Published in final edited form as:

J Mol Biol. 2007 July 20; 370(4): 609–619. doi:10.1016/j.jmb.2007.04.028.

Structural Analysis of Lac Repressor Bound to Allosteric Effectors

Robert Daber, Steven Stayrook, Allison Rosenberg, and Mitchell Lewis*

Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 37th and Hamilton Walk, Philadelphia, Pennsylvania 19104-6059

SUMMARY

The *lac* operon is a model system for understanding how effector molecules regulate transcription and are necessary for allosteric transitions. The crystal structures of the *lac* repressor bound to inducer and anti-inducer molecules provide a model for how these small molecules can modulate repressor function. The structures of the apo repressor and the repressor bound to effector molecules are compared in atomic detail. All effectors examined here bind to the repressor in the same location and are anchored to the repressor through hydrogen bonds to several hydroxyls of the sugar ring. Inducer molecules form a more extensive hydrogen bonding network compared to anti-inducers and neutral effector molecules. The structures of these effector molecules suggest that the O6 hydroxyl on the galactoside is essential for establishing a water mediated hydrogen bonding network that bridges the N-terminal and C-terminal subdomains. The altered hydrogen bonding can account in part for the different structural conformations of the repressor, and is vital for the allosteric transition.

INTRODUCTION

Proteins that perform specific metabolic tasks are often regulated to meet the needs of the organism. In many instances, regulating the flux through a pathway is achieved by adjusting the concentration of an enzyme that controls the rate determining step. Transcriptional regulation is one of the most effective ways to modulate enzyme concentrations. In both prokaryotic and eukaryotic organisms, transcription is frequently controlled by repressors and activators. These molecules either directly or indirectly monitor the accumulation or diminution of a metabolite; they respond like a molecular switch, turning transcription on or off. Effector molecules are the chemical signals that convey the metabolic state of the cell to the genetic machinery.

In *E. coli*, the genes required for lactose utilization are negatively regulated [1]. The *lac* repressor, which is constitutively expressed, binds to an upstream cis-activated operator and consequently blocks transcription of the genes necessary for the cell to utilize lactose as an energy source. In this particular case, the negative regulation is relieved in the presence of a particular effector, allolactose, which binds to the repressor and activates the expression of the genes necessary for lactose metabolism. Understanding how effector molecules alter the properties of the repressor at the molecular level is essential for establishing a detailed understanding of gene regulation. Here we describe the structures of the apo repressor as well

*Corresponding author: email: E-mail: lewis@mail.med.upenn.edu, Phone: 215-898-0949, Fax: 215-573-2503.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

as the repressor bound to three effector molecules, and propose a molecular mechanism for how the effector molecules alter the conformation of the repressor, which in turn attenuates the rate of transcription of the *lac* operon.

The repressor and the operator are the two key macromolecular components of the molecular switch that regulates lactose metabolism. The *lac* repressor is a protein with a modular structure composed of four distinct structural units [2,3]. The N-terminal portion of the sequence (residues 1–49), or ‘headpiece’ domain, contains a canonical helix-turn-helix motif (HTH) that is essential for recognizing the operator. Connecting the headpiece to the body of the repressor is a hinge-helix (residues 50–58). In the absence of DNA, the hinge region is unfolded, allowing residues 1–61 to move freely with respect to the rest of the protein [4]. However, when the operator is present, the hinge-helix becomes ordered and binds to the central portion of the operator in the minor groove [3]. The core of the repressor (residues 62–331) is structurally divided into two subdomains. At the interface of these two domains is the effector binding site that is biologically responsible for monitoring the concentration of a metabolite. The C-terminus of the repressor has an α -helix (residues 340–357) that self-associates into a four-helix bundle, allowing the repressor to form a tetramer.

The repressor binds to an operator that is located between the *lacI* gene and the beginning of the *lacZ* gene and prevents transcription of the structural genes of the operon [1]. The operator sequence is pseudo-symmetric, possessing an approximate dyad axis through the central base pair [5]. Key amino acids located on the HTH motif recognize bases in the major groove of the operator, and, amino acids on the hinge helix interact with the bases in the minor groove of the operator, conferring further specificity of operator binding [3]. This binding to the minor groove of the operator also distorts the conformation of the operator [3].

Effector molecules alter the affinity of the repressor for the operator. The natural inducer of the repressor molecule is allolactose (Figure 1a), an analog of lactose that is created by a side reaction of β -galactosidase [6]. When the inducer is present, the repressor binds less efficiently to the operator, which allows RNA polymerase to recognize its promoter and transcribe the genes necessary for lactose utilization. In an effort to unravel the switching mechanism, Monod *et. al.* and others identified several gratuitous effector molecules that affect the synthesis of β -galactosidase [7]. Some of these compounds were able to mimic the natural inducer allolactose, while others bound but failed to induce, indicating that binding to the repressor was not sufficient to cause the allosteric response [8–11]. Muller-Hill extended the list of effector molecules and classified them as inducers or anti-inducers based upon how they modulated bacterial growth [8], and Riggs performed extensive *in vitro* binding assays in an attempt to understand how a group of chemically similar compounds can produce such distinct phenotypes [17]. A classification scheme arose whereby effectors were defined as inducers when the binding lowered the repressor’s affinity for the operator and as anti-inducers when their binding stabilized the repressor-operator complex. These observations were subsequently quantified by Barkley *et. al.* by measuring the dissociation rates of the repressor-operator complex in the presence of different effectors, as well as the equilibrium association constants for a wide number of effector molecules [12]. A particularly effective gratuitous inducer is 1-isopropyl- β -D-thiogalactoside (IPTG) (Figure 1B). In the presence of IPTG, the *lac* repressor binds to its operator with an affinity that is lowered 1000-fold ($K_d = 10^{-10}$ M) [12]. All other inducers analyzed also destabilize the repressor-operator complex by 1000-fold, suggesting a single mechanism of induction [8]. It was also shown that all inducers bind with higher affinity to the free repressor than to the repressor-operator complex [8]. Anti-inducers show mixed results with some binding more tightly to the free repressor while others bind more tightly to the repressor-operator complex [8]. Anti-inducers also vary in their ability to stabilize the repressor-operator complex. Unlike inducers which all destabilize the repressor-operator complex by 1000-fold, anti-inducers only stabilize the complex 1.7 to five fold [9]. The most

potent anti-inducer, orthonitrophenyl- β -D-fucoside (ONPF) (Figure 1C), is not a naturally occurring molecule and has no known biological function in *E. coli*. There is also a neutral effector molecule, ONPG, which binds to the repressor without altering the stability of the repressor-operator complex, yet produces the same phenotype as anti-inducers [8]. Despite the classification of numerous compounds, the lack of high resolution structural information precluded our understanding of how these effector molecules function. Here we address two fundamental questions: how do these effector molecules alter the ability of the repressor to bind the operator, and what discriminates inducers from anti-inducers?

RESULTS

Previously determined structures demonstrate that the repressor adopts two distinct conformations that correspond to the induced and repressed states [3]. In the presence of the operator and the anti-inducer ONPF, the repressor adopts a structure that is subtly different from the structure of the repressor bound to inducer. It is clear that induction changes the orientation of the N-terminal subdomain relative to the C-terminal subdomain. This structural rearrangement alters both the intramolecular interactions between the N-terminal and C-terminal subdomains of the monomer and the intermolecular interactions between the two N-terminal subdomains [3]. These structures also illustrate that effector molecules bind in a pocket that is located at the interface between the N-terminal and C-terminal subdomains, approximately 40 Å from the operator binding site. Unfortunately, the resolution of these structures was not sufficient to elucidate the molecular mechanism for allosteric regulation. To better understand the allosteric mechanism and the role of effector molecules, we crystallized and solved the structures of a dimeric *lac* repressor in the absence of any effector (apo) as well as in three complexes: the repressor bound to the gratuitous inducer, IPTG; the repressor bound to an anti-inducer ONPF in both the presence and the absence of operator DNA; and the repressor bound to the neutral effector ONPG in the presence of operator DNA.

Structures of the repressor in the induced conformation

To better understand how inducers stabilize the induced conformation, the structure of the repressor was determined in the presence and absence of effector molecules. Our original structures of the repressor in the induced conformation were limited in resolution and it was not possible to accurately position the inducer and perform a detailed analysis of inducer binding or determine the interactions that stabilize the induced conformation. To improve the diffraction, a dimeric repressor was created by inserting a stop codon before the C-terminal α -helix [13]. The dimeric repressor binds to both operator and inducer with the same affinity as the wild-type repressor, but the dimer has fewer conformational degrees of freedom, and consequently is better ordered and diffracts to a higher resolution.

Crystals of the repressor were grown in the presence of the gratuitous inducer, IPTG. Data were collected to the diffraction limit (2.0 Å) and the structure solved as described in the Methods section. An unbiased view of the electron density was obtained by solvent flattening and non-crystallographic symmetry averaging. We also crystallized the dimeric repressor in the absence of ligand so that we could establish the effect of the inducer on the free repressor. The structure of the dimeric apo repressor was solved by molecular replacement using data with *d* spacings greater than 2.5 Å. The structure agrees with that previously reported [14]. A third crystal structure was solved where the repressor was crystallized in the presence of an anti-inducer ONPF. These crystals were isomorphous with the crystals of the repressor bound to the inducer. The phases obtained by molecular replacement were improved and bias removed by symmetry averaging and solvent flattening. A summary of the crystallographic data and refinement statistics are presented in Table 1.

The structure of the repressor bound to IPTG is essentially identical to our previously published structure [13,15]. However, with the higher resolution data, it is clear that the inducer molecule was originally not positioned correctly in the electron density. With improved resolution, the electron density unambiguously demonstrates how the inducer binds to and stabilizes the repressor (Figure 2A). The density reveals a number of specific interactions between the hydroxyls on the sugar and the amino acid side chains of the repressor which establish a detailed hydrogen bonding network. Figure 2B illustrates how the galactoside is anchored to the repressor. All of the oxygen atoms of the galactose ring form hydrogen bonds either by directly interacting with the amino acid sidechains or through water mediated bonding. The O2 and O3 hydroxyls of the effector form direct hydrogen bonds with sidechain residues R197, N246 and D274 of the C-terminal subdomain. The binding is further stabilized by a water mediated hydrogen bond between the O4 hydroxyl of the sugar, the nitrogen of A75, and the sidechain of N246. The O5 ring oxygen atom also participates in hydrogen bonding by forming a water mediated bond to the sidechain of N125. A number of water mediated hydrogen bonds are formed between the O6 hydroxyl of the inducer and the side chains of amino acids S69, D149 and N125, which are located in the N-terminal subdomain of the repressor, and serine residues 191 and 193 of the C-terminal subdomain. The water mediated hydrogen bonds that form through the O6 hydroxyl crosslink the N- and C-terminal subdomains and create an extensive bonding network that is not observed in any of the other structures presented here. Additional interactions are formed by the isopropyl group which nestles into a hydrophobic pocket formed by the sidechains of I79, F161, F293, and L296. These apolar interactions are likely to provide a favorable environment for substituents on the C1 position of the galactoside.

The binding pocket in the apo structure, in contrast, is rather unremarkable and is filled with a few water molecules that form a limited hydrogen bonding network. Although the structures of the apo repressor and repressor-IPTG complex are quite similar, there are modest differences. The overall rms deviation of the main chain atoms is 1.2Å suggesting that the inducer must discreetly alter the structure. This finding is consistent with previous reports indicating that the free repressor undergoes a conformational change upon binding IPTG [16–18]. The disparity in the two structures is primarily localized to the N-terminal subdomain. The rms deviation between these two structures, observed by superposition of the main chain atoms, is less than 0.7Å for the C-terminal subdomains, while the N-terminal subdomains deviate by more than 1.5Å. The divergence between these two structures results in an altered hydrogen bonding pattern, and the repressor forms what appears to be a more stable structure due to the hydrogen bonding network created by the inducer. The hydrogen bonds that stabilize the conformation of the repressor were examined in the absence of water molecules and inducer. The total number of intramolecular hydrogen bonds increases nearly 10% in the repressor-inducer structure. There is virtually no difference in the hydrogen bonding patterns or the number of hydrogen bonds in the C-terminal subdomains. The increase in bonding is localized to the N-terminal subdomain and the interface between the two subdomains. Thus, IPTG either directly or indirectly stabilizes the induced conformation by establishing an extended hydrogen bonding network. Compared to the previously published apo structure, this structure does not contain a glycerol molecule in the binding pocket. Glycerol has been shown to function as a neutral sugar [12], and its absence in the structure presented here is a result of eliminating glycerol as a cryo-protectant for data collection.

In the absence of the operator the structure of the anti-inducer-repressor complex is remarkably similar to the inducer-repressor structure. A superposition of the main chain atoms demonstrates that the conformation of the repressor is essentially identical to the structure of the repressor bound to the inducer. This is consistent with the observation that the crystals of the repressor-anti-inducer complex are virtually isomorphous with the crystals of the inducer-repressor complex. While the quality of these crystals was significantly reduced, as judged by the limit of their diffraction, a number of crystals were examined and the binding of ONPF

was unmistakable in the difference electron density maps. ONPF is bound to the repressor in the same orientation as IPTG with the O2 and O3 hydroxyls of the fucose ring hydrogen bonded to residues R197, N246 and D274. This orientation places the C1 substituent nitrophenyl group in the same hydrophobic pocket as the isopropyl group of IPTG. The electron density for the nitrophenyl substituent is visible although diffuse, suggesting that the ring can adopt multiple conformations and probably does not contribute significantly to the binding. Consistent with this notion, the temperature factors for the nitrophenyl ring are over twice that of the fucose ring. The absence of the O6 hydroxyl on the fucose ring limits the number of hydrogen bonds as compared to that of IPTG binding (Figure 3). Based upon the observation that ONPF binds to the repressor in the same orientation as IPTG and that this orientation is different from previously reported structures [3], we re-examined the structure of the ONPF-repressor-operator complex.

Structure of the repressor in the repressed conformation

In the presence of the operator the repressor adopts a conformation that is significantly different from the apo repressor or the repressor bound to effector molecules. The difference between the induced and the repressed conformation is a consequence of changes in the orientation of the N-terminal subdomain relative to the C-terminal subdomain. This structural rearrangement alters both the intramolecular interactions of the monomer and the intermolecular interactions between the two N-terminal subdomains.

Prompted by the structure of ONPF bound to the repressor in the absence of DNA, we re-examined the ONPF-repressor-operator complex. The structure of the repressor is consistent with what was previously reported, but a more detailed analysis revealed that ONPF was previously modeled into the repressor in the wrong orientation. The structure presented here demonstrates that ONPF binds in the same orientation whether or not operator DNA is present. However, when the repressor is bound to the operator, the density of the nitrophenyl group of the anti-inducer is well ordered. The same hydrogen bonds are seen to anchor the fucoside ring to the repressor, as seen for ONPF in the induced state (Figure 3B). Since the anti-inducer interacts with the repressor in a fashion independent of repressor conformation, it appears that ONPF itself is not responsible for stabilizing the repressed state. Instead, the structures presented here suggest that the operator DNA is responsible for the conformational transition. The structures also demonstrate how ONPF acts as a competitive inhibitor for IPTG binding, since both effectors bind to the repressor in identical fashions. These findings are consistent with data showing the competitive nature of ONPF for IPTG binding [12], as well as with data demonstrating that the binding of ONPF to the free repressor does not change the conformation of the repressor like IPTG does [16–18]. When comparing the hydrogen bonding pattern of the repressor-operator-ONPF complex with the repressor-IPTG complex, we see that both effectors anchor the repressor in identical fashions, but differ in the hydrogen bonding around the C6 substituent. ONPF is both chemically and structurally similar to IPTG, however, as a fucoside it does not have the O6 hydroxyl and consequently cannot form the water mediated hydrogen bonding network that crosslinks the N- and C-terminal subdomains as observed with the inducer. The cross linking of the N- and C-terminal subdomains is therefore not observed in the repressed state.

The neutral effector molecule, ONPG, has chemical properties that are similar to both inducer molecules and anti-inducer molecules. ONPG contains an O6 hydroxyl like IPTG and a nitrophenyl group similar to ONPF (Figure 4A). The structure of this ternary complex was also created by co-crystallizing the repressor with the symmetric operator. The diffraction data were collected and the structure was solved by molecular replacement. The initial phases were refined by symmetry averaging and solvent flattening. Again the crystal quality was adequate, and it was possible to visualize the effector molecule in the binding site although the nitrophenyl

group was sufficiently diffuse. ONPG, like the other effector molecules, binds to the repressor by forming hydrogen bonds between the O2 and O3 hydroxyls on the galactoside and the sidechains of R197, N246 and D274 of the repressor (Figure 4B). When bound to the repressor-operator complex, ONPG does not form the water mediated hydrogen bonding pattern observed with the inducer, suggesting that the water mediated cross linking is important for induction. Moreover, the O6 hydroxyl can crosslink the two subdomains only if the rest of the effector forms stabilizing interactions. Since the nitrophenyl ring is diffuse, we surmise that the fit is not ideal and the O6 hydroxyl is not positioned appropriately to induce the conformational change.

DISCUSSION

The structure of the repressor provides a framework to examine Miller's mutational data [19]. There are several mutant repressors that bind normally to the operator, but do not respond to inducer. These 'super repressors' with the I^s phenotype are incapable of normal induction. While the wild type repressor produces blue colonies on indicator plates that contain IPTG and X-gal, cells that contain mutant 'super repressor' molecules produce white colonies: these repressor molecules either do not bind to the inducer or do not release the operator when the inducer is bound. In both instances these mutant repressors prevent the production of β -galactosidase. The location of these mutations are scattered throughout the linear sequence of the repressor, although there are some regions of the molecule that are more sensitive than others. When the I^s mutations are mapped onto the structure it becomes clear that the mutations that produce the I^s phenotype are localized to the effector binding site or cover the surface of the repressor at the interface of the N-terminal subdomains [20].

The majority of the amino acids that line the effector binding site produce the I^s phenotype when they are mutated. When the phenotype of the 4000 amino acid replacements were tested *in vivo* on plates that contain IPTG, there appeared to be some positions in the binding site that were more critical than others [19]. At some positions amino acid substitutions produce repressor molecules that display the I^s phenotype irrespective of the substitution, while other positions are more tolerant of substitutions and these repressors have a phenotype that is indistinguishable from the wild-type repressor. Figure 5 illustrates that the residues that are most affected by change cluster into the C-terminal subdomain of the effector binding pocket. In contrast, those positions that display a weak I^s phenotype are distributed at the top of the effector binding pocket, corresponding to residues of the N-terminal subdomain. These I^s mutations are consistent with our observation that all inducer molecules are anchored to the repressor through bonds to specific residues that are localized to the C-terminal subdomain.

Mutations in the repressor that produce the I^s phenotype are either directly involved with inducer binding or cluster at the monomer-monomer interface between the N-terminal subdomains, suggesting that the signal is transmitted from the effector site and through the dimer interface to the DNA binding domains. Again, this is consistent with the structural data which show a rearrangement of the N-terminal dimer interface between the induced and repressed conformations. Based upon these structures it is clear that the binding of inducer alters the equilibrium between the induced and the repressed conformation, and this structural rearrangement reduces the repressor's affinity for the operator by destabilizing the repressor-operator complex. Thus, the role of the inducer is to promote a conformational change in the repressor.

In contrast to previously published structures, the higher resolution structures presented here demonstrate that the effectors were previously modeled in the reversed orientation. These structures also show that inducers, anti-inducers and neutral effector molecules all bind to the repressor by creating hydrogen bonds to the sugar ring. In all four structures, the O2 and O3

hydroxyls of the effectors create hydrogen bonds to residues R197, N246, and D274. Removing any one of the sugar hydroxyls from the effector or epimerizing them to alternate stereo chemical configurations prevents the effector from binding [11,12]. Since these hydrogen bonds are common to all three effector molecules, it is improbable that these hydrogen bonds are responsible for the effectors classification as inducers, anti-inducers, or neutral. The primary difference between inducers and anti-inducers are the substituent groups on the first and sixth carbon of the sugar ring.

Inducer molecules differ from anti-inducer or neutral effector molecules with respect to their ability to mediate the intricate hydrogen bonding network that is necessary to stabilize the induced conformation. A distinction between inducers and most anti-inducers is the substituent group on the sixth carbon of the sugar ring. β -galactosides have a hydroxyl group on the sixth carbon. IPTG uses the hydroxyl to form a water-mediated hydrogen bonding network that crosslinks the N- and C-terminal subdomains. This network stabilizes the induced conformation and prevents the repressor from adopting an alternate conformation that is necessary for binding to the operator. The most potent anti-inducer is a fucoside which does not have a hydroxyl group and consequently the same hydrogen bonding pattern is not observed. The importance of the hydroxyl is further supported by comparing the effector function of o-nitrophenyl-1-thio- β -D-fucoside (T-ONPF) with o-nitrophenyl-1-thio- β -D-galactoside (T-ONPG) (Figure 6). The two molecules are essentially identical with the exception of the O6 hydroxyl. Functionally, T-ONPG is an inducer while T-ONPF is an anti-inducer [12]. This mechanism is also consistent with the targeted molecular dynamic studies conducted by Matthews *et. al.* that mapped out the steps the repressor takes during the conformational transition from the induced to repressed state [21]. One of the first steps in the transition is the formation of a hydrogen bond between residue D149 of the N-terminal subdomain and S-193 of the C-terminal subdomain [21]. These two residues are both directly involved in the hydrogen bonding network mediated by the O6 hydroxyl found in IPTG, suggesting that the O6 hydroxyl is critical in aiding one of the first steps in transmitting the allosteric signal through the N-terminal subdomain to the DNA binding domain. Miller also demonstrated that D149 and S193 are critical residues, since almost all mutations at either position resulted in a repressor that could no longer respond to inducer [13].

Although the O6 hydroxyl is necessary for an effector molecule to act as an inducer, it is by no means sufficient. ONPG and several anti-inducers possess the O6 hydroxyl but do not induce, suggesting that substituents at the C1 position, the other position where effectors vary in composition, modulate both the ability of the effector to bind as well as its ability to disrupt the repressor-operator complex. Inducer molecules with different C1 substitutions display a wide range of binding constants that range from $2.4 \times 10^1 \text{M}^{-1}$ to $1.3 \times 10^6 \text{M}^{-1}$. By comparison, the association constant for the natural inducer, allolactose, is $1.7 \times 10^6 \text{M}^{-1}$. The binding efficiency is related to the concentration of the inducer required for half inhibition of the equilibrium binding of the repressor to the operator. Figure 7 is a log-log plot illustrating the linear relationship between the binding affinity of the galactoside and its inducing ability. A potent inducer must have the appropriate chemistry to generate the hydrogen bonding network responsible for induction and sufficient binding energy to stabilize the repressor inducer complex. What differentiates the potency of these inducers is the size of the substituent and the flexibility of its linkage to the galactose ring.

As illustrated in Figure 7, the ability of an effector to induce is directly related to its affinity for the repressor. To a first approximation, the binding affinity is related to the size of the C1 substituent. Galactose without any substituent is a weak inducer and binds to the repressor with relatively low affinity. Substitution of a methyl group increases the binding affinity, as well as its ability to induce. The addition of an isopropyl group increases the binding affinity still further, as does the addition of a butyl group, though larger substituents do not bind as well,

suggesting that there is an optimum substituent size for the repressor binding pocket. Phenyl β -D-galactoside and phenyl β -D-thiogalactoside are anti-inducer molecules, and these larger effector molecules, which possess the C6 hydroxyl, may fail as inducers if the binding of the galactoside is compromised. This can be caused by unfavorable interactions of the C1 substituent with the repressor. The induced conformation can only be stabilized when the galactose ring can be 'anchored' to the repressor in a defined conformation, creating the hydrogen bonding network and crosslinking the N- and C-terminal domains. With each inducer capable of forming the hydrogen bonding crosslink, the differences in potency are expected to be due to the differences in binding affinities.

Effector molecule flexibility is important for efficient binding and function. Altering the connection between the functional group and the galactose ring can drastically alter the properties of the effector. Most of the effector molecules analyzed by Barkley *et al.* were either β -D-galactosides or thio- β -D-galactosides. In nearly all instances the thio linkage resulted in effector molecules that bound more tightly to the repressor. In some circumstances, replacement of the oxygen glycosidic bond with a thio glycosidic bond converted a non-inducing species into an inducer. Nitrophenyl-1- β -D-galactoside (ONPG) is a non-inducer or neutral effector, while nitrophenyl-1-thio- β -D-galactoside (T-ONPG) is a potent inducer. The increased affinity of the thio- β -D-galactosides might be related to an increase in flexibility between the galactoside and nitrophenyl rings. NMR studies have shown that replacement of an oxygen glycosidic bond with a thio glycosidic bond in lactose analogues increases the flexibility of the molecule [22]. The increased flexibility allows these molecules to adopt conformations that increase their binding affinity without jeopardizing the geometry of the galactoside required for stabilizing the induced conformation. Flexibility can also be achieved by the addition of a methylene group between the galactose ring and the C1 substituent group. For example, the anti-inducer phenyl- β -D-galactoside can be transformed into an inducer, benzyl- β -D-galactoside, by the addition of a methylene group. In addition to flexibility, the chemical nature of the substitution affects the ability of the effector molecule to induce.

CONCLUSION

A pattern emerges pertaining to effector ligands of the *lac* repressor when the binding data are considered in the context of the structural studies. For an effector to function as an inducer it must be capable of establishing the water mediated hydrogen bonding network. This complex hydrogen bonding pattern is responsible for stabilizing the induced conformation by bridging the N-terminal subdomain with the C-terminal subdomain. Since the C-terminal domains are relatively fixed for each monomer, this crosslinking leads to a rearrangement of the dimer interface of the N-terminal subdomains and hence an altered binding affinity to the operator. The structure becomes more compact, consistent with the observation that the sedimentation coefficient of the repressor increases about 3% in the presence of saturating concentrations of inducer. All functional inducers are galactosides, suggesting that the O6 hydroxyl is critical for induction. Consequently, fucosides, which do not possess the hydroxyl group, cannot induce. While this hydroxyl is a necessary prerequisite for induction, it is not sufficient: the functional groups attached to the C1 position on the galactoside are also important and provide additional binding energy to stabilize the complex. Moreover, if the substituent on the C1 position does not provide optimal binding, then a galactoside will fail to induce. The size and the chemical nature of the functional group is critical, illustrated by the potency of isopropyl- β -D-thiogalactoside compared to either methyl- β -D-thiogalactoside or phenyl- β -D-thiogalactoside. The natural inducer, allolactose, contains the O6 hydroxyl and is expected to induce by the same mechanism as IPTG.

Structures of the repressor bound to effector molecules provide some insight into the requirements for effector molecules; however, the observations described here only provide

an initial platform for understanding the vital components for functional inducers and anti-inducers. In essence, both specificity and flexibility are necessary for creating an effector molecule that can stabilize the induced conformation of the repressor. Clearly, a structure of the repressor-IPTG-operator complex is crucial element that is still missing: the repressor bound to inducer must associate with the DNA in some fashion, but we surmise that the complex is tenuous and non-specific, as our repeated attempts to crystallize this complex have yielded no results.

Structures in the absence of operator, whether bound to inducer or anti-inducer, maintain the same repressor conformation that we observed in the absence of an effector molecule. The interaction energy derived from the binding of the repressor to its operator is clearly sufficient to drive the repressor towards the alternate or repressed conformation. It is therefore likely that the inducer molecule provides opposing interaction energy by establishing the hydrogen bonding network that is sufficient to overcome the energy gained by binding to the DNA.

METHODS

Protein expression, purification, and crystallization

The repressor protein was produced using a T7 polymerase-driven expression system. We cloned the first 331 residues of the *lac* repressor into this expression system in order to eliminate the tetramerization domain and produce a functional dimeric repressor. The protein was expressed with an N-terminal 6-His tag, and was initially purified by nickel chromatography (Qiagen). The eluted protein was dialyzed and the histidine tag was cleaved by incubating the repressor with TEV protease (Life Technologies) overnight. The repressor was further purified by placing the dialyzed sample back on a washed nickel column and collecting the flow through. The protein was dialyzed into 0.2 M KCl, 0.2 M Tris (pH 7.4), 1 mM EDTA, and 0.3 mM DTT and concentrated to $\sim 15 \text{ mg ml}^{-1}$. The co-crystal structure used a self-complementary 21 nucleotide “ideal” *lac* operator, with the sequence 5'-GAATTGTGAGCGCTCACAATT-3'. The oligonucleotides were purchased from IDT and HPLC-purified. After annealing, this perfectly symmetric operator is 20 base pairs long and has a 5'-G overhang.

The repressor bound to both inducer and anti-inducer was crystallized by hanging drop vapor diffusion from solutions of 1.3–1.5 M NaAcetate, 0.1 M NaCitrate, pH = 5.6. Crystals grew over a period of several weeks to approximately $0.3 \times 0.3 \times 0.7 \text{ mm}$. The apo crystals belong to space group $I4_122$ with one monomer per asymmetric unit and a solvent content of 50%. All of the repressor-effector complexes crystallized under roughly the same conditions and in a related space group $P4_12_12$ (Table 1). The repressor-ONPG complex was crystallized by mixing 10 mg ml^{-1} *lac* dimer with a 1.3-fold molar excess of DNA and a 10-fold excess of ONPG (Sigma). Crystals were grown in 10% polyethylene glycol 400, 2.0 M ammonium sulfate, 0.1 M hepes (pH 7.5), 14% glycerol, and belong to space group R32 with cell dimensions $a = 251.4 \text{ \AA}$, $c = 204.8 \text{ \AA}$. There are 1.5 dimeric repressor/DNA/ONPF complexes in the asymmetric unit and the solvent content is 75%.

X-ray data collection and structure determination

X-ray data were collected to their diffraction limit and were processed using DENZO [23]. The structure was determined by molecular replacement using AmoRe [24]. The initial phases were improved by non-crystallographic symmetry (NCS) averaging and solvent flattening using DPHASE (G. Van Duyne, personal communication). The resulting electron density maps were fit using COOT [25]. The structure was refined using CNS [26]. Refinement included simulated annealing, restrained isotropic temperature factor refinement, a bulk solvent

correction, and monitoring of the free R-factor with 5% of the data omitted. NCS restraints were imposed. Figures were prepared using Molscript [27], Coot[25], and Pymol [28].

Acknowledgments

This work was supported by a grant NIH GM44617 (M.L.) The Protein Data Bank accession codes for the atomic coordinates are: 2PE5 the Lac Repressor bound to ONPG in repressed state, 2PAF the structure of the Lactose Repressor bound to anti-inducer ONPF in induced state, and 2P9H the structure of the Lactose Repressor bound to IPTG

References

1. Jacob F, Monod J. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J Mol Biol* 1961;3:318–356. [PubMed: 13718526]
2. Friedman AM, Fischmann TO, Steitz TA. Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* 1995;268:1721–1727. [PubMed: 7792597]
3. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer.[see comment]. *Science* 1996;271:1247–1254. [PubMed: 8638105]
4. Wade-Jardetzky N, Bray RP, Conover WW, Jardetzky O, Geisler N, Weber K. Differential mobility of the N-terminal headpiece in the lac-repressor protein. *Journal of Molecular Biology* 1979;128:259–264. [PubMed: 372550]
5. Gilbert W, Maxam A. The Nucleotide Sequence of the lac Operator. *Proceedings of the National Academy of Sciences* 1973;70:3581–3584.
6. Jobe A, Bourgeois S. lac Repressor-operator interaction. VI. The natural inducer of the lac operon. *Journal of Molecular Biology* 1972;69:397–408. [PubMed: 4562709]
7. Monod J, Cohen-Bazire G, Cohn M. *Biochim Biophys Acta* 1951;7:585. [PubMed: 14904456]
8. Muller-Hill B, Rickenberg HV, Wallenfels K. Specificity of the Induction of the Enzymes of the Lac Operon in *Escherichia coli*. *J Molec Biol* 1964;10:303–318. [PubMed: 14235589]
9. Jobe A, Bourgeois S. Lac repressor-operator interaction. VIII. Lactose is an anti-inducer of the lac operon. *J Molec Biol* 1973;75:303–313. [PubMed: 4580678]
10. Jarrett HW. Temperature dependence of DNA affinity chromatography of transcription factors. *Analytical Biochemistry* 2000;279:209–217. [PubMed: 10706790]
11. Riggs AD, Newby RF, Bourgeois S. lac Repressor-Operator Interaction II. Effect of Galactosides and Other Ligands. *J Mol Biol* 1970;51:303–314. [PubMed: 4320936]
12. Barkley MD, Riggs AD, Jobe A, Bourgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry* 1975;14:1700–1712. [PubMed: 235964]
13. Bell CE, Lewis M. The Lac repressor: a second generation of structural and functional studies. *Current Opinion in Structural Biology* 2001;11:19–25. [PubMed: 11179887]
14. Bell CE, Barry J, Matthews KS, Lewis M. Structure of a variant of lac repressor with increased thermostability and decreased affinity for operator. *Journal of Molecular Biology* 2001;313:99–109. [PubMed: 11601849]
15. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. *Nature Structural Biology* 2000;7:209–214.
16. Laiken SL, Gross CA, Von Hippel PH. Equilibrium and kinetic studies of *Escherichia coli* lac repressor-inducer interactions. *Journal of Molecular Biology* 1972;66:143–155. [PubMed: 4557195]
17. Oshima Y, Matsuura M, Horiuchi T. Conformational change of the lac repressor induced with the inducer. *Biochemical & Biophysical Research Communications* 1972;47:1444–1450. [PubMed: 4557176]
18. Matthews HR, Thielmann HW, Matthews KS, Jardetzky O. NMR studies of the binding of an inducer and an anti-inducer to the lac repressor. *Annals of the New York Academy of Sciences* 1973;222:226–229. [PubMed: 4522430]
19. Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH. 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. *Journal of Molecular Biology* 1994;240:421–433. [PubMed: 8046748]
20. Pace HC, Kercher MA, Lu P, Markiewicz P, Miller JH, Chang G, Lewis M. Lac repressor genetic map in real space. *Trends in Biochemical Sciences* 1997;22:334–339. [PubMed: 9301333]
 21. Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Science* 2003;12:2523–2541. [PubMed: 14573864]
 22. Montero E, Garcia-Herrero A, Asensio KH, Ogawa JS, Santoyo-Gonzalez F, Canada F, Jimenez-Barbero J. The conformational Behavior of Non-Hydrolyzable Lactose Analogues: The Thioglycoside, Carbaglycoside, and Carba-Iminoglycoside Cases. *Eur J Org Chem* 2000:1945–1952.
 23. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. *Meth Enzymol* 1997;276:307–326.
 24. Navaza J. AMoRe: an Automated Package for Molecular Replacement. *Acta Cryst* 1994;D50:157–163.
 25. Emsley PCK. Coot: model-building tools for molecular graphics. *ACTA CRYSTALLOGRAPHICA SECTION D-BIOLOGICAL CRYSTALLOGRAPHY* 2004;60:2126–2132.
 26. Brunger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 1992;355:472–475. [PubMed: 18481394]
 27. Kraulis P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991;24:946–950.
 28. DeLano, WL. The PyMOL Molecular Graphics System, DeLano Scientific. San Carlos; CA, USA:

Figure 1.
Structure of the effector molecules (A) is the natural inducer allolactose. (B) gratuitous inducer IPTG and (C) is the anti-inducer ONPF.

Figure 2.

Inducer binding site. (A) Illustrates the fit of IPTG to the difference electron density where the phases were calculated from the atomic model of the repressor. The inducer and the molecules were omitted from the structure factor calculations. The difference electron density was contoured at 3.5σ . (B) Illustrates the binding of the inducer to the repressor. The view is rotated $\sim 90^\circ$ from 2A to better illustrate the detailed hydrogen bonding network. The dark blue portions of the structure correspond to residues in the C-terminal domain while the light blue corresponds to the N-terminal portion of the structure. The inducer and the water mediate hydrogen bonds stabilize this conformation of the repressor. The green dashed lines illustrate the water mediated hydrogen bonding network that links the N-terminal and C-terminal subdomains

Figure 3. Illustrates the binding of the anti-inducer, ONPF, the repressor in the absence and the presence of the operator. (A) In the presence of DNA, the anti-inducer forms a ternary complex with the repressor primarily by establishing hydrogen bonds between the O2 and O3 hydroxyls of the fucoside and residues R197, N246 and D274 of the repressor and the nitrophenyl group hydrogen bonds to N146. (B) In absence of DNA, the anti-inducer is also bound to the repressor by hydrogen bonding to the fucoside but the nitrophenyl group does not appear to be ordered or adopt the same conformation.

A

B

Figure 4.

Three related nitrophenyl galctosides. (A) The top figure, nitrophenyl-1-B-D-galactoside (ONPG) is a non-inducer or neutral effector. In the middle figure the oxygen glycosidic bond has been replaced with a thio glycosidic bond, which converted this non-inducing species into an inducer nitrophenyl-1-thio-B-D-galactoside (T-ONPG). The bottom image illustrates the importance of the hydroxyl the effector function o-nitrophenyl-1-thio-B-D-fucoside (T-ONPF) is an anti-inducer,

Figure 5.

Illustrates that mutations of the amino acids that line the effector binding site that produce the I^s phenotype. The most effected amino acids cluster into the C-terminal subdomain of the binding pocket (grey). The I^s mutants that are more tolerant are distributed at the top of the binding pocket and reside in the N-terminal subdomain (yellow). These observations are consistent with our observation that all inducer molecules are anchored to the repressor through by specific residues that are localized to the C-terminal subdomain.

Figure 6.

Illustrates that the binding affinity of inducers is related to their ability to induce the repressor from the operator. The equilibrium binding data measured by Bourgeois [8] are graphed on a log-log plot. The value of kappa defined by Bourgeois [8] is a measure of induction. For these galactosides there is a direct correlation between binding affinity and induction.

Figure 7.

The schematic representation illustrates that the repressor exists in two distinct conformations corresponding to the induced and the repressed states. Binding of the inducer alters the conformation of the repressor. Direct and water mediated hydrogen bonding bridge the N- and C-terminal sub domains, which in turn stabilizes a conformation that has a reduced affinity for the operator

Table 1Crystallographic data and refinement statistics¹

	apo	IPTG	ONPF	ONPG
Resolution (Å)	45.0–2.5	20.0–2.0	17.0–3.5	20.0–3.5
Space Group	I 41 2 2	P 41 21 2	P 41 21 2	R 32
Cell Dimensions	a= 91.7 c =222.3	a= 77.5 c =211.58	a= 77.54 c =212.45	a= 253.1 c =203.1
# Observations	64,464	392,952	42,614	283,645
# Unique (I > σ1)	9,861 (99.2%)	36,802(83.2%)	8685(83.2%)	31,033
Redundancy	6.8	10.7	4.9	9.1
I/σ	12.1 (7.3) ²	16.3 (3.3) ²	6.1 (3.3) ²	7.5 (4.1)
Completeness (%)	99.4 (99.0)	83.5 (47.3)	98.3 (99.4)	99.5 (99.4)
R _{merge} ³ (%)	13.4 (36.2)	13.6 (26.4)	16.6 (33.2)	17.4 (45.4)
Refinement statistics				
# Protein atoms	2027	4060	4070	8573
# Waters	6	710	25	70
R-factor ⁴	18.7 (33.2)	17.17 (33.2)	19.8 (31.7)	23.4 (26.2)
Free R-factor	23.4 (33.8)	23.8 (33.8)	28.4 (37.8)	28.6 (33.8)
Rmsd bonds (Å)	0.02	0.017	0.007	0.05
Rmsd angles (°)	3.1	1.7	1.9	4.1

¹ all data were collected on a Rigaku generator and image data collected with a Mar Image plates² Numbers in parentheses refer to the highest resolution shell only.³ R_{merge} = $\sum I(I_h - \langle I \rangle_h) / \sum I_h$, where $\langle I \rangle_h$ is the average intensity over symmetry equivalents.