeply appreciate my family and friends for all support and fun.

Curriculum Vitae

Jūlija Romaņuka was born on Mach 28, 1981 in Voronezh, Soviet Union (now Russia). After finishing secondary school in Valka (Latvia) in 1999, she entered Riga Technical University (Latvia) where she obtained her Bachelor Degree in Engineering. In December 2003, she got a M.Sc. degree in Bioinformatics at Chalmers University of Technology (Gothenburg, Sweden) under the supervision of Prof. Dr. M. Billeter, studying the conformation of the uteroactive polypeptide Kalata B2 by NMR spectroscopy. From February 2004 to December 2008 she worked on her PhD thesis in the NMR department of Bijvoet Center for Biomolecular Research, Utrecht University, under the supervision of Prof. Dr. Rolf Boelens and Prof. Dr. Robert Kaptein. In February 2009 she started working at Shell, Rijswijk, the Netherlands.

List of publications

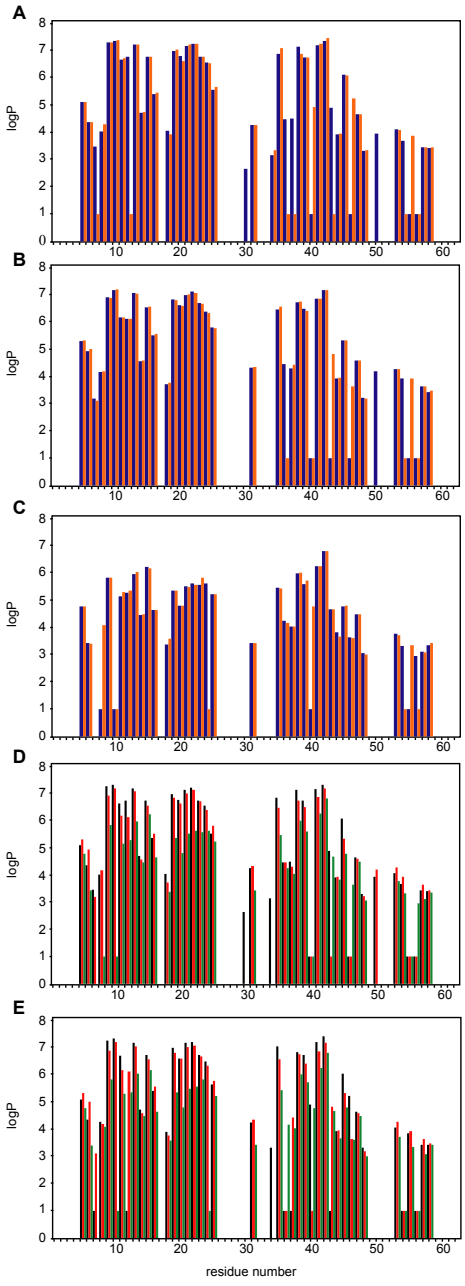
Romanuka J., Folkers G.E., Gnida M., Wienk H., Kaptein R., Boelens R. (2009). On the nature of allosteric transitions of the Lac repressor. *manuscript in preparation*.

Romanuka J., Folkers G.E., Biris N., Wienk H., Tischenko E., Bonvin A.M., Kaptein R., Boelens R. (2009). Specificity and affinity of Lac repressor for the auxiliary operators *O2* and *O3* is explained by the structures of their protein-DNA complexes. *submitted for publication*.

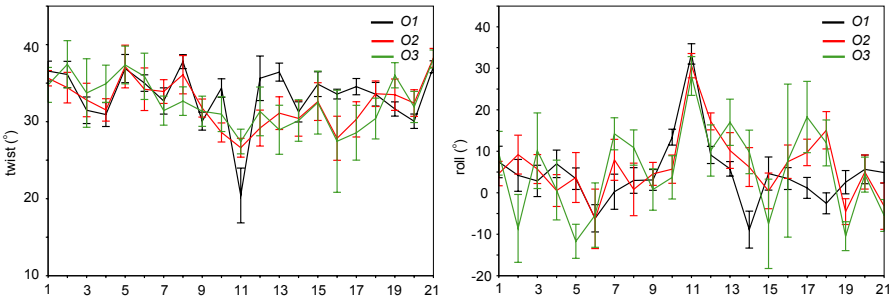
Romanuka J., Van den Bulcke H., Kaptein R., Boelens R., Folkers G.E. (2009). Novel strategies to overcome expression problems encountered with toxic proteins: application to the production of Lac repressor proteins for NMR studies. *submitted for publication*.

Nair S.S., **Romanuka J.**, Billeter M., Skjeldal L., Emmett M.R., Nilsson C.L., Marshall A.G.(2006). Structural characterization of an unusually stable cyclic peptide, kalata B2 from Oldenlandia affinis. *Biochim. Biophys. Acta*, **10**, 1568-1576

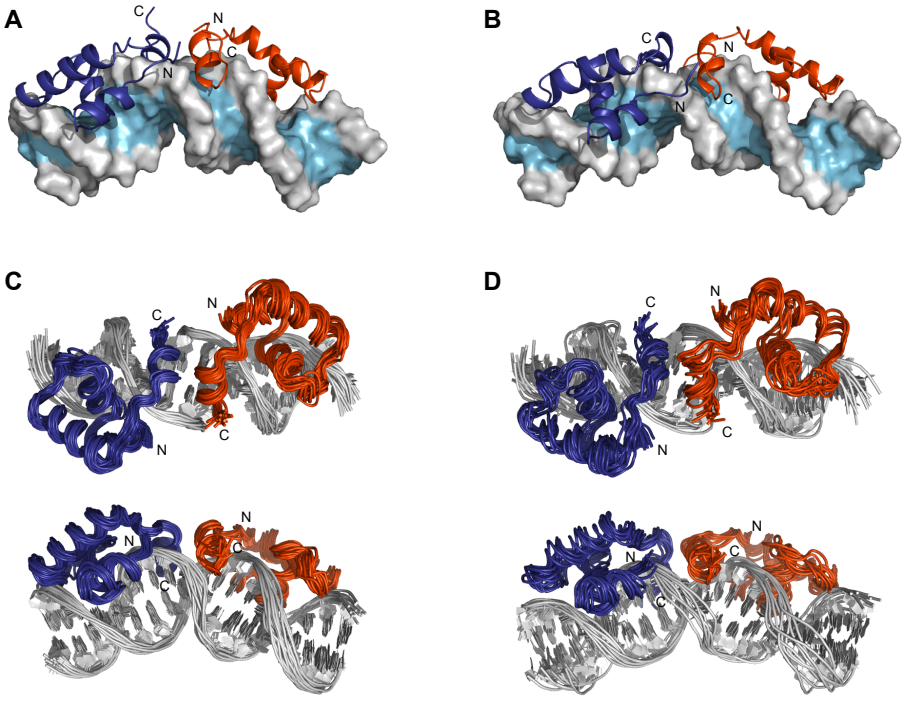
Full color figures



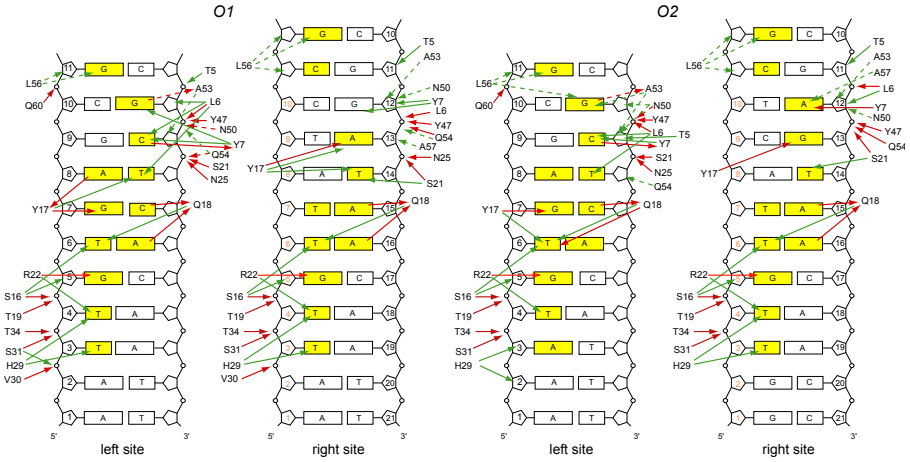
Chapter 2, figure 3. *Per residue protection factors obtained from amide H/D exchange experiments for the lac repressor headpiece bound to the natural operators O1 (A), O2 (B) and O3 (C). In (A, B, C) blue bars refer to the residues of the left subunit, whereas the orange ones to the right. (D) Protection factors for the left monomer for each complex. (E) 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 logP = 1 was arbitrarily given to residues not included in the analysis due to peak overlap.*



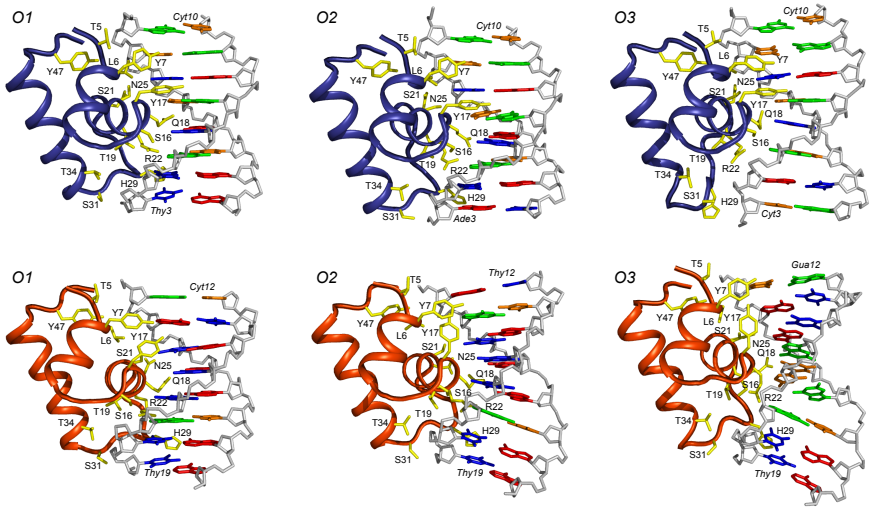
Chapter 2, figure 4. *Inter base pair DNA helical parameters of the lac operators O1 (black), O2 (red), and O3 (green) in complex with dimeric HP62.*



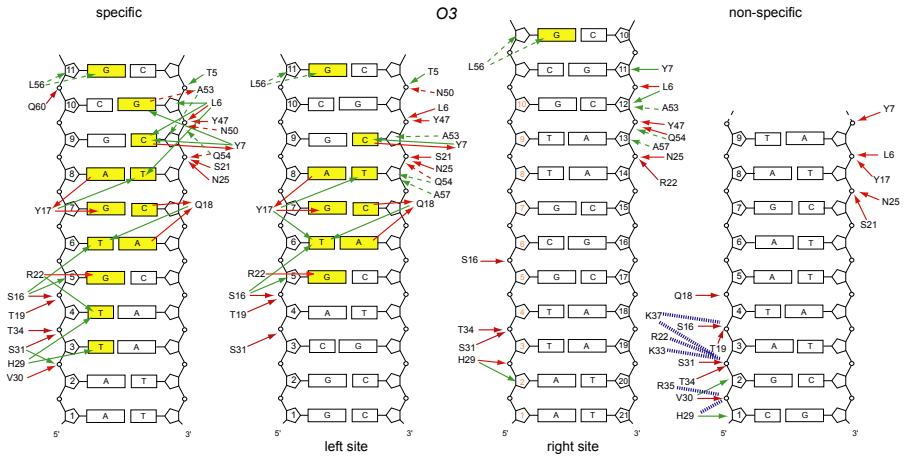
Chapter 2, figure 5. *Three-dimensional structures of the HP62V52C-O2 (A) and HP62V52C-O3 (B) complexes. The left and right Lac HP subunits are colored dark blue and dark orange, respectively. The ensemble of the 20 lowest energy structures of the HP62V52C-O2 complex (C) and the ensemble of the 10 lowest energy structures of the HP62V52C-O3 complex shown in two different views rotated by 90°. For clarity, only the backbone atoms of amino acids Val4-Lys59 and DNA heavy atoms of bp 2-20 are shown.*



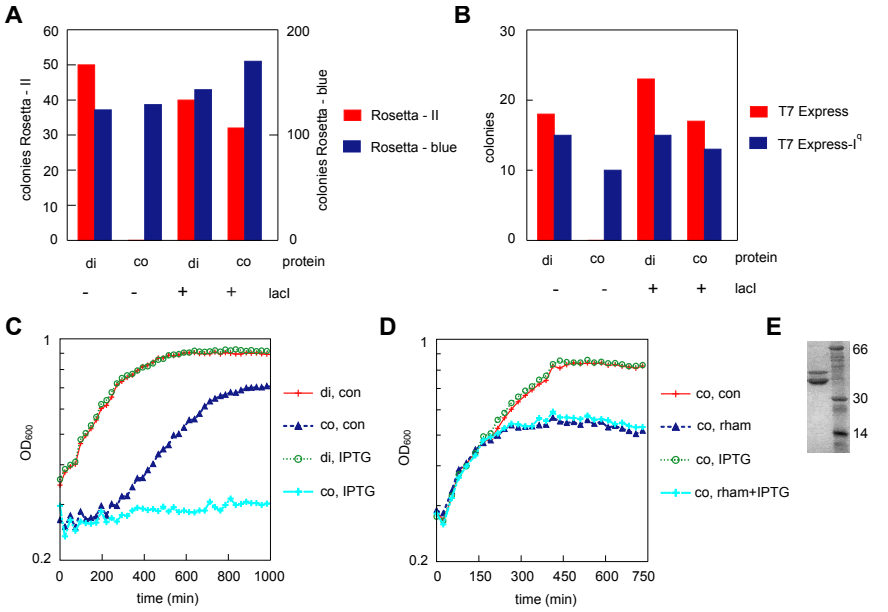
Chapter 2, figure 6. *Comparison of the protein-DNA contacts in the HP62V52C-O1 and HP62V52C-O2 complexes. The bases that are recognized specifically are colored in yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively.*



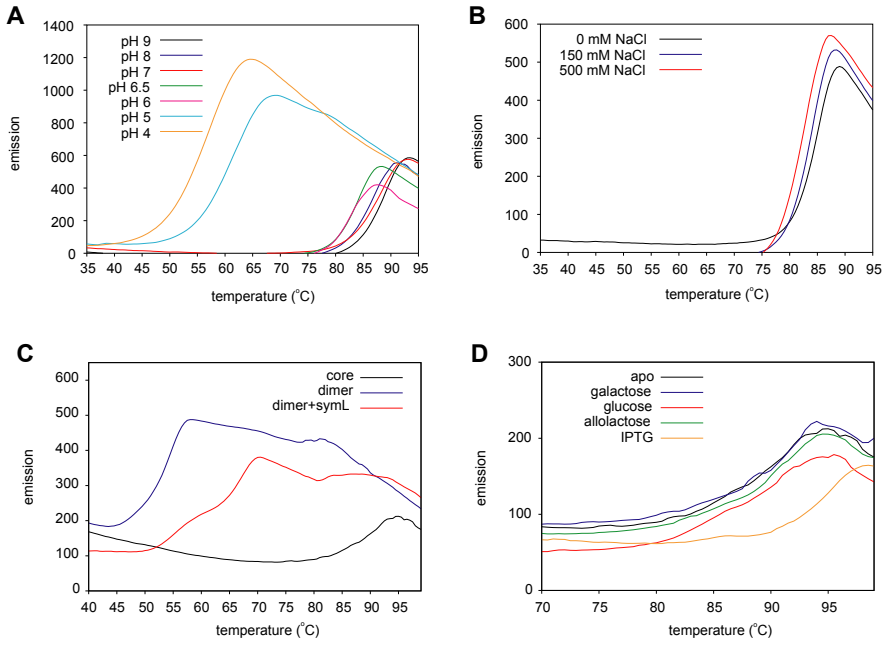
Chapter 2, figure 7. *Protein-DNA interactions observed in the HP62V52C-O1, HP62V52C-O2 and HP62V52C-O3 complexes. The left subunit of each complex is shown in dark blue, whereas the right subunit is shown in dark orange. Residues involved in protein-DNA interactions are shown in yellow.*



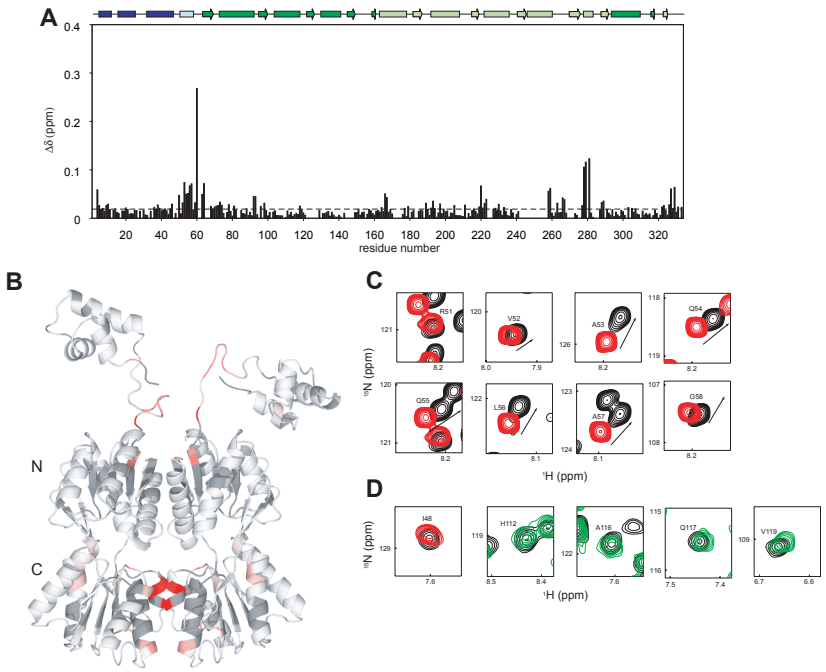
Chapter 2, figure 8. *Comparison of the protein-DNA contacts in the HP62V52C-O3 and HP62V52C-NOD complexes. The bases that are recognized specifically are colored in yellow. The solid and dashed lines indicate interactions in the major and minor grooves, respectively. Red and green lines indicate hydrogen bonding and hydrophobic contacts, respectively.*



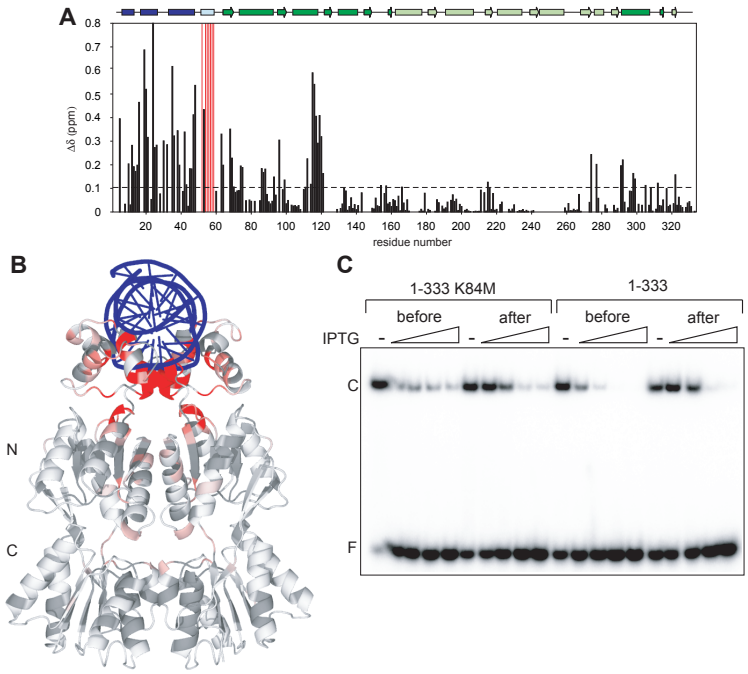
Chapter 3, figure 4. *(A) Number of colonies obtained after transformation of pET-LICHIS expression constructs, that lack (+) or contain (-) the full length LacI gene in the vector, for either the dimeric K84M Lac repressor mutant (di) or the core domain (co) in E. coli Rosetta-II (DE3) pLysS or Rosetta-blue. (B) As in A but using the E.coli strains T7-Express or T7-Express-I^q. (C) Growth of Rosetta-II (DE3)-pLysS cultures transformed with the LacI containing pET-LICHIS expression plasmids for the dimeric Lac repressor or the core. OD_{600} is measured at regular intervals, if OD_{600} is above 0.35, IPTG (final concentration 1 mM) or medium (con) is added to the cultures. Note that the OD_{600} is measured in microplates where the path length is not 1 cm. The here obtained plate reader values can be converted by the following empirical relation: OD_{600} (1 cm) = -0.65+4.5×OD_{600}(plate reader). (D) As in C but using E.coli KRX, an E.coli K12 strain containing the T7 RNA polymerase gene under control of the rhamnose repressor. Induction is performed by the addition of 0.1 % rhamnose and or 1 mM IPTG. (E) Co-purification of the LacI gene product with Lac60-333K84M.*



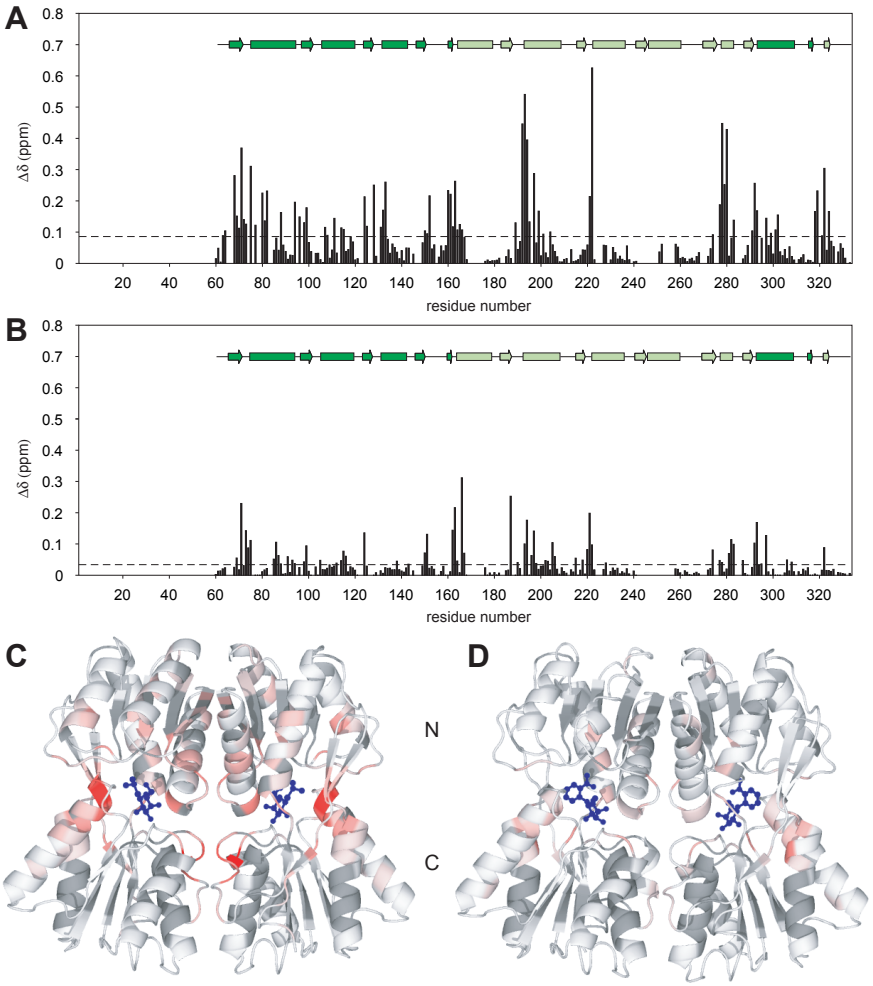
Chapter 3, figure 6. *Melting experiments. (A) Lac60-333K84M T_m dependence on the buffer pH. (B) Lac60-333K84M T_m dependence on the salt concentration. (C) Comparison of the melting temperature for the apo Lac333K84M, Lac60-333K84M and Lac333K84M bound to DNA. (D) Influence of different ligands on the stability of Lac60-333K4M.*



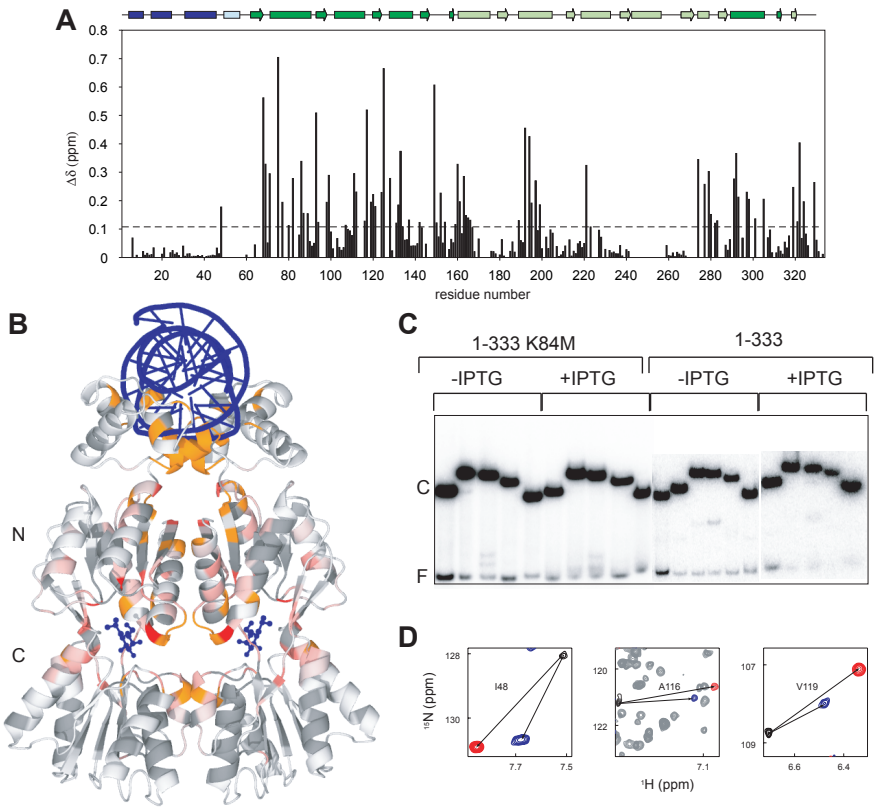
Chapter 4, figure 1. *NMR analysis of the apo form of Lac repressor dimer. (A) Histogram of combined chemical shift changes ($\Delta\delta$) between the isolated domains (HP (residues 1-62) and core domain (residues 60-333)) and the intact Lac repressor dimer. Secondary structure elements of the different modules of the Lac repressor are shown at the top of the histogram: headpiece is shown in dark blue, hinge - light blue, the N-subdomain (residues 62-161 and 293-320) is shown in dark green, C-subdomain (residues 162-289 and 321-329) in light green. (B) $\Delta\delta$ values that are larger than average are mapped onto the model of the free Lac repressor dimer as a gradient of white-red. Residues that were not assigned either due to overlap or incomplete backbone exchange are shown in darker shade of gray. (C, D) Comparison of ^{15}N-TROSY spectra of isolated headpiece (red), isolated core (green) and intact Lac repressor dimer (black). Arrows point in the direction of alpha helix formation.*



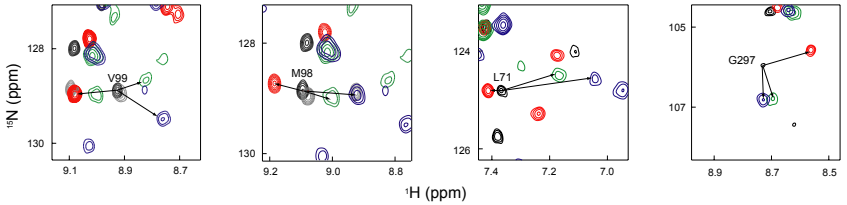
Chapter 4, figure 2. *DNA binding to the Lac repressor dimer. (A) The bar diagram shows the differences in chemical shift for backbone resonances of the free and the protein bound to the idealized symmetric operator (SymL). The hinge region is show in red to emphasize formation of the hinge helix. (B) Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to symmetrical operator and ONPF (1EFA) as a gradient of white-red. (C) Electrophoretic mobility shift assay of Lac repressor binding to the SymL operator, performed in the presence of 0.2 nM wild type Lac dimer or 3.1 nM K84M in the presence or absence of 1, 10, 100, 1000 µM IPTG, added either prior to or after protein-DNA complex formation. The free probe and the protein-DNA complex are indicated with F and C respectively.*



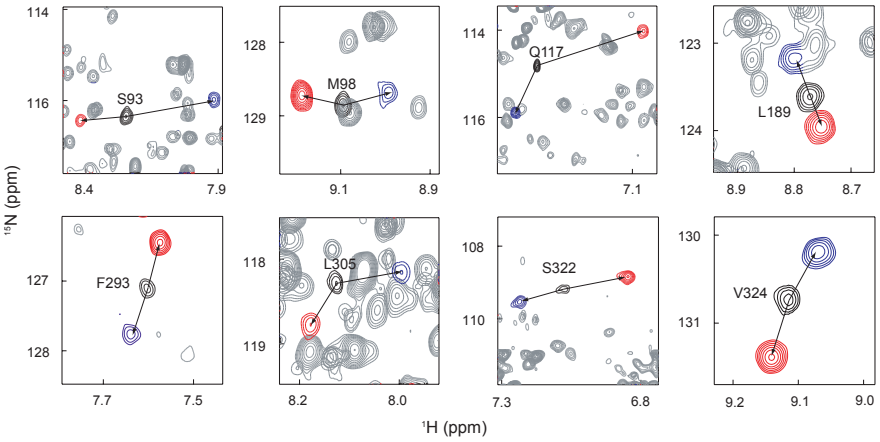
Chapter 4, figure 3. *Lac repressor core dimer binding to inducer IPTG or anti-inducer ONPF. (A, C) The bar diagram shows differences in chemical shift between backbone resonances of the free and IPTG bound Lac repressor core domain. Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to IPTG (2P9H) as a gradient of white-red. The ligands are shown in blue ball-stick representation. (B, D) The bar diagram shows the differences in chemical shift for backbone resonances of the free and ONPF bound Lac repressor core domain. Chemical shift changes are mapped onto the crystal structure of the Lac repressor dimer bound to ONPF (2PAF) as a gradient of white-red. (E) Comparison of* ^{15}*N-TROSY spectra of the free core domain (black), the Lac repressor dimer bound to operator DNA (red), the core domain bound to IPTG (blue) and the core domain bound to ONPF (green).*



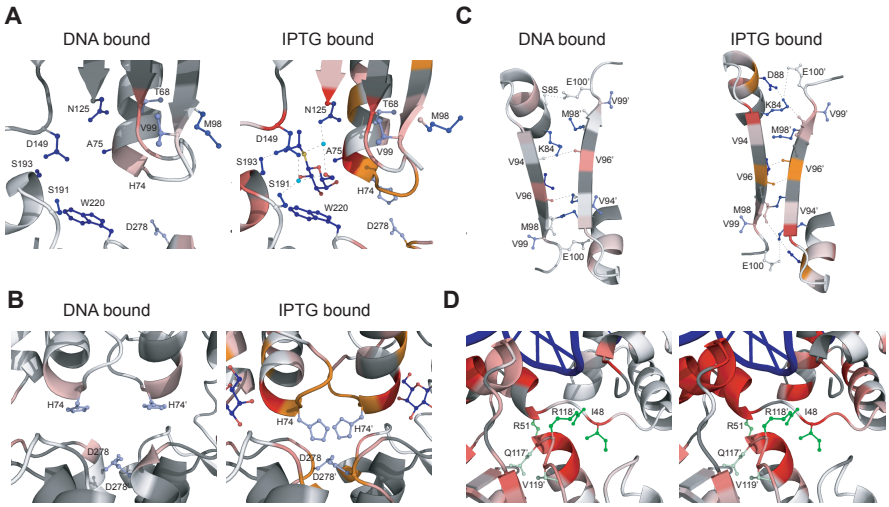
Chapter 4, figure 5. *Ternary complex of the Lac repressor dimer. (A) Chemical shift changes between repressor-SymL and repressor-SymL-IPTG complex. (B) Chemical shift changes plotted on a model for the ternary complex as a gradient of white-red. Residues that were affected but could not be assigned either due to the line broadening or large signal shift are shown in orange. (C) Electrophoretic mobility shift assay using different circularly permuted SymL lac operator fragments, showing bending of the operator DNA in the ternary complex. The free probe and the protein-DNA complex are indicated with F and C respectively. (D) Comparison of* ^{15}*N-TROSY spectra of the free Lac repressor dimer (black), Lac repressor dimer bound to DNA (red) and the ternary complex (blue).*



Chapter 4, figure 4. *Comparison of ^{15}N-TROSY spectra of the free core domain (black), the Lac repressor dimer bound to operator DNA (red), the core domain bound to IPTG (blue) and the core domain bound to ONPF (green).*



Chapter 4, figure 6. *Comparison of three conformations of the Lac repressor dimer. Overlay of ^{15}N-TROSY spectra of the free Lac repressor dimer (black), Lac repressor dimer bound to DNA (red) and the ternary complex (blue).*



Chapter 4, figure 9. *Comparison of the crystallography, mutational and chemical shift data. The chemical shift changes of the backbone amides are plotted on the corresponding structures as a gradient of white-red. The following structures were used: Lac repressor bound to the symmetrical operator (1EFA) and Lac repressor core domain bound to IPTG (2P9H). Side chains of the important residues are shown. The strength of the I^s and I^- phenotypes is shown as gradient of blue and green, respectively. Hydrogen bonds are shown as dashed lines. (A) Chemical shift changes in the inducer binding site: upon DNA binding to the unliganded Lac repressor dimer (left) and upon binding of IPTG to the protein-DNA complex (right). (B) Comparison of interactions near the inducer binding site in the structures of the repressor bound to DNA (left) and IPTG (right). (C) Structural and chemical shift changes at the dimerization interface. (D) Interactions at the interface between the DNA binding domain and N-subdomain of the core. Left panel shows chemical shift differences between the unliganded and DNA bound Lac repressor dimer. Right panel shows chemical shift differences between the isolated headpiece bound to operator and core versus the intact protein bound to DNA. It is clear that the helix-turn-helix motif binds similarly to the isolated headpiece and in the intact dimer. The loop connecting the helical bundle to the hinge and the hinge helix are affected by the presence of the core showing formation of the interdomain contacts.*