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 Boulevard, The University of Kansas Medical Center, Kansas City, Kansas 66160, Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, and School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia

Received March 14, 2008; Revised Manuscript Received May 2, 2008

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. When DNA is released, LacI shows a 20 Å increase in length that was previously attributed to unfolding of 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 exhibit 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.

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 engineering. Functionally significant polymorphisms may occur at nonconserved positions, which are often termed "specificity determinants" (e.g., refs *1* and *2*). Furthermore, functionally important positions are not limited to ligand binding sites (e.g., ref *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[1] 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 a high affinity for operator DNA, and binding the effector IPTG weakens 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 *Escherichia coli* not seen for LacI (*4*). One possibility is that DNA binding is enhanced

[†] Support to L.S.-K. was provided by NIH Grant P20 RR17708 from the Institutional Development Award program of the National Center for Research Resources and NIH Grant GM079423. Support to J.T. 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 DE-FG02-05ER64026.

* To whom correspondence should be addressed. E-mail: lswint-kruse@kumc.edu. Phone: (913) 588-0399. Fax: (913) 588-7440.

[‡] The University of Kansas Medical Center.

[§] Current address: Department of Biochemistry and Cell Biology, MS 140, Rice University, Houston, TX 77005.

[||] University of Utah.

[⊥] Current address: Institute of Chemical Kinetics and Combustion, Novosibirsk 630090, Russia.

[@] University of Sydney.

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

10.1021/bi800443k CCC: $40.75
Published on Web 07/11/2008

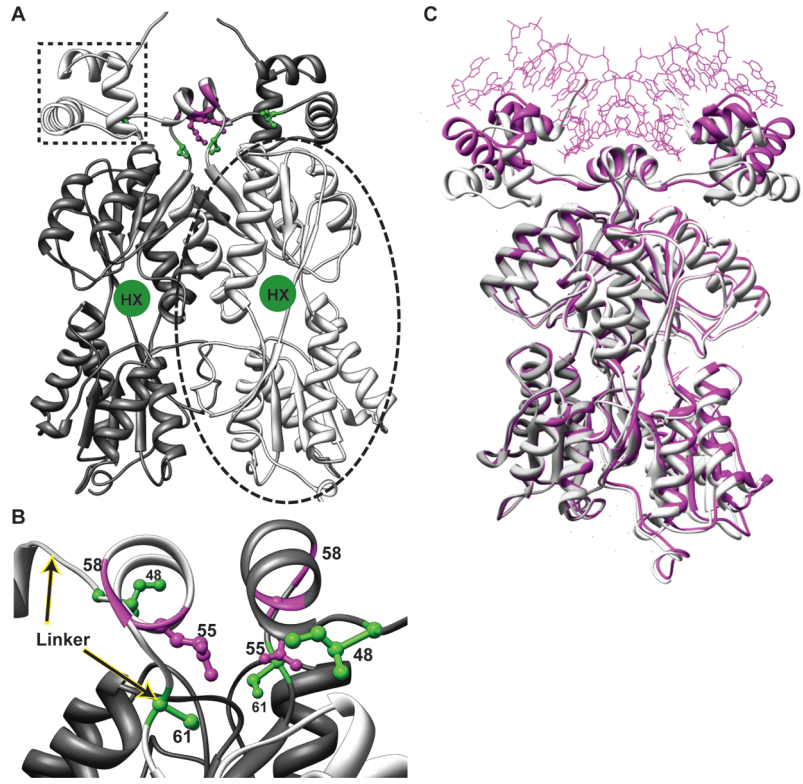

FIGURE 1: Model of the LLhP homodimer and positions of specificity determinants. (A) The model of apo-LLhP was constructed as described in Materials and Methods and rendered with Chimera (*50*). One monomer is colored light gray and the other dark gray. One DNA-binding domain is indicated with the dotted 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 corepressor hypoxanthine is indicated with a green circle in a cleft between the N- and C-subdomains (which are at the top and bottom of the regulatory domain, respectively). (B) The linkers of the LLhP dimer are shown, with the beginning and end indicated with arrows. The orientation is rotated ~90° relative to panel 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 structure. (C) The model for apo-LLhP (white) was aligned with the magenta structure of corepressed PurR bound to DNA [Protein Data Bank entry 1wet (*28*)] using Chimera (*50*). The corepressor is not shown. 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.

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 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 monitored only 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 eight 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 those of LacI and (b) to determine 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 LacI does. 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 those of 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.

MATERIALS AND METHODS

Purification of LLhP Variants. Design and construction of the LLhP coding sequence are reported in ref *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 a SDS−PAGE gel. *E. coli* strains normally used to express LacI for purification were not suitable for LLhP. The protein could not be expressed 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 copurify 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 prevent the epigenetic shutdown of LLhP expression seen previously (*4*). All growths utilized cells from a fresh transformation. Instead of plating the cells, we used them to directly inoculate a 500 mL LB culture contained in 2 L flasks. Cells were grown at 37 °C for approximately 24 h prior to centrifugation at 11000*g*. 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 medium, and stored at −20 °C.

The protocol for LLhP purification was based upon that for LacI (e.g., ref *9*), modified to use buffers in which PurR maintains high activity (e.g., ref *10*). For purification, cell paste from 3 L of culture was incubated on ice for ∼15 min. 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 7700*g* 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 7700*g* for 40 min and resuspended in 35 mL of breaking buffer. Resuspended precipitate was dialyzed 20−30 min against 1 L of breaking buffer for each of three buffer exchanges. A subsequent 30 min centrifugation at 7700*g* 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 from 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. Via 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, 10000; Vivascience, Stonehouse, U.K.) so that it could be loaded into the 2 mL injection loop of an Amersham fast performance liquid chromatography (FPLC) system. 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 0.5−1 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, 10000), which was centrifuged at 2600*g* and room temperature. In this step, temperature is very important, since highly concentrated LLhP protein precipitates at colder temperatures like 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 ∼2 mL using a VIVASPIN 20 concentrator and loaded onto a 5 mL heparin column that was previously equilibrated with buffer A. After the sample was washed with 10 column volumes of 88% buffer A and 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 A. Artigues at The University of Kansas Medical Center using an ESI FTICR mass spectrometer (LTQ FT, ThermoFinnigan, Waltham, MA). LLhP samples required extensive desalting in 0.1% formic acid (five dilutions with a water/formic acid mixture) to produce a reasonable mass value. The expected molecular mass of an LLhP monomer is 38351.64 Da; measured mass values were 38359.8 and 38359.4 Da. The expected molecular mass of a LacI monomer is 38590 Da for the 109A polymorphism (*12*) or 38620 Da 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 (L. Swint-Kruse and H. 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 6.90×10^4 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 coefficient 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 corepressors,

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*).

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 [*lacO*^{1}, *lacO*^{sym}, and *lacO*^{disC} (Table 1)] (*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 $MgCl_2$, and 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, the operator DNA concentration was at least 10-fold higher than the K_d (*20*); $3-5 \times 10^{-8}$ M was typically used. Activities were determined for LLhP, I48S, I48V, Q55T, and Q55S in the absence of hypoxanthine using *lacO*^{1}. To achieve 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, corepressor was always added to protein after DNA to maintain an activity of >90%. Adding hypoxanthine to the protein prior to DNA reduced the activity of LLhP to ~60%. We surmise that binding corepressor 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 radiolabeled DNA. If LacI inducer IPTG was present, its concentration was 1 mM.

In assays for quantitating the DNA binding affinities, the concentration of DNA ligand was at least 10-fold below the K_d and the protein concentration was varied (*20*). DNA binding affinity was measured in the absence and presence of $2-4 \times 10^{-5}$ M corepressor, which was either hypoxanthine or guanine. Although the hypoxanthine concentration was only 2-fold higher than the affinity of PurR for hypoxanthine (*21, 22*), operator capture experiments indicate that changes in DNA binding affinity induced by corepressor 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.

Igor Pro (Wavemetrics) was used to fit data for affinity assays using the equation

$$Y_{obs} = Y_{max} \times \frac{[protein]^n}{K_d^{\,n} + [protein]^n} + c \qquad (1)$$

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 two separate preparations, we are able to detect 3−5-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 (Figure 3 of the Supporting Information), experiments were performed on the same filter and with the same dilution of radiolabeled DNA, 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 ref *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 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 the cross section, R_c, and extrapolated $I(0)_c$ values were calculated using the GNOM option for cross-sectional analysis of elongated (rodlike) 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

Table 2: Operator Binding by LLhP Variants[a]

	lacO^{1}			*lacO*^{sym}			*lacO*^{disC}		
	low-affinity[b] K_d ($\times 10^{-11}$ M)	high-affinity[c] K_d ($\times 10^{-11}$ M)	allosteric ratio[d]	low-affinity K_d ($\times 10^{-11}$ M)	high-affinity K_d ($\times 10^{-11}$ M)	allosteric ratio	low-affinity K_d ($\times 10^{-11}$ M)	high-affinity K_d ($\times 10^{-11}$ 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[f]	>10000	>10000	NR[f]
G58T	>10000	>10000	NR[f]	>10000	>10000	NR[f]	>10000	>10000	NR[f]
S61C	≥20000	15 ± 4.3	~1000	>10000	>10000	NR[f]	NBD[g]	66 ± 5.1	>1000
S61M	≥30000	28 ± 3.4	~1000	>10000	>10000	NR[f]	NBD[g]	126 ± 9.8	>1000
LLhP Hepes buffer[h]	1780 ± 110	13 ± 1.9	136 ± 12						

[a] DNA binding data were determined from three or four independent measurements, using two different preparations of protein. Reported errors represent one standard deviation. The buffer used for DNA binding measurements consisted of 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 1.5×10^{-12} M. [b] For LLhP variants, low-affinity conditions were in the absence of hypoxanthine. For LacI, low-affinity conditions were in the presence of 1 mM IPTG. [c] For LLhP variants, high-affinity conditions included a saturating concentration of hypoxanthine (see Materials and Methods). For LacI, no effector was present. [d] Low-affinity binding divided by high-affinity binding. [e] Values from ref *9*. LacI measurements were also repeated in parallel with LLhP and showed similar results. [f] No response to the addition of hypoxanthine. [g] Very little, if any, binding detected. [h] The same buffer that was used in solution scattering experiments: 0.12 mM Hepes-KOH (pH 7.6), 200 mM KCl, 5% glycerol, 1 mM EDTA, and 0.3 mM DTT.

to protein concentration in milligrams per milliliter 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, 10000; Vivascience) 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 min 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 (204100 L mol^{-1} cm^{-1}) and assuming 100% yield in annealing. Analysis of the forward scattering, $I(0)$, values normalized by protein concentration, in milligrams per milliliter, 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 ratios of the normalized $I(0)$ values for LLhP, the LLhP−*lacO*^{1} complex, and the LLhP−*lacO*^{1}−HX complex to a lysozyme standard were within error of the expected value of 1.0, specifically 1.0 ± 0.2, 0.8 ± 0.2, and 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 94223 for LLhP and 106084 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 ref *23*. The Porod estimated volumes were all ~100000 Å^3.

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 MODELER (*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 midway between that of the PurR−DNA−corepressor complex 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

Table 3: Effects of DNA and Effector Binding in a Comparison of Experimental and Model Structural Parameters for LLhP Chimera and R3

	experimentally derived parameters					model parameters			
sample	concn ($\times 10^4$ M)	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]	−	−	−

[a] Total quality estimate defined in GNOM (*49*), equal to 1.0 for an ideal solution. [b] Model of apo-LLhP (see Materials and Methods). [c] Model of DNA-bound LLhP (see Materials and Methods). [d] Parameters taken from ref *23*. This 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] Parameters taken from ref *23*. This 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] Parameters taken from ref *23*. This is the model for the R3−*lacO*^{1} complex taken without modification from the 1lbg crystal structure (chains A, B, E, and F). Note that this latter structure is missing 4 bp from the DNA sequence (17 bp 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	nd[b]	80 ± 2.7
S61M	59 ± 3.4	nd[b]	117 ± 12

[a] The midpoints of the operator capture experiments were determined from three or four independent measurements, using two different preparations of protein. Reported errors represent one standard deviation. The buffer consisted of 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 of 1.2−3.3 × 10^{-8} M I48S, 2 × 10^{-9} to 1.2 × 10^{-8} M I48V, 1.5−2.5 × 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).

coordinates for the 22-mer *lacO*^{sym} DNA sequence from the NMR structure 1cjg (*38*), truncating base pair 22 to match the size of 21-mer *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 the Corepressor. Allosteric response can be quantitated as the ratio of affinities for DNA in the presence and absence of corepressor. 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 corepressor. A fixed concentration of LLhP protein (see the footnotes of 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 a zero net change in the amount of radiolabeled DNA retained on the filter and thus is silent in the experiment.

Protein was first incubated for 25−30 min with ~2 × 10^{-12} M DNA ligand in the filter binding buffer described above. Corepressor hypoxanthine was added to LLhP−DNA samples (varied between 10^{-9} and 10^{-4} M) and incubated for an additional 25−30 min. Since corepressor binding enhances the affinity of LLhP for DNA, an increased concentration of corepressor resulted in an increased amount of LLhP−DNA complex formed (*21*). The amount of radiolabeled operator bound to LLhP and retained on the nitrocellulose filter was quantified using a Fuji phosphorimager. The resulting data were fit with Igor Pro (Wavemetrics) to a modified version of eq 1, with [protein] replaced with [corepressor] and K_d replaced with $[\text{corepressor}]_{mid}$. The latter parameter is the concentration of corepressor 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 nonspecific 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 results of our previous studies of LacI. Functionally, the 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 corepressor and (b) by the amount of corepressor required to shift the protein from the low- to high-affinity state.

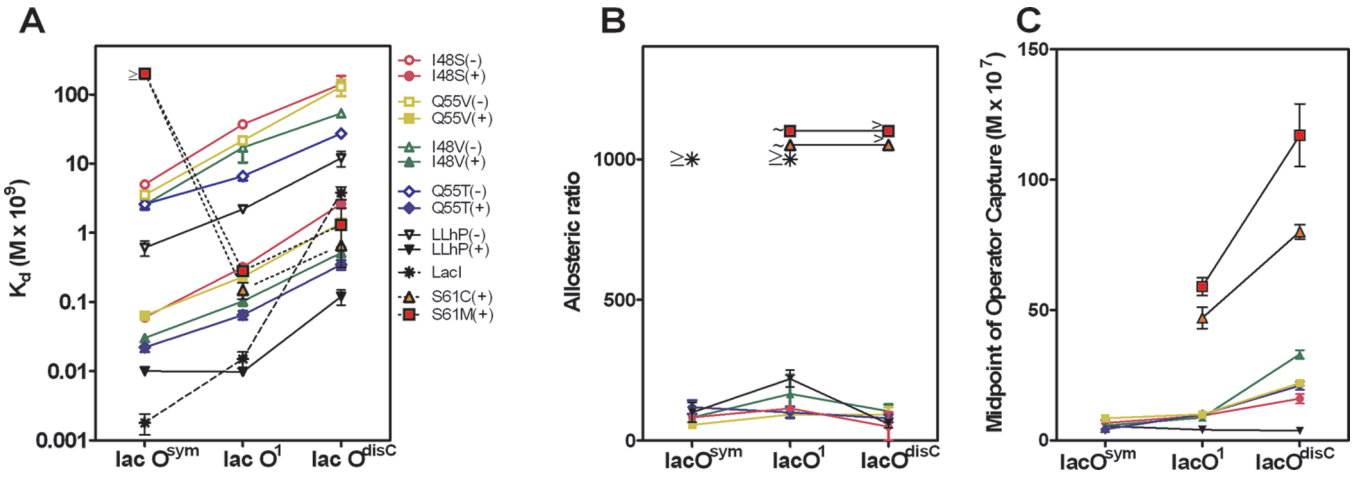

FIGURE 2: Trends in K_d and allosteric response for LLhP variants binding to alternative *lac* operators. The symbols and colors in the legend are used for all three panels. (A) + indicates the presence of saturating hypoxanthine. (A) Affinities of LacI and LLhP variants for $lacO^1$, $lacO^{sym}$, and $lacO^{disC}$. LacI (dashed black line) shows stronger discrimination among the three operators than LLhP does (solid black line). With the exception of the LLhP−HX−$lacO^{sym}$ complex, 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 the 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.

Operator Binding of LLhP Variants; Comparison with That of 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. The operator $lacO^1$ is a naturally occurring operator that is present in vivo (*19*). The operator $lacO^{sym}$ is an "optimized" version of $lacO^1$, with perfect symmetry of the tight-binding, proximal half-site and no central base pair (*17, 18*). The operator $lacO^{disC}$ 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 have the following rank order: $lacO^{sym} < lacO^1 < lacO^{disC}$ (*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 (*9*).[2]

Because of their opposite allosteric response, HX-bound LLhP (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 and Figure 1 of the Supporting Information). LLhP binding to $lacO^1$ is weakened >200-fold upon the loss of corepressor hypoxanthine (HX). Differences between LacI and LLhP emerged when the operator was $lacO^{sym}$ (Table 2 and Figure 1 of the Supporting Information). LacI binds $lacO^{sym}$ nearly 10-fold more tightly than $lacO^1$ (*9, 16, 17*). However, the 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 and Figure 1 of the Supporting Information). 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 an ~1 order of magnitude difference among the three operator sequences (Figure 2A, solid black line), whereas LacI binding differs nearly 3 orders of magnitude (Figure 2A, dashed black line).

Next, DNA binding experiments were carried out for eight 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 and Figure 2 of the Supporting Information). In each case, the K_d rank order for alternative DNA sequences is as follows: $lacO^{sym} < lacO^1 < lacO^{disC}$. Results for the four 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 that 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 variants at positions 48 and 55 and unmodified LLhP, and for all DNA operators (Table 2 and Figure 2B). Thus, these substitutions appear to alter the overall affinity but not the selectivity for alternative DNA ligands. The only instance in which the parallel pattern is broken is for unmodified LLhP+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 weaken binding to any of the three operators (Table 2 and Figure 3 of the Supporting Information), consistent with their poor

[2] In LLhP binding assays, the pH and ionic strength of the current buffer conditions (see the footnotes of Table 2) match those of previous LacI studies (e.g., ref *9*). Under these conditions, the weak K_d associated with the LacI−IPTG species 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.

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 nonspecific 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 exhibit fairly good (albeit reduced) affinity for $lacO^1$ and $lacO^{disC}$ in the presence of hypoxanthine (Table 2 and Figure 4 of the Supporting Information). 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 was the fact that 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 (Figure 4B of the Supporting Information). 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 monodisperse and free of interparticle interference effects, as required for accurate determination of structural parameters. Satisfactory conditions were identified for apo-LLhP, the LLhP−$lacO^1$ complex, and the LLhP−HX−$lacO^1$ complex 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 the allosteric response is comparable (Table 2).

Results of $P(r)$ analyses are shown in Table 3 and Figure 3. The general features of the $P(r)$ analyses 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 observed for the LacI "R3" dimer,[3] 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 concentrations. For concentrations of ≤3.7 mg/mL, the LLhP

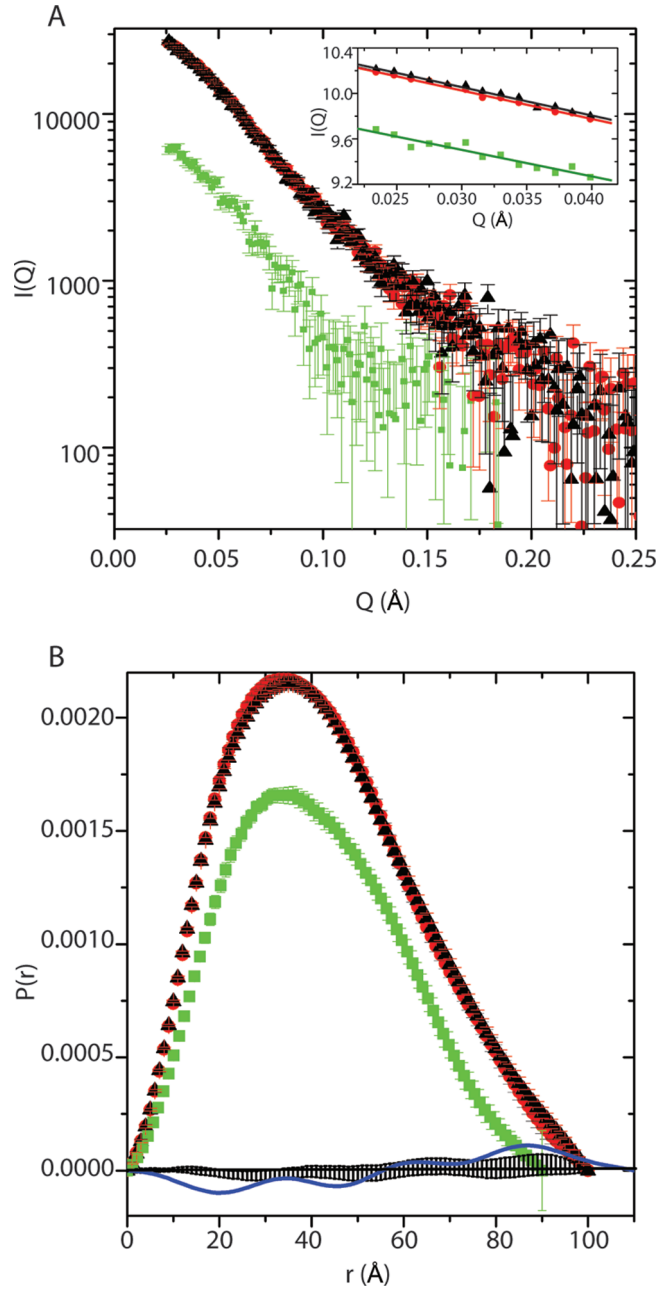

FIGURE 3: Solution X-ray scattering by LLhP. (A) $I(Q)$ vs Q for apo-LLhP (green), the LLhP−$lacO^1$ complex (red), and the LLhP−$lacO^1$−HX complex (black). The inset shows a Guinier plot of the low-Q data and the linear fits obtained (green for apo, red for LLhP−$lacO^1$, and black for LLhP−$lacO^1$−HX). 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) $P(r)$ vs r (large symbols) calculated using the data from panel A, using the same key. Difference plots are also shown with small symbols near the origin for the $P(r)$ profiles of the LLhP−$lacO^1$ complex 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 ref *23*) (blue).

[3] 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.

samples were monodisperse as determined by Guinier (Figure 3A, inset) and $I(0)$ analyses (see Materials and Methods).

Notably, apo-LLhP is significantly less elongated than the the apo-LacI R3 dimer. R_g and d_{max} values for apo-LLhP are 4−5 and 20 Å smaller, respectively, 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 Materials 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 R_g and d_{max} values much larger than those of 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 those of the apo-LacI dimer are consistent with apo-LLhP maintaining a compact hinge helix in the linker.

When LLhP binds *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 the LLhP−*lacO*^{1} complex causes no measurable differences in R_g, d_{max}, or the $P(r)$ profiles (Table 3 and Figure 3B, small black symbols near the origin). This observation again contrasts with 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 on the basis of 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−corepressor complex) than LacI does (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 subdomain 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 of regulatory subdomain reorientations upon hypoxanthine binding to the LLhP−DNA complex.

Functional Aspects of the Allosteric Response in LLhP and Its Variants. 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 and 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 levels, with corepressor added in increasing concentrations. As hypoxanthine binds, the protein acquires increased affinity for radiolabeled 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 and 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 Figures 1B and Figure 5 of the Supporting Information. For "wild-type" LLhP, curves are superimposable for three different operator sequences, with a midpoint around 4.4×10^{-7} M hypoxanthine. Variants at positions 48 and 55 show midpoints similar to those of 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,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 the 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 underappreciated 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 the 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., refs *44* and *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 1 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 corepressed 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 the crystal structures of truncated apo-PurR and the PurR−DNA−corepressor complex are compared (*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 by comparing structures of apoprotein and the ternary complex; the relevant changes must be seen by comparing the structures of the ternary and (unavailable) binary PurR−HX complex. (c) Exchanging the PurR and LacI linker sequences greatly affects the allosteric conformational change in the PurR regulatory domain. All three scenarios have important implications: the first two for designing effective studies of allostery and 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 the LLhP−*lacO*^{1} species 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 S61C and S61M substitutions 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 corepressed 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 an opportunity to 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 the "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 would not 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 position 61 might have interesting in vitro properties. However, when new repressors are engineered 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 for the repressors to be effective in vivo; even though S61C and S61M have reasonable DNA affinity for *lacO*^{1} with an 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. This work provides an 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 data set 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 exhibit 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 via SDS−PAGE, so these proteins are in vast excess of the single

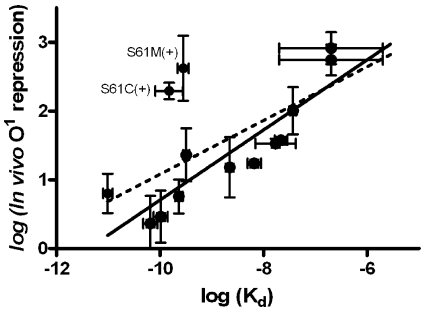

FIGURE 4: In vivo repression vs in vitro DNA binding affinity for LLhP variants. In vivo data are reported in ref *4*. Data shown in this plot include measurements made in both the presence and absence of corepressor for LLhP variants in the current study. The dashed line reflects the best fit (linear 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 in this plot. However, the G58 variants show good correlation between poor repression and weak DNA binding.

genomic binding site (*4*). Instead, this work shows that these proteins require more hypoxanthine 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 corepressor, the concentration could be constrained at a level too low for corepression of the S61 LLhP variants. Since the operator capture midpoints for S61C and S61M are only 5−10-fold different from those of 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 100−200-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., refs *47* and *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.

ACKNOWLEDGMENT

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

SUPPORTING INFORMATION AVAILABLE

DNA binding curves for unmodified LLhP and operator capture for unmodified LLhP (Figure 1), DNA binding curves for LLhP variants at sites 48 and 55 (Figure 2), DNA binding curves for LLhP variants at site 58 (Figure 3), DNA binding curves for LLhP variants at site 61 (Figure 4), and operator capture by LLhP variants (Figure 5). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Pei, J., Cai, W., Kinch, L. N., and Grishin, N. V. (2006) Prediction of functional specificity determinants from protein sequences using log-likelihood ratios. *Bioinformatics 22*, 164–171.
2. Kalinina, O. V., Novichkov, P. S., Mironov, A. A., Gelfand, M. S., and Rakhmaninova, A. B. (2004) SDPpred: A tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. *Nucleic Acids Res. 32*, W424–W428.
3. Pawlyk, A. C., and Pettigrew, D. W. (2002) Transplanting allosteric control of enzyme activity by protein-protein interactions: Coupling a regulatory site to the conserved catalytic core. *Proc. Natl. Acad. Sci. U.S.A. 99*, 11115–11120.
4. Tungtur, S., Egan, S., and Swint-Kruse, L. (2007) Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins: Struct., Funct., Bioinf. 68*, 375–388.
5. Weickert, M. J., and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem. 267*, 15869–15874.
6. Barkley, M. D., Riggs, A. D., Jobe, A., and Burgeois, S. (1975) Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry 14*, 1700–1712.
7. Meng, L. M., and Nygaard, P. (1990) Identification of hypoxanthine and guanine as the co-repressors for the purine regulon genes of *Escherichia coli*. *Mol. Microbiol. 4*, 2187–2192.
8. Swint-Kruse, L., Larson, C., Pettitt, B. M., and Matthews, K. S. (2002) Fine-tuning function: Correlation of hinge domain interactions with functional distinctions between LacI and PurR. *Protein Sci. 11*, 778–794.
9. Zhan, H., Swint-Kruse, L., and Matthews, K. S. (2006) Extrinsic Interactions Dominate Helical Propensity in Coupled Binding and Folding of the Lactose Repressor Protein Hinge Helix. *Biochemistry 45*, 5896–5906.
10. Moraitis, M. I., Xu, H., and Matthews, K. S. (2001) Ion concentration and temperature dependence of DNA binding: Comparison of PurR and LacI repressor proteins. *Biochemistry 40*, 8109–8117.
11. Lu, F., Brennan, R. G., and Zalkin, H. (1998) *Escherichia coli* purine repressor: Key residues for the allosteric transition between active and inactive conformations and for interdomain signaling. *Biochemistry 37*, 15680–15690.
12. Bairoch, A., and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res. 28*, 45–48.
13. Holmquist, B., and Vallee, B. L. (1973) Tryptophan quantitation by magnetic circular dichroism in native and modified proteins. *Biochemistry 12*, 4409–4417.
14. Riggs, A. D., Bourgeois, S., Newby, R. F., and Cohn, M. (1968) DNA binding of the lac repressor. *J. Mol. Biol. 34*, 365–368.
15. Wong, I., and Lohman, T. M. (1993) A double-filter method for nitrocellulose-filter binding: Application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci. U.S.A. 90*, 5428–5432.
16. Falcon, C. M., and Matthews, K. S. (2001) Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. *Biochemistry 40*, 15650–15659.

17. Simons, A., Tils, D., von Wilcken-Bergmann, B., and Muller-Hill, B. (1984) Possible ideal lac operator: *Escherichia coli* lac operator-like sequences from eukaryotic genomes lack the central G•C pair. *Proc. Natl. Acad. Sci. U.S.A. 81*, 1624–1628.
18. Sadler, J. R., Sasmor, H., and Betz, J. L. (1983) A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. U.S.A. 80*, 6785–6789.
19. Gilbert, W., and Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. U.S.A. 70*, 3581–3584.
20. Swint-Kruse, L., and Matthews, K. S. (2004) Thermodynamics, protein modification, and molecular dynamics in characterizing lactose repressor protein: Strategies for complex analyses of protein structure-function. *Methods Enzymol. 379*, 188–209.
21. Lu, F., Schumacher, M. A., Arvidson, D. N., Haldimann, A., Wanner, B. L., Zalkin, H., and Brennan, R. G. (1998) Structure-based redesign of corepressor specificity of the *Escherichia coli* purine repressor by substitution of residue 190. *Biochemistry 37*, 971–982.
22. Choi, K. Y., and Zalkin, H. (1992) Structural characterization and corepressor binding of the *Escherichia coli* purine repressor. *J. Bacteriol. 174*, 6207–6214.
23. Taraban, M., Zhan, H., Whitten, A. E., Langley, D. B., Matthews, K. S., Swint-Kruse, L., and Trewhella, J. (2008) Ligand-induced Conformational Changes and Conformational Dynamics in the Solution Structure of the Lactose Repressor Protein. *J. Mol. Biol. 376*, 466–481.
24. Svergun, D. I. (1991) Mathematical methods in small-angle scattering data analysis. *J. Appl. Crystallogr. 24*, 485–492.
25. Guinier, A., and Fournet, G. (1955) *Small-Angle Scattering of X-rays*, pp 128−129, John Wiley and Sons, New York.
26. Porod, G. (1951) Die Roentgenkleinwinkel-Steuung von Dicht-gepackten Kolloiden Systemen, I Teil. *Kolloid Z. Biol. 124*, 83–111.
27. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRYSOL 195: A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr. 28*, 768–773.
28. Schumacher, M. A., Glasfeld, A., Zalkin, H., and Brennan, R. G. (1997) 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. 272*, 22648–22653.
29. Rai, B. K., Madrid-Aliste, C. J., Fajardo, J. E., and Fiser, A. (2006) MMM: A sequence-to-structure alignment protocol. *Bioinformatics 22*, 2691–2692.
30. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol. 234*, 779–815.
31. Shindyalov, I. N., and Bourne, P. E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng. 11*, 739–747.
32. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187–217.
33. Swint-Kruse, L., and Brown, C. S. (2005) Resmap: Automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics 21*, 3327–3328.
34. Swint-Kruse, L., Elam, C. R., Lin, J. W., Wycuff, D. R., and Shive Matthews, K. (2001) Plasticity of quaternary structure: Twenty-two ways to form a LacI dimer. *Protein Sci. 10*, 262–276.
35. Schumacher, M. A., Choi, K. Y., Lu, F., Zalkin, H., and Brennan, R. G. (1995) Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell 83*, 147–155.
36. Swint-Kruse, L. (2004) Using networks to identify fine structural differences between functionally distinct protein states. *Biochemistry 43*, 10886–10895.
37. Bell, C. E., and Lewis, M. (2000) A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol. 7*, 209–214.
38. Spronk, C. A., Bonvin, A. M., Radha, P. K., Melacini, G., Boelens, R., and Kaptein, R. (1999) The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure 7*, 1483–1492.
39. Lin, S., and Riggs, A. D. (1975) A comparison of lac repressor binding to operator and to nonoperator DNA. *Biochem. Biophys. Res. Commun. 62*, 704–710.
40. Chen, J., Alberti, S., and Matthews, K. S. (1994) Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. *J. Biol. Chem. 269*, 12482–12487.
41. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science 271*, 1247–1254.
42. Mowbray, S. L., and Björkman, A. J. (1999) Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. *J. Mol. Biol. 294*, 487–499.
43. Vigil, D., Blumenthal, D. K., Taylor, S. S., and Trewhella, J. (2006) Solution scattering reveals large differences in the global structures of type II protein kinase A isoforms. *J. Mol. Biol. 357*, 880–889.
44. Yoshikuni, Y., and Keasling, J. D. (2007) Pathway engineering by designed divergent evolution. *Curr. Opin. Chem. Biol. 11*, 233–239.
45. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) The evolvability′ of promiscuous protein functions. *Nat. Genet. 37*, 73–76.
46. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (2003) Parallel evolution of ligand specificity between LacI/GalR family repressors and periplasmic sugar-binding proteins. *Mol. Biol. Evol. 20*, 267–277.
47. Bashton, M., and Chothia, C. (2007) The Generation of New Protein Functions by the Combination of Domains. *Structure 15*, 85–99.
48. Poelwijk, F. J., Kiviet, D. J., and Tans, S. J. (2006) Evolutionary potential of a duplicated repressor-operator pair: Simulating pathways using mutation data. *PLoS Comput. Biol. 2*, e58.
49. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr. 25*, 495–503.
50. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.

BI800443K