Supplementary Information

for

Novel insights from hybrid LacI/GalR proteins: Family-wide functional attributes and biologically-significant variation in transcription repression

Sarah Meinhardt[a], Michael W. Manley, Jr.[a], Nicole A. Becker[b], Jacob A. Hessman[a], L. James Maher III[b], and Liskin Swint-Kruse[a]*

[a]Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City KS 66205, USA

[b]Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine, 200 First Street, SW, Rochester MN 55905, USA

Supplementary Table I. Sequence alignment of LacI/GalR proteins used to create chimeras. Note that, because of differing start positions, alignment numbering does not match LacI numbering used in the manuscript. Experimental details about chimera construction are below the sequence alignment.

In the alignment, the LacI DNA binding domain is highlighted green. The LacI linker is highlighted yellow (LacI positions 45-61). Conserved "YPAL" linker residues are highlighted in red and correspond to LacI positions 47, 49, 53, and 56. Note that *E. coli* CytR lacks the "AL" of this motif (see Discussion), whereas the TreR "FPAM" motif appears to function similarly to YPAL.

The start of each regulatory domain used to create a chimera (LacI position 62) is highlighted with cyan. The positions of the "E230K" mutation in LLhG and LLhS are highlighted in magenta. (Without this mutation, the chimeras are toxic to *E. coli*). The amino acids deleted in LacI"-11" are highlighted in gray.

The yellow highlights in the LacI and PurR regulatory domains indicate the positions that interact with the linker to form an interface.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    5          15         25         35         45         55
GalR            -----MATIK DVARLAGVSV ATVSRVINNS P---KASEAS RLAVHSAMES LSYHPNANAR
GalS            -----MITIR DVARQAGVSV ATVSRVLNNS T---LVSADT REAVMKAVSE LDYRPNANAQ
LacI            ---MKPVTLY DVAEYAGVSY QTVSRVVNQA S---HVSAKT REKVEAAMAE LNYIPNRVAQ
TreR            --MQNRLTIK DIARLSGVGK STVSRVLNNE S---GVSQLT RERVEAVMNQ HGFSPSRSAR
PurR            -----MATIK DVAKRANVST TTVSHVINKT R---FVAEET RNAVWAAIKE LHYSPSAVAR
RbsR            -----MATMK DVARLAGVST STVSHVINKD R---FVSEAI TAKVEAAIKE LNYAPSALAR
FruR            ------MKLD EIARLAGVSR TTASYVINGK AKQYRVSDKT VEKVMAVVRE HNYHPNAVAA
CytR            --------MK DVALKAKVST ATVSRALMNP D---KVSQAT RNRVEKAARE VGYLPQPMGR
AscG            -----MTTML EVAKRAGVSK ATVSRVLSGN G---YVSQET KDRVFQAVEE SGYRPNLLAR
CelR            MERRRRPTLE MVAALAGVGR GTVSRVINGS D---QVSPAT REAVKRAIKE LGYVPNRAAR

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    65         75         85         95        105        115
GalR            ALAQQTTETV GLVV-----G DVSDPFFGAM VKAVEQVAYH TGNFLLIGNG -YHNEQKERQ
GalS            ALATQVSDTI GVVV-----M DVSDAFFGAL VKAVDLVAQQ HQKYVLIGNS -YHEAEKERH
LacI            QLAGKQSLLI GVAT-----S SLALHAPSQI VAAIKSRADQ LGASVVVSMV ERSGVEACKA
TreR            AMRGQSDKVV AIIV-----T RLDSLSENLA VQTMLPAFYE QGYDPIMMES -QFSPQLVAE
PurR            SLKVNHTKSI GLLA-----T SSEAAYFAEI IEAVEKNCFQ KGYTLILGNA -WNNLEKQRA
RbsR            SLKLNQTHTI GMLI-----T ASTNPFYSEL VRGVERSCFE RGYSLVLCNT -EGDEQRMNR
FruR            GLRAGRTRSI GLVI-----P DLENTSYTRI ANYLERQARQ RGYQLLIACS -EDQPDNEMR
CytR            NVKRNESRTI LVIV-----P DICDPFFSEI IRGIEVTAAN HGYLVLIGDC -AHQNQQEKT
AscG            NLSAKSTQTL GLVV---TNT LYHGIYFSEL LFHAARMAEE KGRQLLLADG -KHSAEEERQ
CelR            TLVTRRTDTV ALVVSENNQK LFAEPFYAGI VLGVGVALSE RGFQFVLATG -RSGIEHER-

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   125        135        145        155        165        175
GalR            AI-EQLIRHR CAALVVHAKM IPDA-----D L--ASLMKQM PGM-V-LINR ILPGFEN-RC
GalS            AI-EVLIRQR CNALIVHSKA LSDD-----E L--AQFMDNI PGM-V-LINR VVPGYAH-RC
LacI            AV-HNLLAQR VSGLIINYPL DDQ------D AIAVEAACTN VPA-L-FLDV SDQTP-I-NS
TreR            HL-GVLKRRN IDGVVLFGFT G--------I TEEMLAHWQS S-L-V-LLAR D--AK-G-FA
PurR            YL-SMMAQKR VDGLLVMCSE YPE------P LLAMLEEYRH IPM-V-VMDW GEAKA-DFTD
RbsR            NL-ETLMQKR VDGLLLLCTE THQ------P SREIMQRYPT VPT-V-MMDW APFD--GDSD
FruR            CI-EHLLQRQ VDAIIVSTSL PPEHP----F -YQRWANDPF P-I-V-ALDR ALDRE-H-FT
CytR            FI-DLIITKQ IDGMLLLGSR LP-------F DASIEEQRNL PPM-V-MANE FAPELEL-PT
AscG            AI-QYLLDLR CDAIMIYPRF LSVD-----E IDDIIDAHSQ -PI-M-VLNR RLRKNSS-HS
CelR            -LGGYLAGQH VDGVLLLSLH RDDPLPQMLD EAGVPYVYGG RPLGVPEEQV SYVDIDNIGG
```

```
            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               185        195        205        215        225        235
GalR        IALD-DRYGA WLATRHLIQQ GHTRIGYLC- SNHSISDAED RLQGYYDA-- LAESG-IAAN
GalS        VCLD-NLSGA RMATRMLLNN GHQRIGYLS- SSHGIEDDAM RKAGWMSA-- LKEQD-IIPP
LacI        IIFS-HEDGT RLGVEHLVAL GHQQIALLAG PLSSVSAR-L RLAGWHKY-- LTRNQ-IQPI
TreR        SVCYDDEGAI KILMQRLYDQ GHRNISYLGV PHSDVTTGKR RHEAYLAF-- CKAHK-LHPV
PurR        AVIDNAFEGG YMAGRYLIER GHREIGVIPG PLERNTGA-G RLAGFMKA-- MEEAM-IKVP
RbsR        LIQDNSLLGG DLATQYLIDK GHTRIACITG PLDKTPAR-L RLEGYRAA-- MKRAG-LNIP
FruR        SVVGADQDDA EMLAEELRKF PAETVLYLGA -LPELSVSFL REQGFRTA-- WKDDP-REVH
CytR        VHID-NLTAA FDAVNYLYEQ GHKRIGCIAG PEEMPLCH-Y RLQGYVQA-- LRRCG-IMVD
AscG        VWCD-HKQTS FNAVAELINA GHQEIAFLTG SMDSPTSI-E RLAGYKDA-- LAQHG-IALN
CelR        GRQATQRLIE TGHRRIATIA GPQDMVAGVE RLQGYREA-L LAAGMEYDET LVSYGDFTYD

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               245        255        265        275        285        295
GalR        DR-LVTFGEP DE-SGGEQAM TELLGRGRN- -FTAVACYND SMAAGAMGVL NDNGIDVPGE
GalS        ES-WIGAGTP DM-PGGEAAM VELLGRNLQ- -LTAVFAYND NMAAGALTAL KDNGIAIPLH
LacI        ---AEREGDW SA-MSGFQQT MQMLNEGIV- -PTAMLVAND QMALGAMRAI TESGLRVGAD
TreR        ----AALPGL AM-KQGYENV AKV----ITP ETTALLCATD TLALGASKYL QEQRIDT---
PurR        ES-WIVQGDF EP-ESGYRAM QQILSQPHR- -PTAVFCGGD IMAMGALCAA DEMGLRVPQD
RbsR        DG-YEVTGDF EF-NGGFDAM RQLLSHPLR- -PQAVFTGND AMAVGVYQAL YQAELQVPQD
FruR        -FLYANSYER EAAAQLFEKW LETH-----P MPQALFTTSF ALLQGVMDVT LRRDGKLPSD
CytR        PQ-YIARGDF TF-EAGSKAM QQLLDLPQP- -PTAVFCHSD VMALGALSQA KRQGLKVPED
AscG        EK-LIANGKW TP-ASGAEGV EMLLERGAK- -FSALVASND DMAIGAMKAL HERGVAVPEQ
CelR        SGVAAMRELL DR-APDVDAV FAASDLMG-- -LAALRVLR- --ASGRRVPE DVAVVGYDDS

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               305        315        325        335        345        355
GalR        -ISLIGFDDV LVSRYVRPRL TTVRYPIVTM ATQAAELALA LADNRPLPEI TNVFSPTLVR
GalS        -LSIIGFDDI PIARYTDPQL TTVRYPIASM AKLATELALQ GAAGNIDPRA SHCFMPTLVR
LacI        -ISVVGYDDT EDSSCYIPPL TTIKQDFRLL GQTSVDRLLQ LSQG-QAVKG NQLLPVSLVK
TreR        -LQLASVGNT PLMKFLHPEI VTVDPGYAEA GRQAACQLIA QVTG-RSEPQ QIIIPATLS-
PurR        -VSLIGYDNV RNARYFTPAL TTIHQPKDSL GETAFNMLLD RIVNKREEPQ SIEVHPRLIE
RbsR        -IAVIGYDDI ELASFMTPPL TTIHQPKDEL GELAIDVLIH RITQPTLQQQ RLQLTPILME
FruR        -LAIATFGDN ELLDFLQCPV LAVAQRHRDV AERVLEIVLA SLDEPRKPKP GLTRIKRNLY
CytR        -LSIIGFDNI DLTQFCDPPL TTIAQPRYEI GREAMLLLLD QMQGQHVGSG SRLMDCELII
AscG        -VSVIGFDDI AIAPYTVPAL SSVKIPVTEM I--QIGRLIF MLDGGDFSPP KTFSGKLIRR
CelR        TVAEHAEPP- -MTSVNQPTE LMGREMARLL VDRITGETTE PVRLVLETHL MVRESG....

            ....|....| ....|....| ....|....| ....|
               365        375        385        395
GalR        RHSVSTPSLE ASHHATSD-- ---------- -----
GalS        RHSVATRQNA AAITNSTNQA M--------- -----
LacI        RKTTLAPNTQ TASPRALADS LMQLARQVSR LESGQ
TreR        ---------- ---------- ---------- -----
PurR        RRSVADGPFR DYRR------ ---------- -----
RbsR        RGSA------ ---------- ---------- -----
FruR        RRGVLSRS.. .......... .......... .....
CytR        RGSTRALP-- ---------- ---------- -----
AscG        DSLIAPSR-- ---------- ---------- -----
CelR        .......... .......... .......... .....
```

Chimera construction.

The coding regions for the seven regulatory domains and full-length LacI were ligated into the multi-cloning site of the pGemT vector (Promega), which interrupts a gene for β-galactosidase. Colonies were screened for white color, grown overnight in 2xYT media, and the plasmid DNA was purified. Samples were then sequenced to confirm proper cloning (KUMC Biotechnology Support Facility or Northwoods DNA, Inc., Solway MN) and any errors in the coding region were corrected with site-directed mutagenesis (Quikchange, Stratagene/Agilent Technologies).

Final construction of the chimeras was accomplished in one of three ways. Chimeras LLhF, LLhT, and LLhC were made similarly to LLhP and LLhG (1,2), by first substituting the homologous regulatory domains for that of LacI on the plasmid pLS1. Briefly, pLS1 and pGemT plasmids encoding the regulatory domains were digested with *Bsu*36I and the other appropriate restriction enzyme (Supplementary Table 2). The appropriate fragments were separated by agarose gel electrophoresis, extracted from the gel, ligated, and transformed into *E. coli.* The entire coding regions for the chimeric repressors were sequenced and subsequently subcloned from pLS1 onto the low copy plasmid pHG165.

LLhE was cloned by Bio-Means, Inc. (Sugarland, TX), using the pGemT plasmid containing the CelR regulatory domain and a pHG165 plasmid containing full length LacI. LLhS, LLhR, and LLhA were constructed using an *in vivo* recombination method outlined by Jones (3). For this procedure, LLhG/pHG165a was used as vector. The strategy was to replace the GalR regulatory domain with that of GalS, RbsR, or AscG. To that end, the coding sequence of pHG165a and the sequence for LLhG amino acids 1-61 (equivalent to LacI 1-61 on pLS1) were amplified with appropriate primers (Supplementary Table 2); the large vector fragment excluded only the LLhG regulatory domain. Fragments for the three homologous regulatory domains were amplified from the pGemT plasmids; the primers used (Supplementary Table 2) also added an extra 20-25 base pair overlap with homology to the LLhG/pHG165a vector fragment. Residual LLhG/pHG165a and pGemT plasmids were linearized by digestion with either *Dpn*I or *Age*I so that they did not contaminate subsequent transformations. Fragments for vector and insert were then mixed, using 1-2.5µl of each, and transformed into 50µl of either DH5α Max Efficiency cells (Invitrogen) or XL1-blue cells (Agilent). Recombined pHG165 plasmids were purified from the cells, and the coding regions for all chimeras were fully sequenced to confirm construction (Northwoods).

Supplementary Table II. Primers used in chimera construction and mutagenesis.

A. PCR primers used in chimera construction

chimera	regulatory domain positions	primer sequence	primer name	restriction site
LLhF	62-334	GCTTCGTGCCGGCCGCACACGTTCTATTGG	CRA_NaeI Forward Primer	NaeI
		GCTACCTCAGGTTATTAGCTACGGCTGAGCACG	CRA Reverse2 Primer	
LLhT	63-315	CGTGGGCAAAGCAGCGCTGTGGTCGCCATC	TreR_AfeI	AfeI
		GCCAGGTACCTCAGGTCATCAGGACAGGGTGGCGG	TreR Reverse	
LLhC	68-341	CGTAATGAAAGCGCTACCATTCTGGTGATTG	CytR_Afe Forward	AfeI
		GCTACCTCAGGTTATTAAGGTAACGCGCGTGTTGATCCC	CytR Reverse 2 Primer	
LLhE	65-340	GTCACCCGACGTAGCGCTACCGTAGCCCTG	CelRAfe1 For	
		GCTACCTCAGGTTATCACCCGGATTCCCGCACCATCAAATG	CelRBsu361Rev	
LLhS	60-346	GCAACTCAGGTTAGCGCTACCATTGGCG	GalS-Afe1 Forward Primer	
		CTGCGCCCTGAGGTTATTACATCGCCTGAT	GalS-Bsu361 Reverse Primer	
		AACAACTGGCGGGCAAACAGAGCGACACCATTGGCGTGGTG	GalS-Afe1-RecombFOR	
		GAGGGGACGACGACAGTATCGGCCTGAGGTTATTACATCGC	GalS-RecombREV	
LLhR	60-330	GCCTCAAACTCAATCAAAGCGCTACCATTGGC	RbsR-Afe1forward primer	
		CTGCGCCCTGAGGTTACTAAGCCGAACCGC	RbsR-BSU361 Reverse primer	
		AACAACTGGCGGGCAAACAGAGCCATACCATTGGCATGTTG	RbsR-AfeI-RecombFOR	
		GAGGGGACGACGACAGTATCGGCCTGAGGTTACTAAGCCGA	RbsR-RecombREV	
LLhA	61-337	GAGGGGACGACGACAGTATCGGCCTCAGGTTATTATCGCGAAGGAGCAATGAG	AscG-Recomb REV	
		AACAACTGGCGGGCAAACAGAGCCAGACGCTGGGGCTGGTAGT	AscG-Reg Recomb FOR (2)	
General		CCGATACTGTCGTCGTCCCCTC	Chimer-RecombFOR	
		GCTCTGTTTGCCCGCCAGTTGTTGTG	LLh-RecombRev	
LacI		GCGGGCAGTGAGCCTAAGGCAATTAATG	Lac-Bsu	
		CATTAATTGCCTTAGGCTCACTGCCCGC	rev Lac-Bsu	
		GCATCGGAATTCCACCATCGAATGGTGCAAAACCTTTCG	LacI Forward	
		GCTAGGAATTCTCATCACTGCCCGCTTTCCAGTCGG	LacI Reverse	
		GCTAGGAATTCTCATCACAGCTGCATTAATGAATCGGC	LacI -11 Reverse	

B. Mutagenesis primers used to modify pHG165 and in chimera construction (to revert cloning sites, correct PCR mistakes, add "E230K", etc.)

CRA_Mut Forward	GTTTCCCGCCGAGACGGTGCTTTATCTTGGTGCG
CRA_Mut Reverse	CGCACCAAGATAAAGCACCGTCTCGGCGGGAAAC
LhC_Mut1	GCAAACAGAGCCGCACCATTCTGGTGATTGTCC
rev LhC_Mut1	GGACAATCACCAGAATGGTGCGGCTCTGTTTGC
LhC_Mut 2	CGGAGCTGGAGCTGCCTACAGTTCATATCGACAATCTGAC
rev LhC_Mut2	GTCAGATTGTCGATATGAACTGTAGGCAGCTCCAGCTCCG
pHG-O1out	CGTATGTTGTGTGGCGTGGTACTCATAACAATTTC
rev pHG-O1out	GAAATTGTTATGAGTACCACGCCACACAACATACG
RbsR-FSFor	GCCAGTACCAATCCTTTCTATTCA
RbsR-FSRev	TGAATAGAAAGGATTGGTACTGGC
RbsRmut1For	GGGCAAACAGAGCCATACCATTGGC
RbsRmut1Rev	GCCAATGGTATGGCTCTGTTTGCC
RbsRBsumutF	CCGCTGCGTCCACAGGCCGTC
RbsRBsumutR	GACGGCCTGTGGACGCAGCGG
Lac 109	CGTCGAAGCCTGTAAAACGGCGGTGCACAATC
rev Lac 109	GATTGTGCACCGCCGTTTTACAGGCTTCGACG
GalS-E230K	GGCGGCGATGGTTAAACTGCTGGGGCGC
rev GalS-E230K	GCGCCCCAGCAGTTTAACCATCGCCGCC
L-deletion/stop	GGT GAA TGT GAA ACC ATA ACG TTA TAC GAT GTC
L-secondstop	GGA AGC GGC GAT GGC GTA GCT CAA TTA C
GC - L-deletion/stop	GAC ATC GTA TAA CGT TAT GGT TTC ACA TTC ACC
GC - L-secondstop	GTA ATT GAG CTA CGC CAT CGC CGC TTC C

Supplementary Table III. Primers for Random Mutagenesis and Creation of Looping Constructs.
The “nnn” codons are in different colors for emphasis.
Yellow backgrounds indicate primers with synonymous codons.

LacI Linker Primers common to all chimeras

```
         45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E  L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  L  L  I  G
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTG
  GGCGGAGCTCNNNTACATTCCCAAC Lh-46RDM
      GAGCTCAATTACNNNCCCAACCGCGTGGCA  LLhPI48RDMFor
           CAATTACATTCCCNNNCGCGTGGCAC Lh-50RDM
           CAATTACATTCCCAACNNNGTGGCACAACAACTG LhG-RDM51
                  ATTCCCAACCGCNNNGCACAACAACTG pLLhP52RDM
                     CCCAACCGCGTGGCANNNCAACTGGCG Lh-54RDM
                       CAACCGCGTGGCACAANNNCTGGCGGGCAAACAG LhG-55RDM
                                GGCACAACAACTGGCGNNNAAACAGTCG Lh-58RDM
                                 GCACAACAACTGGCGNNNAAACAGTCG Lh-58RDM(2)
                                GGCACAACAACTGGCGNNNAAACAGAGC LhG-58RDM
```

LLhF

```
         45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E  L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  R  S  I  G
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGAGCCGTTCTATTGGTCTTGTGAT
                                              CTGGCGGGCNNNCAGTCGCGTTCTATTG LhF-59RDM
                                              CTGGCGGGCAAANNNTCGCGTTCTATTG LhF-60RDM
                                              CTGGCGGGCAAACAGNNNCGTTCTATTG LhF-61RDM
                                                     GCAAACAGTCGNNNTCTATTGGTCTTG LhF-62RDM
```

LLhT

```
         45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E  L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  K  V  V  A  I  I  V  T
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGAGCAAAGTGGTCGCCATCATTGTTAC
                                              CTGGCGGGCNNNCAGTCGAAAGTGGTC LhT-59RDM
                                              CTGGCGGGCAAANNNTCGAAAGTGGTC LhT-60RDM
                                              CTGGCGGGCAAACAGNNNAAAGTGGTC LhT-61RDM
                                                     GCAAACAGTCGNNNGTGGTCGCCATC LhT-62RDM
```

```
LLhS
          45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E   L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  D  T  I  G
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGAGCGACACCATTGGCGTGGTGGT
                                                      GGCAAACAGAGCNNNACCATTGGCGTG LhS-62RDM

LLhC
          45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E   L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  A  T  I  L  V  I  V
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGAGCGCTACCATTCTGGTGATTGT
                                                CTGGCGGGCNNNCAGAGCGCTACCATTC  LhC-59RDM
                                                CTGGCGGGCAAANNNAGCGCTACCATTC  LhC-60RDM
                                                   GCGGGCAAACAGNNNGCTACCATTCTG  LhC-61RDM
                                                      GGCAAACAGAGCNNNACCATTCTGGTG  LhC-62RDM
LLhE
          45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E   L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  D  T  V  A  L
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGGACACCGTAGCCCTGGTGGT
                                                CTGGCGGGCNNNCAGTCGGACACCGTAG  LhE-59RDM
                                                CTGGCGGGCAAANNNTCGGACACCGTAG  LhE-60RDM
                                                   GCGGGCAAACAGNNNGACACCGTAGCCC  LhE-61RDM
                                                   GCGGGCAAACAGTCGNNNACCGTAGCCC  LhE-62RDM

LLhA
          45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
M  A  E   L  N  Y  I  P  N  R  V  A  Q  Q  L  A  G  K  Q  S  Q  T  L  G  L
ATGGCGGAGCTCAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGAGCCAGACGCTGGGGCTGGTAGTG
                                                      AAACAGAGCNNNACGCTGGGGCTGGTAG  LhA-62RDM
```

LacI Y282D mutagenesis for looping strains (mutation is underlined)

5′-GACGATAC$_2$GA$_2$GACAGCTCATGT<u>GAC</u>ATC$_3$GC$_2$GT$_2$A$_2$C$_2$AC$_2$ATCA$_3$CAG

5′-CTGT$_3$GATG$_2$TG$_2$T$_2$A$_2$CG$_2$CG$_3$AT<u>GTC</u>ACATGAGCTGTCT$_2$CG$_2$TATCGTC

LacI Y282D PCR confirmation in looping strains

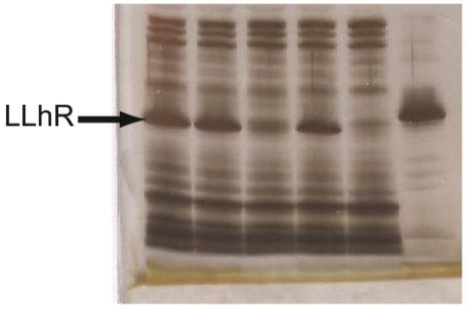
5′-A$_2$G$_2$CGACTG$_2$AGTGC$_2$ATG

5′-GA$_3$C$_2$TGTCGTGC$_2$AGCTG

Supplementary Figure 1. DNA pull-down assay shows that chimera variants are expressed as active proteins *in vivo*. Details about each panel are described on the next page.

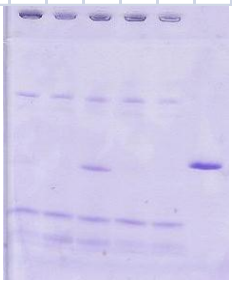
A:

LLhR	+	+	+	+	–	Purified LacI
$lacO^1$	+	–	–	–	–	
$lacO^{sym}$	–	+	–	–	+	
O^{non}	–	–	+	–	–	
$lacO^2$	–	–	–	+	–	



B:

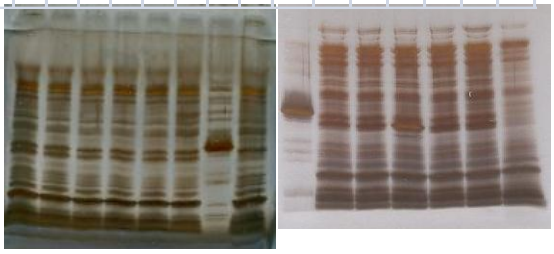
LLhA	+	+	+	+	-	Purified LacI
$lacO^1$	+	-	-	-	+	
$lacO^{sym}$	-	-	+	-	-	
O^{non}	-	+	-	-	-	
$lacO^2$	-	-	-	+	-	



C:

LLhC	+	+	+	+	+	+	-	+	-	-	-	-	-
CL1	-	-	-	-	-	-	-	-	+	+	+	+	-
$lacO^1$	+	-	-	-	-	-	+	-	+	-	-	-	+
$lacO^{sym}$	-	-	+	-	-	-	-	-	-	+	-	-	-
O^{non}	-	+	-	-	-	-	-	-	-	-	+	-	-
$lacO^2$	-	-	-	-	-	-	-	+	-	-	-	+	-
O^e	-	-	-	+	-	-	-	-	-	-	-	-	-
O^i	-	-	-	-	+	-	-	-	-	-	-	-	-
O^{disC}	-	-	-	-	-	+	-	-	-	-	-	-	-

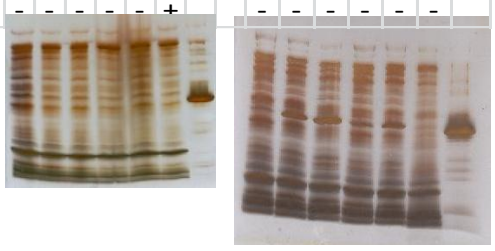
Purified LacI | Purified LacI



D:

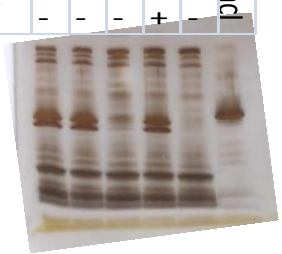
LLhE	+	+	+	+	+	+	+	-	-	-	-	-	-
3mut	-	-	-	-	-	-	-	+	+	+	+	-	
$lacO^1$	+	-	-	-	-	-	-	+	-	-	-	+	
$lacO^{sym}$	-	-	+	-	-	-	-	-	+	-	-	-	
O^{non}	-	+	-	-	-	-	-	-	-	+	-	-	
$lacO^2$	-	-	-	-	-	-	-	+	-	-	+	-	
O^e	-	-	-	+	-	-	-	-	-	-	-	-	
O^i	-	-	-	-	+	-	-	-	-	-	-	-	
O^{disC}	-	-	-	-	-	+	-	-	-	-	-	-	

Purified LacI | Purified LacI



E:

LLhF	+	+	+	+	-	Purified LacI
$lacO^1$	+	-	-	-	+	
$lacO^{sym}$	-	+	-	-	-	
O^{non}	-	-	+	-	-	
$lacO^2$	-	-	-	+	-	



Supplementary Figure 1, continued.

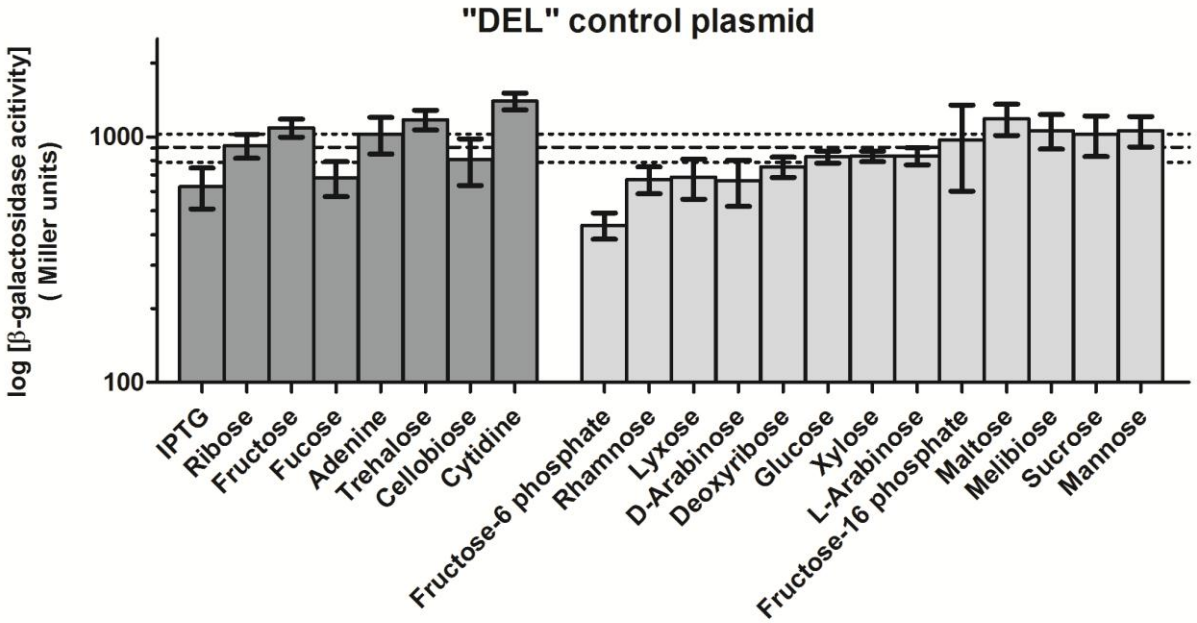
DNA pull-down assays of crude cell extracts were performed for all chimeras and their variants, in order to ascertain that protein is expressed *in vivo* and capable of binding operator DNA such as $lacO^{1}$, $lacO^{sym}$, or $lacO^{2}$ immobilized to magnetic beads. Most variants showed very high protein levels bound to all operators. The ability to bind *lacO* DNA indicates that the chimera construction or mutation does not structurally disrupt the repressor. As referenced in the text of the manuscript, we previously estimated that bands detectable by Coomassie stain correspond to >2500 repressors per cell. (All gels were stained with Coomassie prior to silver staining).

(**A)** LLhR is shown as an example of a repressor variant with high protein expression. The arrow indicates the protein band observed in SDS-PAGE for LLhR; this band is also visible with Coomassie stain. Note that nonspecific DNA (O^{non}) does not pull down much repressor (a faint band is observed in lane 3 when compared to the no-repressor control in lane 5).

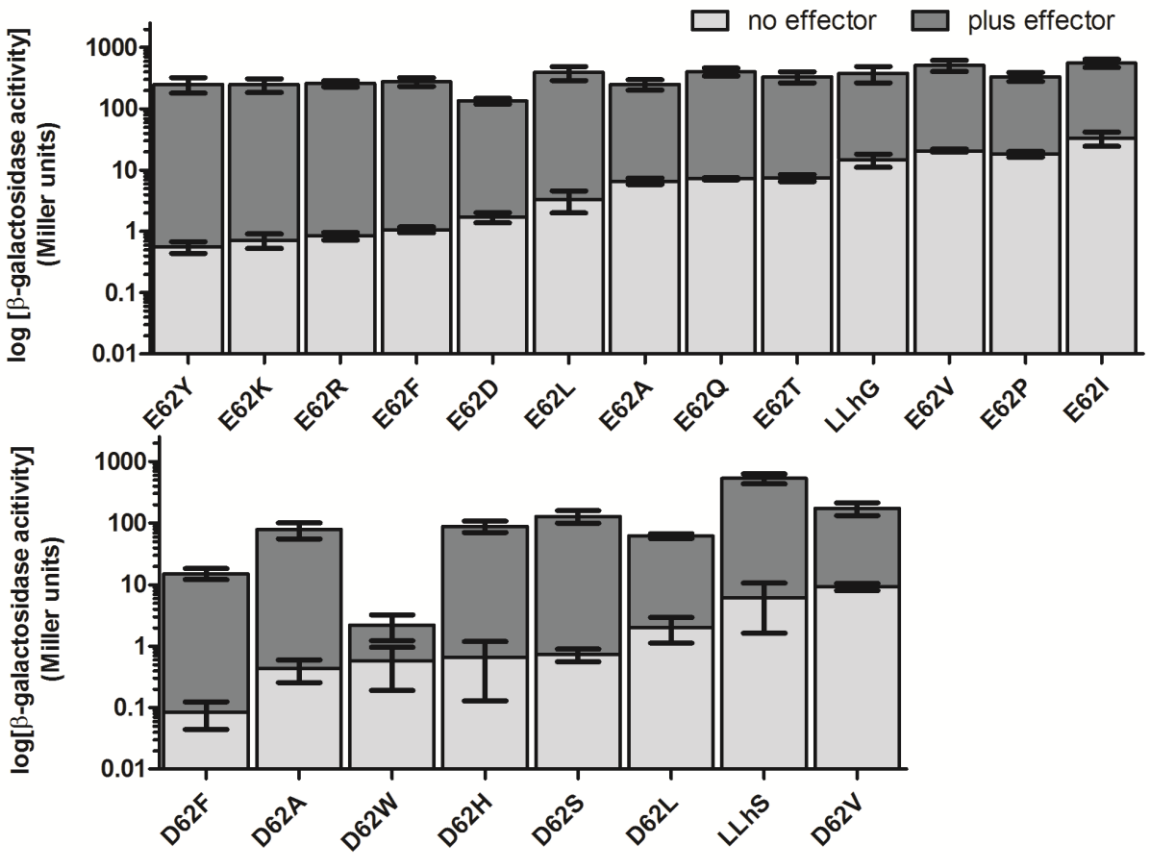
(B) LLhA showed high levels with $lacO^{sym}$, which indicates that binding to other operators must be extremely weak. As a comparison, note that – even in the presence of inducer IPTG – wild-type LacI is pulled down at high levels (data not shown).

(**C)** LLhC showed only nonspecific binding to any operator (compare to the last lane of the first gel, which has no repressor). (**D**) LLhE did not show detectable protein with any operator. The pull-down assay does not discriminate between "no protein expression" and "very poor binding", and we cannot rule out either possibility. However, all mutated variants of LLhC and LLhE showed high levels of protein expression when captured by $lacO^{1}$ and/or $lacO^{sym}$. Panel **(C)** shows the "CL1" variant of LLhC (Q55I) as an example, and panel **(D)** shows the LLhE "3mut" variant (I48V, Q55A, Q60R). The gels shown in **(C)** and **(D)** also show results for operators O^{e}, O^{i} and O^{disC} , which were performed for a separate project.

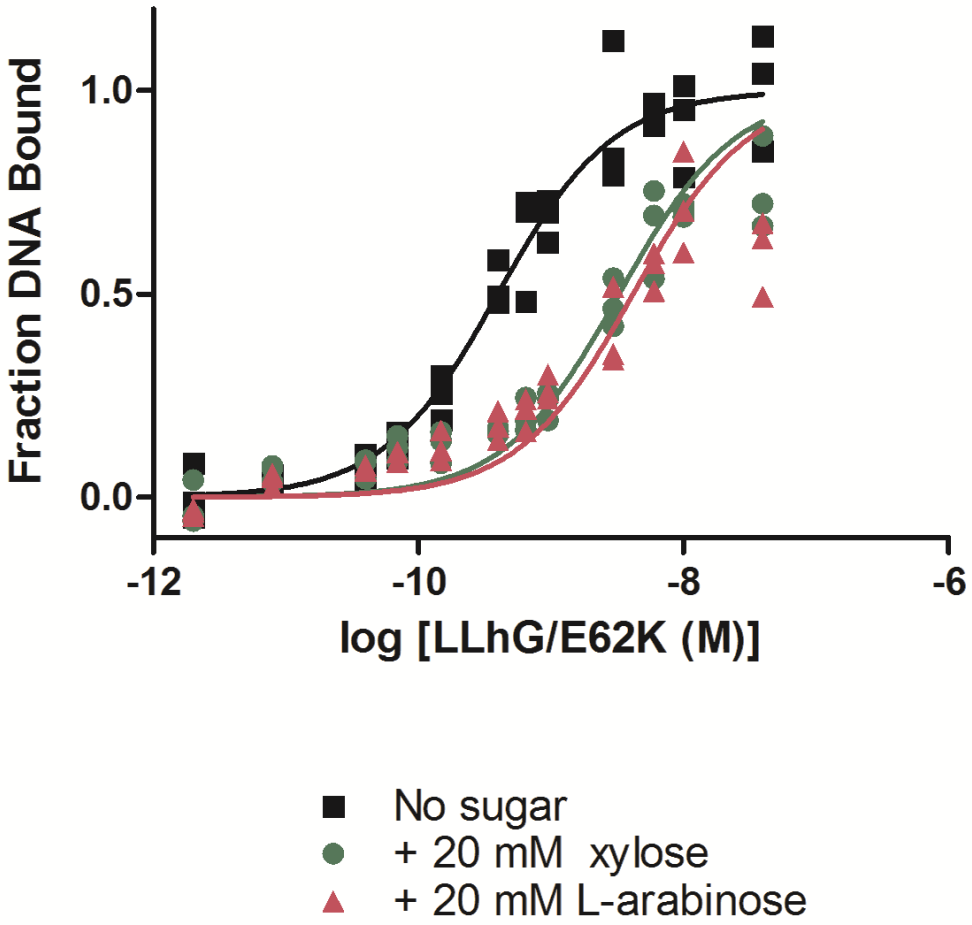
(E) Surprisingly, LLhF showed two strong bands in the pull-down assays, which were present in nearly equimolar amounts. The smaller band could be a proteolysis product of LLhF or result from a hetero-protein interaction.



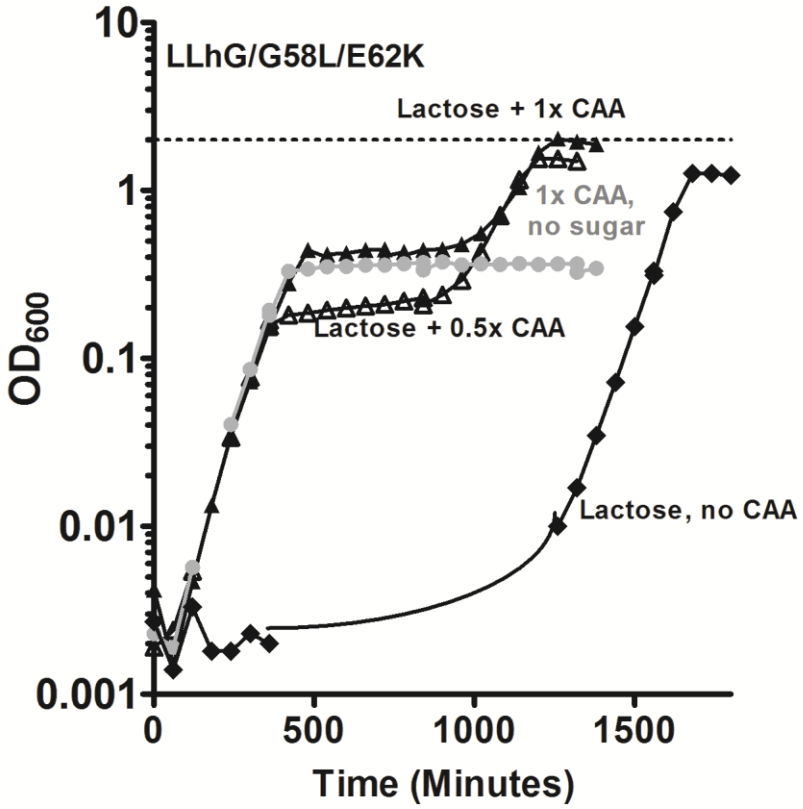
Supplementary Figure 2. The horizontal dashed line indicates the average β-galactosidase value determined for the "DEL" control plasmid in the absence of any effector; the flanking dotted lines indicate one standard deviation of the mean. The dark gray bars on the left show average values for DEL determined in the presence of natural effectors (or their upstream metabolites) and the gratuitous inducer for LacI, IPTG. The light gray bars on the right show values for DEL in the presence of other potential gratuitous effectors. All error bars depict one standard deviation from the mean. All effector concentrations are listed in Table 2. Effectors that diminished activity might do so by competing with the ONPG substrate for the β-galactosidase enzyme. Cytidine might enhance β-galactosidase activity by creating more favorable media/cellular conditions (energetic or nutritional) for *lacZ* mRNA transcription and enzyme expression.



Supplementary Figure 3. Mutation of position 62 enhances repressor function of both LLhG and LLhS. Site-directed, random mutagenesis of position 62 was used to identify amino acid substitutions that either enhance or diminish repression of the lac operon by LLhG (top panel; data from Meinhardt and Swint-Kruse, 2008 has been un-normalized in this plot) and LLhS (bottom panel). The light bars in the front series depict β-galactosidase activities determined in the absence of inducer. The dark bars in the rear series depict β-galactosidase activity determined in the presence of inducer (fucose). Error bars indicate one standard deviation from the mean. Although mutagenesis of position 62 enhanced repression in both chimeras, comparing the rank order of substitutions shows that individual amino acids can elicit different functional effects in the two homologs (note the 62A mutation in particular.) The fold-effect of a mutational change can also differ: Relative to D at position 62, V changed 10-fold in LLhG but only 1.5-fold in LLhS; F changed 20-fold in LLhG and 100-fold in LLhS. All liquid culture values were in agreement with plate assays, except LLhS/D62W, which showed more repression in plate assays.



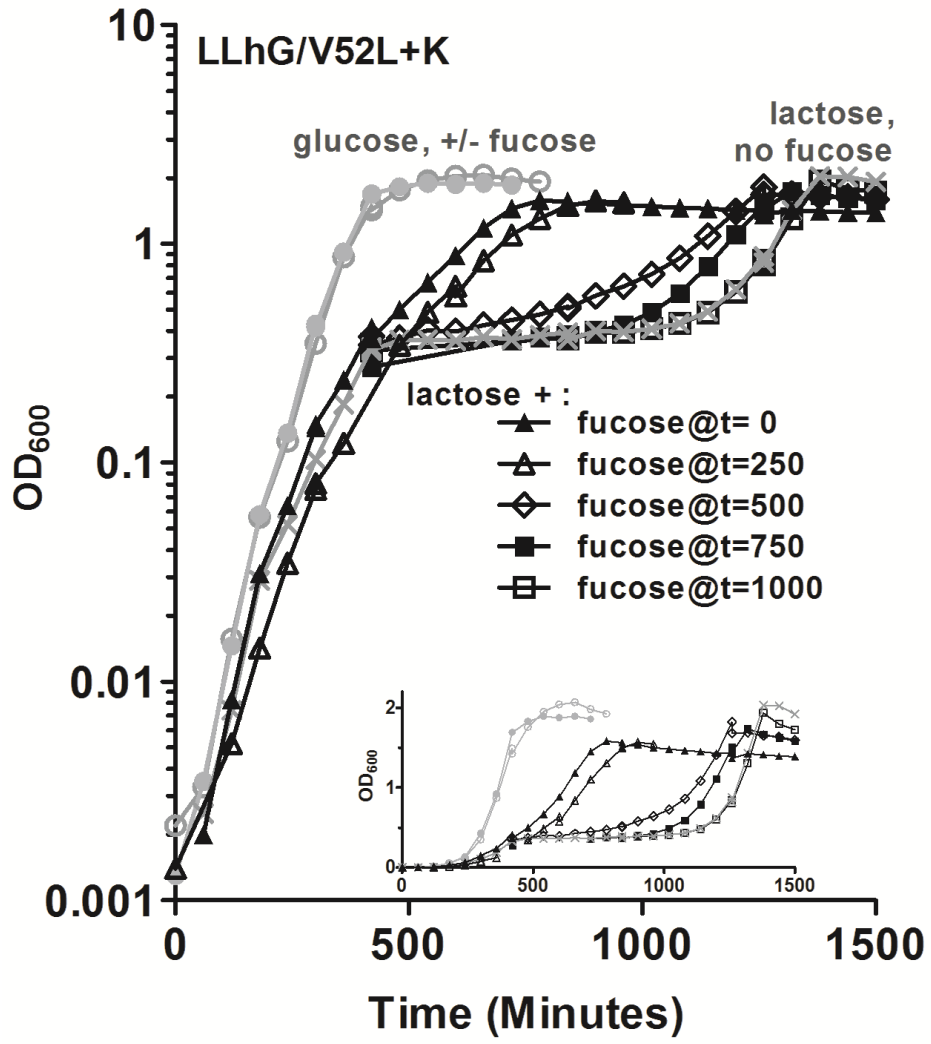
Supplementary Figure 4. Xylose and L-arabinose allosterically reduce the binding affinity of LLhG/E62K for *lacO*[1] DNA. DNA binding assays were carried out under equilibrium conditions for purified LLhG/E62K using purification and binding conditions described in (4). Black squares show binding in the absence of sugar (K_d =4x10^{-10} M). In the presence of 20mM xylose (green circles) and 20 mM L-arabinose (pink triangles), the DNA binding affinity is reduced by at least an order of magnitude, with $K_d \geq$ 4x10^{-9} M. The solid lines indicate the best fit of the data to the binding equation described in (4).



Supplementary Figure 5. The first growth phase is from metabolism of caseamino acids (CAA). When grown on lactose minimal media, bacterial cultures expressing moderate/strong chimeria variants show two distinct growth phases. The example shown (solid black triangles) is for LLhG/G58L/E62K. The first growth phase also occurs for cultures with control plasmids that do not express repressor protein, in minimal media that lacked added sugar (data not shown). Thus, the minimal MOPS media must provide another carbon source.

One ingredient was 0.2 % caseamino acids ("CAA"). To determine whether CAA was the carbon source of the first growth phase, cultures of a LLhG/G58/E62K were grown under four different conditions: (i) with 1X CAA (0.2%) but no sugar (gray circles); (ii) with 1X CAA (0.2%) and lactose (black triangles); (iii) with 0.5X CAA (0.1%) and lactose (open triangles); and (iv) with lactose but no CAA (black diamonds). The single growth phase for condition i closely tracked the first growth phase in condition ii. Cutting the CAA concentration in half (condition iii) essentially halved the height of the first plateau. Eliminating the CAA (condition iv) eliminated the first growth phase. Clearly, the first growth phase was fueled by the CAA.

Condition iv (lactose, no CAA) also had the effect of greatly increasing the lag prior to the second growth phase. Glucose cultures also grew much more slowly in the absence of CAA (data not shown), which probably reflected the general burden of synthesizing the necessary amino acids for growth.



Supplementary Figure 6. In lactose minimal media, bacterial cultures expressing moderate/strong repressors exhibit biphasic growth curves. We tested whether the second growth phase was similar that resulting from induction of the lac operon by using the second inducer for GalR and LLhG variants, D-fucose (see Table 1 in the manuscript for references). This sugar is not metabolized by 3.300 *E. coli* cells and does not support culture growth (data not shown).

For cultures with the slow-growing variant LLhG/V52L/E62K, fucose was added to the lactose minimal media at different time points, which diminished the lag before the second growth phase in a time dependent manner. Gray "X"s are used to depict a growth curve on lactose minimal media with no fucose; this curve overlaps that with D-fucose added at t=1000. Gray circles: Glucose minimal media +/- D-fucose at time zero. Other addition times are indicated on the figure. Surprisingly, fucose addition also appeared to slightly reduce the plateau height of the second growth phase in a fucose-addition-time dependent manner (inset). Given their chemical similarities, D-fucose might inhibit other proteins involved in metabolizing galactose.

Supplementary Figure 7. β-galactosidase activity at the end of growth assays. Left cuvettes: Little to no activity is seen at the end of glucose growth assays. Right cuvettes: The substrate MUG is actively hydrolyzed to fluorescent product at the end of lactose growth assays, which indicates that transcription of the *lac* operon is up-regulated in these cultures. For each pair of cultures, the OD_{600} is >1.5.

Supplementary References

1. Meinhardt, S. and Swint-Kruse, L. (2008) Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives. *Proteins*, **73**, 941-957.
2. Tungtur, S., Egan, S.M. and Swint-Kruse, L. (2007) Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins*, **68**, 375-388.
3. Jones, D.H. (1994) PCR mutagenesis and recombination in vivo. *PCR Methods Appl*, **3**, S141-148.
4. Tungtur, S., Skinner, H., Zhan, H., Swint-Kruse, L. and Beckett, D. (2011) In vivo tests of thermodynamic models of transcription repressor function. *Biophysical Chemistry*, **159**, 142-151.