

## RESEARCH ARTICLE

# A quantitative understanding of lac repressor's binding specificity and flexibility

Zheng Zuo, Yiming Chang and Gary D. Stormo\*

Department of Genetics and Center for Genomic Sciences and Systems Biology, School of Medicine, Washington University, St. Louis, MO 63108, USA

\* Correspondence: stormo@genetics.wustl.edu

Received January 25, 2015; Revised March 21, 2015; Accepted April 8, 2015

**Lac repressor, the first discovered transcriptional regulator, has been shown to confer multiple modes of binding to its operator sites depending on the central spacer length. Other homolog members in the LacI/GalR family (PurR and YcjW) cannot bind their operator sites with similar structural flexibility. To decipher the underlying mechanism for this unique property, we used Spec-seq approach combined with site-directed mutagenesis to quantify the DNA binding specificity of multiple hybrids of lacI and PurR. We find that lac repressor's recognition di-residues YQ and its hinge helix loop regions are both critical for its structural flexibility. Also, specificity profiling of the whole lac operator suggests that a simple additive model from single variants suffice to predict other multivariant sites' energy reasonably well, and the genome occupancy model based on this specificity data correlates well with *in vivo* lac repressor binding profile.**

**Keywords:** lac repressor; binding flexibility; Spec-seq; ionic strength

## INTRODUCTION

Lac repressor, as one classical transcriptional regulator in *Escherichia coli* [1,2], is a homodimer protein and therefore would be presumed to bind its cognate operator site in palindromic and perfectly symmetric fashion. However it was discovered that the *in vivo* lac operators are approximately symmetric and carry a few mismatches between their left and right half-sites[3]. Our previous work[4] showed that the lac repressor binds to the wild-type lac operator in an intrinsic asymmetric fashion. But that work only focused on the inner, asymmetric part (−4 to +4) of the operator, and did not include the outer operator regions (−10 to −5, +5 to +10) that were presumed to be symmetric in terms of sequence specificity (Figure 1A).

Here we designed additional randomized dsDNA libraries to cover the entire operator site (−10 to +10; Figure 1B), and measured the relative binding energy for all single variants and adjacent double variants. Additionally, we varied the ionic strength of the binding buffer as it has been shown that affinity is affected by the salt

concentration [5,6] and some studies suggest that ionic strength can even have a significant impact on transcription factors' binding specificity[7].

If the binding energy to any particular site can be derived by summing the mismatched energy costs compared to the preferred consensus sequence, we can say this means perfect additivity. Very often, this assumption is violated at high-energy plateau, but found to be generally good estimation for lower-energy binding sites[8,9]. For basic helix-loop-helix (bHLH) proteins [10] it was shown that nearly all of the multivariant sites have lower energy than predicted from the sum of the single variants' energies, which we can interpret as that the protein can compensate for the energy loss for multivariant sites. However in our previous work, we found that for CG spacer R2 library, all of the tested double variants have higher energy values and bind with lower affinity than the additive prediction from single variants, usually by at least 1 kT. There could be various interpretations for this result. Here we did Spec-seq for the whole lac operator including all the possible single and adjacent double variants of *O1* operator, thus it is possible

to know this “additivity violation” property across the whole operator site.

To our knowledge lac repressor is the only example known to be able to bind operator sites with variable spacers in LacI/GalR family [11] so far, which we call “binding flexibility” within *E. coli*. Two other LacI family members, PurR and YcjW, were shown to have even spacer operator sites natively and cannot bind with equal high affinity in an extended conformation like lac repressor. Though there are some hypotheses for the selective advantages of such unique configuration evolutionarily [12], the structural mechanism is still elusive. In this work, we used site-directed mutagenesis approach to mutate and swap some positions in lacI and PurR we suspected to be important for the binding flexibility. For each individual hybrid protein, its specificity profile under different spacer lengths was quantified by Spec-seq. Thus we obtain a quantitative understanding of lac repressor’s structural flexibility. Most interestingly, it was found that when we swap lac repressor’s recognition di-residues YQ and hinge-helix loop region into PurR, the mutant form PurR (p4) can bind its operators with multiple spacer lengths similar to lac repressor.

## RESULTS

Figure 1A describes our current understanding about lacI and PurR binding to their operators in a schematic model. For lac repressor, it can adopt three different configurations binding to the operators with 2 bp, 3 bp, and 4 bp spacing in the middle (L2L’, L3R, R’4R) with similar affinities (Table 1). Its hinge helices always recognize and kink the central CG dinucleotide. The YQ di-residues at positions 17–18 recognize CTC motif (2–4) by default, but prefer ATA in the extended conformations. The wild-type lac operator O1 (also O2 and O3) adopts the L3R conformation and therefore are recognized differently between its left and right half-sites. For PurR, the extended conformation L3R is prohibited or decreased by at least 3 kT compared to the L2L’ format (Table 2).

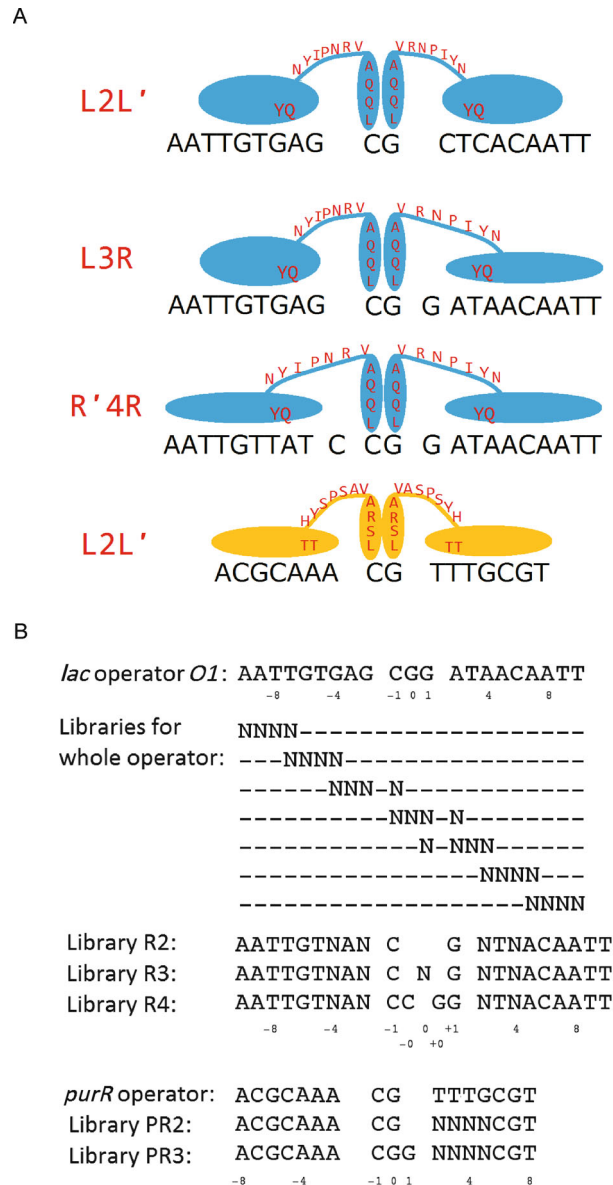
Figure 1B lists all the randomized dsDNA libraries used in this study. To get the specificity profile across the whole lac operator region, we designed 7 tandem overlapping “NNNN” degenerate dsDNA libraries with total diversity no more than 2,000, which covers all the possible single variants and adjacent double variants of O1 site. The R2, R3, and R4 libraries were designed to target the central asymmetric regions with different spacers and cover 3 key configurations (L2L’, L3R, and R’4R). purR’s PR2 and PR3 libraries have smaller sizes (512 total), and were primarily used to test if any hybrid PurR protein can have some structural flexibility to bind to the extended operator as lacI does.

### Lac repressor’s whole operator site is intrinsically asymmetric, both in sequence and specificity level

Figure 2 shows the energy logos produced by regression analysis of our measured binding energy for all those single and double variants compared to wild-type lac operator O1. Firstly, this result is consistent with our previous work that the central core region has asymmetric motif between its left and right half sites, primarily due to their different spacings to the central CG di-nucleotide (–1, 0) recognized by the hinge helices. Also, it is notable that the TGT motif at positions –7 to –5 are significantly higher than that of ACA at positions +5 to +7, though they are symmetric in sequence and recognized by the same set of residues structurally. The energy logo shown is from the standard buffer (1× NEB buffer 4). We also performed the binding reaction at higher ionic strengths (2× and 4× NEB buffer 4) but found essentially no change in specificity (Supplemental Figures 1 and 2) within our measurement precision. While it has been shown previously that higher ionic strengths give rise to lower binding affinities or association constants [6,13], our results show that there is almost no change in specificity, suggesting the ionic effects alter interactions with the DNA backbone exclusively.

### Most of observed additivity violations in lac operator are either neutral or compensatory

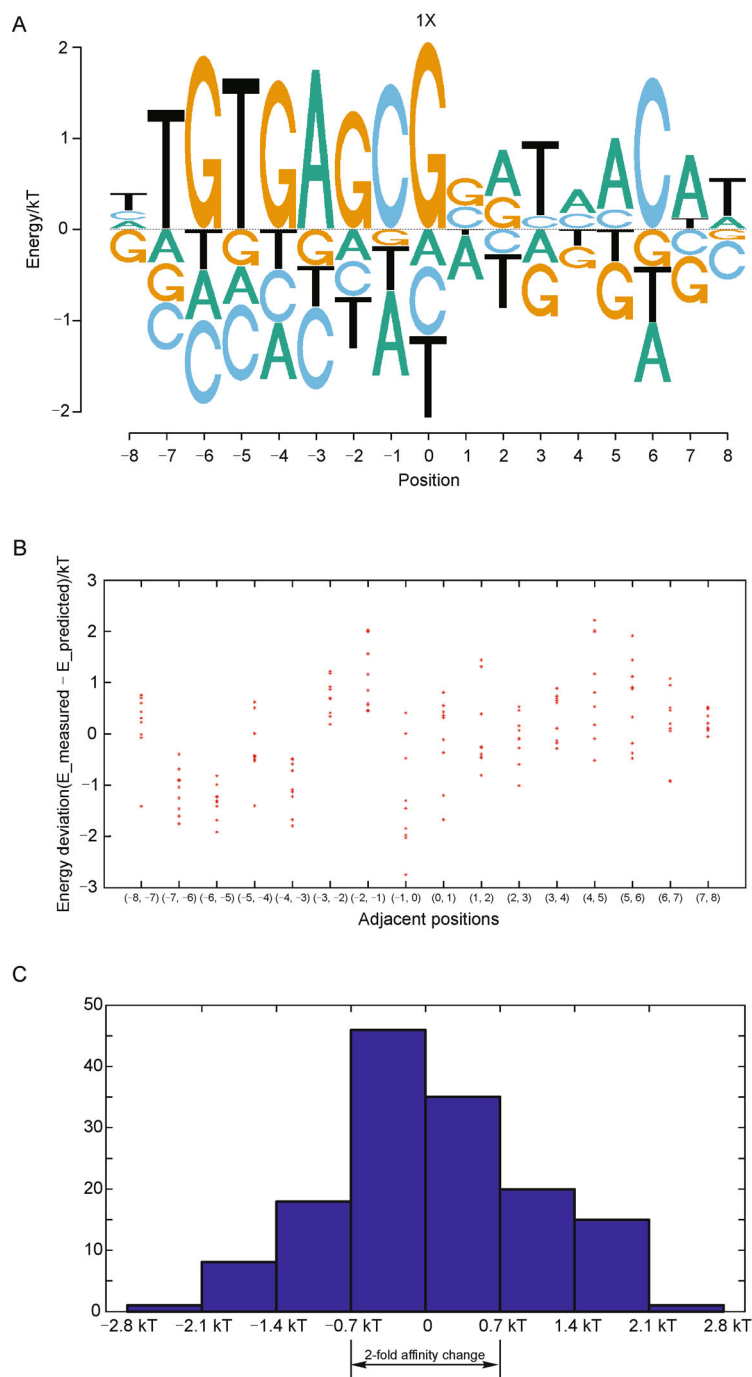
For lac operator O1 from positions –8 to +8, there are totally  $3^2 \times 16 = 144$  adjacent double variants and they are all included in our measurements. For each adjacent double variant, the difference between the observed binding energy and the value calculated by its two single variants can be used as indicator for “additivity violation”. If this energy deviation value is negative, i.e., the measured binding energy has lower value than the predicted number, we can call this “compensatory”, otherwise it is “anti-compensatory”. Figure 2B shows the energy deviation vs. variant pair position for all those 144 double variants. Clearly most of variant pairs have no more than 1 kT energy deviation from the additive model. Furthermore most of the compensatory deviations from additivity occur because of the non-specific binding plateau. The sum of the two single mutants exceeds that plateau, so the double mutant has reduced energy compared to the sum. For position –2, which has only small energy increases for single mutants, all of the adjacent double mutants have large positive increases over the sum, often approaching the non-specific plateau. The right half-site is more mixed, with combinations of both positive and negative deviations from additivity, but most of them being modest in size (Figure 2B and Table S1). Figure 2C is the histogram counting the number of variant pairs with different energy deviation levels.



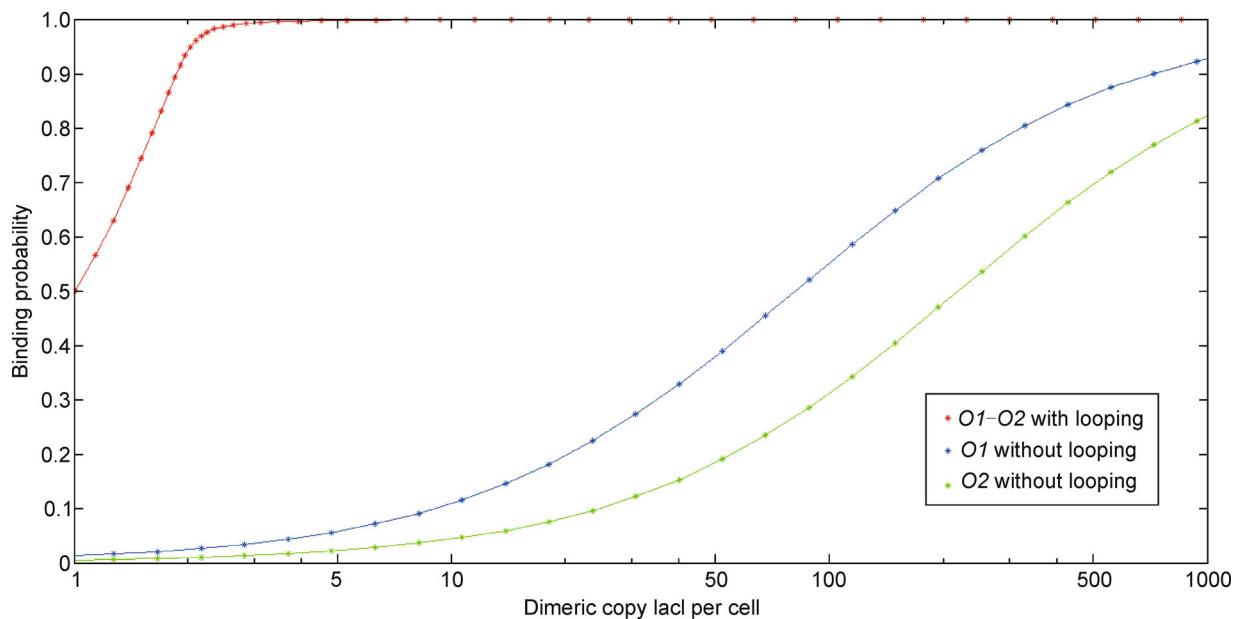
**Figure 1. LacI and PurR's DNA binding models and randomized libraries for Spec-seq runs.** (A) Schematic models for lacI and PurR binding (Blue for lacI and yellow for PurR). Following our previous work's nomenclature, L2L' and R'4R represent symmetric binding conformations with 2 bp CG spacer and 4 bp CCGG spacer in the middle respectively, whereas L3R format carries 3 bp CGG spacer format and binds its left and right half sites differently. Wild-type PurR can only bind defined spacer format L2L'. (B) Libraries design in this study. To cover the whole *lac* operator region, 7 overlapping degenerate DNA libraries were adopted so that all single and adjacent double variants of the wild-type *O1* operator site were included inside; Libraries R2, R3, and R4 were designed to test different lac repressor hybrids (m1–m8)'s binding specificities under 3 different spacer lengths (CG, CGG, and CCGG); Libraries PR2, PR3 were designed to test PurR hybrids (p1–p4)'s binding specificities under 2 spacer formats (CG and CGG). Note: For PurR p4, since it has recognition residues changed, the left half site of PR2 and PR3 is modified from CAAA to CGAA.

Clearly, most of them fall within the 0.7 kT deviation bounds, which corresponds to 2-fold affinity difference. This result clearly shows that, for lac repressor, having

knowledge for all of its single variants' energy levels allows us to usually predict its double variants' binding energy correct within 2-fold affinity accuracy.



**Figure 2. Specificity profiling the whole *lac* operator.** (A) Energy logo for the whole *lac* operator under 1× NEB buffer 4 condition. Energy logo was produced based on the linear regression of all the wild-type *O1* operator's single and adjacent double variants' energy values. (B) Measured vs. predicted energy deviation for all double variant pairs. The energy value of each adjacent double variant can either be measured directly or predicted by summing up the contributions of each individual base variation. Each double variant is plotted in panel B based on its position in the operator and energy deviation value between measured and predicted numbers. (C) Distribution of measured vs. predicted energy for all double variants. All energy deviation values in panel B were re-plotted in histogram form, and it is clear that most of adjacent double variants follow "additivity assumption" reasonably well.



**Figure 3. Modeling of *lac* operator occupancy.** Based on the whole *lac* repressor's specificity profile, an occupancy model across *E. coli* genome is used to predict the binding probability of operator *O1*, *O2* without looping, and *O1–O2* under looped condition (detailed in Supplementary Materials).

### Binding energy model for *lac* repressor can be used to predict its *in-vivo* occupancy level

One important motivation to study transcription factor's specificity is to understand how each *cis*-regulatory element inside living cells gets bound by the corresponding transcription factor (TF) with reasonable occupancy level. For *lac* repressor, given the energy matrix derived from our measurement, it is possible to get reasonable good knowledge of the binding energy for every possible site. Assuming all specific sites can neither carry more than 4 mismatches to consensus site nor have binding energy more than 7 kT, it's estimated that there are no more than 300 specific binding sites. Figure 3 depicts the predicted occupancy level for *O1* and *O2* sites in three situations, i.e., looped *O1–O2*, *O1* without looping, and *O2* without looping. It looks very similar to the one illustrated by von Hippel [14], though we need to introduce some additional simplified assumptions, i.e., non-specific energy level around 11 kT, low copy number of *lac* repressors per cell, and the looping's synergistic effect. Detailed modeling descriptions can be found in Supplementary materials.

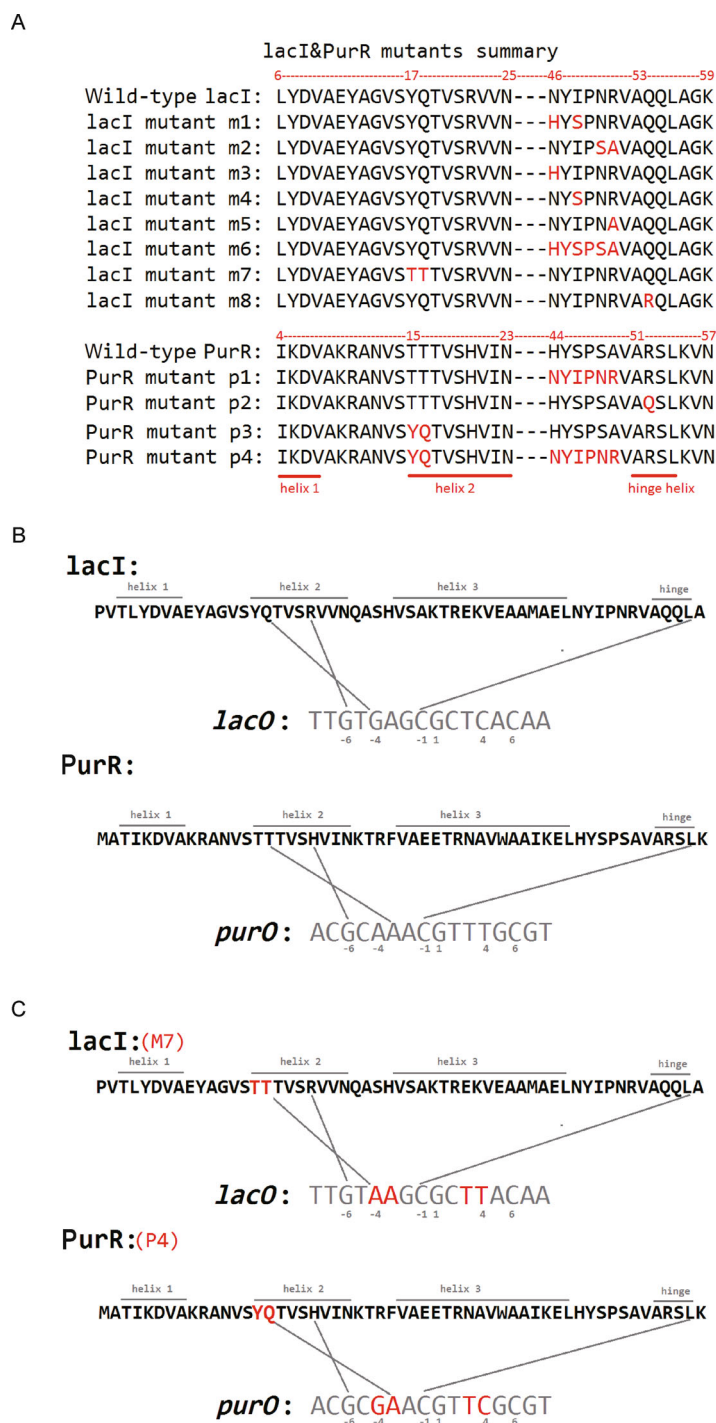
### Recognition di-residues YQ and the hinge helix loop of *lac* repressor are required to confer its structural flexibility

Given existing structures for *lacI* and PurR complexed

with their corresponding operator fragments [15–17], we suspected there are three possible regions responsible for *lac* repressor's unique structural property: the hinge helices; the hinge helices loop connecting to the helix-turn-helix (HTH) DNA binding domain; and the recognition di-residues YQ contacting the bases 2–4. Using *purR* as the control homolog gene, we built series of *lacI* and PurR hybrid proteins to swap these regions one by one, and in specific combinations, between the two proteins (Figure 4A).

For *lacI* hybrid mutants from m1 to m5, each of them has either one or two residues in the linker region mutated to match the corresponding PurR region, whereas m6 has the whole linker switched to match PurR's. The m7 mutant changes the recognition residues YQ for operator position 2–4 to TT, in which case we expect it would bind CTT motif at bases 2–4 instead of CTC by wild-type protein [18]. One interesting fact for *lacI*'s hinge helix is that it has sequence AQQL, instead of the ARXL format for most other LacI family TFs. It was speculated that the weak interaction between Q54 and N25 contributed to the loss of binding energy of wild-type *O1* compared to the perfectly symmetric L2L' site [4, 19]. So m8 mutant is used to test this hypothesis.

We did Spec-seq experiments using randomized DNA libraries R2, R3, and R4 to cover various *lac* operator sites with different spacer lengths. Ideally, if the mutant protein showed significantly different relative energy levels under various spacers compared to the wild-type



**Figure 4. Construction of new lacI and PurR mutants and the newly evolved operators.** (A) lacI and PurR mutants used in this study. Wild-type lacI and PurR DNA binding domains were aligned based on sequence homology. Various hybrid mutants were constructed to target potentially critical regions for lac repressor's DNA binding flexibility (m1–m8, p1–p4). (B) Operators recognized by wild-type lacI and PurR. For wild-type lacI and PurR, the leucine residues in the hinge helix region coordinate the central CG dinucleotide in the operator site, whereas YQ residues in lacI and TT residues in PurR recognize the GA and AA (–4 to +3) sites respectively. (C) Alternative operator sites recognized by mutant lacI m7 and PurR p4.

**Table 1. Properties of wild-type and hybrid mutants lacI.**

Mutant property		L2L'	L3R	R'4R
Wild-type lacI		0	0.5 kT	0.3 kT
M3	Loop region N46H	0	-0.5 kT	-0.4 kT
M4	Loop region I48S	0	0.6 kT	2.5 kT
M5	Loop region R51A	0	0.3 kT	1.6 kT
M7	Recognition residues Y17T, Q18T	0	1.8 kT	3.0 kT
M8	Hinge helix AQQL → ARQL	0	0	0.8 kT

For M7, due to motif recognition change, the L2L', L3R, and R4R' changed to AAG CG CTT, AAG CGG TTA, and TAG CCGG CTAA respectively. M1, M2, and M6 have no visible shifted bands.

**Table 2. Properties of wild-type and hybrid mutants PurR.**

Mutant property		Preferred site overall L2L'	Preferred site in CGG spacer	L3L'
Wild-type		CAAA CG TTTG 0	CAAA CGG TTGC 1 kT	CAAA CGG TTTG 3.3 kT
P1	Loop region HYSPSA → NYIPNR	CAAA CG aTTT 0	CAAA CGG TTGC 0.8 kT	CAAA CGG TTTG 1.5 kT
P2	Hinge helix ARSL → AQSL	CAAA CG aTTG 0	CAAA CGG TTGC 1.9 kT	CAAA CGG TTTG 4.6 kT
P4	Loop region HYSPSA → NYIPNR Recognition residues T15Y, T16Q	CGAA CG TTCG 0	CGAA CGG TTCG (L3L') 0.4 kT	CGAA CGG TTCG (L3L') 0.4 kT

For P4, variant site ACG CAAA CGG TCGC CGT has energy of 0.8 kT. P3 has no visible shifted band.

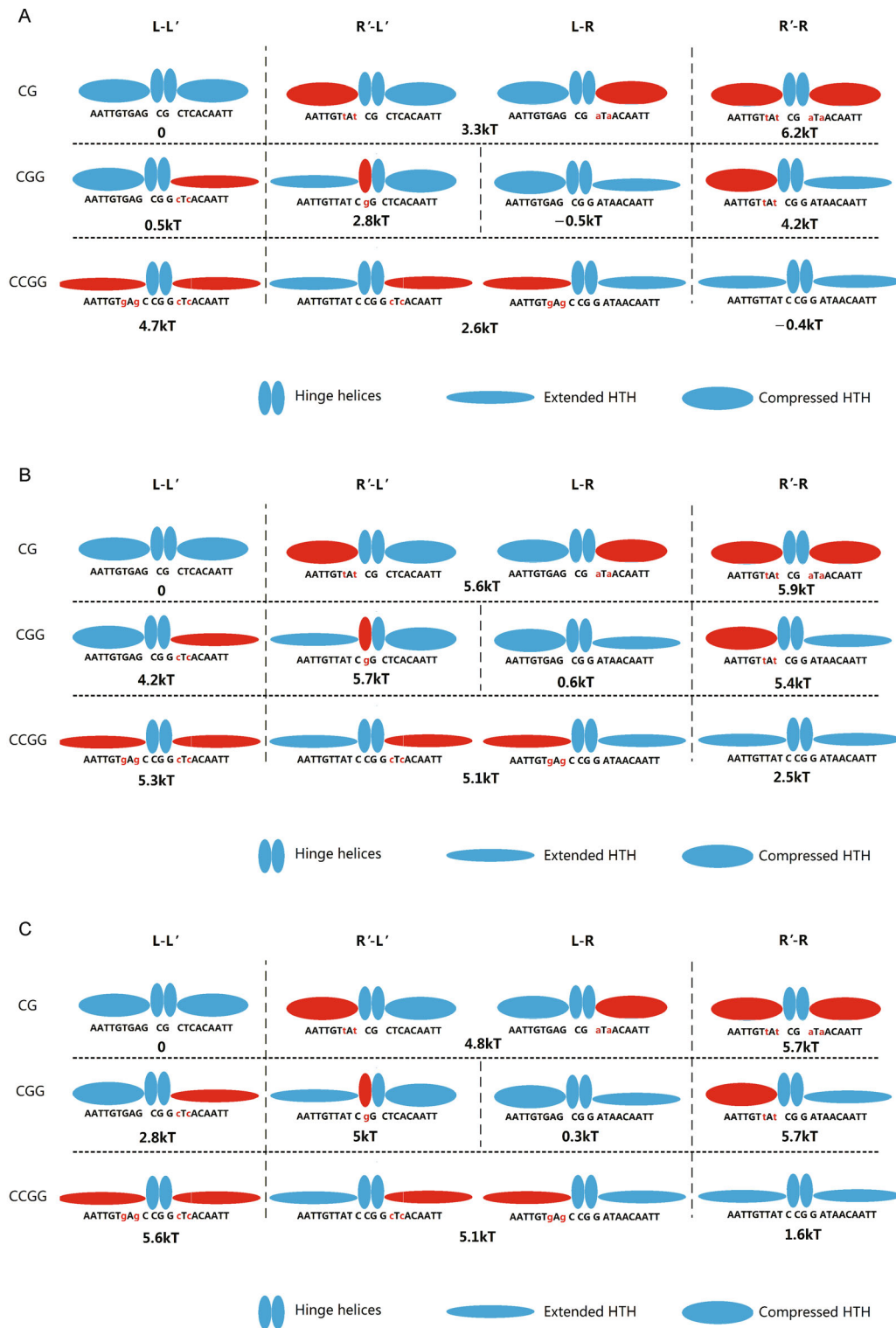
one, then we can infer its “structural flexibility” was disrupted and thus the underlying mutated residue must be critical for the multi-modal binding property. No visibly shifted DNA bands were obtained for m1, m2, and m6, probably because double mutants and whole-loop swap in the lacI linker region significantly disrupted lac repressor's normal binding conformation, whereas for other lacI mutants we successfully separated bound and unbound fragments for sequencing.

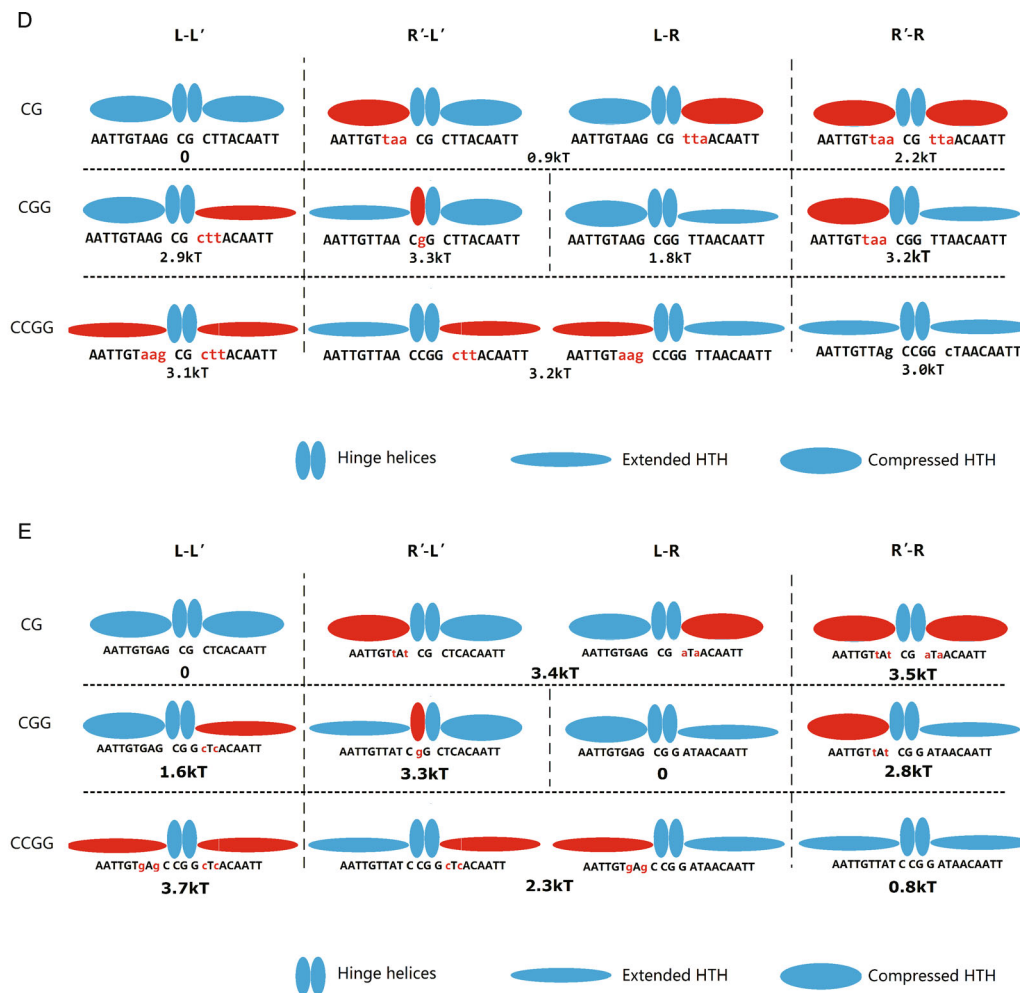
Figure 5 shows the schematic models depicting different binding states for five lacI mutants (m3, m4, m5, m7, m8). Compared to the wild-type case, a few things are clearly noticeable. Firstly for the m3 protein, the L3R variant becomes the best binding site, even better than the L2L' by 0.5 kT. We suspect that N46H mutation decreases the overall affinity under CG spacer, rather than increase the CGG case because the R'4R site also binds better than to the CG spacer. Secondly, for mutants m4 and m5 the R'4R conformation have significantly increased binding energy by more than 1 kT compared to wild-type one (2.5 kT and 1.6 kT respectively), strongly suggesting that the wild-type linker is unique in its ability to support extended conformation under 4 bp CCGG spacer. As for mutant

m7, initially it was expected once we switch the recognition residues YQ to TT, we would not only be able to recognize the *purR* central motif AAG-CG-CTT under 2 bp CG spacer, but also evolve some new motif under extended conformations. But surprisingly the best L3R site (AAG-CGG-TTA) is 1.8 kT worse than its L2L' counterpart (AAG-CG-CTT), and therefore we conclude TT residues cannot facilitate lac repressor's structural flexibility like the YQ does. More interestingly, for m8 mutant we observed slightly increased relative binding energy of L3R compared to wild-type lac repressor (0 kT vs. 0.5 kT). Even though this is not very significant, it is still plausible to speculate that arginine residue can compensate the energy cost better in extended conformation than glutamine in wild-type lac repressor. Table 1 listed all those important variants' (L2L', L3R, and R'4R) relative energy levels for each mutant.

### Hybrid PurR can bind its operator sites in multiple spacer formats similarly

Our previous work[4] has shown that wild-type PurR does not have the binding flexibility of lacI, so here we



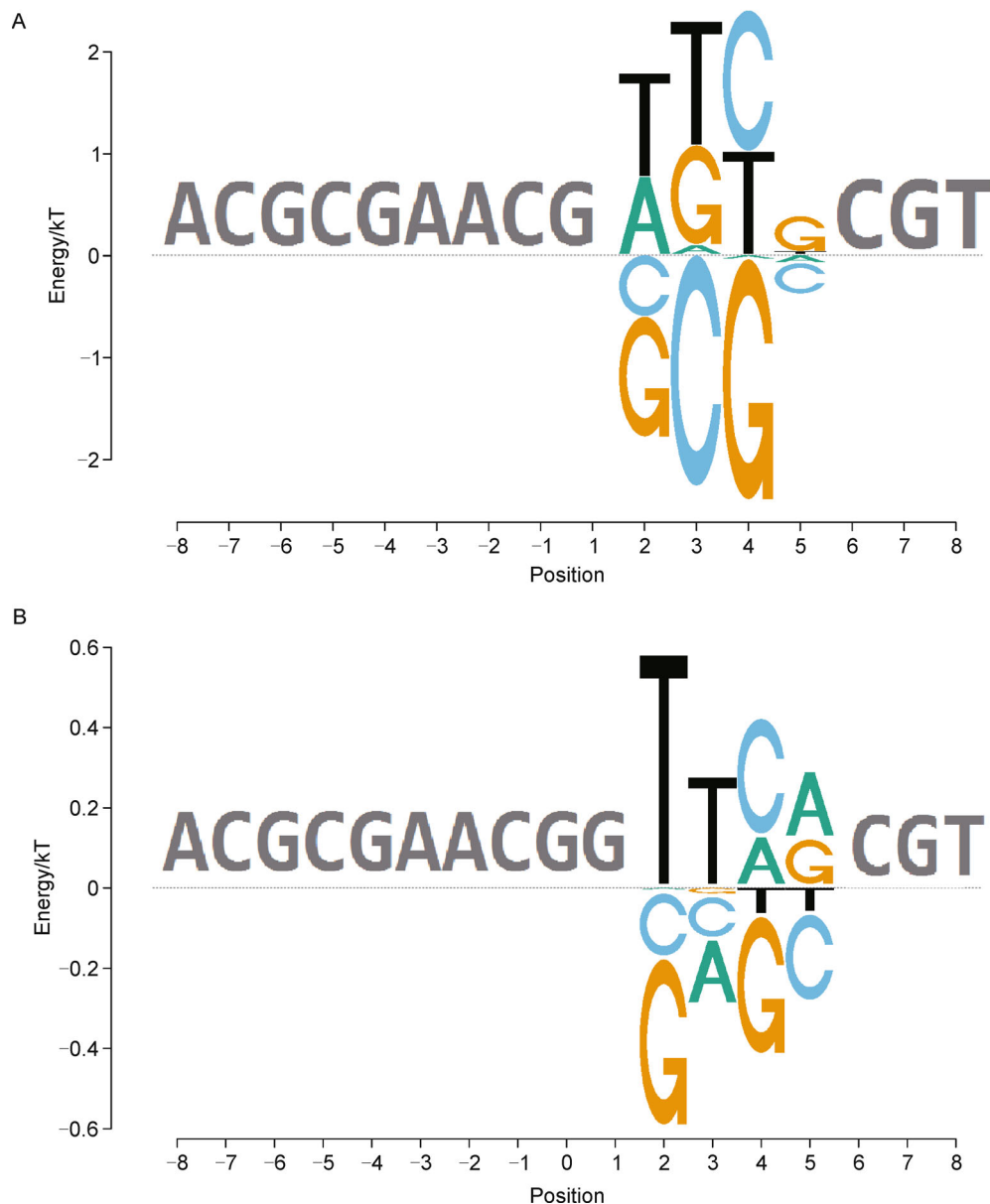


**Figure 5. Energy models for lacI mutants m3, m4, m5, m7, and m8.** (A) Energy models for lacI mutants m3. (B) Energy models for lacI mutants m4. (C) Energy models for lacI mutants m5. (D) Energy models for lacI mutants m7. (E) Energy models for lacI mutants m8.

constructed four PurR hybrid mutants replacing its original residues by ones from lacI, and tested their specificity profiles under 2 bp and 3 bp spacer formats (libraries PR2 and PR3). For wild-type PurR, the preferred binding site under 3 bp CGG spacer CAAA-CGG-TTGC is at least 1 kT worse than its L2L' counterpart and can be considered as mismatched variants under 2 bp spacer instead of a novel, extended motif. We observed similar phenomena in p1 and p2 hybrids, as shown in Table 2. They have the same preferred binding site in CGG spacer as wild-type, which are significantly worse than the L2L' sites (0.8 kT and 1.9 kT).

But interestingly for mutant p4, when we swapped its original recognition di-residues and hinge helix loop both

by lac repressor's, the optimal binding site in CGG spacer performed significantly better (0.4 kT worse than its CG spacer counterpart, Table 2). Figure 6A and 6B are the energy logos produced by regression of all the single and double variants compared to the optimal binding sites under CG and CGG spacers respectively. Noticeably, the core TC motif at position 3–4 is preserved, though there are some quantitative differences. This result strongly suggests that p4 mutant acquired the capability to bind its DNA with multiple spacers as lac repressor does, though there was no new alternative motif for the extended conformation. Table 2 summarized the property and binding energy of important variant sites for each PurR construct.



**Figure 6. Hybrid PurR mutant p4's specificity profiles.** (A) Energy logo of PurR p4 mutant under CG spacer. (B) Energy logo for PurR p4 mutant under CGG spacer. In normal and extended spacer formats (CG and CGG), p4 mutant of PurR shared similar but not identical sequence preferences (TTCG and TTCA in 2 to 5 region), which suggests a common mode of recognition and DNA binding flexibility.

## DISCUSSION

In the early days of study on lac repressor, Riggs et al [20] began the measurement of the binding energy or dissociation constant under different ionic strength, pH, and temperature. It was found as ionic strength increases, e.g., from 0.01 M to 0.1 M KCL, the absolute affinity for lac repressor to its *O*I operator can decrease up to two

orders of magnitude. Our result here (Figure 2A and Figures S1–2) indicated that for lac repressor-operator interaction, the specificity is mostly mediated by the hydrogen bond formation between the bases and some critical residues in the recognition helices of lac repressor, which is insensitive to the salts concentration, even though ionic strength can significantly modulate the protein-DNA backbone electrostatic interactions, which are primarily non-specific[21].

To our knowledge it is the first time to get the “additivity violation” profile across the whole operator region with reasonably good accuracy. Noticeably, most of positions are either neutral, or take compensatory forms. Only for positions (−3, −2), it shows strong anti-compensatory additivity profile. The exact biophysical origin is unclear yet, but that for the left-half site deviations from the wild-type G at position −2 are tolerated alone with only minimal cost, but any further deviation causes a large energy increase, essentially up to non-specific binding to that half-site.

For lac repressor, three primary factors determine its successful positioning onto lac operator, i.e., its dimer repressor specificity profile, low copy number inside cell, and DNA looping. Thus one would ask how other gene regulatory systems facilitate their correct positioning. Among other bacterial gene regulatory systems, DNA looping is a common but not universal phenomenon[22]. Possibly for some global regulators like PurR, it binds to hundreds of sites across genome and does not necessarily require stringent regulation for each individual site, so looping is not necessarily required [23]. Alternatively, for those local operon regulators including the YcjW studied by us, it still could achieve correct operator positioning by compensating with higher TF copy number without looping.

So far all identified and predicted lac repressor binding sites are asymmetric with 3 bp spacer, within *E. coli* and among all other bacterial species. Our current work proved that its multi-modal binding critically depends on both the hinge helix loop and the YQ recognition residues. Intuitively we could think the lac repressor hinge helix linker is exceptionally flexible allowing HTH extension beyond normal format, and the YQ di-residues can stabilize this extended conformation. The evolutionary origin of this property is still elusive, i.e., we don't know which part evolved first, the operator site or the TF itself. It is likely that the ancestral form of lac repressor acquired this property coincidentally first by random mutation, and because of some selective advantage like better induction capability or minimized crosstalk by other TFs, its operator sites switched from the conventional 2 bp spacer format to current 3 bp ones universally. For other LacI/PurR family TFs, since there is no systematic profiling of their specificity profiles experimentally until now, we cannot exclude the possibility that there is some other TF having similar property like lac repressor.

## MATERIALS AND METHODS

### Construction for the *lacI* and *purR* mutants

DHFR control plasmid (provided with NEB PURExpress

kit) was chosen as the original backbone vector since it carries T7 promoter/terminator sequence for protein expression. We replaced DHFR plasmid's original coding fragments by wild-type *lacI* and *purR* genes using Clontech InFusion system first. To get each individual mutant variant clone, two reverse oriented PCR primers carrying the desired codon change (*lacI*-m\*-forward/reversed or *purR*-p\*-forward/reversed) were used to amplify and linearize the original wild-type clone vector. At last, InFusion cloning can also be used to recircularize the linearized plasmid fragments and produce mutant clones (Agilent XL10 competent cells used). All constructed vectors were verified by Sanger sequencing.

### Spec-seq experiments

All basic design and implementation were same as described in our previous work[4]. All protein in this work were made by *in vitro* transcription/translation by NEB PURExpress system. FAM-labelled dsDNA oligonucleotides were separated by electrophoresis mobility assay (EMSA) on 7% acrylamide TBE gels running at 200 V (30 mins, 0°C). After gel fragment was cut, extracted, and purified, each dsDNA library was amplified and barcoded simultaneously by specific set of Illumina Index primers PE1.0-genetics and indexed PE2.0 (listed in Supplementary Materials).

For Spec-seq runs quantifying the whole *lac* operator region, 3 different binding buffer conditions were used, i.e., 1×, 2×, and 4× NEB buffer 4 (50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 1 mM DTT, pH 7.9 at 25°C) with everything else being the same (100 ng dsDNA fragments, 400 ng lac repressor protein in 15 uL reaction systems).

For Spec-seq runs quantifying various *lacI* and *purR* mutants, 1× NEB buffer 4 was chosen as the default binding buffer. 100 ng FAM-labelled dsDNA was added into each 15 uL binding reaction system. Protein of interest was titrated by 2-fold increase per lane (from left to right) starting with 50 ng per 15 uL reaction. All relevant gel figures can be found in Supplementary Materials.

## SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-015-0044-z.

## ACKNOWLEDGEMENTS

We thank members of the Stormo Lab for useful comments on the work. We thank the anonymous reviewers in our previous paper[4] who provided helpful insights on lacI's structural property. Funding for this work was from NIH grants HG000249 and HG005970.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Zheng Zuo, Yiming Chang, and Gary D. Stormo declare they have no conflict of interest, conform with the Helsinki Declaration of 1975, as revised in 2000 (5) concerning Human and animal Rights, and follow out policy concerning Informed Consent as shown on Springer.com.

## REFERENCES

- Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, 3, 318–356
- Lewis, M. (2005) The lac repressor. *C. R. Biol.*, 328, 521–548
- Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the *lac* operator. *Proc. Natl. Acad. Sci. USA*, 70, 3581–3584
- Zuo, Z. and Stormo, G. D. (2014) High-resolution specificity from DNA sequencing highlights alternative modes of lac repressor binding. *Genetics*, 198, 1329–1343
- Mossing, M. C. and Record, M. T. Jr. (1985) Thermodynamic origins of specificity in the lac repressor-operator interaction: Adaptability in the recognition of mutant operator sites. *J. Mol. Biol.*, 186, 295–305
- Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M. and Record, M. T. Jr. (1997) Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.*, 267, 1186–1206
- Hart, D. J., Speight, R. E., Cooper, M. A., Sutherland, J. D. and Blackburn, J. M. (1999) The salt dependence of DNA recognition by NF- $\kappa$ B p50: a detailed kinetic analysis of the effects on affinity and specificity. *Nucleic Acids Res.*, 27, 1063–1069
- Benos, P. V., Bulyk, M. L. and Stormo, G. D. (2002) Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Res.*, 30, 4442–4451
- Stormo, G. D. (2013) Modeling the specificity of protein-DNA interactions. *Quant. Biol.*, 1, 115–130
- Maerkl, S. J. and Quake, S. R. (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science*, 315, 233–237
- Novichkov, P. S., Laikova, O. N., Novichkova, E. S., Gelfand, M. S., Arkin, A. P., Dubchak, I. and Rodionov, D. A. (2010) RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res.*, 38, D111–D118
- Daber, R. and Lewis, M. (2009) Towards evolving a better repressor. *Protein Eng. Des. Sel.*, 22, 673–683
- Record, M. T. Jr, deHaseth, P. L. and Lohman, T. M. (1977) Interpretation of monovalent and divalent cation effects on the lac repressor-operator interaction. *Biochemistry*, 16, 4791–4796
- von Hippel, P. H. (2014) Increased subtlety of transcription factor binding increases complexity of genome regulation. *Proc. Natl. Acad. Sci. USA*, 111, 17344–17345
- Schumacher, M. A., Choi, K. Y., Zalkin, H. and Brennan, R. G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science*, 266, 763–770
- Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, 271, 1247–1254
- Romanuka, J., Folkers, G. E., Biris, N., Tishchenko, E., Wienk, H., Bonvin, A. M., Kaptein, R. and Boelens, R. (2009) Specificity and affinity of Lac repressor for the auxiliary operators *O*<sub>2</sub> and *O*<sub>3</sub> are explained by the structures of their protein-DNA complexes. *J. Mol. Biol.*, 390, 478–489
- Milk, L., Daber, R. and Lewis, M. (2010) Functional rules for lac repressor-operator associations and implications for protein-DNA interactions. *Protein Sci.*, 19, 1162–1172
- Kalodimos, C. G., Boelens, R. and Kaptein, R. (2004) Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. *Chem. Rev.*, 104, 3567–3586
- Riggs, A. D., Suzuki, H. and Bourgeois, S. (1970) Lac repressor-operator interaction: I. Equilibrium studies. *J. Mol. Biol.*, 48, 67–83
- von Hippel, P. H. (2004) Completing the view of transcriptional regulation. *Science*, 305, 350–352
- Coumac, A. and Plumbridge, J. (2013) DNA looping in prokaryotes: experimental and theoretical approaches. *J. Bacteriol.*, 195, 1109–1119
- Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñoz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., García-Sotelo, J. S., López-Fuentes, A., et al. (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Sensor Units). *Nucleic Acids Res.*, 39, D98–D105