Published in final edited form as:
Biochemistry. 2008 August 5; 47(31): 8058–8069. doi:10.1021/bi800443k.

Subdividing repressor function: DNA binding affinity, selectivity, and allostery can be altered by amino acid substitution of nonconserved residues in a LacI/GalR homologue†

Hongli Zhan‡, **Marc Taraban**§, **Jill Trewhella**§,‖, and **Liskin Swint-Kruse**‡,*

‡*Department of Biochemistry and Molecular Biology, MSN 3030, 3901 Rainbow Blvd., The University of Kansas Medical Center, Kansas City, KS 66160*

§*Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA*

‖*School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia*

Abstract

Many mutations that impact protein function occur at residues that do not directly contact ligand. To understand the functional contributions from the sequence that links the DNA-binding and regulatory domains of the LacI/GalR homologues, we have created a chimeric protein (LLhP), which comprises the LacI DNA-binding domain, the LacI linker, and the PurR regulatory domain. Although DNA binding site residues are identical in LLhP and LacI, thermodynamic measurements of DNA binding affinity show that LLhP does not discriminate between alternative DNA ligands as well as LacI. In addition, small-angle scattering experiments show that LLhP is more compact than LacI: Upon DNA release, LacI shows a 20Å increase in length that was previously attributed to unfolding the linker. This change is not seen in apo-LLhP, even though the linker sequences of the two proteins are identical. Together, results indicate that long-range functional and structural changes are propagated across the interface that forms between the linker and regulatory domain. These changes could be mediated *via* the side chains of several linker residues that contact the regulatory domains of the naturally-occurring proteins, LacI and PurR. Substitution of these residues in LLhP leads to a range of functional effects. Four variants show altered affinity for DNA, with no changes in selectivity or allosteric response. Another two result in proteins that bind operator DNA with very low affinity and no allosteric response, similar to LacI binding nonspecific DNA sequences. Two more substitutions simultaneously diminish affinity, enhance allostery, and profoundly alter DNA ligand selectivity. Thus, positions within the linker can be varied to modulate different aspects of repressor function.

Keywords

lactose repressor protein; purine repressor protein; DNA-binding affinity; allostery; specificity; selectivity; domain linker; protein engineering

In proteins, many amino acid polymorphisms are neither severe nor silent. Instead, protein function is modified, with biological outcomes that are important for both evolution and protein

†Support to LSK was provided by NIH Grant P20 RR17708 from the Institutional Development Award program of the National Center for Research Resources and GM079423. Support to JT was provided by an ARC Federation Fellowship (FF0457488) and the scattering experiments were performed at the University of Utah supported by U. S. Department of Energy, Grant No. DE-FG02−05ER64026

*To whom correspondence should be addressed: lswint-kruse@kumc.edu 913−588−0399 phone 913−588−7440 fax.
Current addresses. HLZ: Department of Biochemistry and Cell Biology, MS 140, Rice University, Houston, TX 77005. MT: Institute of Chemical Kinetics and Combustion, Novosibirsk 630090, Russia.

engineering. Functionally significant polymorphisms may occur at nonconserved positions, which are often referred to as "specificity determinants" (*e.g.* (1,2)). Furthermore, functionally important positions are not limited to ligand binding sites (*e.g.* (3)). Because linker sequences can impact any aspect of function that requires two or more functional domains, linkers are candidate regions for containing specificity determinants.

At least four positions act as specificity determinants in the linkers of the LacI/GalR homologues (4). To study functional contributions from these residues, we recently engineered a novel transcription repressor (LLhP) using the LacI[a] linker sequence to join the LacI DNA-binding domain and the PurR regulatory domain (Figure 1A,B) (4). LacI and PurR are two naturally-occurring homologues in the LacI/GalR family of bacterial transcription repressors (5). These homodimers bind to specific "operator" sites on DNA upstream from regulated genes. DNA-binding affinity is modulated when small-molecule effectors bind each regulatory domain (two per dimer) and facilitate a structural change. LacI has high affinity for operator DNA, and binding the effector IPTG diminishes DNA-binding (6). PurR has an opposite allosteric mode; binding either guanine or hypoxanthine (HX) enhances DNA-binding (7).

The chimera LLhP is a strong *in vivo* repressor for the LacI DNA binding site *lacO*^{1}. Like PurR, LLhP repression is enhanced by the addition of hypoxanthine. However, LLhP exhibits moderate toxicity in *E. coli* not seen for LacI (4). One possibility is that the DNA binding is enhanced for additional operator-like sequences – perhaps the chimera acquires affinity for genomic sites that adversely affect bacterial growth. Previous structure/function comparisons (4,8) led us to hypothesize that the toxic phenotype might arise from changing interactions between the LacI linker sites 48, 55, 58, and 61 (Figure 1B) and the PurR regulatory domain. These linker sites are specificity determinants: (a) They are not conserved across the LacI/GalR family (and indeed differ between LacI and PurR); and (b) They contribute to LLhP function, as shown by altered *in vivo* repression when amino acid residues were substituted (4).

The *in vivo* LLhP experiments only monitored repression from *lacO*^{1} and did not ascertain the precise functional changes that arose from domain recombination or amino acid substitution of specificity determinants. Among other variables, changes in function may result from alterations to operator DNA affinity, DNA selectivity, or allosteric response to a small molecule effector. Here, we report results for purified LLhP and 8 mutational variants. Variants were characterized by determining the binding affinities for three different DNA ligands in the absence and presence of hypoxanthine. Our goals were (a) to compare the LLhP DNA-binding function and conformational changes to that of LacI; and (b) to discriminate which aspect of function (affinity, allostery, or DNA-selectivity) is affected by substitution at nonconserved linker residues.

Our results show that "wild-type" LLhP exhibits greater promiscuity in binding alternative DNA sequences than does LacI. LLhP also exhibits a smaller allosteric response than LacI. Both the changed DNA recognition and diminished allostery correlate with the loss of conformational flexibility observed in LLhP by small-angle X-ray solution scattering. Mutations in the LLhP linker affect various aspects of function: Amino acid substitutions at positions 48 and 55 alter DNA affinity, but DNA selectivity and allostery are similar to unmodified LLhP. Two variants at position 58 bind operator DNA very weakly and without allosteric response, which is very similar to LacI binding to nonspecific DNA. Finally, amino acid variation at position 61 results in complex changes in DNA selectivity, affinity, and allosteric response to effector. Thus, amino acid polymorphisms in the LLhP linker specificity determinants impact a range of functional aspects.

[a]Abbreviations: LacI, lactose repressor protein; PurR, purine repressor protein; IPTG, isopropyl-β-D-thiogalactoside; HX, hypoxanthine.

Materials and Methods

Purification of LLhP variants

Design and construction of the LLhP coding sequence is reported in (4). The genes for LLhP variants are under the constitutive *lacI*^{q} promoter; protein expression could be visualized in the lysis supernatant of DH5α and 3.300 *E. coli* upon Coomassie staining of SDS-PAGE. *E. coli* cell strains normally used to express LacI for purification were not suitable for LLhP: The protein would not express in TB1 cells and was toxic to BLIM cells. DH5α expressed the most LLhP protein, but these cells have low levels of endogenous LacI that might co-purify with LLhP. However, thermodynamic parameters are the same for LLhP purified from either DH5α or 3.300 cells, which indicates that contamination is not a problem. In addition, the thermodynamic properties of LLhP variants differ significantly from those of LacI. The protocol for growing *E. coli* 3.300 and DH5α for protein purification was modified to minimize the number of bacterial generations and avoid the epigenetic shutdown of LLhP expression seen previously (4). All growths utilized cells from a fresh transformation. Instead of plating the cells, they were used to directly inoculate a 500 mL LB culture contained in 2 L flasks. Cells were grown at 37°C for approximately 24 hours prior to centrifugation at 11,000xg. Cell paste was resuspended in breaking buffer (5% glycerol, 12 mM Hepes-KOH, pH 7.6, 50 mM KCl, 1 mM EDTA, and 0.3 mM DTT), with 25 mg of lysozyme for cells grown in 6 L of media, and stored at −20°C.

The protocol for LLhP purification was based upon that for LacI (*e.g.* (9)), modified to use buffers in which PurR maintains high activity (*e.g.*(10)). For purification, cell paste from 3 L of culture was incubated on ice for ~15 mins. Additional breaking buffer was added to bring the final volume to 200 ml, with fresh DTT at 0.3 mM. Following digestion of genomic DNA by DNaseI, the crude cell extract was centrifuged at 7,700xg for 50 min. Supernatant was subjected to ammonium sulfate precipitation (37% saturation). After incubation for 45 min at 4°C, the precipitate was collected by centrifugation at 7,700xg for 40 min and resuspended in 35 ml of breaking buffer. Resuspended precipitate was dialyzed 20−30 minutes against 1 L of breaking buffer for each of three buffer exchanges. A subsequent 30 minute centrifugation at 7,700xg precipitated insoluble protein. Supernatant was loaded on a phosphate cellulose column (~30 ml; Whatman P11; Fisher Chemical Co.) that was pre-equilibrated with breaking buffer. LLhP was eluted from the column in gradient buffer (5% glycerol, 12 mM Hepes-KOH, 1 mM EDTA, and 0.3 mM DTT) using a gradient of 50 mM KCl, pH 7.6 to 400 mM KCl, pH 8.0, followed by extensive washing with 50 mM KCl gradient buffer. LLhP eluted at approximately 250−300 mM KCl. By SDS-PAGE, LLhP was >90% pure at this step. The protein could be aliquoted and frozen at −80 °C.

For high protein concentrations, the volume of LLhP eluted from the phosphocellulose column was decreased by concentration in a VIVASPIN 20 concentrator (MWCO: 10,000; Vivascience; Stonehouse, UK) so that it could be loaded into the 2 mL injection loop of an Amersham FPLC. LLhP protein was further purified by size exclusion chromatography using an S200 sizing column (Amersham Biosciences, Piscataway, NJ) in 12 mM Hepes-KOH, 200 mM KCl, 5% glycerol, 1 mM EDTA, and 0.3 mM DTT. The flow rate was 0.5 to 1 ml/min. At this point, the protein could be aliquoted and frozen without diminishing activity. The maximum concentration for maintaining activity upon freezing is 1 mg/mL. Instead of or after freezing, higher concentrations of LLhP (up to 18 mg/mL) were achieved by concentration in a VIVASPIN 20 concentrator (MWCO: 10,000), which was centrifuged at 2,600xg at room temperature. In this step, temperature is very important, since highly concentrated LLhP protein precipitates at colder temperatures similar to the naturally-occurring homologue PurR (11).

With the LLhP variants, we substituted a heparin column (GE Healthcare) for the S200 sizing column, which increased both purity and DNA binding activity. For this column, proteins were

dialyzed into buffer A (~12 mM Hepes, pH 7.6, 50 mM KCl, 5% glycerol, 1 mM EDTA, and 0.3 mM DTT) with three buffer exchanges. The LLhP-containing fraction from the phosphocellulose column was concentrated to about 2 ml using a VIVA spin 20 concentrator and loaded onto a 5-ml heparin column that was previously equilibrated with buffer A. After washing with 10 column volumes of 88% buffer A/12% buffer B (12 mM Hepes, pH 7.6, 500 mM KCl, 5% glycerol, 1 mM EDTA, and 0.3 mM DTT), LLhP was eluted using a 20 column volume gradient of 12% to 50% buffer B. Inactive LLhP eluted during the wash step; the active fraction of LLhP variants eluted near the beginning of the gradient.

Mass spectrometry was used to confirm that LLhP is the purified protein. Analysis was carried by Dr. Antonio Artigues at KUMC using ESI FTICR mass spectrometer (LTQ FT, ThermoFinnigan; Waltham, MA). LLhP samples required extensive desalting in 0.1% formic acid (5 dilutions with water/formic acid) to obtain a reasonable mass value. The expected molecular weight of an LLhP monomer is 38351.64 Daltons; measured mass values were 38359.8 and 38359.4. The expected molecular weight of a LacI monomer is 38590 for the 109A polymorphism (12) or 38620 for 109T. Since the LLhP gene was fully sequenced and showed no unexpected mutations, the difference in the expected and measured mass values may be due to residual contamination of bound salts; the DNA-binding domain of LacI is very highly charged and known to bind some anions extremely tightly (Swint-Kruse and Zhan, unpublished observations).

Extinction coefficient for LLhP

The extinction coefficient of LLhP was determined by magnetic circular dichroism (MCD) spectropolarimetery (13), using a Jasco J-500C spectropolarimeter attached to a 1.13 T electromagnet. The concentration of tryptophan in the protein sample was calculated from the MCD value and correlated with LLhP absorbance at 280 nm. As expected, the extinction coefficient of LLhP was found to be $6.90 \times 10^4\ M^{-1}cm^{-1}$, the same as that calculated for PurR, which is consistent with the fact that all tryptophan residues reside in the regulatory domain of LLhP. The same extinction co-efficient was utilized for all LLhP variants.

Preparation of hypoxanthine and guanine

Hypoxanthine (HX) was dissolved in filter binding buffer (10 mM Tris-HCl, 150 mM KCl, 5% DMSO, 1 mM EDTA, and 0.3 mM DTT) at a stock concentration of 4 mM with gentle heating (40 °C). Guanine could not be dissolved in the same buffer, even with shaking at 37° C overnight. Instead, guanine was dissolved in 1 M KOH to a final concentration of 40 mM. The stock solution was diluted 1:2000 during the subsequent experiment, which resulted in a final KOH concentration of 5×10^{-4} M. Since the "wild-type" LLhP *lacO*^{1}-binding affinities were apparently identical for the two co-repressors, we continued with only hypoxanthine for experiments with LLhP and DNA variants.

Thermodynamic characterization of DNA-binding affinities

Nitrocellulose filter binding assays were employed to measure the activity and DNA-binding affinity of LLhP (14,15). These experiments utilized three different operator DNA sequences (Table 1: *lacO*^{1}, *lacO*^{sym}, and *lacO*^{disc}) (16-19). Double-stranded DNA was annealed from single-stranded oligonucleotides (Integrated DNA Technology, Coralville, IA) in polynucleotide kinase buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6). The resulting 40-mer DNA was labeled with [^{32}P]-ATP using polynucleotide kinase; a Nick column (Amersham Biosciences, Uppsala, Sweden) was used to remove unincorporated free nucleotide.

In activity assays to determine the percent of LLhP capable of binding DNA ligand, operator DNA concentration was at least 10-fold higher than the K_d (20); 3 to 5×10^{-8} M was typically

used. Activities were determined for LLhP, I48S, I48V, Q55T, and Q55S in the absence of hypoxanthine using *lacO*^{1}. In order to reach experimentally accessible conditions for S61C and S61M, activities were determined in the presence of hypoxanthine. Independent, duplicate purifications of these seven LLhP variants had activities that ranged from 85 to 99%. At least one high activity preparation (>95%) was obtained for each variant. Activities could not be ascertained for G58L and G58T, since stoichiometric conditions are not accessible using the filter binding assay. Evidence that these proteins are folded is derived from the partial DNA-binding curves observed in affinity assays.

In samples for either activity or affinity DNA-binding measurements, co-repressor was always added to protein after DNA to maintain activity above 90%. Adding hypoxanthine to the protein prior to DNA reduced activity of LLhP to ~60%. We surmise that binding co-repressor in the absence of DNA causes LLhP to aggregate at the higher protein concentrations needed for activity assays. Although the experimental design had no effect on determined K_d values, we always employed conditions with the highest activity. After rapid filtration through nitrocellulose filter paper (Schleicher and Schuell, Keene, NH, and Whatman, Sanford, ME), the amount of radioactively-labeled DNA-protein complex was quantitated using a Fuji phosphorimager. Some LLhP affinity measurements were made in parallel with LacI binding measurements, using the same nitrocellulose filter and sample of radio-labeled DNA. If present, LacI inducer IPTG concentration was 1 mM.

In assays to quantitate the DNA-binding affinities, the concentration of DNA ligand was at least 10-fold below the K_d and protein concentration was varied (20). DNA-binding affinity was measured in the absence and presence of $2–4\times10^{-5}$ M co-repressor, which was either hypoxanthine or guanine. Although the hypoxanthine concentration was only two-fold higher than the affinity of PurR for hypoxanthine (21,22), operator capture experiments indicate that changes in DNA-binding affinity induced by co-repressor binding are complete for LLhP, I48S, I48V, Q55T, and Q55V (see below). Since S61M and S61C had higher midpoints in operator capture experiments (see below), additional measurements of DNA affinity were performed for these variants using 4×10^{-4} M hypoxanthine. The two hypoxanthine concentrations showed no significant difference in the affinities for DNA.

The program Igor Pro (Wavemetrics, CA) was used to fit data for affinity assays using the equation:

$$Y_{obs} = \left(Y_{max} * \frac{[\mathrm{Prot}]^n}{K_\mathrm{d}^n + [\mathrm{Prot}]^n} \right) + c \quad (1)$$

where Y_{obs} and Y_{max} are the radioactivity retained at a specific protein concentration and the measured radioactivity when all of the DNA is bound to repressor, respectively, K_d is the equilibrium dissociation constant, and c is the background radioactivity in the absence of protein. The value of the Hill coefficient, n, was either fixed at 1 or allowed to float, in which case the values were ~1. Because the activity of the protein is verified to be generally >90% and K_d is determined three independent times using protein from 2 separate preparations, we are able to detect three to five-fold changes between the averaged K_d values (*i.e.* Table 2). When complete binding curves could not be obtained for reliable K_d determination (*e.g.* Supplementary Figure 3), experiments were performed on the same filter and with the same dilution of radio-labeled DNA, in order to allow direct comparison of the quantity of bound DNA.

Small-angle X-ray solution scattering and construction of LLhP models

X-ray scattering data were collected at 11°C using the small-angle instrument at the University of Utah (described in (23)); scattering data were reduced to $I(Q)$ versus Q and analyzed as

previously described (23). $I(Q)$ is the scattered X-ray intensity per unit solid angle and Q is the amplitude of the scattering vector, given by $4\pi(\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the scattered X-rays (1.54 Å).

The $P(r)$ analyses were performed using the program GNOM (24) with corrections for the slit geometry of the scattering instrument. $P(r)$ is the frequency of vector lengths connecting small-volume elements within the entire volume of the scattering particle, weighted by their scattering contrast. For a uniform scattering density object, $P(r)$ goes to zero at the maximum dimension, d_{max}, of the object. Radius of gyration values (R_g) were calculated as the second moment of $P(r)$. Estimates of the radius of gyration of cross section, R_c, and extrapolated $I(0)_c$ values were calculated using the GNOM option for cross-sectional analysis of elongated (rod-like) particles (24). R_c is the scattering contrast-weighted, root-mean-square distance of all elemental areas from the center of the cross-sectional area of a rod-shaped particle, and $I(0)_c$ normalized to protein concentration in mg/ml is proportional to mass-per-unit length (25).

Wild-type LLhP and 21-mer *lacO*^{1} DNA were used for solution scattering experiments. LLhP was concentrated in a VIVASPIN 20 concentrator (MWCO: 10,000; Vivascience; Stonehouse, UK) in a buffer that allowed sufficiently high, monodisperse protein concentrations: 0.12 mM Hepes-KOH, pH 7.6, 200 mM KCl, 5% glycerol,1 mM EDTA, and 0.3 mM DTT. DNA and hypoxanthine stock solutions were made using the filtrates from concentrating the protein solution. Complexes of LLhP, *lacO*^{1}, and hypoxanthine were prepared by adding DNA to a final 1:1 stoichiometry and/or hypoxanthine to a final concentration of 1.4×10^{-4} M. The final concentration of each complex was adjusted using the protein filtrate, and the filtrate served as an exact solvent blank in the scattering experiments. Solutions of binary and ternary complexes were routinely incubated for 30 minutes at room temperature before measurement.

The need for DTT in the samples prohibited accurate concentration determination by UV extinction. Instead, we employed quantitative amino acid analysis, as described previously (23). The concentrations of *lacO*^{1} DNA were determined prior to annealing from the absorbance at 260 nm using known molar absorption coefficients for single-stranded DNA (204,100 L/(mol•cm) and assuming 100% yield in annealing. Analysis of the forward scattering, $I(0)$, values normalized by protein concentration, in mg/ml, volume and mean contrast was used to check that the scattering particles were monodisperse and consistent with the expected stoichiometry as previously described (23); the ratio of the normalized $I(0)$ values for LLhP, LLhP-*lacO*^{1}, LLhP-*lacO*^{1}-HX to a lysozyme standard were within error of the expected value of 1.0; specifically 1.0 ± 0.2, 0.8 ± 0.2, 0.8 ±0.2, respectively. The errors are based on propagated counting statistics, the estimated error in protein concentration based upon repeated amino acid analyses, and in the calculated contrast factors (23). As an additional check on the monodispersity of the samples, estimates for the volume of the scattering particles were obtained using the Porod invariant (26) and compared to expected particle volumes; expected values of 94,223 for LLhP and 106,084 for LLhP complexed with *lacO*^{1} were determined using the relevant amino acid and DNA sequences with standard atomic volumes for nucleic acids and proteins as described in (23). The Porod estimated volumes were all ~100,000 Å^3.

The program CRYSOL (27) was used to predict scattering profiles from structure coordinates. Since no crystal structures are available for LLhP, we modeled the apo- and DNA-bound forms using available structures of LacI and PurR. Apo-LLhP was modeled by mapping the LLhP monomer sequence onto the 1wet monomer for PurR (28) using the MMM web-server (29), which aligns the sequences and utilizes MODELLER (30) to build a structural model. The homodimer was reconstructed using CE/CL (8,31); and the resulting structure was energy minimized to remove all atom clashes using Charmm (32). To check the validity of the final apo-model (Figure 1A), the monomer-monomer interface between the regulatory domains was compared to that of 1wet using Resmap (33). The C-subdomain interface is in very good

agreement with PurR structures (34). The N-subdomain interface appears to be mid-way between that of the PurR-DNA-co-repressor and apo-PurR (28,34,35); similar results were obtained from energy minimization of LacI in the absence of DNA (36). This model was used to predict a scattering profile for apo-LLhP.

In the absence of DNA constraints, the DNA-binding domains of the apo-LLhP model shift as compared to DNA-bound PurR (Figure 1C; (28)) or LacI. Thus, we could not simply add the DNA coordinates to this model to obtain the DNA-bound structure. Instead, we used a previous alignment of LacI and PurR structures that was based on their respective DNA coordinates (8). Various combinations of DNA, DNA-binding domains, linkers, and regulatory domains were used to simulate solution scattering data. The best agreement with experimental data was obtained from combining coordinates from (a) the LacI DNA-binding domains and linkers of the 1efa crystal structure (amino acids 2−61) (37) with (b) the coordinates for the 22-mer $lacO^{sym}$ DNA sequence from the NMR structure 1cjg (38), truncating basepair 22 to match the size of 21mer $lacO^{1}$ used in solution scattering, and (c) the regulatory domain coordinates of PurR from the co-repressed structure 1wet (amino acids 60−340) (28).

Quantitation of allosteric response to co-repressor

Allosteric response can be quantitated as the ratio of affinities for DNA in the presence and absence of co-repressor. Alternatively, operator capture experiments provide a second means of observing the allosteric response. These experiments employ the nitrocellulose filter-binding assay and take advantage of the fact that DNA-binding is enhanced in the presence of co-repressor. A fixed concentration of LLhP protein (see footnotes to Table 4) was chosen so that the final conditions of the operator capture experiment corresponded to 70−90% bound DNA in the affinity assay determined in the presence of hypoxanthine. This maximized the magnitude of the detected signal, but also allowed some of the “low-affinity” LLhP-DNA complex to form. However, effector binding to this complex should result in zero net change in the amount of radio-labeled DNA retained on the filter, and thus is silent in the experiment.

Protein was first incubated 25−30 minutes with ~2×10^{-12} M DNA ligand in the filter binding buffer described above. Co-repressor hypoxanthine was added to LLhP-DNA samples (varied between 10^{-9} and 10^{-4} M) and incubated another 25−30 minutes. Since co-repressor binding enhances LLhP affinity for DNA, increased concentration of co-repressor resulted in an increased amount of LLhP-DNA complex formed (21). The amount of radio-labeled operator bound to LLhP and retained on the nitrocellulose filter was quantified using a Fuji phosphorimager. The resulting data were fit with the program Igor Pro (Wavemetrics, CA) to a modified version of equation 1, with *[Prot]* replaced by *[co-repressor]* and K_d replaced by $[co\text{-}repressor]_{mid}$. The latter parameter is the concentration of co-repressor for which 50% operator is captured.

Results

To determine the functional impact of domain recombination and polymorphisms at linker specificity determinants in LLhP, we designed experiments to: (a) measure DNA binding affinity and selectivity, (b) assess the structural response that occurs when effector binding allosterically switches LLhP between states with low- and high-affinity for operator DNA, (c) quantify how linker polymorphisms alter allosteric response, and (d) evaluate binding of LLhP variants to non-specific DNA.

To that end, we first purified “wild-type” LLhP, I48S, I48V, Q55T, Q55V, G58L, G58T, S61C, and S61M. Next, we utilized three DNA sequences for which LacI has measurable affinity ($lacO^{1}$, $lacO^{sym}$, and $lacO^{disC}$; Table 1) and determined whether the affinities for LLhP variants parallel LacI in their response to altered DNA. Structural aspects of allostery were assessed

using small-angle scattering experiments and compared to our previous studies of LacI. Functionally, allosteric response in LLhP and its variants was quantitated in two different ways: (a) As the ratio between the DNA affinities in the absence and presence of co-repressor; and (b) by the amount of co-repressor required to shift the protein from the low- to high-affinity state.

Operator binding of LLhP variants; comparison with LacI

The simplest explanation of how *in vivo* repression might be altered in LLhP variants is *via* changes in DNA-binding properties. A substitution might alter the strength of the protein-DNA binding interaction, as reflected in the measured K_d. In addition, the variant might have changes in DNA selectivity, which could alter the rank order with which a protein binds similar sequences. We therefore compared LLhP and LacI binding affinities for $lacO^1$, $lacO^{sym}$, and $lacO^{disC}$ (Table 1) in the presence and absence of their effectors. $lacO^1$ is a naturally-occurring operator that is present *in vivo* (19). $lacO^{sym}$ is an "optimized" version of $lacO^1$, with perfect symmetry of the tight-binding, proximal half-site and no central base pair (17,18). $lacO^{disC}$ is derived by symmetrizing the distal half-site of $lacO^1$ and adding another central base pair (16). The K_d values for LacI binding show the rank order of: $lacO^{sym} < lacO^1 < lacO^{disc}$ (9, 16). Addition of IPTG diminishes the K_d for all of these sequences to a value greater than 10^{-7} M, the upper limit of the filter binding experiments[b] (9).

Because of their opposite allosteric response, LLhP+HX and LacI are comparable states, with high affinity for DNA. The K_d of LLhP+HX (or guanine, data not shown) for $lacO^1$ is very similar to that of LacI without inducer (Table 2, Supplementary Figure 1). LLhP binding to $lacO^1$ is diminished >200-fold upon the loss of co-repressor hypoxanthine (HX). Differences between LacI and LLhP emerged when the operator was $lacO^{sym}$ (Table 2, Supplementary Figure 1). LacI binds $lacO^{sym}$ nearly 10-fold more tightly than $lacO^1$ (9,16,17). However, affinity of LLhP+HX is the same for $lacO^{sym}$ and $lacO^1$, although LLhP *without* hypoxanthine binds $lacO^{sym}$ more tightly than it binds $lacO^1$. Surprisingly, LLhP+HX binds $lacO^{disC}$ ~30-fold more tightly than LacI (Table 2, Supplementary Figure 1). Comparing the rank orders of these three operators binding to LacI and LLhP+HX shows that the chimera does not discriminate between the altered sequences as well as the natural protein: LLhP+HX exhibits ~1 order of magnitude difference between the three operator sequences (Figure 2A, solid black line), whereas LacI binding differs nearly three orders of magnitude (Figure 2A, dashed black line).

Next, DNA-binding experiments were carried out for 8 LLhP variants with linker polymorphisms – I48S, I48V, Q55T, Q55V, G58L, G58T, S61C, and S61M. Relative to unmodified LLhP, the two substitutions at positions 48 and 55 diminish affinity (Table 2, Supplementary Figure 2). In each case, the K_d rank order for alternative DNA sequences is: $lacO^{sym} < lacO^1 < lacO^{disC}$. Results for the 4 variants parallel each other and roughly parallel the binding affinities of unmodified LLhP; the difference between the affinities for alternative operator sequences remains smaller than seen with LacI (slopes of the lines in Figure 2A). Loss of hypoxanthine lowers the affinity of these variants, but the allosteric ratios are within 2-fold for all 48/55 variants and unmodified LLhP, and for all DNA operators (Table 2 and Figure 2B). Thus, these substitutions appear to alter overall affinity but not selectivity for alternative DNA ligands. The only instance in which the parallel pattern is broken is for unmodified LLhP

[b] In LLhP binding assays, the pH and ionic strength of the current buffer conditions (see footnotes in Table 2) match previous LacI studies (*e.g.* (9)). In these conditions, the weak K_d associated with LacI-IPTG binding to DNA cannot be accurately measured by the filter binding technique. To fully form the LacI-$lacO^1$-IPTG complex for solution scattering experiments (Table 3; (23)), we chose different buffer conditions that simultaneously enhance formation of the ternary complex and allow the highly concentrated sample to remain monodisperse. For similar reasons, LLhP solution scattering experiments and binding experiments were performed in the buffer noted in the last row of Table 2.

+HX binding to *lacO*^{sym}; one possible explanation is that the "wild-type" chimera has reached a limit of how tightly it can bind any DNA sequence.

Both substitutions at LLhP position 58 greatly diminish binding to any of the three operators (Table 2, Supplementary Figure 3), consistent with their poor *in vivo* repression (4). Only partial binding curves are obtained, and we do not detect any response to hypoxanthine. This behavior is similar to that of LacI when binding to non-specific DNA sequences – low affinity with no response to inducer (39). In contrast, substitutions at LLhP position 61 had surprising *in vitro* behaviors. S61C and S61M did not repress well *in vivo* (4), but show fairly good (albeit reduced) affinity for *lacO*^{1} and *lacO*^{disC} in the presence of hypoxanthine (Table 2, Supplementary Figure 4A). In addition, the allosteric ratio of S61M and S61C for these two operators was ~1 order of magnitude larger than for the other LLhP variants (Table 2 and Figure 2B). Most striking, the 61 variants do not have a measurable K_d when binding to *lacO*^{sym}, nor do they show a detectable allosteric response to hypoxanthine when binding this operator (Supplementary Figure 4B). A primary difference in *lacO*^{sym} and the other two sequences is the diminished spacing between the two half-sites (Table 1). The 61 mutants might require increased spacing for high-affinity binding and allosteric response.

Structural comparison of LLhP and LacI allosteric responses

To evaluate potential structural changes in LLhP that occur upon binding *lacO*^{1} and hypoxanthine, we performed small-angle X-ray scattering, experiments, similar to our earlier studies of LacI and its complexes with DNA and IPTG (23). We chose the "wild-type" LLhP variant because it exhibits the tightest DNA binding in the absence of hypoxanthine; meaningful interpretation of results requires complete formation of the LLhP-*lacO*^{1} complex. We chose the *lacO*^{1} operator over *lacO*^{sym} because of the slightly larger allosteric response seen for "wild-type" LLhP (Table 2); if allostery is associated with domain movements, we reasoned that this condition would provide the best chance to observe them. Scattering data were acquired for apo-LLhP and in complex with *lacO*^{1}, with and without hypoxanthine. As described previously (23), $I(0)$, Guinier, and $P(r)$ analyses were used in combination with the protein concentration dependence of the scattering data to establish conditions for which the samples were mono-disperse and free of inter-particle interference effects, as required for accurate determination of structural parameters. Satisfactory conditions were identified for apo-LLhP, LLhP-*lacO*^{1}, and LLhP-HX-*lacO*^{1} at sufficiently low protein concentrations (Table 3). Samples of LLhP bound to effector hypoxanthine in the absence of DNA were consistently aggregated; thus no data are presented. We also determined the K_d values for DNA binding in the required Hepes buffer; although binding affinity is diminished, the magnitude of allosteric response is comparable (Table 2).

Results of $P(r)$ analyses are shown in Table 3 and Figure 3. The general features of the $P(r)$ are consistent with the expected homodimer shape, and the LLhP-*lacO*^{1} complexes show the expected increase in $I(0)$ due to the DNA. As was observed for the LacI "R3" dimer[c], apo-LLhP has a well defined radius of gyration of cross section, R_c, and mass per unit length. Both parameters are independent of protein concentration in the range 2.6 − 6.1 mg/ml, whereas R_g shows concentration dependent increases above 3.7 mg/ml. These results indicate that LLhP, like the LacI dimer, forms some sort of end-to-end association with increasing concentration. For concentrations ≤ 3.7 mg/ml, the LLhP samples were monodisperse as determined by Guinier (Figure 3A, insert) and $I(0)$ analyses (see Materials and Methods).

[c]Wild-type LacI has a C-terminal tetramerization domain that, in the R3 variant, is substituted with the GCN4 dimerization domain (40). The LacI dimer is the functional unit for DNA binding, and R3 has extremely similar affinity and allosteric response as a dimer within a LacI tetramer (23). DNA binding experiments such as those presented in Table 2 are designed so that the intrinsic binding affinity of a LacI dimer is measured. Chimera LLhP is a dimer.

Notably, apo-LLhP is significantly less elongated than the apo-LacI R3 dimer. R_g and d_{max} values for apo-LLhP are respectively 4−5 Å and 20 Å smaller than those for apo-R3. The LLhP R_c value is larger by ~2 Å (Table 3). To aid in the interpretation of the scattering data, we constructed models for apo-LLhP and its DNA complex (see Material and Methods). The R_g and d_{max} values measured for apo-LLhP are very similar ($\chi^2 = 0.76$) to those calculated from either the protein component of the PurR-DNA crystal structure (1wet; (28)) or the homology model of apo-LLhP. This result is in stark contrast to solution scattering results for apo-LacI, for which we measured much larger R_g and d_{max} values than the protein component of the LacI-DNA complex in the 1efa crystal structure (37). We were able to model the conformation of apo-LacI by allowing the linker's hinge helix to unfold and extend (23). The smaller R_g and d_{max} values measured for apo-LLhP compared to the apo-LacI dimer are consistent with apo-LLhP maintaining a compact hinge-helix in the linker.

Upon LLhP binding *lacO*^{1}, R_g increases ~2 Å and d_{max} increases ~10 Å, consistent with the addition of the DNA. Likewise, the model of the LLhP-*lacO*^{1} complex and the crystallographic structure of PurR bound to its cognate operator (1wet; (28)) show good agreement with the data, although our model has a somewhat improved fit (χ^2 value of 0.88 compared to 1.00), perhaps due to matching the experimental DNA sequence. All models give predicted R_g and d_{max} values that agree with the measured data within experimental uncertainty. Notably, all models and structures that fit the experimental data have a compact hinge helix within the linker sequence (Figure 1B), similar to the hinge helix observed in the LacI-*lacO*^{1} complex.

Another observation in the current work is that hypoxanthine binding to LLhP-*lacO*^{1} causes no measurable differences in R_g, d_{max}, or the $P(r)$ profiles (Table 3; Figure 3B, small black symbols near the origin). This observation again contrasts the results for the LacI dimer (23), which showed a small but statistically significant 3% increase in R_g and a corresponding redistribution of vector lengths in P(r) upon binding effector IPTG that was consistent with small domain rearrangements within the LacI dimer (23). For comparison, the LacI difference $P(r)$ plot is superimposed on the LLhP plots (Figure 3B; small blue symbols). The outcome for LLhP was not expected based on the available PurR and LacI crystal structures (28,35, 37,41,42). In these structures, PurR shows a larger reorientation of the regulatory subdomains (apo- compared to the PurR-DNA-co-repressor complex) than does LacI (either apo- or IPTG-bound compared to the LacI-DNA complex), which translates into a larger change predicted for the PurR $P(r)$ profile. The simplest interpretation of the results is that the intact LLhP does not exhibit large structural changes upon binding hypoxanthine. Alternatively, the DNA-binding domains in LLhP might make compensatory motions that mask the regulatory sub-domain reorientations.

Taken together, the structural information derived from the scattering data suggests that LLhP is a less flexible structure than LacI: The same linker sequence remains compact in the context of the LLhP chimera upon removal of operator DNA, but appears to be extended in the context of LacI. Further, we do not see evidence for regulatory subdomain reorientations upon hypoxanthine binding to the LLhP-DNA complex.

Allosteric response in LLhP and its variants; Functional Aspects

As mentioned above, the allosteric ratio determined from K_d in the presence and absence of effector provides one measure of allosteric response (Table 2, Figure 2B). However, affinity experiments are performed with saturating effector and do not address the question of how much effector is required for the allosteric change. The latter can be monitored with operator capture experiments (21): These experiments utilized concentrations of protein and DNA fixed at subsaturation, with co-repressor added in increasing concentrations. As hypoxanthine binds, the protein acquires increased affinity for radio-labeled DNA, resulting in more protein-DNA complex that is retained on the nitrocellulose filter. The midpoint of an operator capture curve

reflects both the concentration of effector required as well as the free energy needed to make the allosteric switch.

Midpoint values are listed in Table 4 for unmodified LLhP and variants that had measurable values of K_d; binding curves are shown in Supplementary Figure 1B and Supplementary Figure 5. For "wild-type" LLhP, curves are super-imposable for three different operator sequences, with a mid-point around 4.4×10^{-7} M hypoxanthine. Variants at 48 and 55 show midpoints similar to LLhP on $lacO^1$ and $lacO^{sym}$, whereas midpoints for $lacO^{disc}$ exhibit some divergence (Figure 2C). Since this operator has the lowest affinity of the three examined, possible explanations are that $lacO^{disC}$ does not provide as much driving energy for making the allosteric switch, or that subtle protein structural differences are more apparent with weaker binding. Strikingly, S61C and S61M require up to 10-fold more hypoxanthine to accomplish the allosteric change for either $lacO^1$ or $lacO^{disC}$ (Figure 2C). Additional experiments are required to determine whether the altered allosteric responses (Figure 2B and C) are due to allosterically-diminished, intrinsic hypoxanthine binding or to altered energetics of the protein allosteric change.

Discussion

A growing body of evidence supports the idea that the sequences linking domains play important roles beyond a simple joining function. For example, the sequences linking functional domains in the cAMP-dependent protein kinase can dramatically affect structure and hence function of this protein (43). Therefore, it is not surprising to find specificity determinants – functionally important amino acids that are not conserved in a protein family – in regions linking protein functional domains. What has been under-appreciated is the range of ways these sites can contribute to function. Here, we report that significant functional changes arise both from domain recombination that creates the LLhP chimera and from amino acid polymorphisms at linker specificity determinants. Since all residues that directly contact DNA should be the same in LacI and the LLhP variants, the changed interfaces between the linkers and the regulatory domains must convey the different functional responses.

Promiscuity in DNA binding

A striking difference between LLhP and LacI is that the chimera shows enhanced promiscuity for alternative DNA sequences (Figure 2A). This phenomenon may be related to diminished flexibility of the LLhP linker seen in solution scattering experiments. The apo-LLhP linker remains compact upon removal of DNA, whereas apo-LacI shows a dramatic extension, likely due to unfolding of the hinge helix (23). We hypothesize that residual structure in LLhP might "lock" the protein into a conformation that has a moderately high affinity for a range of DNA sequences – this would enhance binding to "poor" operators such as $lacO^{disC}$. The correlation between protein flexibility and discrimination between alternative sequences might be a general feature of the LacI/GalR proteins: Promiscuous DNA-binding was also seen for a LacI variant with the $V52C_{ox}$ mutation, which constrains the juxtapositions of the linkers by adding a disulfide bond between the hinge helices (16).

The promiscuity seen for LLhP DNA-binding may also indicate that "reverse evolution" occurred upon domain recombination. One theory as to how new proteins evolve is that the protein function proceeds from more promiscuous to more specific (*e.g.* (44,45)). Members of the LacI/GalR family appear to have arisen by gene duplication (46), which must have been followed by sequence divergence at specificity determinants. The LacI linker sequence would have of course evolved in the context of the LacI regulatory domain to have extremely high affinity and selectivity for the natural *lac* operators. Thus, while LLhP has the expected gross function, domain recombination might have eliminated the features that allow a high degree of discrimination between operator sequences. Amino acid substitution at one position

explored in the current work – site 61 – appears to restore discrimination between alternative operators, albeit with a very different outcome than for LacI.

Allosteric response to effector binding

For LLhP, the change in DNA-affinity upon effector binding is an order of magnitude smaller in the chimera than in LacI (Figure 2B). Correlation between the magnitude of allosteric response and conformational flexibility might be a general feature of the LacI/GalR family. The other conformationally-constrained example – disulfide-linked LacI $V52C_{ox}$ – also shows an allosteric response that is much smaller than that of LacI (16).

LLhP is co-repressed like PurR, and the magnitude of allostery for LLhP is roughly equivalent to that of PurR (10). Thus, we were surprised that the LLhP solution scattering did not detect the large subdomain rearrangements that are seen when comparing the crystal structures of truncated, apo-PurR to that of the PurR-DNA-co-repressor complex (28,35,42). Compensatory motions of the LLhP DNA-binding domains could possibly mask evidence in the scattering data for reorientation of the regulatory subdomains. However, if this very special circumstance is not the case, the lack of evident subdomain motions must draw upon one or more of the following explanations: (a) Truncation of the DNA-binding domains, needed to obtain the apo-PurR structure, dramatically alters the subdomain juxtapositions in the PurR regulatory domain; (b) The allosteric conformational change of PurR cannot be inferred from comparing structures of apo-protein and the ternary complex – the relevant changes must be seen by comparing the structures of the ternary and (unavailable) binary PurR-HX complex; or (c) Exchanging the PurR and LacI linker sequences greatly affects the allosteric conformational change of the PurR regulatory domain. All three scenarios have important implications: the first two for designing effective studies of allostery, the third as an unexpected consequence of protein engineering.

The dimensions obtained from solution scattering data do not indicate any significant structural change in the linker/regulatory domain interface when LLhP-$lacO^1$ binds hypoxanthine. This observation is consistent with the LLhP mutagenesis data, since several mutations have equivalent effects on the low- and high-affinity conditions (I48S, I48V, Q55T, and Q55V). Of course, some positions might be unchanged in the allosteric transition, whereas others are affected. Indeed, the substitutions S61C and S61M do not show state-equivalence in functional effects, which leads to a larger magnitude of allosteric response. The interface in LacI also appears to persist in the induced complex (23); thus, this may be a common feature of many LacI/GalR proteins, regardless of whether they are induced or co-repressed upon binding effector.

Linker polymorphisms and protein engineering

One goal of protein engineering is to identify and utilize patterns from the sequence/function relationship of natural homologues to rationally create novel functions. Given the impact of amino acid substitution at linker positions, these sites provide opportunity engineer novel repressors for biotechnology. In this study, we compared two substitutions at each linker site to the third variant of “wild-type” LLhP. The current pairs of “mutations” act alike and quite differently from “wild-type”, even though their side chains are not chemically similar, except perhaps S61C and S61M. In hindsight, we chose pairs of substitutions with similar *in vivo* phenotypes, raising the question of how other substitutions at the same sites alter function. Would the affects of other substitutions at positions 48 and 55 be limited to changes in affinity? One certainly wouldn't expect all the variations at position 61 to have the same, complicated responses of S61C and S61M. Given the unexpected behaviors of these two variants, other substitutions at 61 might have interesting *in vitro* properties. However, when engineering new repressors for biotechnology, the final outcome is that the variants at position 61 do not repress

in vivo. All aspects of function must be carefully balanced in order to be effective *in vivo* – even though S61C and S61M have reasonable DNA affinity for *lacO*^{1} with enhanced allosteric response, other factors abolish their value as repressors.

Reconciliation of in vivo and in vitro functions

In vivo characterization of repression is a fast way to assess the functions of large numbers of protein variants. The current work provides opportunity to benchmark the *in vivo* assay, to facilitate interpretation of future experiments. Comparison of *in vivo* repression (4) and *in vitro* affinity for the LLhP variants leads to the plot shown in Figure 4. The whole dataset shows poor correlation between *lacO*^{1} affinity and *in vivo* repression (dashed line, $R^2 = 0.40$). However, removing the points for S61C(+HX) and S61M(+HX) greatly improves the correlation (solid line, $R^2 = 0.79$) between the two functional assays. The 58 variants also show good correlation, with very little repression and unmeasurable K_d values for *lacO*^{1}.

The S61 outliers are not due to the trivial explanation of altered *in vivo* protein concentrations: We see large quantities of full-length, soluble protein with Coomassie stain on SDS-PAGE, so these proteins are in vast excess of the single genomic binding site (4). Instead, the current work shows that these proteins require more hypoxanthine in order to make the allosteric change (Figure 2C). We cannot find any reported values for *in vivo* hypoxanthine concentrations in *E. coli*. However, hypoxanthine is a metabolite of adenine and therefore concentrations are almost certainly tightly regulated: The hypoxanthine concentration is probably never zero, yet even in the presence of additional exogenous co-repressor, the concentration could be constrained at a level too low for co-repression of the S61 LLhP variants. Since the operator capture midpoints for S61C and S61M are only 5- to 10-fold different from the other variants, cells might not allow the full allosteric "switch" for any of the LLhP variants. In other words, *in vivo* hypoxanthine concentrations might fall in a narrow range that allows some incomplete allosteric response for most LLhP variants but is too low to allow S61C or S61M binding. Indeed, the *in vivo* allosteric response was only 2−10 fold for all LLhP variants (4), much smaller than the 100- to 200-fold response measured *in vitro* for *lacO*^{1} (Table 2).

Conclusion

Both gene duplication and domain recombination, followed by sequence divergence, are hypothesized to be mechanisms by which new protein functions evolve (*e.g.* (47,48)). Here, we show that both domain recombination and amino acid substitutions in nonconserved amino acids can result in a protein that generally retains its original function – repression from *lacO*^{1} – and at the same time acquires altered functional aspects – promiscuity in DNA binding or altered selectivity for alternative operators. Such changes might impact the fitness of an *E. coli* bacterium expressing these proteins. In addition, these linker specificity determinants provide opportunities to rationally engineer new proteins. To that end, we must now determine whether the observed results are unique to the LacI/LLhP linker sequence or are general properties of the nonconserved linkers found in the LacI/GalR family.

Acknowledgements

We thank Mr. Sudheer Tungtur for assistance growing cells for protein expression and assistance creating the model shown in Figure 1; Dr. Antonio Artigues (KUMC) for mass spectrometry of analysis of wild-type LLhP; and Drs. Graham Palmer and Yury Kamensky for use of the MCD (Rice University). Dr. Clare Woodward provided many helpful discussions about the solution scattering experiments. Dr. Kathleen S. Matthews graciously allowed HLZ to perform experiments in her laboratory at Rice University.

References

1. Pei J, Cai W, Kinch LN, Grishin NV. Prediction of functional specificity determinants from protein sequences using log-likelihood ratios. Bioinformatics 2006;22:164–171. [PubMed: 16278237]
2. Kalinina OV, Novichkov PS, Mironov AA, Gelfand MS, Rakhmaninova AB. SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. Nucl Acids Res 2004;32:W424–428. [PubMed: 15215423]
3. Pawlyk AC, Pettigrew DW. Transplanting allosteric control of enzyme activity by protein-protein interactions: coupling a regulatory site to the conserved catalytic core. Proc Natl Acad Sci USA 2002;99:11115–11120. [PubMed: 12161559]
4. Tungtur S, Egan S, Swint-Kruse L. Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. Proteins: Struct Func Bioinf 2007;68:375–388.
5. Weickert MJ, Adhya S. A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem 1992;267:15869–15874. [PubMed: 1639817]
6. Barkley MD, Riggs AD, Jobe A, Burgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry 1975;14:1700–1712. [PubMed: 235964]
7. Meng LM, Nygaard P. Identification of hypoxanthine and guanine as the co-repressors for the purine regulon genes of Escherichia coli. Mol Microbiol 1990;4:2187–2192. [PubMed: 2089227]
8. Swint-Kruse L, Larson C, Pettitt BM, Matthews KS. Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR. Protein Sci 2002;11:778–794. [PubMed: 11910022]
9. Zhan H, Swint-Kruse L, Matthews KS. Extrinsic Interactions Dominate Helical Propensity in Coupled Binding and Folding of the Lactose Repressor Protein Hinge Helix. Biochemistry 2006;45:5896–5906. [PubMed: 16669632]
10. Moraitis MI, Xu H, Matthews KS. Ion concentration and temperature dependence of DNA binding: comparison of PurR and LacI repressor proteins. Biochemistry 2001;40:8109–8117. [PubMed: 11434780]
11. Lu F, Brennan RG, Zalkin H. Escherichia coli purine repressor: key residues for the allosteric transition between active and inactive conformations and for interdomain signaling. Biochemistry 1998;37:15680–15690. [PubMed: 9843372]
12. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucl Acids Res 2000;28:45–48. [PubMed: 10592178]
13. Holmquist B, Vallee BL. Tryptophan quantitation by magnetic circular dichroism in native and modified proteins. Biochemistry 1973;12:4409–4417. [PubMed: 4750252]
14. Riggs AD, Bourgeois S, Newby RF, Cohn M. DNA binding of the lac repressor. J Mol Biol 1968;34:365–368. [PubMed: 4938552]
15. Wong I, Lohman TM. A double-filter method for nitrocellulose-filter binding: Application to protein-nucleic acid interactions. Proc Natl Acad Sci USA 1993;90:5428–5432. [PubMed: 8516284]
16. Falcon CM, Matthews KS. Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. Biochemistry 2001;40:15650–15659. [PubMed: 11747440]
17. Simons A, Tils D, von Wilcken-Bergmann B, Muller-Hill B. Possible ideal lac operator: Escherichia coli lac operator-like sequences from eukaryotic genomes lack the central G X C pair. Proc Natl Acad Sci U S A 1984;81:1624–1628. [PubMed: 6369330]
18. Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. Proc Natl Acad Sci U S A 1983;80:6785–6789. [PubMed: 6316325]
19. Gilbert W, Maxam A. The nucleotide sequence of the lac operator. Proc Natl Acad Sci U S A 1973;70:3581–3584. [PubMed: 4587255]
20. Swint-Kruse L, Matthews KS. Thermodynamics, protein modification, and molecular dynamics in characterizing lactose repressor protein: strategies for complex analyses of protein structure-function. Methods Enzymol 2004;379:188–209. [PubMed: 15051359]

21. Lu F, Schumacher MA, Arvidson DN, Haldimann A, Wanner BL, Zalkin H, Brennan RG. Structure-based redesign of corepressor specificity of the Escherichia coli purine repressor by substitution of residue 190. Biochemistry 1998;37:971–982. [PubMed: 9454587]
22. Choi KY, Zalkin H. Structural characterization and corepressor binding of the Escherichia coli purine repressor. J Bacteriol 1992;174:6207–6214. [PubMed: 1400170]
23. Taraban M, Zhan H, Whitten AE, Langley DB, Matthews KS, Swint-Kruse L, Trewhella J. Ligand-induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein. J Mol Biol 2008;376:466–481. [PubMed: 18164724]
24. Svergun DI. Mathematical methods in small-angle scattering data analysis. J Appl Cryst 1991;24:485–492.
25. Guinier, A.; Fournet, G. Small-Angle Scattering of X-rays. John Wiley and Sons; New York: 1955. p. 128-129.
26. Porod G. Die Roentgenkleinwinkel-Steuung von Dichtgepackten Kolloiden Systemen, I Teil. Kolloid Z Biol 1951;124:83–111.
27. Svergun DI, Barberato C, Koch MHJ. CRYSOL — a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 1995;28:768–773.
28. Schumacher MA, Glasfeld A, Zalkin H, Brennan RG. The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. J Biol Chem 1997;272:22648–22653. [PubMed: 9278422]
29. Rai BK, Madrid-Aliste CJ, Fajardo JE, Fiser A. MMM: a sequence-to-structure alignment protocol. Bioinformatics 2006;22:2691–2692. [PubMed: 16928737]
30. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815. [PubMed: 8254673]
31. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747. [PubMed: 9796821]
32. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
33. Swint-Kruse L, Brown CS. Resmap: automated representation of macromolecular interfaces as two-dimensional networks. Bioinformatics 2005;21:3327–3328. [PubMed: 15914544]
34. Swint-Kruse L, Elam CR, Lin JW, Wycuff DR, Shive Matthews K. Plasticity of quaternary structure: twenty-two ways to form a LacI dimer. Protein Sci 2001;10:262–276. [PubMed: 11266612]
35. Schumacher MA, Choi KY, Lu F, Zalkin H, Brennan RG. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. Cell 1995;83:147–155. [PubMed: 7553867]
36. Swint-Kruse L. Using networks to identify fine structural differences between functionally distinct protein states. Biochemistry 2004;43:10886–10895. [PubMed: 15323549]
37. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol 2000;7:209–214. [PubMed: 10700279]
38. Spronk CA, Bonvin AM, Radha PK, Melacini G, Boelens R, Kaptein R. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure 1999;7:1483–1492. [PubMed: 10647179]
39. Lin S, Riggs AD. A comparison of lac repressor binding to operator and to nonoperator DNA. Biochem Biophys Res Commun 1975;62:704–710. [PubMed: 1120077]
40. Chen J, Alberti S, Matthews KS. Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. J Biol Chem 1994;269:12482–12487. [PubMed: 8175655]
41. 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. Science 1996;271:1247–1254. [PubMed: 8638105]
42. Mowbray SL, Björkman AJ. Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. J Mol Biol 1999;294:487–499. [PubMed: 10610774]

43. Vigil D, Blumenthal DK, Taylor SS, Trewhella J. Solution scattering reveals large differences in the global structures of type II protein kinase A isoforms. J Mol Biol 2006;357:880–889. [PubMed: 16460759]
44. Yoshikuni Y, Keasling JD. Pathway engineering by designed divergent evolution. Curr Opin Chem Biol 2007;11:233–239. [PubMed: 17353138]
45. Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. Nature genetics 2005;37:73–76. [PubMed: 15568024]
46. Fukami-Kobayashi K, Tateno Y, Nishikawa K. Parallel evolution of ligand specificity between LacI/GalR family repressors and periplasmic sugar-binding proteins. Mol Biol Evol 2003;20:267–277. [PubMed: 12598694]
47. Bashton M, Chothia C. The Generation of New Protein Functions by the Combination of Domains. Structure 2007;15:85–99. [PubMed: 17223535]
48. Poelwijk FJ, Kiviet DJ, Tans SJ. Evolutionary potential of a duplicated repressor-operator pair: simulating pathways using mutation data. PLoS computational biology 2006;2:e58. [PubMed: 16733549]
49. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Cryst 1992;25:495–503.
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612. [PubMed: 15264254]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

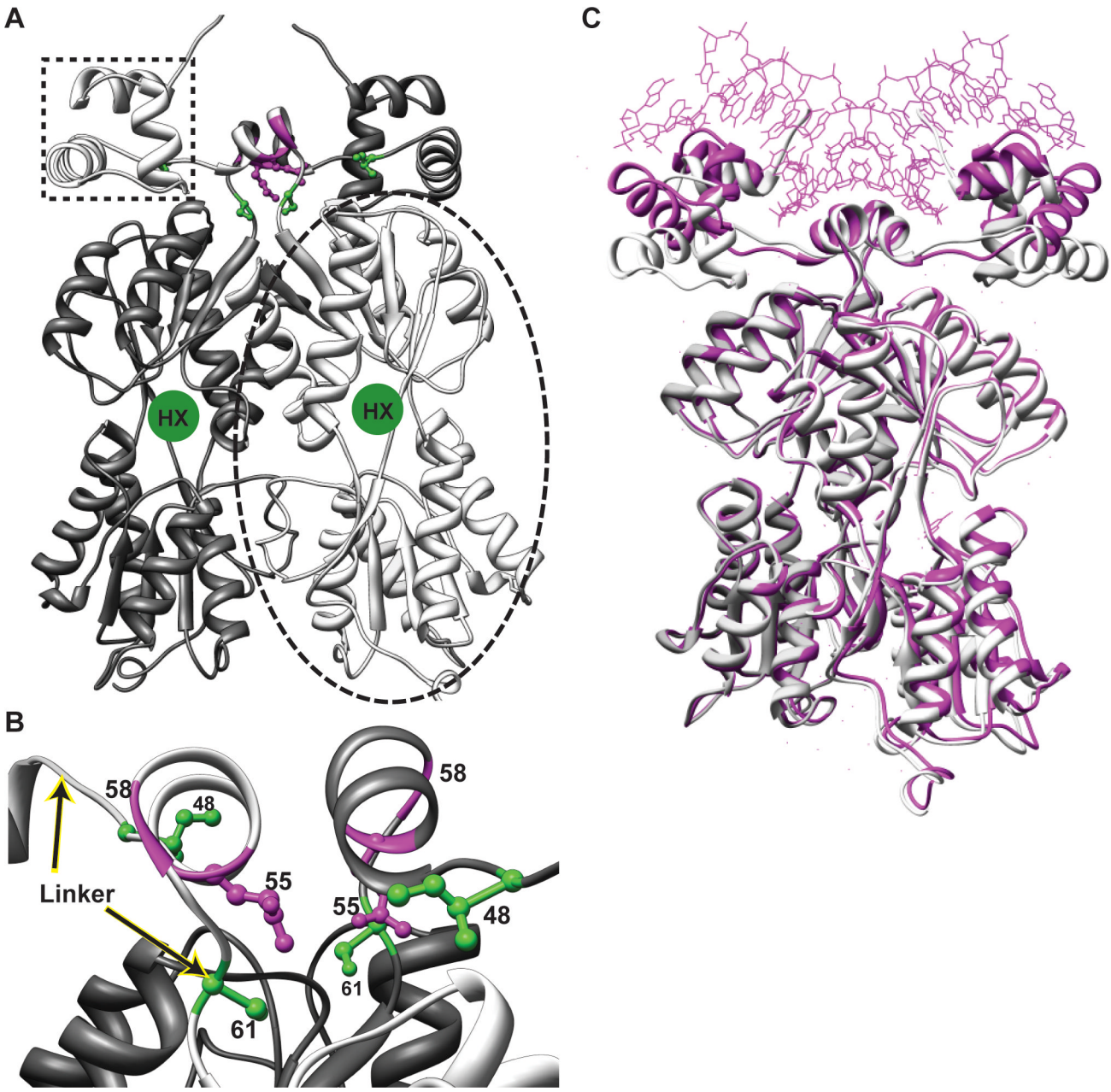

Figure 1. Model of the LLhP homodimer and positions of specificity determinants

(A) The model of apo-LLhP was constructed as described in Materials and Methods and rendered with Chimera (50). One monomer is colored light gray and the other is dark gray. One DNA-binding domain is indicated with the dotted box; the two DNA-binding domains of a dimer are required for binding operator DNA. One regulatory domain is indicated within the dashed oval. The binding site for co-repressor hypoxanthine is indicated with a green circle in a cleft between the N- and C-subdomains (which are respectively at the top and bottom of the regulatory domain.) **(B)** The linkers of the LLhP dimer are shown, with the beginning and end indicated with arrows. The orientation is rotated ~90° relative to (A). The central helical region of the linker is often called the “hinge helix”. The positions of linker specificity determinants for wild-type sites I48, Q55, and S61 are shown in ball-and-stick representation in green and magenta. Gly58 is indicated with a magenta ribbon for the backbone. None of the linker

specificity determinants directly contact DNA in either the LacI or PurR structures. **(C)** The model for apo-LLhP (white) was aligned with the magenta structure of co-repressed PurR bound to DNA (1wet; (28)) using Chimera (50). Co-repressor is not shown in this figure. Note that the DNA-binding domains of apo-LLhP have moved relative to those of the DNA-bound PurR. Intriguingly, molecular dynamics simulations of the LacI DNA-binding domain show similar motions (8). Flexibility in the region spanning amino acids 45−50 is required for this motion, suggesting a possible mechanism for altering DNA-binding affinity by polymorphisms at position 48.

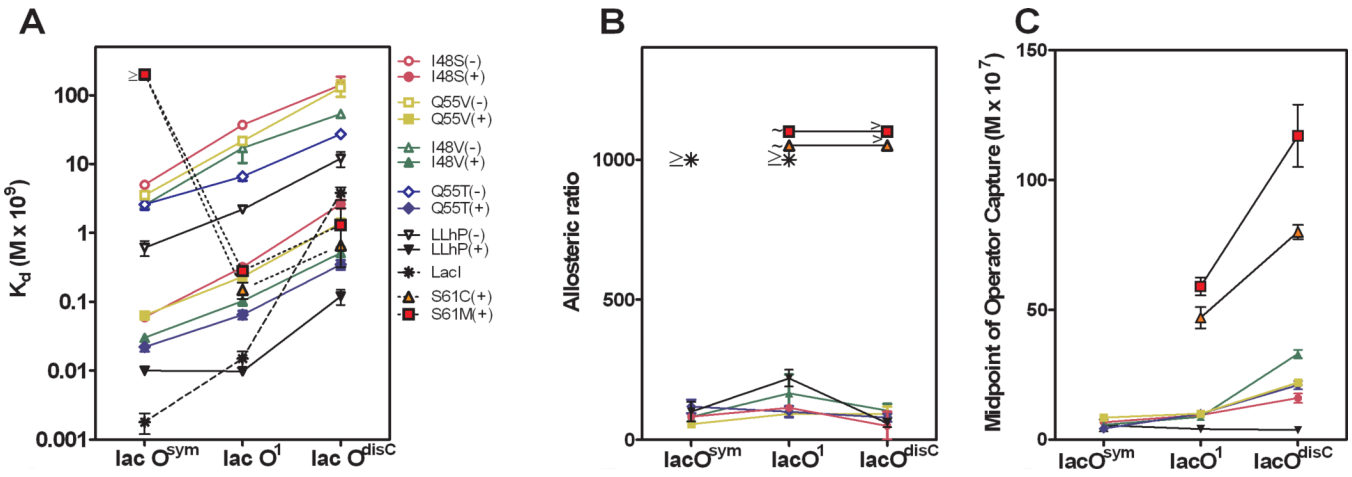

Figure 2. Trends in K_d and allosteric response for LLhP variants binding to alternative *lac* operators

The symbols and colors in legend shown are used for all three panels. For panel (A), "+" indicates the presence of saturating hypoxanthine. (**A)** Affinities of LacI and LLhP variants for $lacO^1$, $lacO^{sym}$, and $lacO^{disC}$. LacI (dashed black line) shows steeper discrimination between the 3 operators than does LLhP (solid black line). With the exception of LLhP+HX +$lacO^{sym}$, unmodified LLhP, the 48 variants, and the 55 variants have different affinities but the same selectivities (solid, parallel lines). The LLhP variants at position 61 show a different pattern of DNA selectivity (dotted lines). **(B)** Allosteric ratios of LLhP variants report the magnitude of change in K_d when the repressor binds hypoxanthine. Data for S61C and S61M are slightly offset to aid visual inspection. **(C**) Midpoints of operator capture for LLhP variants reflect the amount of hypoxanthine required to elicit the full allosteric change.

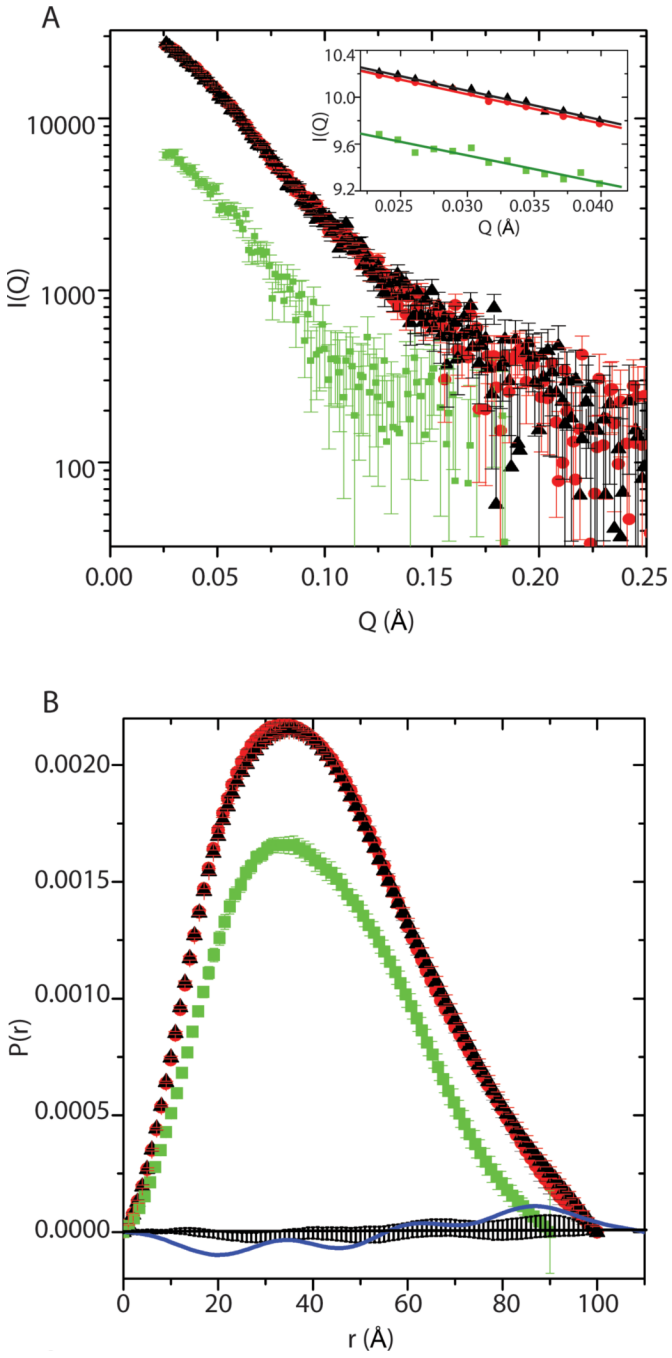

Figure 3. Solution X-ray scattering by LLhP

(A) *I(Q)* versus *Q* for apo-LLhP (green), LLhP-*lacO*^{1} (red) and LLhP-*lacO*^{1}-HX (black). The **inset** shows a Guinier plot of the low-Q data and the linear fits obtained (apo is green; LLhP-*lacO*^{1} is red; LLhP-*lacO*^{1}-HX is black). The scattering data are shown normalized to constant monitor counts, and for the Guinier plots, the apo-LLhP data are multiplied by 3 to allow for easier display. **(B)** *P(r)* versus *r* (large symbols) calculated using the data in (A), using the same key. Difference plots are also shown in small symbols near the origin Y=0 for the *P(r)* profiles of LLhP-*lacO*^{1} with and without hypoxathine (black). For comparison, a difference plot is also shown for the LacI R3 dimer bound to *lacO*^{1} with and without IPTG (from (23); blue).

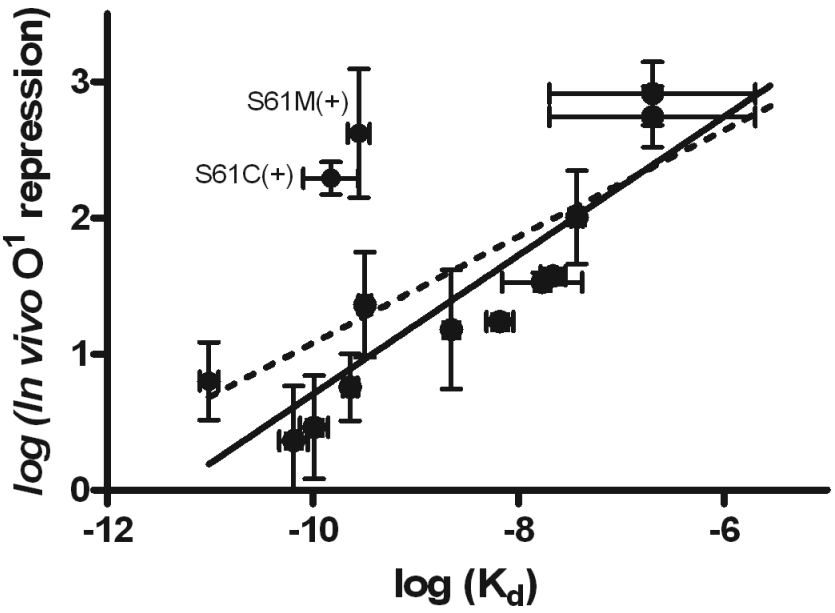

Figure 4. ***In vivo*** **repression versus** ***in vitro*** **DNA-binding affinity for LLhP variants**

In vivo data are reported in (4). Data shown on this plot include measurements made in both the presence and absence of co-repressor for LLhP variants in the current study. The dashed line reflects the best fit (linear regression) of the entire data set; $R^2 = 0.40$. The solid line reflects the best fit when S61C(+) and S61M(+) are not included; $R^2 = 0.79$. Since K_d values for G58T and G58V could not be reasonably estimated, these data are not included on this plot. However, the G58 variants show good correlation between poor repression and weak DNA binding.

Table 1

Sequences of LacI DNA operator ligands[a].

	5′	1	2	3	4	5	6	7	8			9	10	11	12	13	14	15	16
lacO^{1}		T	T	G	T	G	A	G	C		G	G	A	T	A	A	C	A	A
lacO^{sym}		T	T	G	T	G	A	G	C			G	C	T	C	A	C	A	A
lacO^{disC}		T	T	G	T	T	A	T	C	C	G	G	A	T	A	A	C	A	A

[a] *lacO*^{1} is one of the naturally occurring operators found in the *lac* operon (19) and controls transcription in the *in vivo* assays described previously (4). *lacO*^{sym} and *lacO*^{disC} are engineered variants (16-18).

Table 2

Operator binding by LLhP variants[a].

	lacO[1]			*lacO*[sym]			*lacO*[disC]		
	Low affinity[b] K_d (x 10^{-11} M)	**High affinity[c] K_d (x 10^{-11} M)**	**Allosteric ratio[d]**	**Low affinity K_d (x 10^{-11} M)**	**High affinity K_d (x 10^{-11} M)**	**Allosteric ratio**	**Low affinity K_d (x 10^{-11} M)**	**High affinity K_d (x 10^{-11} M)**	**Allosteric ratio**
LacI[e]	>10000	1.5 ± 0.4	>1000	>10000	0.18 ± 0.06	>10000	>10000	380 ± 80	>100
LLhP	220 ± 20	0.98 ± 0.09	220 ± 30	61 ± 15	1.0 ± 0.10	60 ± 15	1200 ± 300	12 ± 3	100 ± 35
I48S	3700 ± 220	32 ± 2.0	110 ± 10	500 ± 34	6.0 ± 0.20	83 ± 7	~14000±4600	260 ± 36	~50
I48V	1700 ± 660	10 ± 1.4	170 ± 70	260 ± 18	3.1 ± 0.20	82 ± 8	5300 ± 560	51 ± 11	100 ± 25
Q55T	660 ± 86	6.5 ± 0.94	100 ± 20	260 ± 43	2.2 ± 0.25	119 ± 24	2700 ± 250	35 ± 5.3	79 ± 14
Q55V	2200 ± 260	23 ± 1.6	93 ± 13	350 ± 37	6.4 ± 0.80	56 ± 9	13000 ± 3400	140 ± 20	92 ± 27
G58L	>10000	>10000	NR[f]	>10000	>10000	NR	>10000	>10000	NR
G58T	>10000	>10000	NR	>10000	>10000	NR	>10000	>10000	NR
S61C	≥20000	15 ± 4.3	~1000	>10000	>10000	NR	NBD[g]	66 ± 5.1	>1000
S61M	≥30000	28 ± 3.4	~1000	>10000	>10000	NR	NBD	126 ± 9.8	>1000
LLhP Hepes buffer[h]	1780 ± 110	13 ± 1.9	136 ± 12						

[a] DNA binding data were determined from 3−4 independent measurements, using two different preparations of protein. Reported errors represent one standard deviation. Buffer used for DNA-binding measurements was 10 mM Tris-HCl, pH 7.4, 150 mM KCl, 5% DMSO, 0.1 mM EDTA, and 0.3 mM DTT. DNA concentration was 1.5×10^{-12} M.

[b] For LLhP variants, low affinity conditions were in the absence of hypoxanthine. For Lacl, low affinity conditions were in the presence of 1mM IPTG.

[c] For LLhP variants, high affinity conditions included a saturating concentration of hypoxanthine (see Materials and Methods). For Lacl, no effector was present.

[d] Low affinity binding divided by high affinity binding.

[e] Values from (9). Lacl measurements were also repeated in parallel with LLhP and showed similar results.

[f] NR: No response to the addition of hypoxanthine.

[g] NBD: Very little, if any, binding detected.

[h] The same buffer as solution scattering experiments: 0.12 mM Hepes-KOH, pH 7.6, 200 mM KCl, 5% glycerol, 1 mM EDTA, and 0.3 mM DTT.

Table 3

Comparison of Experimental and Model Structural Parameters for LLhP Chimera and R3: Effects of DNA and Effector Binding

Experimentally derived parameters						Model parameters			
Sample	**conc. (M×10^4)**	**R_c (Å)**	**R_g (Å)**	**d_{max} (Å)**	**Fit[a]**	**R_c**	**R_g**	**d_{max}**	**χ^2**
LLhP Homodimer									
Apo-LLhP	0.24	18.5±0.7	31.1±0.5	90	0.97	18.7[b]	30.2[b]	90[b]	0.76[b]
LLhP-*lacO*^{1}	0.70	-	32.7±0.3	100	0.98	-	-	-	-
LLhP-*lacO*^{1}-HX	0.70	-	33.0±0.2	100	0.98	-	32.7[c]	100[c]	0.88[c]
LacI Homodimer									
Apo-R3	0.32	16.8±0.3	35.8±0.5	110	0.96	16.7[d]	29.8[d]	90[d]	1.8[d]
						-	34.0[e]	110[e]	1.0[e]
R3-*lacO*^{1}	0.72	-	34.6±0.3	110	0.96	-	34.1[f]	110[f]	1.3[f]
R3-*lacO*^{1}-IPTG	0.80	-	35.6±0.2	110	0.97	[g]	-	-	-

[d,e,f] Parameters taken from reference (23)

[a] Total quality estimate defined in GNOM (49), equal to 1.0 for an ideal solution. Model parameters are given for

[b] Model of apo-LLhP (see Materials and Methods).

[c] Model of DNA-bound LLhP (see Materials and Methods).

[d] is the model of the apo-LacI dimer taken from the one homodimer of the tetrameric LacI crystal structure (1lbg chains A and B; (41))

[e] is the LacI R3 dimer model optimized by refinement against the scattering data that shows the extended configuration (as opposed to the folded hinge-helix) for the linker sequence connecting the DBD and regulatory domains

[f] is the model for R3-*lacO*^{1} taken without modification from the 1LBG crystal structure (1lbg chains A, B, E, and F). Note that this latter structure is missing 4 base pairs from the DNA sequence (17 basepairs instead of 21) and 11 amino acid residues from the protein, resulting in a somewhat higher than ideal χ^2 value.

[g] No high-resolution structures are available for the LacI-*lacO*^{1}-IPTG complex.

Table 4

Operator capture by LLhP and variants ($\times 10^{-7}$ M hypoxanthine)[a].

	lacO^{1}	*lacO*^{sym}	*lacO*^{disC}
LLhP	4.1 ± 0.7	5.5 ± 2.3	3.7 ± 0.8
I48S	9.5 ± 0.7	6.7 ± 0.6	16 ± 1.8
I48V	8.8 ± 0.6	5.7 ± 0.5	33 ± 1.6
Q55T	10 ± 0.8	4.3 ± 0.4	21 ± 1.6
Q55V	10 ± 1.3	8.5 ± 0.6	22 ± 0.8
S61C	47 ± 4.1	n.d.[b]	80 ± 2.7
S61M	59 ± 3.4	n.d.	117 ± 12

[a] The midpoints of the operator capture experiments were determined from 3−4 independent measurements, using two different preparations of protein. Reported errors represent one standard deviation. Buffer was 10 mM Tris-HCl, pH 7.4, 150 mM KCl, 5% DMSO, 0.1 mM EDTA, and 0.3 mM DTT. The DNA concentration was ~2×10^{-12} M and hypoxanthine concentrations were varied. Protein concentrations were chosen so that the final conditions were in the range of 70−90% saturation on affinity binding curves determined in the presence of hypoxanthine: concentrations were 1.2 to 3.3×10^{-8} M I48S; 2×10^{-9} to 1.2×10^{-8} M I48V; 1.5×10^{-9} to 2.5×10^{-9} M Q55T; 3×10^{-9} to 3×10^{-8} M Q55V; and 6×10^{-9} M S61C and S61M.

[b] Not determined, because affinity experiments showed no allosteric response and very weak affinity (Table 2).