Published in final edited form as:

Proteins. 2008 December ; 73(4): 941–957. doi:10.1002/prot.22121.

Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: Bioinformatics-based predictions generate true positives and false negatives

Sarah Meinhardt[¶] 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

Abstract

In protein families, conserved residues often contribute to a common general function, such as DNA-binding. However, unique attributes for each homologue (*e.g.* recognition of alternative DNA sequences) must arise from variation in other functionally-important positions. The locations of these “specificity determinant” positions are obscured amongst the background of varied residues that do not make significant contributions to either structure or function. To isolate specificity determinants, a number of bioinformatics algorithms have been developed. When applied to the LacI/GalR family of transcription regulators, several specificity determinants are predicted in the 18 amino acids that link the DNA-binding and regulatory domains. However, results from alternative algorithms are only in partial agreement with each other. Here, we experimentally evaluate these predictions using an engineered repressor comprising the LacI DNA-binding domain, the LacI linker, and the GalR regulatory domain (LLhG). “Wild-type” LLhG has altered DNA specificity and weaker *lacO*^I repression compared to LacI or a similar LacI:PurR chimera. Next, predictions of linker specificity determinants were tested, using amino acid substitution and *in vivo* repression assays to assess functional change. In LLhG, all predicted sites are specificity determinants, as well as three sites not predicted by any algorithm. Strategies are suggested for diminishing the number of false negative predictions. Finally, individual substitutions at LLhG specificity determinants exhibited a broad range of functional changes that are not predicted by bioinformatics algorithms. Results suggest that some variants have altered affinity for DNA, some have altered allosteric response, and some appear to have changed specificity for alternative DNA ligands.

Keywords

lactose repressor protein; galactose repressor protein; allostery; LacI/GalR family; transcription repression; protein engineering

Introduction

Protein families can be identified by their related sequences, which often correlate with similarities in general structures and functions. Conversely, the unique functional attributes of an individual protein must be conveyed by positions that are not conserved in sequence alignments. Identifying these positions (“specificity determinants”) is key to protein engineering and to full use of data generated by the various genome projects. However, identification of specificity determinants is difficult. In sequence alignments, they are obscured

*To whom correspondence should be addressed. lswint-kruse@kumc.edu, 913-588-0399 phone. 913-588-9896 fax.

amongst a background of nonconserved residues that have no structural or functional roles^a. Structure/function studies of individual proteins cannot discriminate between specificity determinants and the conserved residues required for the common function of family members.

Thus, identification of specificity determinants requires a combinatorial approach. To that end, we combined analyses of structural, mutational, and sequence data to hypothesize the locations of specificity determinants in the 18 amino acids that link the DNA-binding and regulatory domains of the LacI/GalR^b proteins (Figure 1: Table 1, pink).¹ Subsequently, the LacI/GalR family was used in the development of two bioinformatics-based predictions of specificity determinants (Table 1, marked with “X”). In the first, Gelfand and co-workers subdivided sequence alignments into orthologue and paralogue groups^c prior to statistical analysis of nonconserved residues. This approach (“SDPpred”)^{2–4} incorporates functional information, since orthologues are assumed to have the same ligand specificity whereas paralogues recognize different ligands. In a second study, Grishin and colleagues attempted to minimize the reliance upon investigator-defined functional subgroups. Their algorithm (“SPEL”)⁵ first simulated evolutionary changes that could lead to observed sequence changes and then compared them to a random model, which might be expected for the sites with no evolutionary constraints. One assumption in both of the bioinformatics studies is that all proteins in a family utilize the same residue locations as specificity determinants.

The primary goal of the current work is to experimentally compare the predictive powers of the three studies described above. A second goal is to begin assessing whether positions and functional outcomes are similar for multiple homologues. If we utilized the naturally occurring homologues for these studies, interpretation of results would be complicated by the fact that each homologue recognizes a different DNA ligand.⁶ Therefore, we engineered a series of chimeras that comprise the LacI DNA-binding domain and linker fused to the regulatory domain of *E. coli* paralogues (Figure 1A,B). Most of the predicted linker specificity determinants do not directly contact DNA. Instead, these side chains interact with the regulatory domain or the linker of the partner monomer (Figure 1C–E). Thus, in the comparison set of chimeric proteins, the amino acids that directly contact DNA are unchanged, whereas linker specificity determinants have unique contexts.

We previously employed a LacI:PuR chimera (LLhP, Figure 1B) to verify that our predicted locations of four specificity determinants are correct.⁷ Here, we assess and compare the bioinformatics predictions. In the LacI/GalR linkers, all studies predict the importance of sites 55 and 58. However, the predictions disagree in regard to positions 48, 52, 59, and 61 (Table 1). One possible source of the discrepancies is that various family members could utilize alternative positions as specificity determinants. For example, substituting site 48 in LacI might alter function, whereas substitutions at the analogous position in GalR might be silent. Since our predictions¹ were strongly influenced by data for LacI and PurR, the LLhP chimera might not provide the most stringent “test-case” of family-wide specificity determinants. Thus, we designed a second chimera (named “LLhG”) using the LacI DNA-binding domain and linker and the GalR regulatory domain (Figure 1B).

Because the LacI/GalR proteins regulate transcription, function of a large number of repressor variants can be monitored using *in vivo* assays. The *in vivo* function of LLhG is clearly different from either LacI or LLhP: Repression of a downstream reporter gene *via lacO^I* is weaker and DNA-binding specificity appears to be altered. The functional contributions to LLhG from

^aWithout evolutionary constraints, these non-important residues are free to vary.

^bLacI: lactose repressor protein; GalR: galactose repressor protein; PuR: purine repressor protein; LLhP: chimera between the LacI DNA-binding domain, LacI linker, and PuR regulatory domain.

^cOrthologues are homologues that carry out the same function in different organisms. Paralogues coexist in the same organism, but carry out different functions.

predicted specificity determinants were gauged by randomly mutating each position and assessing *in vivo* changes in transcription repression. All of the predicted specificity determinants alter function when subjected to mutagenesis, regardless of the prediction method. In addition, we identified specificity determinants at positions 51, 60, and 62 that can be used to restore strong *lacO^I* repression to LLhG. These positions were not predicted by any of the previous bioinformatics studies. Thus, for the linkers of the LacI/GalR proteins, existing algorithms under-predict which non-conserved residues are functionally important.

Materials and Methods

Chimera construction

Primers for mutagenesis were purchased from Integrated DNA Technology (Coralville, IA). DNA sequencing was carried out by Northwoods DNA, Inc. (Solway, MN). LLhG was created by joining the lac DNA-binding domain and linker (residues 1-61) to the GalR core (60-343). LLhG construction paralleled that previously reported for LLhP: Primers 5' GCTGGCGCAGCAGACCTTTAAACGGTCGG 3' and 5' GCTACCTCAGGTTATTAGTCGCTGGTTGCATGATGACTTGC 3' were used to amplify only the GalR regulatory domain from the *E. coli* DH5 α genome, creating a DraI site at position 60, adding an additional stop codon, and creating a Bsu36I site at the end of the gene. The PCR product was TA cloned into pGemT vector (Promega, Madison, WI). White colonies were cultured overnight in 3ml 2xYT; plasmid DNA was purified with QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) or Quantum Prep Plasmid Miniprep Kit (Bio-Rad Laboratories, Inc, Chicago IL). Candidate plasmids were screened with restriction cuts for the appropriate insert and, if positive, sequenced using the SP6 and T7 primers.

The LacI component of LLhG was obtained in the same manner as that described for LLhP⁷: The coding region for the LacI regulatory domain was removed from the pLS1-AfeI plasmid by digestion with AfeI at codon 62 and a Bsu36I site that is downstream of the coding region. The pGemT-GalR plasmid was digested with DraI and Bsu36. Vector and insert fragments were separated by gel electrophoresis and gel purified using Montage Ultra Free column (Millipore Corp., Billerica, MA). Fragments were ligated at 16°C overnight and transformed to DH5a Max or High Efficiency cells (Invitrogen Corp., Calsbad, CA).

The coding region for LLhG did not readily ligate, unless in the presence of 40 mM galactose, which is an inducer of wild-type GalR.⁸ Under these conditions, ligation and further genetic manipulations were successful. The DraI site used to construct LLhG altered the amino acid of position 62 from an E to a K (LLhG numbering; this is position 60 in GalR). Therefore, we restored E62 in LLhG using site-directed mutagenesis. The entire coding region of LLhG was sequenced using the primers 5'GCTCGAGGTCGACGGATCCC 3' and 5' CATCAACATTAAATGTGAGC 3'. Growth in the presence of inducer galactose precluded functional studies of LLhG variants. However, we identified a fortuitous E230K substitution that did not require the presence of galactose and was previously characterized to be necessary for GalR repressosome formation but not for DNA binding.⁹ We thus decided to continue our studies with the E230K versions of LLhG. **All protein variants reported herein contain the E230K substitution.**

Next, we subcloned LLhG onto a modified version of pHG165¹⁰ called pHG165a. Subcloning utilized the EcoRI restriction sites that flank the chimera coding regions and the EcoRI site present on pHG165. This lower-copy plasmid allows reliable measurements of the β -galactosidase assay in liquid culture.¹¹⁻¹³ However, pHG165 contains a *lacO^I* binding site. Our previous work with chimera LLhP on this plasmid had very high repression of the reporter gene in *E. coli* 3.300 cells, and the extra *lacO^I* site on the pHG165 plasmid did not appear to impair that work. Since preliminary experiments with LLhG on high-copy plasmid indicated

that it was not a good repressor of *lacO^I* (as indicated by blue colonies in plate assay) we decided to remove the extra site to eliminate potential competition. This was accomplished using site-directed mutagenesis and the primer pHG-O1out (Supplementary Table 1); the subsequent plasmid was called pHG165a. Subcloning was verified by the formation of appropriate dropout bands upon digestion with SacI and ScaI. Sequences of subcloned genes were verified by sequencing the entire coding region.

All other mutants were made using site-directed mutagenesis and the primers listed in Supplementary Table 1. Random mutants were created as for LLhP.⁷ Mutagenesis was verified by sequencing the full coding region.

Determination of *in vivo* protein levels

To verify expression of full-length, soluble protein, cells from a 3 mL 2xYT overnight culture were lysed and the supernatant was analyzed with SDS-PAGE. In general, LLhG variants exhibited less soluble protein than LLhP, which could be clearly distinguished with Coomassie stain.⁷ We therefore verified the presence of soluble, full-length LLhG variants using DNA-pulldown assays. For this assay, 1 pmol of 5'-biotinylated DNA sequences (Integrated DNA Technologies, Coralville, IA; Supplementary Tables 1 and 2) containing either the naturally occurring *lacO^I* binding site¹⁴, a tight-binding operator *lacO^{sym}*^{15,16}, or a nonspecific binding site called *O^{non}*^{17,18} were coupled to each 1 µl of Streptavidin Magnetic Beads (New England Biolabs, Inc., Ipswich, MA) that had been exchanged into Buffer 1^d. DNA-beads were exchanged step-wise into Buffer 2^e and FB buffer^f. Cells from the 3 mL overnight culture were pelleted and resuspended in 0.1 ml Breaking buffer with 1 µl of 1M DTT added^g, and lysed by freeze/thaw after the addition of 40 µl of 5 mg/ml lysozyme. Supernatant was obtained by centrifugation. Ten (10) µl supernatant were incubated with 50 µl of DNA-labeled beads in FB buffer. The final concentration of immobilized DNA was $\sim 10^{-7}$ M, allowing *lacO^I* binding for even induced LacI and most of the LLhG variants that repress *lacO^I* poorly *in vivo*. Beads were subsequently washed in FB buffer, and finally resuspended in 15–20 µl of 13% SDS with 0.33 M DTT. After heating 10 minutes at 90°C, 1 µl of the final supernatant was subjected to SDS-PAGE and visualized with Coomassie stain.

We verified expression of the appropriately-sized, soluble protein by comparing results from LLhG-expressing bacteria to bacteria without the plasmid encoding the chimeras (data not shown).^h In addition, more LLhG protein was evident when the DNA contained a *lacO^I* binding site than when it contained the nonspecific sequence *O^{non}* (these proteins are expected to weakly bind non-specific DNA²⁰). For each linker position studied in the current work, the pull-down assay was carried out for the two weakest repressor variants. Other representative LLhG variants with a range of repression values were also surveyed, to ensure that changes in repression showed no correlation with *in vivo* protein concentrations. In both samples sets, most LLhG variants showed no change in the amount of soluble protein binding to immobilized *lacO^I* (as determined by comparing band intensities in SDS-PAGE normalized to a loading control; data not shown). Some variants were not efficiently bound by *lacO^I*, but protein was bound by the stronger *lacO^{sym}* binding site; exceptions are noted in Results. We approximated the number of repressors in each *E. coli* cell with the following calculation: A 10 µl aliquot of resuspended 3.300 cells gives rise to $3\text{--}12 \times 10^7$ colonies. Using the 50 ng detection-limit for Coomassie stain, volumes detailed in the protocol above, and a molecular weight of 75250 per

^dBuffer 1 – 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5 M NaCl.

^eBuffer 2 – 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.25 M NaCl.

^fFB buffer¹⁹ - 10 mM Tris-HCl pH 7.4, 150 mM KCl, 10 mM EDTA, 5% DMSO, 0.3 mM DTT.

^gBreaking buffer¹⁹ - 0.2 M Tris-HCl, 0.2 M KCl, 0.01 M MgCl₂, 5% glucose.

^hIn other experiments, we purified the band corresponding to that assigned to LLhG in the pull-down assay and used mass spectrometry to verify that the molecular weight is as expected for LLhG (data not shown; Dr. Antonio Artigues, KUMC).

dimer, we estimate between 3,000 and 13,000 repressor molecules per cell. This value is a lower limit for many of the variants, since many samples (1) show bands that are well-above the Commae detection limit and (2) serial dilutions show that beads are saturated and thus are not capturing all of the available protein.

Phenotypic analysis: Assays of β -galactosidase activity

One of the reasons for choosing a transcription repressor family to study specificity determinants is that their functions allow rapid, *in vivo* functional screening of many variants. These assays are well-established for several LacI/GalR proteins (*e.g.* 13,21–24). We have implemented two versions of repression assays that utilize the *lacO^I* binding site – plate assays provide speed, whereas liquid culture assays are quantitative. In both, low values of reporter gene activity (β -galactosidase) correlate with strong repression.

Both plate and liquid culture assays of β -galactosidase activity were performed as for LLhP⁷ using *E. coli* 3.300 cells (*E. coli* Genetic Stock Center, Yale University). This bacterial strain has an interrupted *lacI* gene but an intact genomic *lacZYA* operon controlled by the operator sequence *lacO^I*.²⁵ Plate assays utilized the blue-white indicator 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (Xgal)¹³ in standard LB plates with 100 μ g/ml ampicillin. White colonies express protein capable of repressing the *lacZYA* operon by binding *lacO^I*. If expressed protein cannot repress transcription, colonies are blue. If present, inducer galactose was 40 mM or inducer fucose was 20 mM. Control experiments utilized 3.300 cells grown in the presence of galactose or fucose with no pHG165a plasmid. These experiments showed that galactose partially inhibited the β -galactosidase colorimetric reaction but fucose did not. Thus, we used fucose for the quantitative, liquid culture assays of β -galactosidase activity.

Liquid culture assays of variants at sites 48, 52, 55, 58, 59, 61, and 62 were performed in minimal media as reported for LLhP.^{7,11–13} Each condition (in either the absence or presence of inducer) was used to generate two samples with 1x and 2x volumes of culture, respectively. The internal control for normalization was LLhG+20 mM fucose; the average daily activity of this sample was set to 100 units and used to normalize all other results. Note that the previously published LLhP results⁷ were normalized to LacI+IPTG. Average values reported for each LLhG variant were determined from 3–6 independent assays; reported errors are standard deviations of the average normalized values. We also assayed repression of LLhG variants on pHG165 with the intact *lacO^I* site. These variants demonstrated statistically equivalent repression to that of the same protein variants on pHG165a (data not shown).

For variants at sites 51 and 60, the liquid culture protocol was modified for 96-well plates, using the same reagents as above but the high-throughput strategy outlined by Griffith and Wolf.²⁶ This allowed quantification of repression by 22 variants in quadruplicate per 96-well plate (Greiner Bio-One UV-Star 96-well plates; OpticsPlanet, Inc., Northbrook, IL), with one plate in the absence of fucose and a second in the presence of fucose. Each quadruplicate measurement was repeated starting with two separate bacterial colonies; the values presented in the figures are the average of 8 normalized determinations; error is the standard deviation. As before, control colonies expressing “wild-type” LLhG were included in each day’s measurements and the (+) fucose values were used to normalize values for all other variants. Normalized values for LLhG in the absence of fucose were in good agreement between the low-and high-throughput methods (5.2 ± 2.1 and 6.2 ± 2.7 , respectively).

While LLhP required a fresh transformation for every assay, LLhG liquid culture assays were consistent using colonies from plates that were up to a week old. A few LLhG variants (noted in the text and figures) showed evidence of toxic function. In these cases, liquid cultures grew more slowly than the controls, with doubling times increased as much as two-fold. Growth rates were not enhanced by the addition of inducers.

Results

Characterization of “wild-type” LLhG function

In structures of representative LacI/GalR proteins, side chains of various linker residues interact with sites on the regulatory domains. Therefore, creation of a chimeric protein provides a new context that might alter the function of the LacI DNA-binding domain. Indeed, for LLhG, the first indication of functional change arose during chimera construction. When trying to ligate the GalR regulatory domain to the LacI DNA-binding domain, colony frequency was extremely low and any product had mutations or truncations not present in the preceding step (genomic amplification of the regulatory domain). Mutations could not be reverted with site-directed mutagenesis. However, when we included galactose in the growth media, we obtained the correct ligation products. Furthermore, colonies expressing LLhG would only grow on media containing GalR inducers galactose and fucose; 10 mM glucose or 0.8% glycerol did not substitute. Therefore, we hypothesize that LLhG is repressing *E. coli* genes essential to growth and must bind a different DNA target sequence than is normally recognized by the non-toxic, full-length LacI. DNA ligand specificity must be altered, even though the DNA-binding site residues are identical for LLhG and LacI.

While interesting, toxicity made work with LLhG very difficult. Thus, we re-examined some of the non-toxic, mutated chimeras identified during the ligation trials. One LLhG construct had a mutation corresponding to GalR E230K (Figure 1A). This variant was previously characterized in GalR as retaining ability to bind DNA but unable to build the higher-order “repressosome” (comprising two GalR dimers, DNA, and heteroprotein HU) required for full regulation of the *gal* operon.^{9,27} Structures of LacI and PurR suggest that GalR position 230 is far from the surface of the regulatory domain that interacts with the linker and is not near the effector binding site (Figure 1A).²⁷ In LacI, the homologous position at Q231 does not participate in the allosteric pathway connecting the effector- and inducer-binding sites.²⁸ Together, the GalR and LacI data suggest that the “E230K”ⁱ variant rescues LLhG toxicity by preventing it from assembling a repressosome on *E. coli* genes not regulated by LacI.

The E230K substitution is present on all LLhG chimera variants reported in the rest of this manuscript. For simplicity in the tables and figures, this mutation is not explicitly noted.

Repression assays confirmed that LLhG has altered function compared to LacI and LLhP. The latter proteins are tight repressors of *lacO*^l, producing white colonies and more than 1000-fold repression in liquid culture assays (reference 7, and data not shown). In contrast, colonies expressing LLhG were blue in *lacO*^l plate assays (Supplementary Figure). Compared to control strains with plasmids lacking a repressor gene (Figure 2, “pHG165a”), 30-fold repression was detected for LLhG in the liquid culture protocol (Figure 2, “E62”; and Figure 3, “LLhG”). LLhG is induced in the presence of GalR inducers⁸ fucose and galactose; induced values are very similar to the “no-repression” control (Figure 2, “E62” dark gray bars and “pHG165a”).

Criteria used to identify specificity determinant positions

The definition of a specificity determinant is: “A position (1) that is not conserved in a sequence alignment, and (2) for which substitution changes function without disrupting the protein’s overall fold.” The meaning of “changed function” has not been rigorously developed in the bioinformatics literature, primarily because these algorithms are limited to predicting locations. Clearly, various authors anticipate that these positions will determine which ligand is recognized by the protein. However, many other aspects of function could be altered by amino acid substitution. For a transcription repressor, function can be subdivided into DNA binding

ⁱThe actual number of GalR position 230 changes in the chimera.

affinity, DNA specificity, effector binding affinity and specificity, allosteric response, and binding to nonspecific DNA sequences.

Since a major goal of the current work is to test the predicted locations of specificity determinants, we chose *in vivo* repression assays. These assays are the aggregate of many functional aspects: Enhanced repression might result from stronger DNA binding affinity, diminished allosteric response, or diminished nonspecific binding (excess nonspecific, genomic DNA can compete with the single operator binding site). Diminished repression might result from weakened affinity for the operator or enhanced affinity for other operator-like or nonspecific sequences. Allosteric response may be assessed by monitoring repression in the presence and absence of effector. Unexpected function – such as altering potential interactions with other proteins – are also reflected in these assays. The *in vivo* repression assay therefore allows detection of specificity determinants that impact a wide range of functional aspects.

In vivo repression assays have two potential drawbacks: Repressor activity can also be changed by misfolded protein or by altered protein concentrations. However, structures available for LacI/GalR family members show that neither the N-linker nor the C-linker has regular secondary structure.^{29–36} Thus, we do not expect substitutions other than G or P (and possibly W) to affect the overall structure of these regions. Instead, we expect that alternative side-chain will have varied interactions with other regions of the protein (*e.g.* the regulatory domain). For the hinge helix, repression changes upon amino acid substitution are compared to helical propensities. Past^{7,19} and current comparisons with helical propensity detect little or no correlation with functional change.

In vivo LLhG protein concentrations were determined with DNA pull-down assays. We assayed the weakest repressors at each linker position and other variants with a range of repression activities. All but two variants (see below) have protein levels detectable by Coomassie stain, indicating *in vivo* levels that are three to four orders of magnitude greater than the single *lacO^I* binding site (see Materials and Methods). Most of the proteins have comparable expression levels, regardless of their repression activities. Although seven LLhG variants show diminished protein, the change does not correlate with the magnitude of functional change: protein concentrations are only ~4-fold less than other variants, whereas repression changes are about 100-fold. Thus, a significant amount of the repression change must be due to altered function.

This result also highlights a circular feature of this assay – variants may show diminished protein in the pull-down assay because their affinities for *lacO^I* are diminished. Therefore, we repeated the assay using the *lacO^{sym}* operator (Supplementary Table 2), which binds to LacI an order of magnitude more tightly than *lacO^I*.^{15–17,19} For six of the variants above with diminished protein levels, *lacO^{sym}* assays increased the amount of protein that could be detected. The seventh variant showed comparable protein with *lacO^I* and *lacO^{sym}* (importantly, still several orders of magnitude in excess of the single *in vivo lacO^I* binding site). The only two variants for which we could not detect soluble protein are K59E and K59P. For these variants, we cannot discriminate between diminished *in vivo* protein concentrations and diminished binding affinity for both *lacO^I* and *lacO^{sym}*.

These results lead us to conclude that, for most of the LLhG variants, diminished repression is due to altered function, most likely diminished *lacO^I* binding. Two other pieces of information support the validity of *in vivo* identification of specificity determinants: (1) The repression data for substitutions at position 52 very closely parallel *in vitro* measurements of *lacO^I* binding affinity for 15 LacI variants (see below).¹⁹ (2) All sites but 61 have substitutions that cause gain of repression. This situation more clearly confirms the designation of “specificity determinant”, since increases in already high LLhG and LLhG/E62K protein

concentrations are very unlikely to affect *in vivo* repression. Site 61 is a confirmed specificity determinant in LLhP (reference 7 and unpublished *in vitro* DNA-binding assays.)

Position 62 is an unpredicted specificity determinant in LLhG

In the process of constructing LLhG, a required restriction site altered codon 62 at the beginning of the GalR regulatory domain to express a lysine (LLhG/E62K). Our compensatory design was to mutate residue 62 back to the E of wild-type GalR. However, we also monitored the function of LLhG/E62K. Like LLhG, the E62K variant is toxic and rescued with the E230K substitution. However, the E62K variant (henceforth with E230K) is a 25-fold stronger repressor for *lacO^I* than LLhG (Figure 2). Position 62 was not predicted to be a specificity determinant by any previous study.^{1,3–5} Additional experiments with position 62 were twofold: (1) This position was varied to test the effects of other substitutions at this site. (2) We used LLhG and LLhG/E62K as “weak” and “strong” repressor backgrounds, respectively, to assay the effects of substitutions at other specificity determinants (see below).

The 13 substitutions at position 62 in LLhG (Figure 2) yield a range of repression values that span 1.5 orders of magnitude. Results show no correlation with side chain chemistry, except for two pairs of comparable residues (K/R, F/Y). D is the only substitution that enhances repression in the *presence* of inducer by more than two-fold. In LacI and PurR (as well as in a third homologue - CcpA), position 62 interacts with other regions of the regulatory domain in a homologue-specific manner (Table 2, underlined; 1,29,30,33,35,36). We hypothesize that the LLhG 62 variants have altered interactions with the GalR regulatory domain. Future mutagenesis of the regulatory domain will test this hypothesis.

Substitution at predicted LacI/GalR specificity determinants in LLhG

At least three studies predicted the presence of specificity determinants in the sequences that link the DNA-binding and regulatory domains in the LacI/GalR proteins (Table 1, “X”) ^{1,3–5}. Alignments of the LacI and GalR linker sequences are shown in Table 1; different colors highlight which amino acids are conserved across the family (green), which additional residues are conserved between the homologues of this study (blue), and which sites were previously shown to be specificity determinants (pink).⁷

The current experiments test the validity of the predictions by correlating amino acid substitution of individual positions with *in vivo* functional change, in the absence and presence of inducer. (Although site 57 is also predicted to be a specificity determinant, we excluded this site because it directly contacts DNA and impacts the binding affinity of PurR³⁷ and LacI²⁴) Results are summarized in Figure 3 through Figure 8. Long-term, results will be incorporated into a database that will inform an underlying assumption of bioinformatics algorithms: “All homologues in the family utilize the same positions as specificity determinants.” Therefore, where overlap exists in the current work, we note similarities and differences for a given residue in the alternative contexts of LLhG, LacI, and LLhP.

N-linker position 48—The sequence and structural interactions of amino acid 48 varies considerably between family members.^{1,29,30,33,35,36} However, only our multidisciplinary study predicted functional contributions from this position. In LLhG variants, most substitutions at position 48 diminished repression as compared to the parent E62 and E62K proteins (Figure 3). Even so, several 48 variants in LLhG/E62K repress more strongly than “wild-type” LLhG. Therefore, the “wild-type” isoleucine at site 48 must be one of the amino acids that allows the tightest repression from *lacO^I*. Leucine also facilitates tight repression – I48L/E62K shows enhanced repression in plate assays relative to LLhG/E62K (Supplementary Figure, white vs. light blue colonies, respectively). In contrast, the I48L substitution abolishes repression in LLhP and greatly diminishes repression in LacI.^{7,24} We previously speculated

that an L side chain could interact with the N-terminal DNA-binding domain and “lock” the repressor in a low affinity state.⁷ Alternatively, I48L might alter DNA specificity (including enhanced nonspecific binding) in LacI and LLhP.

Several variants at site 48 caused significant changes in bacterial growth. In LLhG, I48S and I48V caused liquid cultures to grow so slowly that accurate repression values could not be obtained. Adding inducer did not restore robust growth. Neither of these variants have detectable repression from *lacO*^I in plate assays (Figure 3B). Toxicity might result if these variants have increased non-specific binding, which in LacI is not affected by addition of inducer.³⁸ Results are even more complex for I48N and I48E/E62K. Colonies expressing these variants grow normally without inducer. However, adding fucose – but not the alternative inducer galactose – caused cultures to quit growing after 2 hours. On plate assays under these conditions, I48E/E62K had no colonies and I48N had extremely tiny colonies. One possibility is that these LLhG variants acquire specificity for other (toxic) sites on the *E. coli* genome that have different response to alternative inducers, which might now function as anti-inducers. DNA-dependent allostery has been previously observed in LacI linker variants^{17–19}, and anti-inducers are known for both LacI³⁹ and GalR⁴⁰.

Helix positions 52 and 55—Both positions 52 and 55 have varied sequences and structural interactions in the LacI/GalR family.^{1,33,35,36} Position 52 was predicted to be a specificity determinant by the two bioinformatics studies, whereas 55 was predicted to be a specificity determinant by all three studies. The effects of substitution at positions 52 and 55 were determined in LLhG and LLhG/E62K (Figure 4 and Figure 5). Variants were obtained that either enhanced or diminished repression, with a total range spanning 3.5 orders of magnitude. These residues are part of the linker hinge helix that undergoes a coil-to-helix conformational change when LacI binds *lacO*^{I,31,41–45} and LLhG repression changes might therefore be related to different helical propensities. However, we see little (if any) correlation between repression and helical propensity⁴⁶ for LLhG substitutions at either position 52 or 55.

Instead, with the possible exception of V52L, the rank order of repression by all LLhG variants at site 52 correlate very well with the rank order of DNA binding affinities determined for fifteen purified 52 variants of LacI¹⁹. Thus, altered repression appears to arise from changes in DNA-binding affinity. Exchanging the regulatory domain does not impact position 52, consistent with the structural observation that these side chains only interact with the partner hinge helix. The previous LacI results were interpreted as changes in helix-helix packing that were influenced by the sequence of the DNA ligand bound¹⁹. Several LacI 52 variants also demonstrated diminished allosteric response to inducer, due to increased binding in the presence of inducer IPTG^J. This behavior is recapitulated for many of the same substitutions of LLhG and LLhG/E62K in the presence of inducer fucose, which retain 2- to 50-fold repression compared to most induced variants (Figure 4, dark gray bars).

Two substitutions at position 55 are notable because they do not behave the same in LLhG and LLhG/E62K (Figure 5): Q55I has opposite effects in LLhG and LLhG/E62K, enhancing repression of the former by slightly more than 2-fold and diminishing the latter by 3-fold. Q55L has no effect on LLhG/E62 but diminishes repression of LLhG/E62K more than 30-fold. In addition, Q55I and Q55L *enhance* repression in the presence of inducer, by up to 5-fold (Figure 5); this effect is also seen with Q55M in LLhG. The different outcomes of hydrophobic substitutions could be related to the fact that position 55 can participate in the helix-helix interface that forms within a homodimer (Figure 1D). The high-resolution structure of the hinge helix is unknown for the low-affinity, DNA-bound condition for any of the repressors (such as induced LacI or induced LLhG). However, evidence is accumulating for the formation of

^JIPTG: isopropyl β-D-1-thiogalactopyranoside.

an interface in this complex.^{7,45} The presence of an interface in LLhG again provides a satisfactory explanation for how the hydrophobic mutations (I, L, and M) could facilitate or strengthen interface formation in the induced state, thereby enhancing DNA-binding of this conformation.

C-linker positions 58, 59, and 61—Position 58 is predicted to be a specificity determinant by all three studies. However, predictions disagree as to the importance of sites 59 and 61 (Table 1). Structurally, position 58 is the first residue of the LacI C-linker but the last residue of the PurR hinge helix (Figure 1E). The accompanying change of the C-linker (which has no regular secondary structure) allows position 61 to make interactions with the regulatory domain in PurR that are absent in LacI.^{1,33,35,36} Position 59 has a long hydrophobic interaction with DNA in the LacI complex but not PurR. We also postulated that K59 might make an ionic interaction with the charged side chain of E62 in LLhG.

Since nine different substitutions at position 58 in homologue LLhP essentially abolished repression⁷, we hypothesized that G58 was a unique requirement for *lacO^I* binding. However, in LLHG/E62K, only two of six substitutions at site 58 show comparable effects. Instead, two variants improved repression: G58K repression in the LLhG/E62 background is enhanced 16-fold (Figure 6). We are intrigued that changes at either end of the C-linker (58 or 62) can substantially improve repression. However, the double G58K/E62K variant did not further enhance repression in the high affinity state (Figure 6). Instead, repression (+) fucose is enhanced 10-fold with a concomitant decrease in allostery. G58S has similar behaviors in LLhG variants, with the E62K variant improving repression 4-fold in the presence of inducer. In contrast, G58S in LLhP abolished repression in plate assays.⁷

Position 61 is not very sensitive to substitution in LLhG, again in contrast to the dramatic loss of repression seen in LLhP.⁷ Using the two-fold criteria to define a change, only S61D in LLhG/E62K has a significant effect (Figure 7, dotted boxes). Given the proximity of the two charges in S61D/E62K, we do not find the diminished repression surprising. S61D also abolished repression in LLhP plate assays, and LLhP residue 62 is also K. The LacI to GalR mutation S61T has little effect in LLhG and slightly worsens repression by LLhG/E62K. Previously, we noted that position 61 is affected differently by the same substitutions in LacI and LLhP.⁷ This trend may hold for LLhG, since the S61T mutation dramatically diminishes repression of LLhP. However, the little overlap between other random substitutions is not sufficient to address the question of how the same substitutions behave in different proteins.

Substitutions at position 59 in LLhG or LLhG/E62K have functional effects that range from little change to dramatic loss of repression (Figure 8); none of the current substitutions improve repression. Lost repression indicates that site 59 is a specificity determinant, in agreement with bioinformatics predictions of Gelfand and colleagues (Table 1)^{3,4}. We also hypothesized that a charge-charge interaction between K59 and E62 might contribute to the functional sensitivity of the latter position. However, LLhG E62D improves repression, whereas LLhG K59E and the LacI to GalR substitution K59Q are poor repressors, with values equivalent to the no repressor control (Figure 2; “pHG165a” and Figure 8). Thus, a C-linker charge-charge interaction does not appear to contribute repression. Last, LLhG K59F and K59W caused bacterial cultures to grow slowly, indicating a gain of toxic function.

Evaluation of sites that are not predicted to be specificity determinants

All of the positions predicted to be specificity determinants are true positives. This leads to the question of whether *any* of the nonconserved LacI/GalR linker sites can be varied without altering function. To test this, we mutated sites 51 and 60, which were not predicted by any study to contribute to function. Both positions show low conservation across the LacI/GalR family. Site 51 is the second residue in the central hinge helix and exhibits different interactions

with the regulatory domains of LacI and PurR; however, because it is not sensitive to substitution in LacI or PurR, we did not previously predict it to be a specificity determinant.¹ Site 60 is located in the unstructured C-linker and also shows different interactions with the regulatory domains of LacI and PurR.¹ Again, this site is not reported to be sensitive to substitution in LacI.²⁴ Plate assays of LLhG variants show that repression can be diminished (blue) or enhanced (white) by variation at these positions (data not shown). Results from quantitative liquid culture assays of LLhG variants are presented in Figure 9 and Figure 10.

For position 51, several hydrophobic and aromatic side chains enhance repression up to 5-fold in both the absence and presence of inducer, whereas several small polar side chains diminish repression (Figure 9). Repression of LLhG/E62K/R51G was diminished ~2-fold in the absence of fucose but enhanced in the presence of inducer, causing a loss of allosteric response. We see no correlation with helical propensity of the N2 residue for any of the substitutions.⁴⁶

Notably, LLhG/R51W is not induced by fucose in the liquid culture assay (Figure 9) and showed very little induction in the plate assay (data not shown). In the absence of inducer, this variant produced white colonies in the plate assay (data not shown), and we thus expected the liquid culture value to be smaller than that of E62K, which has very light blue colonies (Supplementary Figure). Instead, the R51W liquid culture value is ~5-fold higher than that of E62K (Figure 9). Based on our experience with epigenetic shutdown of LLhP⁷, we successively re-streaked colonies expressing R51W and found that progressively more colonies turned dark blue (5–10% by the fourth replating; data not shown). Therefore, during the 1.5 day course of the liquid culture measurement, these blue colonies are proliferating and raising the value of the β -galactosidase assay. In light of the tight repression seen in the plate assay and the lack of inducibility, we hypothesize that R51W is locked in a conformation with high affinity not only for *lacO^I* but for other sites on the genome, resulting in selective pressure for bacteria to shutdown production of this LLhG variant. Intriguingly, tryptophan does not occur naturally at position 51, despite the variability (at least 13 amino acids) seen in the ~1000 sequences of 50 LacI/GalR orthologue groups (data not shown and ¹).

At position 60, leucine and methionine diminish repression by 4-fold, whereas several smaller side chains have minimal impact. However, the charged side chain of R enhances repression of LLhG by slightly more than 2-fold, and K enhances repression nearly 20-fold. Positively charged amino acids are also well-tolerated at position 60 in LLhG/E62K: Q60K might enhance repression of LLhG/E62K, and Q60R is within 2-fold of values for LLhG/E62K. We note that these variants have a high charge density in the C-linker, comprising K59, K or R60, S61, and E or K62. One possibility is that positive side chains allow more favorable interactions with negatively-charged DNA. However, note that the C-linkers are fairly remote from the DNA; position 62 is >10Å distant (Figure 1). Instead, we intuit that the high density of positive charge alters the structure of the C-linker or changes interactions with the regulatory domain.

Discussion

Engineering new functions from existing proteins and comprehension of genetic polymorphisms have two common requirements: (1) Identifying which non-conserved sites contribute to function; and (2) Appreciating what kinds of functional changes result from altering the amino acids at specificity determinants. To facilitate the first, many efforts are currently directed towards developing predictive bioinformatics analyses (*e.g.* ^{2–5,47–58}). The LacI/GalR proteins have served as a test family for two of these projects^{2–5}, as well as our multi-disciplinary study.¹ All predictions identified linker positions as possible specificity determinants, but the predictions are only in partial agreement with each other (Table 1). The current results show that the linker sites that contribute to LLhG function comprise all of the previously predicted positions, as well as additional positions identified herein. Thus, we must

raise the question of why prediction methods under-perform in a region that is critical for function.

Each prediction method is probably limited by a different factor, but comparing prediction to experimental data from only 1 or 2 proteins might have impacted all three studies. Our multidisciplinary study¹ relied upon mutagenesis data for only 2 proteins and had the requirement of a structural difference between LacI and PurR. These criteria were probably too stringent; indeed, we speculated that we were missing position 52, which is conserved between LacI and PurR. The two bioinformatics analyses (SDPpred and SPEL) both assume that all family members utilize the same sites as specificity determinants, and compare their predictions to mutagenesis of only LacI. However, some positions might be specificity determinants in only a subset of the homologues. For example, substitution of site 51 impacts function in LLhG but not LacI²⁴ (which also caused us to previously miss the importance of this site¹). Therefore, either (1) position 51 plays a different role in the LLhG chimera, and therefore is difficult to detect with bioinformatics; or (2) the available LacI/PurR data^k is insufficient to detect change. At the very least, these results provide a cautionary note about relying upon limited datasets for understanding the roles of specificity determinants in protein families.

We also deduce that SDPpred and SPEL studies utilized too large a dataset in their analyses of the LacI/GalR family. Examination of the *E. coli* paralogues illustrates this possibility: Of these 16 proteins, 11 have the highly-conserved linker components at positions 47, 49, 53, and 56 (Table 3, top eleven rows). Five paralogues lack these elements and/or have several G and P located in the central “helical” region (Table 3, bottom five rows). Of these, CytR is experimentally and structurally different than LacI and PurR,^{59–63} and we suspect this is true for the other 4 paralogues. However, the CytR, GntR, and IdnR orthologue groups were included in the bioinformatics predictions for the LacI/GalR family. We suggest that these groups should be treated separately; removing their sequences from SDPpred and SPEL analyses might diminish the number of false negative predictions in LacI-like sequences.

We also noted that several of the conserved linker sites overlap with the predictions by a third algorithm called “Evolutionary Trace Analysis” (ETA; Table 3; *e.g.* 48,50,64–66). ETA incorporates structural information with sequence analyses, in order to predict which invariant and “class-specific” sites are important to protein function. Previously, ETA results have been directly compared with results from SDPpred and SPEL. We find the comparison to be uninformative, because the programs appear to identify different subgroup levels within a phylogenetic tree: ETA finds residues that discriminate between large subgroups, whereas SDPpred and SPEL identify sites that discriminate homologues within subgroups. Indeed, a better strategy for predicting specificity determinants may be two-fold: (1) Use ETA to identify major subgroups, such as those that possess or lack conserved linker features; and (2) subsequently predict additional specificity determinants within each subgroup via SDPpred or SPEL. A similar strategy was recently adopted by Valencia and colleagues for predicting functionally-important residues from hierarchical information such as enzymatic classification;⁵² Ye *et al.* also recently realized that evolutionary pressure on functionally-important residues (resulting in their sequence conservation or variation) is differentially exerted across a phylogenetic tree.^{58, 1}

From the combined predictions of SDPpred, SPEL, ETA, and our previous work, only 3 linker sites are not predicted to contribute to function. Early in the current work, we discovered that

^kThe available *in vivo* LacI data does not report whether repression is enhanced, and the PurR study comprised a single substitution.

¹Although Ye *et al.* used the LacI/GalR proteins as a test family for their most recent work, they only considered whether residues that directly contact inducer IPTG are specificity determining positions.

one of these sites – 62 – can be varied to alter LLhG function. We subsequently tested sites 51 and 60. They too can alter function. Therefore, the entire linker region appears to be a functional “hotspot”. This may be a unique feature of the LacI/GalR proteins and predispose them to under-prediction of specificity determinants. Additional assessment of bioinformatics predictions should include regions that have fewer functionally-important residues.

Given the high density of specificity determinants, the linker region must also be an evolutionary hotspot. The LacI/GalR proteins are presumed to have arisen by gene duplication followed by sequence divergence.⁶⁷ Evolutionary fixation will not occur unless the change is large enough to impact bacterial growth, adding a third requirement to the definition of “specificity determinant”. Future experiments will determine how much change in repression of the *lac* operon is required to alter the bacterial life cycle. The current work clearly shows the possibility, since a number of variants impact bacterial growth rates (Figure 3B and Figure 8B).

Despite the partial success of current bioinformatics studies, predicting the location of specificity determinants remains a simpler problem than forecasting the functional outcomes of substitution. While significant evidence indicates the importance of the linker, the range of functional contributions from this region has been under-appreciated. Our previous LLhP study suggests that substitutions in the linker can affect allostery, affinity, and perhaps specificity.⁷ Similar results are presented here for LLhG. Future efforts will be directed towards determining whether a given site contributes to the same aspect of function in different homologues. For example, does variation at position 48 always alter affinity but not specificity?

Finally, comparing effects of the same substitution in LLhG and LLhG/E62K leads to multiple examples of nonadditivity. For example, the individual substitutions that comprise LLhG/E62K/G58K and LLhG/E62K/G58S each enhance repression of the high-affinity state (Figure 6). However, the two substitutions in combination do *not* further enhance repression in the high-affinity condition. Instead, repression is enhanced in the low-affinity condition. Such outcomes suggest the presence of small, functional networks on the common scaffold. These networks are not easily identified from structure alone but may be ascertained by combinatorial mutational strategies.

In conclusion, several existing strategies for identifying specificity determinants appear to under-predict the locations of sites that contribute to LacI/GalR function. Even the union of all the predictions is not sufficient, because all missed the potential for contributions from positions 51, 60, and 62. As noted above, one key for improving these analyses may lie in choosing the appropriate dataset – the entire family is in itself too large. Nonetheless, it is encouraging that no study predicted a false positive in the LacI/GalR linker region. We construe that sites predicted to be functionally important by either SDPpred or SPEL are valid targets for further study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Mr. Sudheer Tungtur for experimental assistance, as well as many helpful discussions. Dr. Nick Grishin and Jimin Pei (UT Southwestern) graciously shared their full prediction set of LacI/GalR specificity determinants. Dr. James McAfee (Pittsburg State University) suggested that high levels of non-specific DNA binding could be toxic to *E. coli*. Drs. Sarah Bondos (Rice University), Aron Fenton (KUMC) and Marina Jeyasingham (KUMC) provided critical feedback on the manuscript. LSK received support from NIH Grant P20 RR17708 from the Institutional Development Award program of the National Center for Research Resources and from GM079423.

References

- 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]
- 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. *Nucleic Acids Res* 2004;32:W424–428. [PubMed: 15215423]
- Mirny LA, Gelfand MS. Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors. *J Mol Biol* 2002;321:7–20. [PubMed: 12139929]
- Kalinina OV, Mironov AA, Gelfand MS, Rakhmaninova AB. Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Sci* 2004;13:443–456. [PubMed: 14739328]
- 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]
- Weickert MJ, Adhya S. A family of bacterial regulators homologous to Gal and Lac repressors. *J Biol Chem* 1992;267:15869–15874. [PubMed: 1639817]
- Tungtur S, Egan SM, 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.
- Majumdar A, Rudikoff S, Adhya S. Purification and properties of Gal repressor:pL-galR fusion in pKC31 plasmid vector. *J Biol Chem* 1987;262:2326–2331. [PubMed: 2950087]
- Geanacopoulos M, Vasmatazis G, Lewis DE, Roy S, Lee B, Adhya S. GalR mutants defective in repressosome formation. *Genes Dev* 1999;13:1251–1262. [PubMed: 10346814]
- Stewart GS, Lubinsky-Mink S, Jackson CG, Cassel A, Kuhn J. pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression. *Plasmid* 1986;15:172–181. [PubMed: 3012611]
- Neidhardt FC, Bloch PL, Smith DF. Culture medium for enterobacteria. *J Bacteriol* 1974;119:736–747. [PubMed: 4604283]
- Bhende PM, Egan SM. Amino acid-DNA contacts by RhaS: an AraC family transcription activator. *J Bacteriol* 1999;181:5185–5192. [PubMed: 10464186]
- Miller, JH. *A Short Course in Bacterial Genetics: A Laboratory Handbook for Escherichia coli and Related Bacteria*. Plainview, N.Y.: Cold Spring Laboratory Press; 1992.
- Gilbert W, Maxam A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci USA* 1973;70:3581–3584. [PubMed: 4587255]
- 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 USA* 1984;81:1624–1628. [PubMed: 6369330]
- Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc Natl Acad Sci USA* 1983;80:6785–6789. [PubMed: 6316325]
- 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]
- Falcon CM, Matthews KS. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry* 2000;39:11074–11083. [PubMed: 10998245]
- 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]
- Lin SY, Riggs AD. Lac repressor binding to DNA not containing the lac operator and to synthetic poly dAT. *Nature* 1970;228:1184–1186. [PubMed: 4321099]
- Swint-Kruse L, Zhan H, Fairbanks BM, Maheshwari A, Matthews KS. Perturbation from a distance: mutations that alter LacI function through long-range effects. *Biochemistry* 2003;42:14004–14016. [PubMed: 14636069]

22. 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]
23. Zhou YN, Chatterjee S, Roy S, Adhya S. The non-inducible nature of super-repressors of the gal operon in *Escherichia coli*. *J Mol Biol* 1995;253:414–425. [PubMed: 7473724]
24. Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Müller-Hill B. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J Mol Biol* 1996;261:509–523. [PubMed: 8794873]
25. Luria SE, Adams JN, Ting RC. Transduction of lactose-utilizing ability among strains of *E. coli* and *S. dysenteriae* and the properties of the transducing phage particles. *Virology* 1960;12:348–390. [PubMed: 13764402]
26. Griffith KL, Wolf RE. Measuring [beta]-Galactosidase Activity in Bacteria: Cell Growth, Permeabilization, and Enzyme Assays in 96-Well Arrays. *Biochem Biophys Res Commun* 2002;290:397–402. [PubMed: 11779182]
27. Geanakopoulos M, Adhya S. Genetic analysis of GalR tetramerization in DNA looping during repressosome assembly. *J Biol Chem* 2002;277:33148–33152. [PubMed: 12065579]
28. Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Sci* 2003;12:2523–2541. [PubMed: 14573864]
29. Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell* 2004;118:731–741. [PubMed: 15369672]
30. Schumacher MA, Seidel G, Hillen W, Brennan RG. Phosphoprotein Crh-Ser46-P Displays Altered Binding to CcpA to Effect Carbon Catabolite Regulation. *J Biol Chem* 2006;281:6793–6800. [PubMed: 16316990]
31. 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]
32. Bell CE, Lewis M. Crystallographic analysis of Lac repressor bound to natural operator O1. *J Mol Biol* 2001;312:921–926. [PubMed: 11580238]
33. 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]
34. 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]
35. Schumacher MA, Choi KY, Zalkin H, Brennan RG. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* 1994;266:763–770. [PubMed: 7973627]
36. 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]
37. Glasfeld A, Koehler AN, Schumacher MA, Brennan RG. The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor groove interactions. *J Mol Biol* 1999;291:347–361. [PubMed: 10438625]
38. 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]
39. Riggs AD, Newby RF, Bourgeois S. lac repressor--operator interaction. II Effect of galactosides and other ligands. *J Mol Biol* 1970;51:303–314. [PubMed: 4320936]
40. Buttin G. Regulatory Mechanisms in the Biosynthesis of the Enzymes of Galactose Metabolism in *Escherichia coli* K 12. I. the Induced Biosynthesis of Galactokinase and the Simultaneous Induction of the Enzymatic Sequence. *J Mol Biol* 1963;7:164–182. [PubMed: 14062650]
41. Spronk CAEM, Slijper M, van Boom JH, Kaptein R, Boelens R. Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. *Nature Struct Biol* 1996;3:916–919. [PubMed: 8901866]

42. Kalodimos CG, Folkers GE, Boelens R, Kaptein R. Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. *Proc Natl Acad Sci USA* 2001;98:6039–6044. [PubMed: 11353825]
43. Ha J-H, Spolar RS, Record MT Jr. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J Mol Biol* 1989;209:801–816. [PubMed: 2585510]
44. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 1994;263:777–784. [PubMed: 8303294]
45. Taraban M, Zhan H, Whitten AE, Langley DB, Matthews KS, Swint-Kruse L, Trehwella 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]
46. Kumar S, Bansal M. Dissecting alpha-helices: position-specific analysis of alpha-helices in globular proteins. *Proteins* 1998;31:460–476. [PubMed: 9626705]
47. Casari G, Sander C, Valencia A. A Method to Predict Functional Residues in Proteins. *Nat Struct Biol* 1995;2:171–178. [PubMed: 7749921]
48. Lichtarge O, Bourne HR, Cohen FE. An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. *J Mol Biol* 1996;257:342. [PubMed: 8609628]
49. Hannenhalli SS, Russell RB. Analysis and prediction of functional sub-types from protein sequence alignments. *J Mol Biol* 2000;303:61–76. [PubMed: 11021970]
50. Mihalek I, Res I, Lichtarge O. Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. *Bioinformatics* 2006;22:1656–1657. [PubMed: 16644792]
51. Donald JE, Shakhnovich EI. Determining functional specificity from protein sequences. *Bioinformatics* 2005;21:2629–2635. [PubMed: 15797914]
52. Pazos F, Rausell A, Valencia A. Phylogeny-independent detection of functional residues. *Bioinformatics* 2006;22:1440–1448. [PubMed: 16551661]
53. Ye K, Anton Feenstra K, Heringa J, Ijzerman AP, Marchiori E. Multi-RELIEF: a method to recognize specificity determining residues from multiple sequence alignments using a Machine-Learning approach for feature weighting. *Bioinformatics* 2008;24:18–25. [PubMed: 18024975]
54. Masha Y, Niv LSRJRHASHW. Identification of GATC- and CCGG-recognizing Type II REases and their putative specificity-determining positions using Scan2S - A novel motif scan algorithm with optional secondary structure constraints. *Proteins: Struc, Func, and Bioinf* 2008;71:631–640.
55. Yin Y, Kirsch JF. Identification of functional paralog shift mutations: Conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proc Natl Acad Sci USA* 2007;104:17353–17357. [PubMed: 17947381]
56. Chakrabarti S, Bryant SH, Panchenko AR. Functional Specificity Lies within the Properties and Evolutionary Changes of Amino Acids. *J Mol Biol* 2007;373:801–810. [PubMed: 17868687]
57. Donald JE, Shakhnovich EI. Predicting specificity-determining residues in two large eukaryotic transcription factor families. *Nucleic Acids Res* 2005;33:4455–4465. [PubMed: 16085755]
58. Ye K, Vriend G, Ijzerman AP. Tracing evolutionary pressure. *Bioinformatics* 2008;24:908–915. [PubMed: 18304936]
59. Pedersen H, Valentin-Hansen P. Protein-induced fit: the CRP activator protein changes sequence-specific DNA recognition by the CytR repressor, a highly flexible LacI member. *EMBO J* 1997;16:2108–2118. [PubMed: 9155036]
60. Jørgensen CI, Kallipolitis BH, Valentin-Hansen P. DNA-binding characteristics of the *Escherichia coli* CytR regulator: a relaxed spacing requirement between operator half-sites is provided by a flexible, unstructured interdomain linker. *Mol Microbiol* 1998;27:41–50. [PubMed: 9466254]
61. Kallipolitis BH, Valentin-Hansen P. A role for the interdomain linker region of the *Escherichia coli* CytR regulator in repression complex formation. *J Mol Biol* 2004;342:1–7. [PubMed: 15313602]
62. Moody CL, Tretyachenko-Ladokhina V, Senear DF, Cocco MJ. 2029-Pos Structural Characterization Of CytR, A Bacterial Gene Repressor, Using NMR. *Biophys J* 2008;94:2029.
63. Tretyachenko-Ladokhina V, Cocco MJ, Senear DF. Flexibility and Adaptability in Binding of *E. coli* Cytidine Repressor to Different Operators Suggests a Role in Differential Gene Regulation. *J Mol Biol* 2006;362:271–286. [PubMed: 16919681]

64. Madabushi S, Yao H, Marsh M, Kristensen DM, Philippi A, Sowa ME, Lichtarge O. Structural clusters of evolutionary trace residues are statistically significant and common in proteins. *J Mol Biol* 2002;316:139. [PubMed: 11829509]
65. Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, Lichtarge O. Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J Biol Chem* 2004;279:8126–8132. [PubMed: 14660595]
66. Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J Mol Biol* 2004;336:1265. [PubMed: 15037084]
67. 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]
68. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng* 1998;11:739–747. [PubMed: 9796821]
69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410. [PubMed: 2231712]
70. 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]
71. Choi KY, Zalkin H. Role of the purine repressor hinge sequence in repressor function. *J Bacteriol* 1994;176:1767–1772. [PubMed: 8132474]
72. 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]

Figure 1.

(A) Representative LacI/GalR structure. Homodimer formation (monomers are represented by light and dark gray ribbons) is required for the LacI/GalR proteins to bind cognate their DNA sequences (blue sticks at the top of the figure).⁶ The protein linker is colored magenta (N-linker, C-linker) and green (hinge helix). The beginning of the linker is marked with an arrow and the last residue is position 62 (magenta spheres). The black spheres show where ligand occupies the binding site of the regulatory domain. Green spheres approximate the location of the LLhG E230K mutation. Position 62 is shown with a magenta space-filling model. The pdb used was that of LacI bound to anti-inducer (1efa; 33). (B) Schematic of chimeric proteins. On the left, the structure of wildtype LacI is depicted in cyan. LLhP (center) comprises the LacI DNA binding domain and linker (cyan ovals and rectangles) and the PurR regulatory domain (large pink rounded rectangles.) LLhG has the LacI DNA binding domain and linker fused to the GalR regulatory domain (large green rounded rectangles). Each chimera has changed interactions between linker specificity determinants and the top surfaces of the regulatory domains. (C) N-linker side chains are shown in magenta with ball-and-stick representation. N-linker specificity determinant 48 is shown with a space filling representation. (D) Hinge helix side chains are shown by sticks on the left helix and ball/stick on the right helix. (E) Side chains

are shown in ball/stick for the left C-linker and by sticks for the right C-linker. All structures in Figure 1 were created with the program UCSF Chimera.⁷²

Figure 2.

Substitutions at LLhG position 62 alter repression from *lacO^I*. Repression levels inversely correlate with the amount of β -galactosidase activity measured – low values correspond to tight repression. Bars labeled “pHG165a” show results for colonies that carried plasmid without cloned repressor. For cells expressing LLhG variants, β -galactosidase activity was determined in the absence (light gray) and presence of 20 mM inducer fucose (dark gray). On this plot, LLhG is designated as “E62”; this variant and E62K are indicated with asterisks (*). Average values are for measurements made on 3–6 different occasions, with two measurements each day. Error bars represent standard deviations of mean values. The upper gray bar depicts a two-fold change around the value for LLhG+inducer. Dotted lines are to aid visual inspection of the graph.

Figure 3.

Substitutions at LLhG position 48. (A) For each variant, β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values shown are for measurements made on 3–6 different occasions, with two measurements each day. LLhG and E62K are indicated with asterisks (*). Error bars represent standard deviations of mean values. The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer. Although the large error on the value for I48L+K does not allow differentiation from the E62K value, plate assays show that the I48L variant represses more tightly (Supplementary Figure). (B) Results for plate assays of additional variants. “B” = blue; “LB” = light blue; “W” = white; “tox” = slows or halts culture growth. I48P might disrupt structure, and liquid culture assays were not performed.

Figure 4.

Substitutions at LLhG position 52. β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values shown are for measurements made on 3–6 different occasions, with two measurements each day. Error bars represent standard deviations of mean values. LLhG and E62K are indicated with asterisks (*). The solid horizontal line corresponds to the Y axis of most other figures and is used to call attention to the fact that variants at position 52 are among the tightest repressor variants that we have identified. The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer.

Figure 5.

Substitutions at LLhG position 55. (A) β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values shown are for measurements made on 3–6 different occasions, with two measurements each day. Error bars represent standard deviations of mean values. LLhG and E62K are indicated with asterisks (*). The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer. (B) Results for plate assays of additional variants. These variants were not further studied because they did not enhance repression; toxicity in liquid culture assays is unknown. “B” = blue; “LB” = light blue; “W” = white.

Figure 6.

Substitutions at LLhG position 58. β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values shown are for measurements made on 3–6 different occasions, with two measurements each day. Error bars represent standard deviations of mean values. LLhG and E62K are indicated with asterisks (*). The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer.

Figure 7.

Substitutions at LLhG position 61. β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values shown are for measurements made on 3–6 different occasions, with two measurements each day. Error bars represent standard deviations of mean values. LLhG and E62K are indicated with asterisks (*). The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer.

Figure 8.

Substitutions at LLhG position 59. (A) β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose. “+K” indicates that the substitution was created on LLhG/E62K. Average values are for measurements made on 3–6 different occasions, with two measurements each day. LLhG and E62K are indicated with asterisks (*). Error bars represent standard deviations of mean values. The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer. (B) Results for plate assays of additional variants. “B” = blue; “LB” = light blue; “W” = white; “tox” = slows or halts culture growth.

Figure 9.

Substitutions at LLhG position 51. β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose using a protocol for 96-well plates. “+K” indicates that the substitution was created on LLhG/E62K. Average normalized values are for measurements made two separate bacterial colonies, each in quadruplicate. Values for LLhG (determined in 96 well plates) and E62K (value repeated from other figures) are indicated with asterisks (*). Error bars represent standard deviations of mean values. The solid horizontal line corresponds to the Y axis of most other figures and is used to call attention to the fact that variants at position 51 are among the tightest repressor variants that we have identified. The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer. See Results for details of anomalies in the R51W value.

Figure 10.

Substitutions at LLhG position 60. β -galactosidase activity was determined in the absence and presence of 20 mM inducer fucose using a protocol for 96-well plates. “+K” indicates that the substitution was created on LLhG/E62K. Average normalized values are for measurements made two separate bacterial colonies, each in quadruplicate. Values for LLhG (determined in 96 well plates) and E62K (value repeated from other figures) are indicated with asterisks (*). Error bars represent standard deviations of mean values. The dotted boxes indicate limits of two-fold change for LLhG and LLhG/E62K in the absence of inducer. The upper gray bar depicts a two-fold change around the value for LLhG+inducer.

Table 1

Linker sequences in LacI, GalR, and LLhG^a

Residue #'s	N-linker					Hinge helix					C-linker							
	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
LacI	L	N	N	Q	P	N	R	V	A	Q	Q	L	A	G	K	Q	S	L
GalR	L	S	N	H	P	N	A	N	A	R	A	L	A	Q	Q	T	T	E
LLhG	L	N	N	Q	P	N	R	V	A	Q	Q	L	A	G	K	Q	S	E
LLhG/E62K	L	N	N	Q	P	N	R	V	A	Q	Q	L	A	G	K	Q	S	K
Predicted and Confirmed Linker Specificity Determinants																		
Swint-Kruse et al ¹				X							X		X ^b	X			X	
SDPpred ^{3,4}								X			X		X	X	X			
SPEL ^{c,5}								X			X		X	X				
LLhG substitutions ^d				X			X	X			X		(X)	X	X	X	X	X

^a Different fonts represent different structural regions of the linker. Green highlights amino acids that are conserved in the LacI/GalR family. Blue highlights residues that are conserved between LacI and GalR. Pink indicates residues previously identified to be specificity determinants.⁷ The striped background calls attention to position 62, the first amino acid of the regulatory domain.

^b Residue 57 makes direct contact with DNA and is known to be a specificity determinant in PurR.³⁷

^c Members of the Grishin lab graciously communicated their complete list of predicted specificity determinants. A cut-off of the first 40 amino acids was used to compare SPEL predictions to the top 40 predictions by SDPpred.

^d Position 57 is marked in parentheses because preliminary, unpublished results for LLhP agree with the findings in footnote b. Since this residue directly contacts DNA, it was not mutated in this study.

Table 2

	A	S	V	M	G	E	R	S	G	V	R	C	K	A	V	H	N	L	A	O	R	V	S
G	A	S	V	M	G	E	R	S	G	V	R	C	K	A	V	H	N	L	A	O	R	V	S
G	Y	T	A	N	L	-	N	N	-	L	N	R	R	Q	A	S	M	M	Q	K	R	V	D
G	N	I	N	N	I	G	Y	H	G	E	-	E	E	K	G	Q	Q	L	R	H	R	C	A

ces alignments were generated from structure comparisons with CE/CL.⁶⁸ L_{act} and GalR alignments were generated by BLAST.⁶⁹ Underlined positions (regions 90–95 and PurR
lue 62). Gray highlights residues that interact with any other linker amino acid; specific interactions are noted in the text.

Table 3
Alignment^a of linker residues for *E. coli* paralogs in the LacI/GalR family and predicted specificity determinants.

LacI Residue #'s	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
	N-linker			Hinge Helix			C-linker ^b											
LACI_ECOLI	L	N	Y	I	P	N	R	V	A	Q	Q	L	A	G	K	Q	S	L
RBSR_ECOLI	L	N	Y	A	P	S	A	L	A	R	S	L	K	L	N	Q	T	H
PURR_ECOLI	L	H	Y	S	P	S	A	V	A	R	S	L	K	V	N	H	T	K
GALS_ECOLI	L	D	Y	R	P	N	A	N	A	Q	A	L	A	T	Q	V	S	D
GALR_ECOLI	L	S	Y	H	P	N	A	N	A	R	A	L	A	Q	Q	T	T	E
YCJW_ECOLI	L	Q	Y	Q	P	N	K	L	A	R	A	L	T	S	S	G	F	D
CSCR_ECOLI	L	N	Y	V	P	D	L	S	A	R	K	M	R	A	Q	G	R	K
TRER_ECOLI	H	G	F	S	P	S	R	S	A	R	A	M	R	G	Q	S	D	K
FRUR_ECOLI	H	N	Y	H	P	N	A	V	A	R	A	L	R	A	G	R	T	R
ASCG_ECOLI	S	G	Y	R	P	N	L	L	A	R	N	L	S	A	K	S	T	Q
RAFR_ECOLI	R	G	Y	R	P	N	T	Q	A	R	R	L	K	T	G	K	T	D
Unusual hinge helix sequence (bold)																		
IDNR_ECOLI	I	N	Y	I	P	N	R	A _c	A	P	G	M	L	L	N	A	Q	S
GNTR_ECOLI	L	G	Y	I	P	N	R	-	A	P	D	I ^d	L	S	N	A	T	S
CYTR_ECOLI	V	G	Y	L	P	Q	P	M	G	R	N	V	K	R	N	E	S	R
No proline at 49																		
MALI_ECOLI	L	G	F	V	R	N	R	Q	A	S	A	L	R	G	G	Q	S	G
EBGR_ECOLI	L	E	Y	K	-	T	S	S	A	R	K	L	Q	T	G	A	V	N
Evolutionary Trace ^e	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
All other predictions (Table 1)							X	X										
Unpredicted specificity determinants ^f																		

^aProtein sequences were identified in Swiss-Prot.⁷⁰ A full-length alignment was first generated with BLAST⁶⁹ and then the linker regions of *E. coli* paralogs were manually optimized to align all conserved residues (gray boxes).

^bThe break on either end of the C-linker to the hinge helix or regulatory domain can vary between homologues.

^cThe location of the gap is unclear, but these sequences cannot align both P49 and A53 with other family members. In combination with pro and gly residues in the central "helix", this difference may reflect a changed role for the linker, as is known for CytR.^{59–61}

^dL and M are the only amino acids that allow function in LacI or PurR.^{1,24,71}

^eThe LacI structure Iefa³³ was analyzed the ETA web-interface Report_Maker⁵⁰. Linker residues were noted that fall in the top 25% of residues predicted to be functionally important.

^fIdentified experimentally in this work.