ALLOSTERY AND APPLICATIONS OF THE LAC REPRESSOR

Matthew Almond Sochor

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

_____

Mitchell Lewis, D.Phil, John Morgan Professor of Biomedical Research and Education

Graduate Group Chairperson

_____

Kathryn M. Ferguson, Ph.D., Associate Professor of Physiology

Dissertation Committee:
Kim A. Sharp, Ph.D., Associate Professor of Biochemistry and Biophysics
Gregory D. Van Duyne, Ph.D., Professor of Biochemistry and Biophysics
Kristen W. Lynch, Ph.D., Associate Professor of Biochemistry and Biophysics
Mark Goulian, Ph.D., Professor of Biology
James Shorter, Ph.D., Associate Professor of Biochemistry and Biophysics
Robert Fairman, Ph.D., Professor of Biology, Haverford College

ALLOSTERY AND APPLICATIONS OF THE LAC REPRESSOR

Dedication

To my grandmother Ruth Sochor.

She raised three successful kids on her own and made the opportunities I've had possible.

Upon being asked how it feels to be a great grandmother, she replied,

"I've always been great. Now you just have to admit it."

Acknowledgements

I'd like to thank my parents for putting up with me throughout the years and generally being wonderful and supportive people. Also, thanks to Bill for being a good older brother and keeping me in my place.

I've met too many characters in the greater Philadelphia area during my time here that have impacted my life in many ways, both positive and negative. Thanks for challenging me to learn new things, from playing the accordion to protein purification, and generally making my 20s a good time. Special thanks to the MRI guys. The drinks, quizzo, and political arguments kept me sane.

A huge thank you to the the various scientists I have worked with over the years. Bob Daber and Leslie Milk both were excellent lab mates and helped get me started. My classmates have provided endless discussion and support over the years, special thanks to Ed Ballister, Josh Allen, Devin Dersh, and Morgan DeSantis. Finally, my collaborators in the Bennett lab made Chapter 5 possible. Special thanks to Adam Wojno, Theodore Drivas, Latha Vasireddy, Jeannette Bennicelli, Daniel Chung and of course Jean Bennett.

Thank you to my beautiful wife Elizabeth, you know I couldn't have even come close to finishing this without both your support and your brain. Finally, I'd like to acknowledge my son Iggy. You have given me perspective on what is important in life.

ABSTRACT

ALLOSTERY AND APPLICATIONS OF THE LAC REPRESSOR

Matthew Almond Sochor

Mitchell Lewis

The lac repressor has been extensively studied for nearly half a century; this long and complicated experimental history leaves many subtle connections unexplored. This thesis sought to forge those connections from isolated and purified components up to functioning lac genetic switches in cells and even organisms. We first connected the genetics and structure of the lac repressor to *in vivo* gene regulation in *Escherichia coli*. We found that point mutations of amino acids that structurally make specific contacts with DNA can alter repressor-operator DNA affinity and even the conformational equilibrium of the repressor. We then found that point mutations of amino acids that structurally make specific contacts with effector molecules can alter repressor-effector affinity and the conformational equilibrium. All results are well explained by a Monod, Wyman, and Changeux model of allostery. We next connected purified *in vitro* components with *in vivo* gene regulation in *E. coli*. We used an *in vitro* transcription assay to measure repressor-operator DNA binding affinity, repressor-effector binding affinity, and conformational equilibrium. Only the repressor-operator DNA binding affinity disagreed with literature values from other *in vitro* experiments, however it did agree with a published value which should hold under *in vivo* conditions. We were able to use our *in vitro* thermodynamic parameters to accurately predict the *in vivo* gene regulation when cell crowding was considered. Finally we developed an autogenously regulated lac repressor for AAV-mediated gene therapy. We were able to improve the gene regulation of the autogenous switch by using multiple operator DNA sites, a tetrameric lac

repressor, and point mutations to the lac repressor. The autogenous switch was shown to function in various cell types and was capable of reversible regulation of luciferase in living mice.

Table of Contents

List of Tables

List of Illustrations

Chapter 1:
The Lac Repressor and Gene Regulation

1.1 Introduction

All life exists in a universe of limited resources. Cells contain a multitude of potential genes that they can express to achieve various aims and there is a significant advantage to the cell to be able to choose which genes it is currently expressing in order to best tailor itself to its environment. A red blood cell expresses hemoglobin in order to maximize oxygen carrying capacity while a neuron does not. Similarly, sodium channels are prevalent in the neuron to transmit electric pulses while they are utterly absent in the red blood cell. Both carry the same genetic material, however they use a large complement of gene regulation techniques to adapt to their respective environments.

The gene regulation of *Escherichia coli* (*E. coli*) is no different. Bacteria must respond to changing environments: temperature changes, the availability of various carbon sources, pH, the presence of other bacteria, etc. Bacteria that most efficiently adapt to their environment will thrive. Those that express useless genes will waste precious resources and ultimately fail to compete. It is no surprise then that a large family of proteins has been identified in *E. coli* and other bacteria that specifically regulate a huge complement of genes related to sugar metabolism. The LacI/GalR family of proteins (Weickert & Adhya, 1992) includes more than 1000 members based upon a 2008 Swiss-Prot BLAST search (Swint-Kruse & Matthews, 2009) and regulates genes responsible for catabolism, toxicity, and nucleotide synthesis. In general, these proteins all have a helix-turn-helix motif that binds one or more specific DNA operator sequences and the affinity of this binding is allosterically modified by the binding of small

effector species (Swint-Kruse & Matthews, 2009). Typically, the effector is related to the genes in question. For example, the lac repressor controls the synthesis of three genes (*lacZ*, *lacY* and *lacA* encoding for β-galactosidase, β-galactoside permease and thiogalactoside transacetylase, respectively) that encode for proteins related to lactose metabolism (Lewis, 2005). The natural effector of the lac repressor is allolactose, a byproduct of lactose catabolism by β-galactosidase (the product of the *lacZ* gene). Allolactose binds the lac repressor and results in an allosteric conformational shift and a loss of affinity for operator DNA. RNA polymerase is then free to transcribe the *lacZ/Y/A* poly-cistronic message and lactose metabolism begins in earnest. This theme of the effector (allolactose), an intermediate of the process to be regulated (lactose metabolism), being sensed by a protein (lac repressor) that is used to change the expression of genes that are metabolically functional (*lacZ/Y/A*) is repeated throughout the LacI/GalR family (Swint-Kruse & Matthews, 2009).

This protein family is particularly fascinating because it combines in one package several essential features of protein biology: DNA binding, small molecule binding, and allostery. Furthermore, the net effect, regulation of a gene or set of genes, is completely general: the lac repressor (and many other repressors and activators in the LacI/GalR family) can be used to regulate any gene that is under control of a promoter with a lac specific DNA operator sequence. This realization has led to a multitude of generalized plasmids that constitutively express a repressor and have a second regulated promoter followed by a multiple cloning site that can be used to insert the researcher's gene of interest (e.g. the pET series for the lac repressor, the pBAD series for the arabinose repressor). The system is so simple that it is not only applicable to the bacteria where it initially was identified. The lac repressor has been used to control

genes in embryonic stem cells (Caron et al., 2005) and even to control gene expression in living mice (Cronin, Gluba, & Scrable, 2001).

So if the lac repressor is such a triumph of translational science, why continue to study it? What is left that is unknown? First and foremost, the lac repressor has essentially been applied as-is. The most commonly used protein is exactly that which was initially found in *E. coli* despite the long held knowledge that the repressor specifically evolved to optimally regulate lactose metabolism and not the multitude of scenarios it is actually utilized in, such as toxic gene repression/over-expression or eukaryotic gene regulation. It is well established that the gene regulation phenotype is highly modifiable through point mutation (Kleina & Miller, 1990) leading to repressors that have less leakiness (useful for toxic gene control), that turn the genes more "on", and that respond to different effectors and different concentrations. An understanding of the physical basis for these phenotypic changes could lead to easier translation of specific lac repressor mutants for specific applications.

The goal of this thesis was to take the current knowledge regarding the lac repressor and to achieve two goals:

(1) to understand everything within a single theoretical framework or model, and

(2) to apply this knowledge to create genetic switches more ideal for their particular applications.

1.2 Structure and Function in the Lac Operon

The history of the study of the lac operon mirrors the history of modern biochemistry and biophysics. The earliest studies were related to the observation that

cells could alter their ability to metabolize various compounds; one such compound was the metabolism of lactose by *E. coli*. The "adaption" of cells from glucose to lactose metabolism resulted in a diauxic growth curve; cells follow the typical exponential growth as they metabolize glucose followed by a lag phase as glucose becomes depleted, then a secondary exponential growth is observed as the cells switch to lactose metabolism (Monod, 1942). The metabolism of lactose was quickly linked to the enzyme β-galactosidase which was observed to become more prevalent when bacteria were growing on lactose versus glucose (Cohn, 1957). Studies measuring the accumulation of radio-labeled galactosides led to the discovery that lactose is actively pumped into the cell via β-galactoside permease which is also more prevalent during lactose growth (Buttin, Cohen, Monod, & Rickenberg, 1956). The final regulated gene of the lac operon was identified through the discovery of thiogalactoside transacetylase (Zabin, Kepes, & Monod, 1962). What remained however was *how* the expression of these three genes were tied to lactose metabolism. Before any actual repressor proteins were identified, a groundbreaking analysis was performed to understand gene regulation. A negative control scheme was described where the product of a gene, acting through the cytoplasm, was able to regulate the synthesis of another set of genes modulated by the presence of (usually metabolic) small molecules (Jacob & Monod, 1961). The cytoplasmic agent however, was unknown. The modulation by small molecules was of particular interest. It was known that other proteins exhibit a property wherein the binding of a small molecule causes a change in functionality at a structurally (-steric) separated (allo-) region (allostery). A "plausible" model of allostery was proposed which would, in time, be shown to be applicable to the lac repressor (Monod, Wyman, & Changeux, 1965). The lac repressor was finally isolated one year later and determined to be the "cytoplasmic agent" responsible for regulation of gene expression predicted by Jacob

and Monod (Gilbert & Müller-Hill, 1966). Shortly thereafter it was determined that the lac repressor reversibly binds to a specific DNA region (deemed the operator) to regulate gene synthesis (Gilbert & Müller-Hill, 1967). Allolactose, a byproduct of lactose metabolism by β-galactosidase, was found to be the effector in the lac operon (Jobe & Bourgeois, 1972, 1973). It was not known however that allolactose worked through an allosteric mechanism (the common prefix allo- means “other” or “different” and is simply coincidental). DNA sequencing techniques were developed which led to the sequencing of both the lac repressor (Farabaugh, 1978) and its operator DNA (O1) sequence (Gilbert & Maxam, 1973). The second and third DNA operators (O2 and O3, respectively) of the lac operon were discovered and found to also play a role in lactose metabolism (Pfahl, Gulde, & Bourgeois, 1979). It was then shown that lac repressor does not prevent binding of RNA polymerase to the promoter, but it does prevent the initiation of transcription (Schmitz & Galas, 1979). At this point it was established that the lac operon consisted of the lac repressor (constitutively expressed by the *lacI* gene) binding to the DNA operators preventing transcription of the downstream poly-cistronic message (*lacZ/Y/A* genes) in a reversible manner dependent upon the natural inducer molecule allolactose or gratuitous inducers such as isopropyl β-D-thiogalactoside (IPTG).

The focus at this point changed from identifying the players to identifying their roles. Central to this focus was the slow problem of determining the three-dimensional structure of the lac repressor alone and bound to its ligands (DNA and effectors). The journey from the identification of the lac repressor to the first crystal structure 30 years later includes a vast array of techniques to probe individual parts of the protein to determine their individual roles in the overall functioning of the repressor.

The first issue was simply understanding what part of the amino acid sequence of the lac repressor was responsible for what property or properties: folding, dimerization, tetramerization, DNA binding, effector binding, and allosteric signaling. Trypsin digestion provided the first clue as to the domain structure of the lac repressor. Tryptic digestion cleaved the 59 amino-terminal (N-terminal) residues and 20 carboxyl-terminal (C-terminal) residues preserving a highly trypsin-resistant core (Files & Weber, 1976). These three regions of the protein proved highly relevant to the structure and function of the lac repressor.

The N-terminal region of the lac repressor was isolated and was determined to be necessary for DNA binding (Geisler & Weber, 1977) and found to specifically bind the known O1 lac operator (Ogata & Gilbert, 1978). Modification of tyrosines with tetranitromethane showed that Y7 and Y17 modifications significantly impaired both specific and non-specific DNA binding, further indicating the role of the N-terminal region of the repressor in DNA binding (Hsieh & Matthews, 1981). Genetic mapping was undertaken on tight DNA binding mutants of the lac repressor and they were found to correlate to two regions: the N-terminus and near amino acids 255 and 295 (Pfahl, 1981). A small angle X-ray scattering (SAXS) study indicated that the DNA binding sites of the tetrameric lac repressor consists of the N-terminal 59 residues and that they lie at opposite ends of an elongated tetramer (Mckay, Pickover, & Steitz, 1982). A genetic study showed that of 131 amino acid substitutions in 38 of the first 62 N-terminal positions, almost all abolished DNA binding and only two, P3Y and S61L, showed tighter DNA binding (Miller, 1984). A nuclear magnetic resonance (NMR) structure of the N-terminal 51 residues showed the now classic helix-turn-helix motif and its structural homology with two other DNA binding proteins *c*I and λ (Kaptein, Zuiderweg, Scheek,

Boelens, & van Gunsteren, 1985) The helix-turn-helix motif has a recognition helix that fits into the major groove of DNA and makes specific contacts with nucleotides and is a major determinant of specificity in DNA binding. The specificity determining residues of the recognition helix were determined to be tyrosine 17 (Y17), glutamine 18 (Q18), and arginine 21 (R21) and they were thought to make specific contacts with positions 4, 5 and 6 of the DNA operator. Mutagenesis of Y17 and Q18 combined with changes at positions 4, 5 and 6 showed that not only could DNA affinity be destroyed, but specificity could be changed entirely (Sartorius, Lehming, Kisters, von Wilcken-Bergmann, & Müller-Hill, 1989). This work would later be extended to more fully screen these positions in both the operator and repressor in order to determine functional rules for specific DNA binding (Milk, Daber, & Lewis, 2010). While two-dimensional NMR demonstrated specific contacts between headpiece residues and DNA (Kaptein, Lamerichs, Boelens, & Rullmann, 1990; Slijper, Bonvin, Boelens, & Kaptein, 1996), eventually an X-ray crystallographic structure of the lac repressor-operator DNA complex was solved that demonstrated exactly how the lac repressor binds DNA (Lewis et al., 1996; Pace, Lu, & Lewis, 1990).

The next piece of the lac repressor was the trypsin-resistant core. Early on it was found that a Y269S mutation and missense mutations at positions 210 and 216 failed to form tetramers (Schmitz, Schmeissner, Miller, & Lu, 1976). A genetic study then linked approximately 2000 missense mutations to varying phenotypes and found most of the trypsin-resistant core to be relatively tolerant to mutation although many interesting phenotypes were still found within this region (Miller & Schmeissner, 1979). Chemical modification of cysteine 281 decreased repressor affinity for inducer, indicating its potential role in sugar binding (Daly, Olson, & Matthews, 1986) and this role was later

confirmed through mutagenesis of C281 (Chakerian & Matthews, 1991). Arginine 197 was also found to be important for inducer binding (Spotts, Chakerian, & Matthews, 1991). Finally, a genetic study, wherein more than 4000 mutations were made along 328 positions of the lac repressor, identified the amino acids spread throughout the linear sequence to be responsible for inducer binding (Markiewicz, Kleina, Cruz, Ehret, & Miller, 1994). The crystal structure revealed an effector binding pocket that correlated well with the genetic data (Daber, Stayrook, Rosenberg, & Lewis, 2007; Lewis et al., 1996). Interestingly, there was a parallel evolution between the core region of the LacI/GalR family and periplasmic sugar-binding proteins of *E. coli* (Fukami-Kobayashi, 2003). The core region was also responsible for dimerization. This dimerization could be disrupted through mutation (Dong et al., 1999; Spott, Dong, Kisters-Woike, & Müller-Hill, 2000) and exhibited plasticity, such that 22 point mutations in the dimer interface could still form homo-dimers (Swint-Kruse, Elam, Lin, Wycuff, & Matthews, 2001). This plasticity was elegantly demonstrated in a study that made a known mutant that disrupted dimerization and looked for compensating mutations in the other monomer, resulting in a hetero-dimeric lac repressor (Daber & Lewis, 2009). There was also a structural realignment upon sugar binding within the core region (Barry & Matthews, 1997). The C-terminal lobe of the core region remains relatively unchanged with and without sugar bound. The more N-terminal lobe, which connects to the N-terminal DNA binding domain, exhibits a significant structural shift when sugar is bound. The dimer interface of the N-terminal lobe is of particular interest. Lysine 84 was found to lie along the dimer interface (Chang, Olson, & Matthews, 1993) and mutating this residue to leucine (K84L) was found to stabilize the dimer (Barry & Matthews, 1999; Bell, Barry, Matthews, & Lewis, 2001; Nichols & Matthews, 1997). Curiously, this stabilizing mutant decreases both DNA and inducer affinity despite not directly being responsible for either

DNA or inducer binding, which points at an allosteric role. Indeed other residues in this region were implicated to have allosteric importance, including alanine 110. The mutation A110T lowered operator affinity and increased IPTG affinity, whereas A110K increased operator affinity while decreasing IPTG affinity (Müller-Hartmann & Müller-Hill, 1996). It was later found that positions 94-98, which are known to make re-arrangements upon sugar binding, also play a significant role in allosteric signaling (Zhan, Camargo, & Matthews, 2010). The transmission of the allosteric signal was beautifully mapped out through a targeted molecular dynamics simulation (Flynn et al., 2003) and confirmed through mutations (Swint-kruse, Zhan, Fairbanks, Maheshwari, & Matthews, 2003). They found L148F increased IPTG affinity and decreased operator affinity (analogous to A110T), whereas S151P had the opposite effect (analogous to A110K). Clearly there is plasticity not only in dimerization but also allostery. It was well elucidated then that the N-terminal domain binds DNA and that that binding is allosterically modified by inducer binding in the cleft of the core region. The allostery could be altered not only in residues within the core but also in the hinge region of the N-terminal DNA binding domain (Falcon & Matthews, 1999, 2001; Falcon, Swint-Kruse, & Matthews, 1997).

The C-terminal segment is also extremely important to the overall function of the lac repressor. Homology studies indicated that the the C-terminal region may have a leucine zipper motif. Leucine to alanine mutations at positions 342, 349 and 356 resulted in dimeric repressors which could bind inducer and operator DNA with reduced affinity (Chakerian et al., 1991). It was then shown that removing 5, 11, 18 and 32 amino acids from the C-terminus also created a dimeric repressor capable of inducer binding and with reduced DNA affinity (Chen & Matthews, 1992). Next, it was postulated that the

necessary heptad repeat of three leucines formed an abbreviated coiled-coil domain resulting in tetramerization. A point mutation was known that resulted in a monomeric lac repressor species (Y282D). The coiled-coil was tested by extending the C-terminus with one and two additional heptads, combined with the Y282D mutation, resulting in a dimeric lac repressor, likely through an extended coiled-coil at the C-terminus (Chen, Surendran, Lee, & Matthews, 1994). The C-terminus was definitively identified as a coiled-coil responsible for tetramerization and many properties of the coiled-coil were soon determined (Fairman et al., 1995; Lewis et al., 1996). Tetramerization of the repressor was of course central to the ability of the lac repressor to bind two DNA operators. As noted previously, impairment of tetramerization typically decreased operator DNA affinity. Obviously, this decrease is not a direct effect, as the N- and C-terminii of the lac repressor are spatially separated; this points to an allosteric mechanism wherein binding of one DNA operator to a repressor dimer within the tetramer increases the affinity of the operator DNA for the second dimer. It was known that the lac repressor bends DNA upon binding (Culard & Maurizot, 1981; Kim, Zweib, Wu, & Adhya, 1989) and it was soon discovered that when the lac repressor binds two DNA operators it creates a looped DNA structure (Kramer et al., 1987). Cooperativity of binding was observed *in vivo* between the O1 and O3 operators (Sasse-dwight & Gralla, 1988), the O1 and O2 operators (Flashner & Gralla, 1988), and between all three operators (Oehler, Eismann, Kramer, & Müller-Hill, 1990). A beautiful study was performed to understand how the spacing of the operators affected the cooperative binding and found a phase dependence to the spacing where repression, and hence operator binding, is maximal (Oehler, Amouyal, Kolkhof, von Wilcken-Bergmann, & Müller-Hill, 1994) and the optimal spacing was determined (Müller, Oehler, & Müller-Hill, 1996).

Overall a picture emerged that showed how a dimer of dimers looped DNA in order to shut off transcription and how the binding of small sugar molecules in a completely different region could result in the release of DNA to turn transcription back on. The lac repressor exhibits allostery in its cooperative binding of DNA and allostery in its connection between DNA and effector binding. While many individual residues were demonstrated to change or disrupt allostery, they fail to explain how its mechanism functions. Amazingly, a conformation based model of allostery was already postulated in 1965 by Monod, Wyman, and Changeux, the so-called MWC model, that wonderfully describes the half-century worth of research into the lac operon that followed. In order to appreciate and understand the MWC model, we must first understand the physics of proteins and their conformations.

1.3 Protein Structure and the Boltzmann

Proteins are chains of amino acids and they are known to fold into specific conformations or structures. A vast amount of work has gone into both experimentally determining and theoretically predicting protein structures from known amino acid sequence (Englander, Mayne, & Krishna, 2012). The picture that has emerged of the so-called “folding problem” is that an energy landscape exists wherein each point on the landscape is a possible conformation of the protein and each point has a given energy. A given protein will sample this landscape as it folds and tend towards lower and lower energies, akin to water running downhill until it reaches a pool at the bottom. Much like the surface of the earth, there can be many valleys, slight and deep depressions, hills and mountains in the energy landscape and just as water will pool in the low spots, proteins will fold into structures which have local energetic minima. Proteins do not fold into a single conformation but instead they can fold into many conformations, some of

which are more energetically favorable than others. Just how energetically favorable depends on a host of factors: solvent, binding partners, temperature, protein sequence, etc. Therefore the energy landscape isn't static for a given amino acid sequence, it is a culmination of both the protein and its environment. Furthermore, proteins have energy all their own in the form of thermal energy. They may be comfortably sitting in a conformation which is a local minimum and then through the thermal jostling of their environment be pushed out of that well to find a new conformation in a new energy well. The thermal energy essentially defines the scale of features that are considered relevant. At low temperatures, minute energetic differences become relevant and new minima are present. At high temperatures, only the deepest energy minima are relevant; the protein has too much thermal energy to be stable in any of the lesser wells.

This abstraction not only leads to a clear intuitive picture of how proteins find their structural conformations and transition between conformations, it also provides a mathematical framework to quantify the relative occupation of each conformation for a given population of protein. The Boltzmann distribution defines the fraction of proteins that will occupy each conformation based upon the energy and degeneracy of that conformation with respect to every other conformation,

$$\frac{N_i}{N} = \frac{g_i \exp\left(\frac{-E_i}{k_B T}\right)}{Z(T)} \qquad (1)$$

Here N_i/N is the fraction of the protein in the i^{th} conformation. The degeneracy of each state, g_i, defines the redundancy of the conformation. This usually is a result of an internal symmetry of the protein; for example, a dimeric protein where each monomer can individually bind a small molecule. The structure of the dimer bound to a single

molecule is degenerate, it could be bound to either monomer and it would look identical, thus $g_i = 2$ in this case. Typically, the degeneracy equals one (non-degenerate) as most conformations are unique. The energy of the i^{th} conformation is E_i and the thermal energy of the system is given by the product of the Boltzmann constant, k_B, and the temperature of the system, T. The energy and the temperature are always positive values, so the exponential is maximized (and hence the fraction maximized) at lower energies. At higher energies the numerator exponentially approaches zero and that conformation is effectively never populated. The denominator is called the partition function and is defined as,

$$Z(T) = \sum_i g_i \exp\left(\frac{-E_i}{k_B T}\right) \tag{2}$$

The partition function essentially is a sum of the energetics of every conformation. For simple energy landscapes that are fully defined by a mathematical function, the partition function becomes an integral over that function. However, for extremely complex (and largely unknown) energy landscapes like that of a protein folding, the partition function will generally include just those states that significantly contribute: the local minima.

An example is helpful to understand the Boltzmann distribution. We can consider a theoretical protein that folds into two primary conformations which we will call T and R and see how the Boltzmann partitions the protein between the two states. First we consider the case where the two conformations have the same energy ($E_T = 2$, $E_R = 2$, $k_B T = 1$). We see that,

$$\frac{N_T}{N}=\frac{N_R}{N}=\frac{e^{-2}}{e^{-2}+e^{-2}}=\frac{1}{2} \tag{3}$$

Each state is equally likely because they have the same energy. Now if the T state has an energy twice that of the R state ($E_T = 4$, $E_R = 2$, $k_BT = 1$), we would expect most but not all of the protein to reside in the R state. Indeed we find,

$$\frac{N_T}{N}=\frac{e^{-4}}{e^{-4}+e^{-2}}=0.119 \tag{4}$$

$$\frac{N_R}{N}=\frac{e^{-2}}{e^{-4}+e^{-2}}=0.881 \tag{5}$$

So roughly 88% of the protein is found in the lower energy, R, conformation. Clearly, this is an approximation as every protein will sample much more than two conformations, however it is generally observed that many proteins will primarily reside in a small subset of conformations which are significantly lower in energy than the other conformations. Indeed, even a four fold difference in energy results in roughly 99.8% of the protein residing in the lower energy conformation. Therefore we typically try to identify the most prominent low energy conformations of a protein and seek to explain its experimental behavior through these conformations.

The Boltzmann is incredibly useful because it reduces a complex abstraction to a simple mathematical formula that tells us how often each structural conformation is occupied. In particular, this foundation will be of use when we seek to understand and to model protein allostery.

1.4 Monod, Wyman and Changeux (MWC) Model of Allostery

Proteins range from exquisitely simple to remarkably complex. The simplest proteins have a single primary conformation and perform a single role: binding a ligand, catalyzing a reaction, binding another protein, etc. More complex proteins have multiple stable conformations and modulate their role. But how do they achieve this modulation?

We first consider a simple, theoretical protein that catalyzes a reaction in a structurally defined pocket on its surface. The rate of catalysis is a culmination of many factors: buffer, amino acid sequence, the conformation of the protein, etc. However, most often these things are relatively fixed for the cell. The protein has a conformation that is defined by its energy landscape. The amino acid sequence is defined by the DNA sequence. The buffer might change (depending on environmental conditions in and around the cell) but most often it will be within a well defined range. So what happens when it would be advantageous for the cell to have the catalysis sometimes fast and sometimes slow?

The most obvious route is to competitively bind the catalytic pocket of the enzyme. If the catabolic precursor cannot bind, because something else is there gumming up the works, the enzymatic rate is slowed. This however requires evolution to come up with a close mimic to the catabolic precursor which is extremely specific and may not even be possible. Despite this difficulty, many examples of competitively binding occur in nature, such as carbon monoxide and oxygen competitively binding to hemoglobin (which is why carbon monoxide is a poison).

A second route is to alter the amino acid sequence. This can be achieved by chemically altering key side chains of the protein. These post-translational modifications

(PTMs) include, but are not limited to, phosphorylation, acetylation, glycosylation and even cleavage of the protein. One drawback to this approach is that a physical alteration of the protein is occurring. Reversing PTMs is another business in and of itself in the cell (e.g. kinases and phosphatases constantly phosphorylating and removing phosphorylation). Many, many examples of PTMs exist in nature (Beltrao, Bork, Krogan, & van Noort, 2013).

The third route is to alter the conformation of the protein. An altered conformation could have an altered catalytic pocket with a slower or faster rate of catalysis. This can be achieved through a variety of mechanisms including the aforementioned PTMs and the binding of small molecules or other proteins. The nature of this approach is that a specific part of the protein does not need to be targeted, instead any process that stabilizes an alternate conformation with different properties is potentially useful. Allostery is the phenomenon wherein a function which occurs in one structurally defined region of a protein is altered from a secondary, spatially separated region of the protein. Allostery is made possible by altering the energy landscape of a protein and changing the population of alternate conformations. Again, many examples of allostery occur in nature (Edelstein & Le Novère, 2013; Lewis, 2013; Nussinov & Tsai, 2013).

Allostery was first successfully modeled by Monod, Wyman and Changeux (Monod et al., 1965) typically shortened to MWC. The MWC model considers a protein with two or more potential conformations which are in equilibrium. Each conformation has its own set of properties related to the function of the protein (binding constants, catalytic constants, etc.). The bulk behavior of the system at equilibrium will be the sum of behaviors of the protein in its Boltzmann defined combination of states. Addition of an

allosteric binding partner changes the relative population of states and changes the bulk property of the system under study. The MWC model boils down allosteric regulation of a protein to this: to alter a protein's function, do something to move more of the protein into a different conformation where the function of interest is different.

To understand this idea, we need to more fully delve into the individual properties of the MWC model. First, the conformational equilibrium is an intrinsic property of the protein given its environment (buffer, temperature, etc.). Considering a hypothetical protein with two conformations, T and R, the conformational equilibrium, L, is defined as,

$$L = \frac{[T]}{[R]} \tag{6}$$

Here hard brackets denote concentrations of that species. The total concentration of the hypothetical protein, P, is then defined as,

$$[P]_{tot} = [T] + [R] = [R](1 + L) \tag{7}$$

The conformational equilibrium is of course fully defined by the energetic landscape of the protein, and hence the Boltzmann distribution. Assuming that these conformations are non-degenerate, we can see,

$$L = \frac{\frac{[T]}{[P]_{tot}}}{\frac{[R]}{[P]_{tot}}} = \frac{\frac{\exp\left(\frac{-E_T}{k_B T}\right)}{Z(T)}}{\frac{\exp\left(\frac{-E_R}{k_B T}\right)}{Z(T)}} = \exp\left(\frac{-(E_T - E_R)}{k_B T}\right) \tag{8}$$

So the conformational equilibrium is simply a function of the energy difference between the two conformations and is an intrinsic property of the protein.

Lets consider now a hypothetical molecule, X, that binds the protein. It binds the T and R conformations with different affinities,

$$K_T = \frac{[TX]}{[T][X]} \tag{9}$$

$$K_R = \frac{[RX]}{[R][X]} \tag{10}$$

Here [TX] and [RX] are the X bound states of conformations T and R, respectively. Again, the binding affinity constants are intrinsic properties of the protein given the environmental conditions.

How does all of this add up to allosterically regulate the protein? An example is most illustrative. We define the hypothetical energy of the system as the following: E_T = 4, E_R = 2, k_BT = 1. Equations (4) and (5) show that roughly 88% of the protein is in the R conformation and 12% in the T conformation. If the R conformation catalyzes a reaction at a rate of 10 per second and the T conformation at a rate of 1 per second and given 100 proteins in a cell, there is an effective rate of,

$$k_{eff} = 0.88 * 10 \frac{rxn}{sec} + 0.12 * 1 \frac{rxn}{sec} = 8.92 \frac{rxn}{sec} \tag{11}$$

The *effective* rate, which is what is measured experimentally, is a bulk property of the system. MWC would model this through the conformational equilibrium parameter,

$$L = \exp\left(\frac{-(E_T - E_R)}{k_B T}\right) = \exp\left(\frac{-(4-2)}{1}\right) = 0.135 \tag{12}$$

Using the MWC solution we typically use concentrations and rates with respect to concentrations. Here we will assume that the total protein concentration is 1 mM, k_R = 10 rxn/sec/mM and k_T = 1 rxn/sec/mM. We can then calculate the concentration of the protein in the R and T conformations,

$$[R] = \frac{[P]_{tot}}{(1+L)} = \frac{1\,mM}{1+0.135} = 0.881\,mM \tag{13}$$

$$[L] = [P]_{tot} - [R] = 0.119\,mM \tag{14}$$

And the effective rate is then,

$$k_{eff} = [R]k_R + [L]k_L$$
$$= 0.88\,mM * 10\frac{rxn}{sec * mM} + 0.12\,mM * 1\frac{rxn}{sec * mM} = 8.92\frac{rxn}{sec} \tag{15}$$

So the MWC route of course arrives at the same conclusion as the Boltzmann route.

This effective rate can be allosterically modified by binding ligand X. This more complicated MWC model is solved by considering the total sums of each species, P and X, and the thermodynamic constants that link them,

$$[P]_{tot} = [R] + [RX] + [T] + [TX] \tag{16}$$

$$[X]_{tot}=[X]+[RX]+[TX] \tag{17}$$

$$L=\frac{[T]}{[R]} \tag{18}$$

$$K_T=\frac{[TX]}{[T][X]} \tag{19}$$

$$K_R=\frac{[RX]}{[R][X]} \tag{20}$$

The general strategy is as before: calculate how much of each species is around and use that to determine the effective catalytic rate.

If the ligand prefers binding to the lower catalytic state, T, we can effectively turn off the protein by adding ligand X. For example, if K_T = 10 mM^{-1}, K_R = 1 mM^{-1}, and $[X]_{tot}$ = 1 mM, then we can calculate the concentration of each species of the protein P by solving the equations (16) - (20). We do this by solving all of the equations for the free concentration of the R conformation, [R], in terms of just known thermodynamic constants and then using [R] to measure the effective catalytic rate. We start by re-organizing the above equations,

$$[T]=L[R] \tag{21}$$

$$[TX]=K_T[T][X]=K_TL[R][X] \tag{22}$$

$$[RX]=K_R[R][X] \tag{23}$$

Plugging these values into equation (17) we get,

$$[X]_{tot}=[X]+K_R[R][X]+K_TL[R][X]=[X](1+[R](K_R+K_TL)) \tag{24}$$

Re-organizing we get the free concentration of the ligand, [X], in terms of thermodynamic constants and [R],

$$[X]=\frac{[X]_{tot}}{1+[R](K_R+K_TL)} \tag{25}$$

We can then insert equations (22), (23), and (25) into (16) to get,

$$[P]_{tot}=[R]+\frac{K_R[R][X]_{tot}}{1+[R](K_R+K_TL)}+L[R]+\frac{K_TL[R][X]_{tot}}{1+[R](K_R+K_TL)} \tag{26}$$

We then simply need to solve for [R]. Start by multiplying the denominator of equation (25) on both sides of the equation,

$$\begin{gathered}[P]_{tot}+[P]_{tot}[R](K_R+K_TL)= \\ [R](1+L)+[R]^2(1+L)(K_R+K_TL)+K_R[R][X]_{tot}+K_TL[R][X]_{tot}\end{gathered} \tag{27}$$

Re-organizing to solve for terms of [R],

$$[R]^2(1+L)(K_R+K_TL)+[R](1+L+([X]_{tot}-[P]_{tot})(K_R+K_TL))-[P]_{tot}=0 \tag{28}$$

We define the following constants and solve the second order polynomial,

$$A=(1+L)(K_R+K_TL)=(1+0.135)(1\,mM^{-1}+10\,mM^{-1}*0.135)=2.67\,mM^{-1} \tag{29}$$

$$B=(1+L+([X]_{tot}-[P]_{tot})(K_R+K_T L))=$$
$$1+0.135+(1\,mM-1\,mM)(1\,mM^{-1}+10\,mM^{-1}*0.135)$$
$$B=1.135 \tag{30}$$

$$C=-[P]_{tot}=-1\,mM \tag{31}$$

$$[R]^2 A+[R]B+C=0 \tag{32}$$

$$[R]=\frac{-B\pm\sqrt{(B^2-4AC)}}{2A}=\frac{-1.135\pm\sqrt{(1.135^2-4*2.67\,mM^{-1}*(-1mM))}}{2*(2.67\,mM^{-1})}$$
$$[R]=0.44\,mM \text{ or } -0.86\,mM \tag{33}$$

Since this is a biological system where [R] is a concentration of a protein we are constrained by reality. Therefore,

$$0\,mM\le[R]\le[P]_{tot} \tag{34}$$

Therefore [R] = 0.44 mM. We can then use the value of [R] to calculate the remaining species,

$$[T]=L[R]=0.135*0.44\,mM=0.06\,mM \tag{35}$$

$$[X]=\frac{[X]_{tot}}{1+[R](K_R+K_T L)}=\frac{1\,mM}{1+0.44\,mM(1\,mM^{-1}+10\,mM^{-1}*0.135)}$$
$$[X]=0.49\,mM \tag{36}$$

$$[RX]=K_R[R][X]=1\,mM^{-1}*0.44\,mM*0.49\,mM=0.22\,mM \tag{37}$$

$$[TX] = K_T[T][X] = 10\,mM^{-1} * 0.06\,mM * 0.49\,mM = 0.29\,mM \tag{38}$$

And we can check the accuracy of our solution by making sure everything adds up,

$$[X]_{tot} = [X] + [RX] + [TX] = 0.49\,mM + 0.22\,mM + 0.29\,mM = 1\,mM \tag{39}$$

$$\begin{aligned}[P]_{tot} &= [R] + [RX] + [T] + [TX] \\ &= 0.44\,mM + 0.22\,mM + 0.06\,mM + 0.29\,mM = 1\,mM\end{aligned} \tag{40}$$

And of course, the entire point of this exercise is to see the effect of adding 1 mM of ligand X on the catalytic rate of protein P,

$$\begin{aligned}k_{eff} &= ([R] + [RX]) * k_R + ([T] + [TX]) * k_T = \\ 0.66\,mM * 10\,&\frac{rxn}{sec * mM} + 0.34\,mM * 1\,\frac{rxn}{sec * mM} \\ k_{eff} &= 6.94\,\frac{rxn}{sec}\end{aligned} \tag{41}$$

The allosteric effect of adding 1 mM ligand X is to change the catalytic rate from 8.92 rxn/sec to 6.94 rxn/sec. This is achieved by altering the *effective* conformational equilibrium of the R and T states. Adding ligand X, which prefers binding to the T state, pushes more of the protein to that state thereby lowering the bulk (effective) catalytic rate.

The MWC model provides a mathematical framework for understanding the link between the fundamental thermodynamic parameters (conformational equilibrium, rate constants, binding constants) and the experimentally measurable values (decreased catalytic rate). One primary drawback of doing these calculations that should

immediately be obvious is that even for simple systems, analytical solutions are complex. For complex systems, such as the lac repressor, the MWC solution either is not used or simplifications are used to model the experimental behavior.

1.5 The MWC Model of Lac Repressor Genetic Regulation

The allosteric response of the lac repressor is classically modeled by MWC. The lac repressor is a tetramer that is better understood as a dimer of dimers. A single dimer of the lac repressor binds a single DNA operator through a pair of its N-terminal helix-turn-helix motifs along with a larger DNA binding interface. The N-terminal domain is linked to a larger clam-shell domain that contains at its center an effector binding domain. Therefore, one dimer of lac repressor binds two effector molecules and a single DNA operator. The dimers are then linked through a C-terminal coiled-coil domain that tetramerizes the protein. A tetramer is therefore capable of binding two distinct DNA operators and four effector molecules.

Two different allosteric effects are found within a lac repressor tetramer. First, the binding of one dimer to one DNA operator increases the affinity of the second dimer to a second DNA operator. However, this is only true when the two DNA operators are linked into a larger DNA strand and furthermore the cooperativity of binding is highly dependent upon the spacing between the two operators. As DNA is not a small molecule, the conformations of the DNA come into play, but its the same old MWC idea: different conformations are populated when binding occurs and higher affinity is a net result.

The second allosteric property found in the lac repressor is the link between effector binding and operator DNA binding. Essentially this allostery occurs within the

individual dimers of the tetramer, and in fact a dimeric lac repressor is a fully functional gene regulator. The dimer can be thought of as having two primary conformations, R and R*. There of course exists a large number of conformations of the lac repressor dimer, but R and R* can be thought of as the two primary minima in the energy landscape. The Boltzmann distribution defines the relative occupancies of the R and R* states at equilibrium and is fully dependent upon the energetic differences between the two states. The conformational equilibrium of the lac repressor, K_{RR^*}, has been experimentally determined to slightly favor the R* state (Daber, Sharp, & Lewis, 2009; Daber, Sochor, & Lewis, 2011).

The two conformations each can bind one operator DNA with an affinity (K_{RO} for the R state; K_{R^*O} for the R* state) and two effector molecules each with same affinity (K_{RE} for the R state; K_{R^*E} for the R* state). For the wild type lac repressor, we define the R state as the conformation with higher operator DNA affinity ($K_{RO} > K_{R^*O}$). The relative affinities of R and R* for effectors determines whether they are classified as inducers or co-repressors. If $K_{RE} > K_{R^*E}$ then the effector prefers binding to the high DNA affinity state and will push more of the repressor into a DNA bound state; these effectors are termed co-repressors (such as orthonitrophenyl-β-D-galactoside (ONPG)). If $K_{RE} < K_{R^*E}$ then the effector binds more strongly to the low DNA affinity state and will decrease operator DNA binding; these effectors are termed inducers and include allolactose and IPTG.

The lac operon of *E. coli* uses three DNA operators (O1, O2 and O3) and a tetrameric lac repressor and the effector is the inducer allolactose which is a short lived metabolic intermediate. This is a highly non-trivial system to model and thus a simplest functional system has been developed which consists of a single operator DNA located

within the DNA promoter sequence (where O1 is located), a dimeric lac repressor, and the effector is the non-hydrolyzable inducer IPTG. This system isolates the allosteric interaction between effector and operator DNA binding.

1.6 Mutations to the Lac Repressor Alters Gene Regulation

Mutational studies of the lac repressor began as a means to discern structurally and functionally relevant regions of the protein. The lac repressor was discovered before most standard biochemical techniques existed to answer these questions such as protein over-expression for purification, protein crystallography, protein NMR and modern PCR based cloning techniques.

The earliest lac repressor mutants were discovered using mutating agents and selection pressure to try to isolate mutants. For example, bacteria could be exposed to N-methyl-N'-nitro-N-nitrosoguanadine and given repeated rounds of forward and backwards selection pressures to identify tighter binding lac operon phenotypes (Gilbert & Müller-Hill, 1966). A more refined approach was later developed using amber codon (UAG) suppression (Belin, 2003) where tRNAs could be used to suppress the amber stop codon and insert desired amino acids in its place. For example, amber codon suppression was used to show that mutation of tyrosine 269 to serine (Y269S) inhibits tetramerization (Schmitz et al., 1976).

The lac repressor was discovered before much of the standard biochemical approaches existed, therefore the literature of the lac repressor often follows the changing state-of-the-art of the time. One of the fascinating results of this history is that studies of the lac repressor exist that would never be repeated for a newly identified protein today. Chief amongst these studies is the series of 15 papers which sought to

link genetic alteration of the lac repressor to phenotype. The end result was point mutations of every amino acid from positions 2 through 329 to 12 or 13 different amino acids (>4000 mutations in total) which were then phenotypically categorized (Suckow et al., 1996). The phenotype I^- indicates mutations where repression is impaired in the absence of inducer and are therefore defective. I^- mutations alter DNA binding, folding or aggregation of the lac repressor. The phenotype I^S indicates mutations that repress but do not respond to inducer (super-repressing). I^S mutations could alter inducer binding, allosteric signaling, or be the result of extremely tight DNA binding. The phenotype I^r indicates mutations that co-repress with what is normally an inducing effector. I^r mutations could be a result of changing DNA affinity, reversing the allostery, or altering inducer binding. Essentially these studies created a map wherein different aspects of the lac repressor allostery was linked to different regions of the amino acid sequence.

When the x-ray crystallographic and NMR structures of the lac repressor were solved (Lewis et al., 1996; Loth, Gnida, Romanuka, Kaptein, & Boelens, 2013), this body of mutational data was mapped onto the three-dimensional structure. In general there is excellent agreement between these two techniques. Mutations to the effector binding pocket typically result in the I^S phenotype because inducer binding is destroyed; the lac repressor can bind operator DNA but not effector. Similarly, I^r mutations typically are found in the effector binding pocket as they involve changing effector affinity. Mutations to the N-terminal DNA binding domain typically result in the I^- phenotypes because operator DNA binding is destroyed; the lac repressor can bind effector but not operator DNA. A large fraction of the amino acid substitutions made were tolerated by the protein. Disruptive amino acids (I^-) were found to lie along the dimer interface and at key

locations of the lac repressor fold. This large scale mutagenesis combined with structure provides a unique confirmation that the experimentally derived structures are extremely relevant to the phenotypes observed in living bacteria. The central limitation of the genetic studies is that they could only identify bulk qualitative phenotypic effects of the mutations. It took a confirmation with the structure to link mutations with thermodynamic binding parameters, however only qualitatively.

Later studies began to identify other point mutations of the lac repressor and quantitatively assess the effects. For example the mutation K84L was identified which creates a more thermostable repressor but impairs both DNA and effector affinities (Barry & Matthews, 1999). This mutation was crystallographically characterized and found to lock dimer interface of the lac repressor in a novel conformation (Bell et al., 2001).

1.7 Aims

It has been well established how the lac repressor functions within *E. coli* to regulate the *lacZ/Y/A* polycistronic message and induce lactose metabolism when it is needed. Furthermore it is well known that the gene regulatory phenotype can be changed by point mutation of the lac repressor or by changing or moving operator DNA sequences.

Several important gaps in knowledge however remain. How do point mutations to the lac repressor affect the thermodynamic parameters of the MWC model? Can we link specific regions of the lac repressor to specific thermodynamic constants? Furthermore, while we know the lac repressor is functional in purified *in vitro* conditions,

do the *in vitro* measured binding values match the *in vivo* measured values? Finally, can we use these known lac repressor genetic switches to treat human diseases?

The specific aims of this thesis seek to address these questions:

1. **Linking Mutation to Thermodynamics** (Chapter 3). We have taken mutations to two structurally specific regions and linked them to changes to specific thermodynamic parameters of the MWC model. Mutations in the specificity region of the DNA binding domain are seen to alter the repressor-DNA affinity and potentially the conformational equilibrium. Mutations in the specificity region of the effector binding pocket are seen to alter the repressor-effector affinities and the conformational equilibrium. The interplay between the changes in the thermodynamic parameters results in a wide range of phenotypically interesting genetic switches ranging from more ideal on-off switches to inverted switches.
2. **Linking *In Vivo* to *In Vitro*** (Chapter 4). We have created a quantitative *in vivo* assay that allows for accurate measurement of both regulated transgene (YFP) and fluorescently tagged lac repressor (lac-mCherry). We measured the thermodynamic binding parameters of the lac genetic switch from an *in vitro* transcription assay and use these values to predict the *in vivo* phenotypic profile. We found that in general our measured *in vitro* thermodynamic values matched those of previous studies, except for repressor-DNA affinity. Our affinity was several orders of magnitude weaker than previous studies. Despite this, our values do recreate the *in vivo* gene regulatory profile when 40% cell crowding is modeled. Furthermore, only our values could create functional genetic regulators at the modeled concentrations of the natural lac operon. This study showed the seminal importance of confirming *in vitro* results with *in vivo* data.

3. **An Autogenously Regulated Lac Repressor for AAV-Mediated Eukaryotic Gene Therapy** (Chapter 5). We have redesigned the lac repressor genetic switch to be autogenously regulated (regulates its own expression) with the goal of creating an AAV-mediated gene therapy regulatory tool. This autogenous switch was shown to work in various cell types in cell culture. We have improved regulatory behavior of the switch through the use of multiple operator DNA sites and the restoration of the tetramerization domain in the lac repressor. We found further improvement by making point mutations to the lac repressor that are known to improve the phenotype in *E. coli*. Finally, we packaged our autogenously regulated lac genetic switch in AAV vectors. These vectors were used to reversibly regulate luciferase expression in the livers of living mice as a function of IPTG.

Chapter 2:

Materials and Methods

2.1 Bacterial Strains and Media

Several bacterial strains were used for prokaryotic experiments and all are derived from Escherichia coli (*E. coli*).

The strain DH5α (*E. coli fhuA2 Δ(argF-lac)U169 phoA glnV44 φ80' Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) was used for cloning, plasmid preparations, and for prokaryotic green fluorescent protein (GFP) regulation assays. This strain does contain some of the lactose operon and therefore is not a clean background within which to perform the prokaryotic GFP regulation assay. It was later replaced with the EPB225 cell line (described below).

The strain BL21(DE3) (*E. coli B dcm ompT hsdS($r_B^-m_B^-$) gal*) was used for recombinant protein expression to make purified lac repressor and lac repressor mutants.

The strain EPB229 (*E. coli F^- λ^- ilvG- rfb-50 rph-1 Δ(lacI-lacA)::frt*) was used for prokaryotic GFP regulation assays. This strain was made by the laboratory of Dr. Mark Goulian and derived from MG1655 (*E. coli F^- λ^- ilvG- rfb-50 rph-1*). This strain is the "wild-type" K-12 strain and has a total deletion of the lac operon allowing for a clean background within which to study our lac genetic regulatory systems.

Liquid media for standard bacterial growth (plasmid preparation, cloning, initial prokaryotic GFP regulation assays) was Luria-Bertain (LB) broth. Similarly, LB-agar was used for growth on solid plates. Later prokaryotic GFP regulation assays with EPB229

cells used MOPS minimal media supplanted with 0.4% glucose. This media provided minimal auto-fluorescent background.

Antibiotics were used for selection of bacterial cells transformed with desired plasmids at the following concentrations:

Ampicillin = 100 μg/mL

Chloramphenicol = 50 μg/mL

Kanamycin = 50 μg/mL

For most prokaryotic GFP regulation assays, a two plasmid selection method was used by using media with both ampicillin and chloramphenicol.

2.2 Eukaryotic Cell Lines and Media

The cell line HEK293T, derived from human embryonic kidney cells and stably expressing the Simian Vacuolating Virus (SV40) which codes for the large T-antigen, was used for Eukaryotic GFP and luciferase regulation assays and also in microscopy. These cells are highly adapted to cell culture conditions and are not thought to be particularly characteristic of healthy kidney cells. This cell line is adherent and was grown on plastic cell culture plates. The culture medium was Dulbecco's modified Eagle's medium (DMEM) containing 20% fetal bovine serum (FBS), L-glutamine and sodium bicarbonate.

The cell line Cos-7, derived from the African green monkey *Cercopithecus aethiops* fibroblast-like kidney cells, was used in Eukaryotic GFP and luciferase regulation assays. This cell line was developed from the CV-1 line by transformation

with SV40 to express the T-antigen. This cell line is adherent and grown on plastic cell culture plates. The culture medium was Dulbecco's modified Eagle's medium (DMEM) containing 20% fetal bovine serum (FBS), L-glutamine and sodium bicarbonate.

The cell line ARPE-19, derived from human retinal pigment epithelial cells, was also used in Eukaryotic GFP and luciferase regulation assays. This is an adherent cell line and was grown on plastic cell culture plates. The culture medium was Dulbecco's modified Eagle's medium (DMEM) supplemented with the F-12 nutrient mixture containing 20% fetal bovine serum (FBS), L-glutamine and sodium bicarbonate.

2.3 Plasmid Preparation

Plasmids were isolated using the NucleoSpin plasmid purification kit (Clontech). For both the repressor and reporter bacterial plasmids a slightly modified protocol was followed as they both are low-copy number plasmids. A larger initial volume of LB is inoculated (8mL) and larger volumes of individual DNA purification buffers are used: 500 µL A1, 500 µL A2, and 700 µL A3 of the Macherey-Nagel quick purification kit. The remainder of the protocol was unchanged.

Large volumes of plasmid were obtained using the Plasmid Plus Maxi Kit (Qiagen). This allowed for endotoxin free plasmid preparation for transient transfections and also for virus preparation from AAV encoding plasmids.

2.4 Cloning

Inverse PCR Mutagenesis

The following protocol was used for point mutations, deletions and small insertions (typically less than 10 base pairs).

The primer design completely defines what the result of the cloning will be. Each primer is broken down into two parts: an annealing region that overlaps with the starting plasmid and an optional un-matched 5' tail. The annealing region is designed to follow two basic rules. First, the melting temperature should be as close to 60C as possible. Second, both the 3' and 5' ends should be either cytosine (C) or guanine (G). The reverse primer is designed to anneal to the left of the region to change/insert/delete and the forward primer is designed to anneal to the right of the region to change/insert/delete.

For a deletion, primers are chosen that skip the region to delete. For a point mutation, a 5' tail is added to one of the primers that matches the desired sequence. Similarly, for an insertion a 5' tail is added to one of the primers that matches the desired insertion. Finally, one of the primers is given a 5' phosphorylation.

Full circle PCR is used with Phusion Hot Start High Fidelity DNA Polymerase (NEB) and the extension time is set to 20 seconds for every kilobase and follows the recommended thermal cycle. PCR follows 15 rounds of amplification and is then held at 4C until ready to proceed. The mixture now contains plasmid (initially isolated from DH5α) and linearized PCR product which has the desired change/insert/deletion. A 10 minute ligation with 2000 Units T4 DNA Ligase (NEB), 10 µL quick ligation buffer (NEB) and 10 µL PCR product is used to ligate the linearized PCR product into closed plasmid. This is followed by a 1 hour digestion to remove the initial plasmid by adding 5 µL NEB #4 Buffer, 25 µL MilliQ, and 20 Units DpnI (NEB). The DpnI enzyme specifically cleaves DNA with the methylation pattern given by DH5α cells, therefore specifically cleaving only the starting plasmid while sparing the ligated PCR product. 2.5 µL of this mixture is

transformed into 50 µL DH5α cells and plated onto LB agar with appropriate selection antibiotic.

Typically >90% of colonies that survive on the plate have the intended change/deletion/insertion. Common difficulties are 5' adenosine (A) or thymine (T) bases will not make it into the final plasmid which is why they are to be avoided if possible in the initial primer design step. It is also commonly seen that the starting plasmid will survive the DpnI digestion so it is recommended to include a DpnI negative control where primers are omitted so only starting plasmid exists.

In-Fusion

This method is used to insert large DNA segments into plasmids. The primer design exactly follows the recommendations of the In-Fusion HD Liquid Kit (Clontech). Long DNA regions (approximately greater than 1 kilobase) were amplified with Phusion Hot Start High Fidelity DNA Polymerase (NEB) following the standard thermal cycle and extension time of 20 seconds per kilobase. Small DNA regions (approximately less than 1 kilobase) were amplified with Vent DNA Polymerase (NEB) following the standard thermal cycle and extension time of 60 seconds per kilobase. Linearized PCR fragments were isolated through 5% agarose gel filtration and the NucleoSpin Extract II (Clonetech) purification kit. Alternatively, restriction enzyme digestion was used when appropriate to linearized DNA fragments as described in the In-Fusion protocol.

Essentially, two linearized fragments are desired: the vector which typically contains the desired new plasmid and the insert to be placed into that plasmid. The vector and insert are designed such that 15 base pairs on each end perfectly match. The In-Fusion enzyme then anneals the insert and vector to create a final plasmid.

The In-Fusion reaction follows the recommended protocol and 2.5 µL is transfected into 50 µL DH5α cells and plated onto LB-agar with appropriate antibiotic. Again >90% of the colonies typically contain the designed insert.

2.5 Lac Repressor Purification

Lac Repressor constructs were cloned into a pBAD expression vector which controls expression of a gene using the arabinose repressor system. All repressor constructs were inserted into the multiple cloning site (MCS) of the pBAD vector and given a 6x-Histidine tag on the C-terminus. Cloning was achieved using in-fusion.

Purification begins by first transforming the desired plasmid into BL21(DE3) cells and plating onto LB supplemented with ampicillin and incubating overnight at 37C. The plate is used to innoculate 100 mL 2xYT medium (Sigma) supplemented with ampicillin in the morning. I picked many colonies (>100) in the innoculation to prevent any potential bias from selecting a single colony. This starter culture is incubated at 37C with shaking (~200 RPM) for 3-4 hours until the flask grows past the mid-log phase and approaches the stationary phase (optical density at 600 nm (OD_{600}) > 1.0). 10mL of the starter culture is then used to innoculate 1L 2xYT medium supplemented with ampicillin. The large culture is incubated at 37C with shaking (~200RPM) until early mid-log growth phase (OD_{600}=0.3-0.4). At this point, the incubator temperature is reduced to 15C to slow bacterial growth. Expression is also induced at this point by adding 10mL of 20% v/v arabinose that has been sterile filtered. The liquid culture is allowed to grow overnight (~16 hours).

In the morning, the culture is removed from the incubator and placed into 1L centrifuge flasks. The cells are centrifuged at 5000g for 10 minutes at 4C to pellet the

cells. The supernatant is removed and cells are resuspended 20 mL in Nickel Lysis Buffer (300 mM NaCl, 50 mM NaH_2PO_4, pH 8.0). Cells are lysed using three passes through a cell homogenizer (Avestin) at ~18,000 PSI and whole cell lysate is immediately stored on ice. Lysate is then centrifuged at 15,000g for 15 minutes at 4C to separate soluble protein from insolubate cell lysates. Supernatant is filtered through filter paper (Whatman #3) at 4C to remove further insoluble materials. A sample of the supernatant is collected for analysis on an SDS-PAGE gel.

A his-tag purification column is prepared by loaded 2 mL His60 Ni Superflow Resin (Clontech) into a Flex-Column (Kontes) fitted with a stopcock to regulate flow rate through the column. All work with the his-tag column is undertaken in refrigeration at 4C. Resin is first equilibrated with 20 mL cold Nickel Lysis Buffer and the stopcock is set to very slowly drip out the buffer such that the nickel beads will form an equilibrated bed of beads. Filtered supernatant is then carefully added to the column and flow again is set to very slowly drip. A sample of the flow through is collected for SDS-PAGE gel analysis. The beads are then washed three times with Nickel Wash Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, 2.5% v/v glycerol, pH 8) and samples of each wash are collected for SDS-PAGE gel analysis. His-tagged lac repressor is then eluted from the column using Nickel Elution Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, 2.5% v/v glycerol, pH 8) and elutant was collected. A sample of elutant was saved for SDS-PAGE gel analysis.

Nickel elutant is then buffer exchanged using an Amicon Ultra centrifuge filter with a 30,000 molecular weight cutoff (Millipore). Elutant is centrifuged at 5,000g for 15 minutes at 4C and the filter flow through is discarded. Typically, depending on yield, 200-300 µL concentrated protein remains above the filter. Protein is carefully

resuspended in ~3-4 mL ice-cold Gel Filtration (GF) Buffer (200 mM Tris pH 7.4, 200 mM Kcl, 10 mM EDTA, 3 mM DTT). Buffer replaced in this fashion reduces the initial buffer to about 7% of its starting concentration. Buffer is replaced three times resulting in < 0.1% of the elutant buffer remaining. Protein is then concentrated a final time to achieve a protein concentration in the 1-10 mg/mL range as measured using absorbence at 280nM (A280) on a spectrophotometer (NanoDrop).

Purified protein can be saved at this step by flash freezing the protein in liquid nitrogen and storing at -80C.

All samples from the nickel purification are boiled in LDS sample buffer (Expedeon) for 10 minutes and run on a 4-12% SDS-PAGE gel (Expedeon) at 110V for 1 hour.

Protein for *in vitro* transcription assay underwent a size exclusion chromatography step. His-tag purified lac repressor in GF buffer was concentrated and loaded onto a Hi Load 16/60 Superdex 75 column and sample was driven by Akta Prime FPLC. Fractions were collected and samples were SDS boiled and run on SDS-PAGE gels to determine purity. Protein was concentrated again using centrifugal 30,000 MW filter and stored in GF buffer. Protein was flash frozen and store at -80C.

2.6 Prokaryotic GFP and YFP Regulation Assays

GFP regulation assay

EPB229 cells were co-transformed with a lac repressor plasmid providing chloramphenicol resistance and a reporter plasmid providing ampicillin resistance and

plated onto LB agarose plates with ampicillin and chloramphenicol. Plates were incubated overnight at 37C and stored at 4C.

The repressor plasmids are all derived from pABD34 plasmid and constitutively express the LacI gene using its native promoter from *E. coli*. The LacI sequence has 10 C-terminal residues truncated to prevent tetramerization. Point mutations to the LacI sequence made using inverse PCR mutagenesis. For repressor quantification experiments a C-terminal mCherry tag was inserted using in-fusion cloning.

Reporter plasmid were derived from the pBR plasmid and have a reporter gene, Green or Yellow Fluorescent Protein (GFP or YFP), under control of the native LacZ/Y/A promoter. The operator sequences of the promoter used were:

O1: 5' AA TT GTG AGC G GAT AAC AA TT 3'

Lsym: 5' AA TT GTG AGC GCT CAC AA TT 3'

In the original assay, only GFP was monitored. Single colonies were picked into 1 mL LB supplemented with ampicillin and chloramphenicol and grown ~ 16 hours at 37C with shaking (~200 RPM). 100 µL of the sample aliquoted into a 96-well clear bottom plate (Corning) and OD_{600} and fluorescence (excite: 488 nm, emit: 525 nm) were measured on Infinite M1000 plate reader (Tecan).

Samples to be assayed are typically picked in triplicate. Fluorescence is normalized by OD_{600} and averaged across replicates.

YFP regulation assay

Reporter plasmid was made as previously reported with the O1 operator sequence (5'-AA TT GTG AGC G GAT AAC AA TT-3') followed by YFP and providing ampicillin (AMP) resistance. Lac repressor was expressed on a second plasmid as previously described providing chloramphenicol (CAM) resistance. A C-terminal mCherry tag was added to the Lac repressor gene after an 11bp linker to create the Lac-mCherry construct.

We double transformed reporter and repressor plasmids into EPB229 cells (F^- Δ(lacI-lacA)::frt). These cells were derived from the MG1655 "wild type" line. Colonies were picked in triplicate into MOPS minimal media with 0.4% glucose, AMP and CAM and grown overnight at 37C with shaking. 50μL of the overnight culture was used to inoculate 1mL fresh MOPS minimal media supplanted with varying amounts of IPTG. We measured optical density at 600nm (OD_{600}), YFP fluorescence (excite: 510 nm emit: 535 nm), and mCherry fluorescence (excite: 585 nm emit: 610 nm) for all wells over a 12 hour period using a TECAN M1000 plate reader in 384 well optical bottom plates (Corning).

In vivo data was normalized for growth by measuring cells in triplicate as they were growing. All data points collected were then fit to a 2^{nd} order polynomial to obtain a curve which is fluorescence as a function of OD_{600}. Positive control was established by co-transforming EPB229 cells with O1 YFP reporter and a CAM plasmid without Lac-mCherry (pABD34). YFP signal was normalized to the polynomial fit from the positive control. Final values for fitting were calculated for cells at approximately mid-log growth phase (0.4 OD_{600}).

2.7 Prokaryotic Lac Repressor Quantification Assay

EPB229 cells were co-transformed with Lac-mCherry and O1 YFP reporter. An individual colony was picked into MOPS minimal media with 0.4% glucose, AMP and CAM and grown overnight at 37C with shaking. 50 µL was innoculated into fresh media and grown to mid-log phase.

Serial dilutions were made over 10 orders of magnitude and each dilution had OD_{600} measured (TECAN M1000 and Ultrospec 2100 pro) and 100 µL plated onto LB agar with AMP and CAM.

Purified Lac-mCherry was quantitated with both a BCA assay and optical A_{280} measurements using a Nano-drop. Dilutions were made over 8 orders of magnitude and 50 µL was loaded into clear bottom 384 well plates in triplicate. mCherry fluorescent measurements (excite: 585 nm emit: 610 nm) were made using various gains to establish linear regimes for the instrument (TECAN M1000).

We established a raw cell count by plating dilutions of a culture of EPB229 cells. We found $1.92e^6$ cells/µL at mid-log growth phase which is similar to standard estimates of $1e^6$ cells/µL for *E. coli*. Aliquots of known cell counts were then used to establish a linear relationship with OD_{600} on our plate reader. Similarly, purified Lac-mCherry of known concentration was used to establish a linear relationship with mCherry fluorescence on our plate reader at a fixed gain.

EPB229 cells were co-transformed with plasmid constitutively expressing Lac-mCherry and a reporter plasmid which has YFP under the control of the natural operator

O1. We measured mCherry fluorescence at a fixed gain and OD_{600} from which we calculated the concentration of Lac-mCherry in the well and the number of cells in the well. The approximate volume of *E. coli* was estimated to be $1x10^{-15}$ L. Multiplying volume of *E. coli* by number of cells allows us to estimate what fraction of the well volume is intracellular.

Calibration of raw mCherry fluorescent signal and OD_{600} was converted to intracellular repressor concentration.

2.8 *In Vitro* Transcription

A reporter plasmid was made with the O1 operator after a T7 promoter. Reporter was linearized to 450bp and purified by spin column purification (Clontech).

MaxiScript T7 kit (Ambion) was used to perform *in vitro* transcription. CTP[α-^{32}P] was incorporated into mRNA transcripts and the water fraction of the standard reaction was supplanted with varying concentrations of Lac-mCherry and IPTG. Transcription was allowed to proceed for 30 minutes at 37C until halted by boiling. Samples were loaded onto polyacrylamide gels and electrophoresis was used to separate free CTP[α-^{32}P] from that incorporated into mRNA. Gels were dried and exposed to radiological plates. Plates were imaged on a Typhoon scanner and bands were quantitated using ImageJ (NIH).

2.9 Transient Transfections

Cells were allowed to grow to approximately 50 to 70% confluence and split to prepare for transient transfection. Cells were allowed to grow in 37C incubator with 5%

CO_2 overnight to attach. Media was changed in the morning and cells were now ready for transient transfection.

Plasmid DNA was prepared by first mixing with Opti-Mem (Invitrogen), vortexing, and letting sit at room temperature for 5 minutes. Lipofectamine LTX (Invitrogen) was then added to the mixture, very briefly vortexed, and then allowed to sit at room temperature for 30 minutes. The prepared solution is good for several hours at room temperature. The final mixture is then added to cells which are then returned to the incubator. Media should be changed after 24 hours. The exact amounts of plasmid DNA, Opti-Mem, and Lipofectamine LTX (and potential additives such as Plus reagent) were determined by the cell type and can be found within the product literature.

2.10 Native Gel Electrophoresis

We used the NativePage (Invitrogen) kit for native gel electrophoresis. The primary advantage of this kit is that it is based upon Blue Native Polyacrylamide Gel Electrophoresis (BN Page) which uses Coomassie G-250 as the molecule to provide charge shift for proteins. Coomassie G-250 binds to proteins providing a net negative charge without denaturing the protein.

Purified protein was thawed and mixed in non-denaturing sample loading buffer. 10 µL of samples and NativeMark (Invitrogen) protein standard were added to wells of NativePage 4-16% Bis-Tris Gels. Gels were loaded within Novex Mini-cell (Invitrogen) gel running boxes.

The interior (between the gels) was filled with NativePage cathode running buffer which contains Coomassie G-250. The exterior was filled with NativePage anode running buffer. Gels were run at 150V for 2 hours, then removed. Gels were placed in

Fix (40% methanol, 10% acetic acid), microwaved on high for 45 seconds, and then shaken on an orbital shaker for 15 minutes to fix the gel. The gels were then switched to Destain (8% acetic acid), microwaved on high for 45 seconds, and placed on an orbital shaker overnight to remove any unbound Coomassie G-250. Gels were then imaged.

2.11 Eukaryotic Gene Regulation Assays

All gene regulation assays start by transiently transfecting the cell line of choice with the appropriate plasmid. Media is supplemented with effector of varying concentration in wells (IPTG for lac repressor experiments and doxycycline for tet repressor experiments) and specific experimental conditions are typically grown side by side in triplicate. Experimental measurement times varied depending upon cell confluence but typically occurs between 24 and 72 hours after transient transfection.

Fluorescence Microscopy

Fluorescence microscopy was performed in several different ways. For live cell microscopy, cells were grown as previously described. Media was switched to phosphate buffered saline (PBS) to reduce auto-fluorescence and cells were imaged on a Nikon Eclipse TE2000-U inverted microscope equipped with green and blue filters and a Nikon Intensilight C-HGFI power source.

High resolution fluorescence microscopy was also performed by growing cells in the Lab-Tek II 8 well Chamber Slides system (Nunc). Cells followed the standard cell culture, transient transfection, and experimental conditions only scaled down for the smaller surface area of the individual wells. To prepare for microscopy, cells were washed three times in DPBS with calcium and magnesium (Corning) and then all media was removed. Cells were fixed for 1 hour at room temperature with 2%

paraformaldehyde (PFA) and then washed three times with DPBS with calcium and magnesium. All liquid was removed and chamber supports were separated from the slide. Cells were immersed in a small droplet of FluoroMount-G (Southern Biotech) supplemented with 4',6-diamidino-2-phenylindole (DAPI) to reduce fluorophore quenching and nuclear staining, respectively. Slide covers were then placed on top of cells and affixed with nail polish. Slides were stored at 4C. Fixed slides were viewed on the a Zeiss AX100 microscope with an X-cite series 120Q light source.

Bulk Fluorescence Assays

Quantitative fluorescence data was achieved by measuring the bulk properties of eukaryotic cells. Two assays were developed which were shown to give approximately the same result.

First, cells were washed three times with DPBS, were lysed with Reporter Lysis Buffer (Promega) and the cells were scraped off of the bottom of the growth plates. 50 µL whole cell lysate was then loaded into wells of a 384 well optical bottom plate (Corning) into the Tecan M1000 and fluorescent emission was read with the following excitation/emission wavelengths: GFP excite: 485nm emit: 510nm, YFP excite: 510 nm emit: 535 nm, and mCherry excite: 585 nm emit: 610 nm.

Second, cell culture plates were washed three times with DPBS and the entire cell culture plate was scanned on the Typhoon scanner. The scanner was set with a 3.5 mm vertical offset to better visualize the cells growing within the wells of the plate. The 473 nm blue excitation laser was used to excite GFP and the >520 nm filter was used to read emission. Images were analyzed using ImageJ (NIH).

Luciferase Assay

As before, cells were washed three times with DPBS, were lysed with Reporter Lysis Buffer (Promega) and the cells were scraped off of the bottom of the growth plates. 5 µL of whole cell lysate was added to 45 µL of luciferase assay buffer (Promega) containing luciferin within a well of a 384 well optical bottom plate (Corning). Plates were immediately loaded into the Tecan M1000 and luciferase signal was measured for each well over time. Peak luciferase signal values were used for quantification.

2.12 Western

Whole cell lysate is first collected from cells of interest as previously described. Whole cell lysate is then mixed with 4x LDS sample buffer (Expedeon) and boiled at 70C for 10 minutes to denature the proteins. Samples were loaded into wells of 10-20% SDS-PAGE gels (Expedeon) and electrophoresed at 110V for 1 hour using the RunBlue SDS Running Buffer (Expedeon).

Gels were removed and washed with MilliQ water and then prepared for semi-dry transfer to Immobilon-P transfer membrane (Millipore). Gels were soaked for 10 minutes in western transfer buffer (25 mM Tris, 200mM glycine, 10% methanol). Transfer membrane is placed in 100% methanol for 15 seconds, moved to MilliQ for 2 minutes, and then to western transfer buffer for 5 minutes. Six pieces of 3mm filter paper (Whatman) are also soaked in western transfer buffer. The transfer stack is prepared as follows: 3 pieces filter paper – equilibrated PVDF transfer membrane – equilibrated gel – 3 pieces filter paper. The sandwich is then loaded onto the semi-dry transfer plates and run at at 15V for 15 minutes.

PVDF membrane is then removed from the sandwich and blocked in Blotto (1g non-fat powdered milk in 20 mL PBS-T(PBS with 0.1% Tween-20)) for 1 hour.

Membrane was then washed three times for 10 minutes each with PBS-T. Membrane was then placed in the primary antibody at 4C overnight on an orbital shaker. The buffer and concentration of primary antibody were determined by the manufacturers recommendation for each antibody. In the morning, the gel was washed three times with PBS-T and placed in the appropriate secondary antibody conjugated to horseradish peroxidase (HRP). The gels were then exposed to chemiluminescent substrate specific to HRP and imaged.

2.13 Mouse Experiments

AAV8 viral vector was prepared using the University of Pennsylvania Vector Core from plasmid DNA.

Individual mice were given tail vein injections of AAV8 at a viral dose of ($1e^{11}$ multiplicity of injection (MOI)) regardless of weight. Weight in grams for each mouse was recorded throughout the experiment. Mice were fed a normal diet and given two weeks to incorporate the virus in their liver tissue.

Gavage protocol was 25 µL of 1M IPTG per 10 gm body weight, two times a day for 3 days. IPTG was gavaged through an oral syringe into the stomach of the mice. Mice were either given normal water or gavaged IPTG throughout the experiment to either induce or repress expression of the transgene.

Luciferase imaging was performed by giving mice interperitoneal injections of luciferin and imaging them on a live mouse scanner. Raw luciferase signal was quantified with ImageJ (NIH).

2.14 Modeling

All functions from text were coded into Matlab (Mathworks) scripts to calculate the MWC model outputs. Standard linear regression minimization functions were used to minimize the difference between experimental and theoretical curves in order to fit for the thermodynamic parameters of interest. Various minimization schemes were used and all were built in Matlab functions such as a genetic algorithm and direct-pattern search.

The error estimates in fit parameters were done by a monte carlo approach. Each individual data point had some random fraction of its error added to it and the data was refit. This process was repeated > 100 times and then the standard deviation of the individual parameters was taken to be the error in that fit parameter. The best fit to the data without error added was taken to be the reported value for the individual parameters.

2.15 Plasmid Maps

The following is a list of all plasmids used throughout this thesis.

Prokaryotic plasmids

pBR series reporter plasmids were derived from giving ampicillin resistance and strict plasmid copy number of approximately 10-20 per cell (Figure 2.1). The reporter cassette contains the natural *lacZ/Y/A* promoter from the lac operon of *E. coli*. Only one operator is present in the plasmid and is located downstream of the promoter in the transcriptional initiation region (Figure 2.2). Regulated genes downstream of the promoter included eGFP, YFP and the polycistronic message of dimeric lac repressor-mCherry fusion

followed by YFP. An 18 base pair spacer which was the 18 base pairs that preceded the LacI-mCherry gene and were:

5' CAA TTC AGG GTG GTG AAT 3'

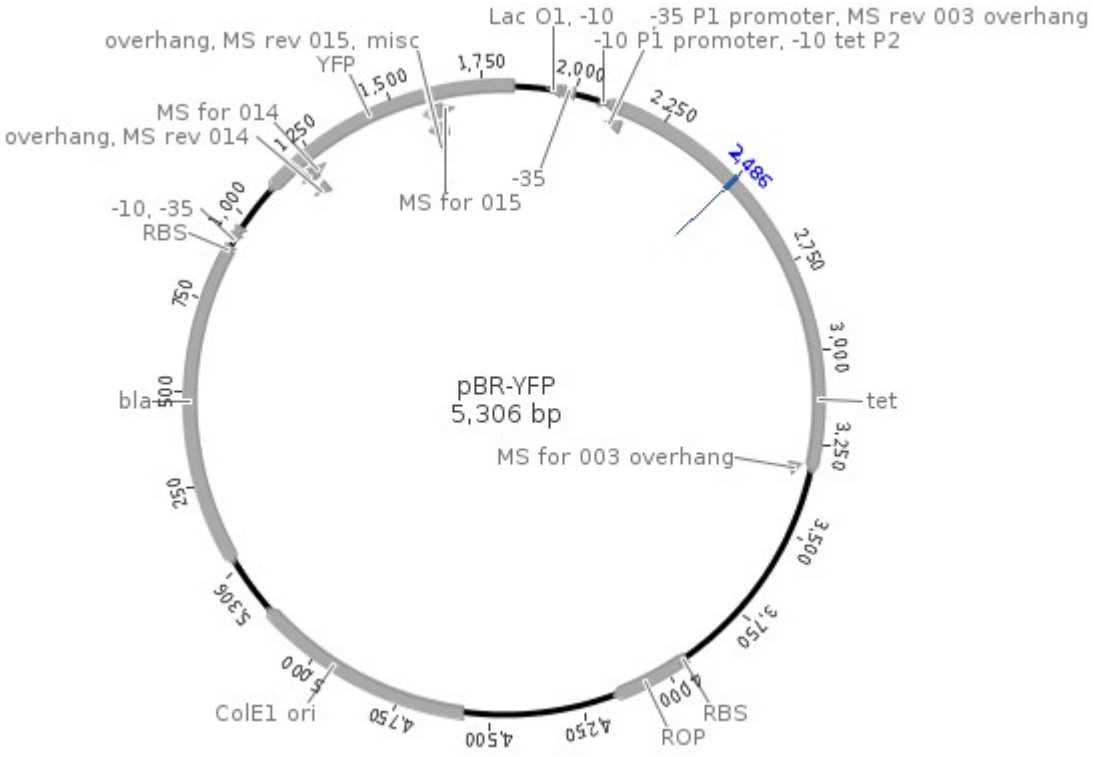


Figure 2.1. Plasmid map of the pBR-YFP plasmid.

Figure 2.2. Close-up of the operator placement within the lacZ/Y/A promoter of the pBR reporter plasmids.

The pLacI series of plasmids provide chloramphenicol resistance and constitutive lac repressor from the natural *lacI* promoter of *E. coli* (Figure 2.3). The *lacI* gene has the 11 C-terminal codons truncated. The lac repressor-mCherry fusion includes an 18 base pair linker with the sequence:

5' GGC TCA GGT CTC GAG TTG 3'

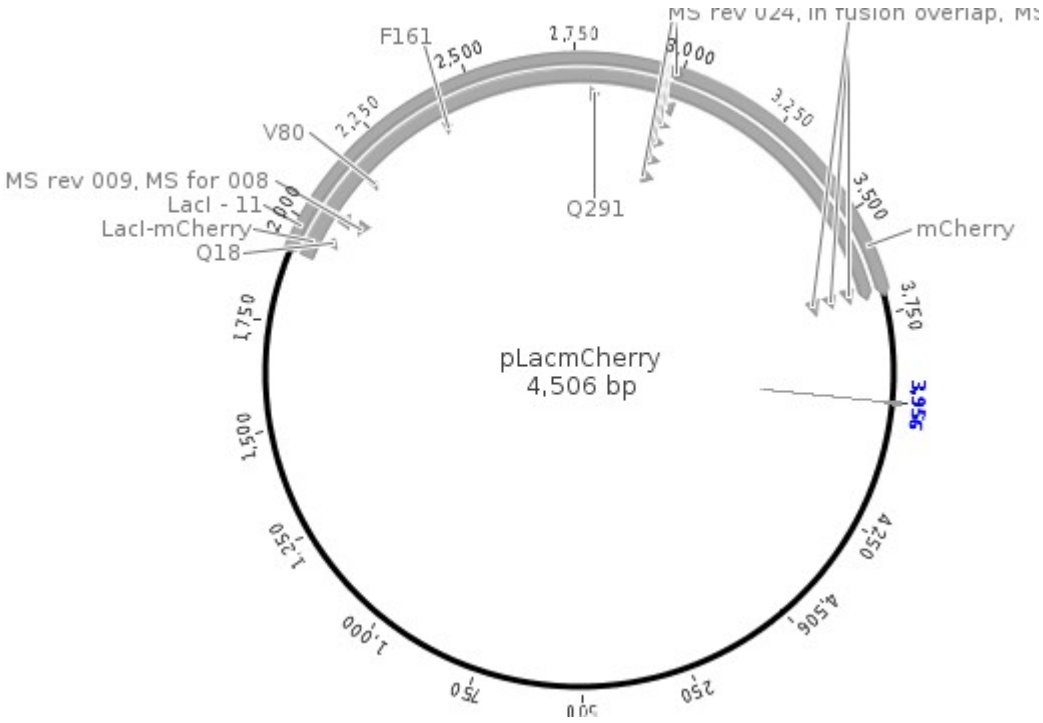


Figure 2.3. Plasmid map of the pLacI series plasmid with the lac repressor-mCherry fusion protein.

The arabinose expression plasmid is pBAD-DEST-49 (Invitrogen) and expression genes were placed within the first multiple cloning site.

Eukaryotic Plasmids

The pIRES plasmid (Clontech) has a minimal CMVI promoter with two multiple cloning sites linked by the IRES sequence. The original IRES plasmid put YFP in the first multiple cloning site and EuLacmCherry in the second multiple cloning site.

The EuLac sequence was taken from the codon optimized sequence for Eukaryotic cells with the splice site fixed (Cronin et al., 2001). We first used a dimeric EuLac sequence with the 11 C-terminal codons removed. Later we used the EuLacTet sequence with the 11 C-terminal codons of wild type *lacI* restored for tetrameric lac repressor. The following is the EuLacTet sequence with the 11 C-terminal amino acids that result in tetramerization highlighted in bold:

5'- ATG AAA CCA GTA ACG TTA TAC GAC GTC GCA GAG TAT GCC GGT GTC TCT TAT CAG ACT GTT TCC AGA GTG GTG AAC CAG GCC AGC CAT GTT TCT GCC AAA ACC AGG GAA AAA GTG GAA GCA GCC ATG GCA GAG CTG AAT TAC ATT CCC AAC AGA GTG GCA CAA CAA CTG GCA GGC AAA CAG AGC TTG CTG ATT GGA GTT GCC ACC TCC AGT CTG GCC CTG CAT GCA CCA TCT CAA ATT GTG GCA GCC ATT AAA TCT AGA GCT GAT CAA CTG GGA GCC TCT GTG GTG GTG TCA ATG GTA GAA AGA AGT GGA GTT GAA GCC TGT AAA GCT GCA GTG CAC AAT CTT CTG GCA CAA AGA GTC AGT GGG CTG ATC ATT AAC TAT CCA CTG GAT GAC CAG GAT GCC ATT GCT GTG GAA GCT GCC TGC ACT AAT GTT CCA GCA CTC TTT CTT GAT GTC TCT GAC CAG ACA CCC ATC AAC AGT ATT ATT TTC TCC CAT GAA GAT GGT ACA AGA CTG GGT GTG GAG CAT CTG GTT GCA TTG GGA CAC CAG CAA ATT GCA CTG CTT GCG GGC CCA CTC AGT TCT GTC TCA GCA AGG CTG AGA CTG GCC GGC TGG CAT AAA TAT CTC ACT AGG AAT CAA ATT CAG CCA ATA GCT GAA AGA GAA GGG GAC TGG AGT GCC ATG TCT GGG TTT CAA CAA ACC ATG CAA ATG CTG AAT GAG GGC ATT GTT CCC ACT GCA ATG CTG GTT GCC AAT GAT CAG ATG GCA CTG GGT GCA ATG AGA GCC ATT ACT GAG TCT GGG CTG AGA GTT GGT GCA GAT ATC TCG GTA GTG GGA TAC GAC GAT ACC GAA GAC AGC TCA TGT TAT ATC CCG CCG TTA ACC ACC ATC AAA CAG GAT TTT CGC CTG

CTG GGG CAA ACC AGC GTG GAC CGC TTG CTG CAA CTC TCT CAG GGC CAG GCG GTG AAG GGC AAT CAG CTG TTG CCA GTC TCA CTG GTG AAG AGA AAA ACC ACC CTG GCA CCC AAT ACA CAA ACT GCC TCT CCC CGG GCA TTG GCT GAT TCA CTC ATG CAG CTG **GCA CGA CAG GTT TCC CGA CTG GAA AGC GGG CAG** -3'

The operator was placed within the CMV\IE\Promoter region of the minimal CMV promoter 13 base pairs downstream of the TATA box. The following is the sequence of the CMV\IE\Promoter region including the Lsym operator in bold:

5'- CGC CCC GTT GAC GCA AAT GGG CGG TAG GCG TGT ACG GTG GGA GGT CTA TAT AAG CAG AGC TCG **AAT TGT GAG CGC TCA CAA TT**G AGC TCG TTT AGT GAA CCG TCA GAT C -3'

The plasmid map of pIRES Lsym YFP IRES EuLacmCh is given in Figure 2.4. The pSW plasmid expresses EuLac followed by the 2A sequence followed by either the GFP or Luciferase gene Figure 2.5. The entire cassette is flanked by ITR repeat regions for inclusion into AAV-viral capsids.

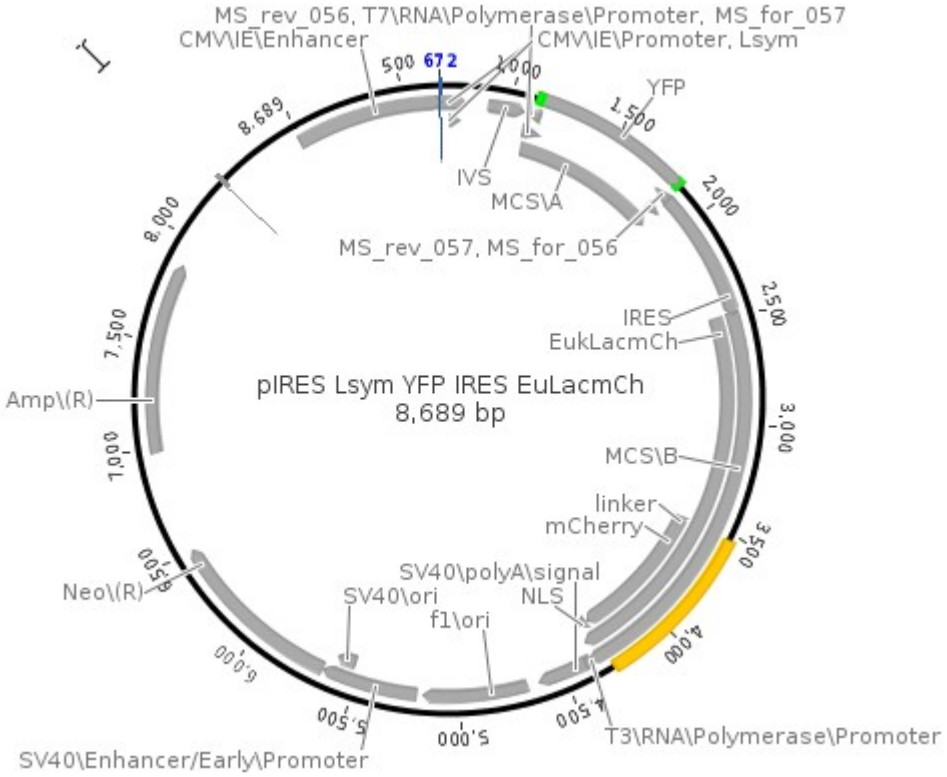


Figure 2.4. Plasmid map of the pIRES plasmid with the minimal CMVI promoter under control of the Lsym operator. The plasmid co-expresses YFP and the lac repressor fused to mCherry.

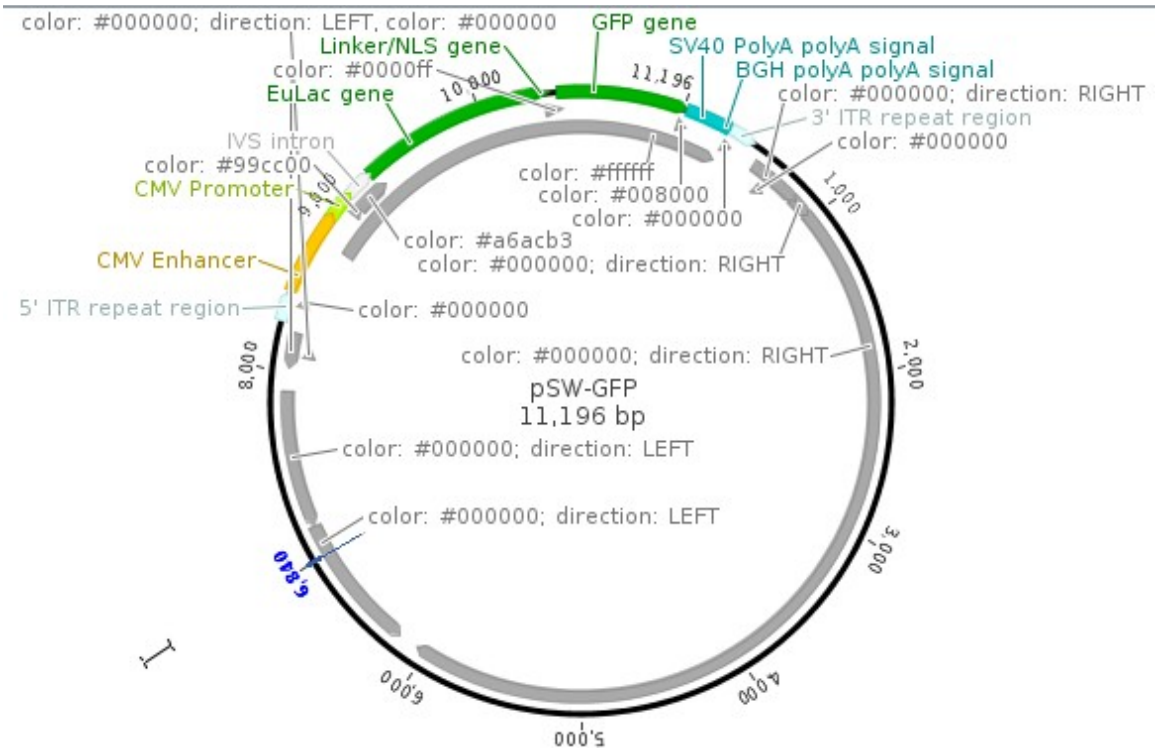


Figure 2.5. pSW-GFP plasmid map. The CMV promoter in this plasmid has one Lsym site in the same location as described above for the pIRES series plasmid.

Chapter 3:
Linking Mutation to Thermodynamics

(Adapted from Daber, Sochor and Lewis (2011) Thermodynamic Analysis of Mutant lac Repressors. *Journal of Molecular Biology. 409, 76-87.)*

3.1 Introduction

The study of the lac repressor, like many proteins, includes a large body of work wherein the repressor itself is mutated and its altered function analyzed. For the lac repressor, this history is complicated by the ever changing experimental techniques and knowledge of the repressor itself. So while a huge body of lac repressor mutations and their phenotypes is present in the literature, there is no direct link between how mutating specific parts of the lac repressor could potentially change the thermodynamic equilibrium and binding parameters specified by the MWC model of allostery. It is a much more stringent test of the MWC model if it can successfully explain not only the behavior of the wild type repressor, but also mutations to specific regions which should directly alter one or more of the thermodynamic parameters while leaving the others unchanged.

As discussed previously, the tetrameric lac repressor is doubly allosteric. First, the tetramer allosterically binds two DNA operators which has the effect of increasing the affinity of the repressor for operator DNA. Second, each dimer allosterically lowers its operator DNA affinity upon inducer binding. These two allosteric mechanisms work in concert to efficiently regulate the lac operon in *E. coli*. The vast majority of studies on the lac repressor take place with a full length tetrameric protein where the effect of both

mechanisms are present. It would be much more ideal to study the two processes in isolation in order to understand their individual effects.

It is relatively straightforward to isolate the allostery of DNA binding by the tetramer. One simply needs to study the properties of the repressor and operator DNA binding in the absence of any effector molecules. The natural lac operon includes three operators: O1 which is located within the promoter, O2 which is 401 base pairs downstream of O1, and O3 which is 92 base pairs upstream of O1. The operators are all pseudo-palindromic and have similar, yet distinct sequences and different affinities for the lac repressor. The last operator that is typically studied is a fully symmetric operator that was created by copying the left half of the O1 operator and removing the central G base, which we call Lsym (Sadler, Sasmor, & Betzt, 1983). The sequences of the four operators is given in Table 3.1.

Name			L6-4	L3-1	P0	R1-3	R4-6		
O1	5' -	AA TT	GTG	AGC	G	GAT	AAC	AA TT	-3'
O2	5' -	AA AT	GTG	AGC	G	AGT	AAC	AA CC	-3'
O3	5' -	GG CA	GTG	AGC	G	CAA	CGC	AA TT	-3'
Lsym	5' -	AA TT	GTG	AGC		GCT	CAC	AA TT	-3'

Table 3.1. The sequences of the three natural operators of the lac operon and the left symmetric operator.

What is immediately obvious is that positions 1-6 of the left side (L1-6) of all of the operators is identical. Affinity is essentially then a function of base pairing with the right side and the spacing (or lack thereof) created by the central guanine base at position 0 (P0).

An elegant study quantitatively established exactly how the lac repressor cooperatively binds its various operators (Oehler et al., 1994). They used a quantitative β-galactosidase assay to quantify repression by either dimeric or tetrameric lac repressor on promoters with various combinations of DNA operators. They first established that dimeric lac repressor and tetrameric lac repressor will repress a single operator in the O1 location (regardless of which operator) with the same efficiency. There is no inherent benefit to binding the first operator through tetramerization. They also established that O1 represses ~4 times more than O2 and ~200 times more than O3. Next they added a second downstream operator and showed an ~10 fold increase in repression for tetramer. There was no improvement for dimeric lac repressor, as expected. There is a similar benefit in placing additional operators upstream in the O3 position for the tetrameric lac repressor. The spacing between the two operators was later shown to be critical with a specific phase of approximately 11 base pairs between repression maximums (Müller et al., 1996). This phasing is understood to be related to the structure of DNA. DNA has the famous double helix structure with a minor and major groove. The helix-turn-helix of the lac repressor binds within the major groove of DNA making specific contacts with bases 4, 5 and 6 of the operator. Since the DNA twists with every additional base and the lac repressor dimers are linked through the coiled-coil tetramerization domain, specific spacing is required between the two operators such that the major groove for positions 4, 5 and 6 (for both the left and right halves) is facing the helix-turn-helix of each dimer of the lac repressor tetramer. Taking an MWC view of tetrameric lac repressor binding two operators requires considering multiple structures of not only the repressor, but also the DNA. Conceptually however, there are two combinations of conformations that matter. In one, only one dimer binds one operator with a mild affinity. In the second conformation, the tetramer and the DNA are twisted

such that an extremely stable pairing exists and both dimers bind both operators with very high affinity. The two sets of conformations are in equilibrium and adding a second operator within the DNA molecule at the correct spacing will alter the effective conformational equilibrium towards the higher DNA affinity state. The net effect is an approximately 10 or higher fold increase in repression.

While the allosteric binding of two DNA operators is useful for bacteria to more completely shut off the lac operon, it is confounding if we want to study the allosteric relationship between effector and DNA binding. If we want to isolate this effect, we must work only with the dimeric lac repressor and a single DNA operator. The next goal is to isolate perturbations to DNA binding from perturbations from effector binding. This is achieved by targeting residues in the lac repressor that contribute to either specific DNA binding or effector binding. A quantitative assay must be employed that measures DNA binding as a function of effector concentration (or vice versa). Finally, a comprehensive MWC allosteric model and solution to the model must be defined. Using all of these tools it should be possible to globally compare wild type lac repressor to mutants and measure how those mutations alter the thermodynamic parameters of the MWC model of allostery.

3.2 The MWC Model of the Dimeric Lac Genetic Switch

First we will establish an MWC model of the dimeric lac repressor binding to effectors and operator DNA (Figure 3.1).

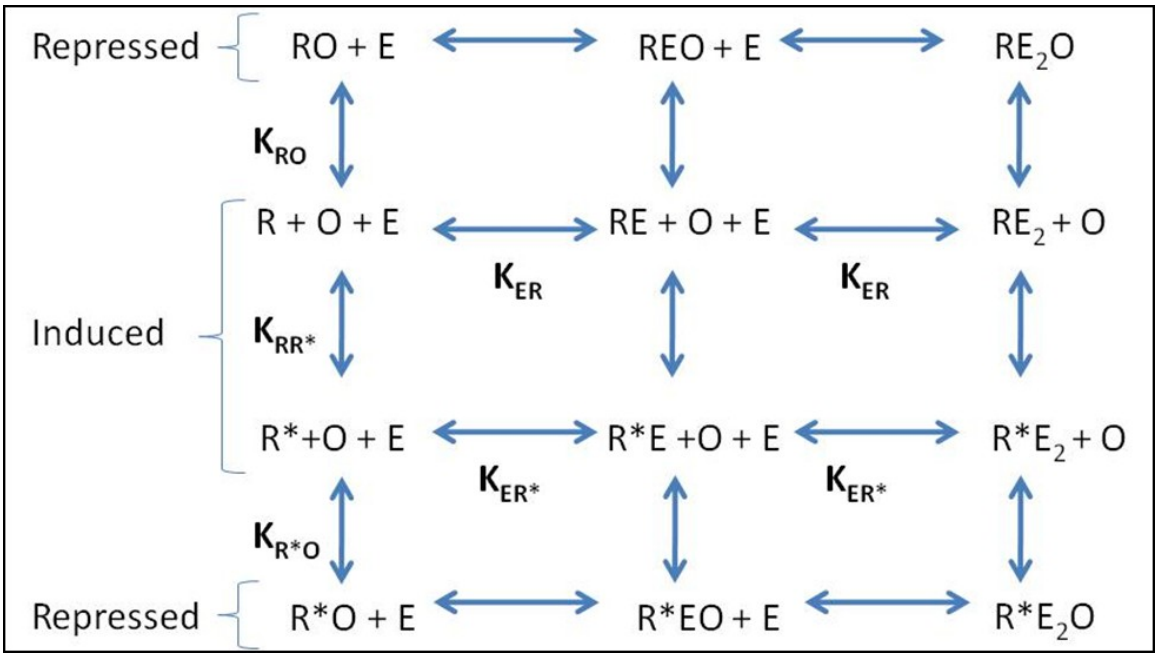


Figure 3.1. MWC model of dimeric lac repressor (R and R*) binding to a single DNA operator (O) and one or two effector molecules (E). The conformational equilibrium between the two conformations of the lac repressor (R and R*) is denoted by K_{RR^*}. The remaining parameters are the equilibrium affinity constants of the two states for operator (K_{RO} and K_{R^*O}) and effector (K_{RE} and K_{R^*E}).

The dimeric lac repressor exists in two primary conformations in this model, R and R*. The R conformation is designated as the higher DNA affinity conformation. Each lac repressor dimer can bind one operator DNA (O) and two effector molecules (E). The model assumes that the dimeric structures of the lac repressor are symmetric. Therefore for each given conformation the two effector binding sites are the same and have the same affinity of for effector. We assume that this symmetry is not broken upon binding of the first effector molecule, hence the binding of the second molecule to the second site has the same affinity as the first.

Experimentally, we would like to measure the output from a promoter regulated by the lac genetic switch. It is assumed that transcription by RNA polymerase from the promoter is linearly related to the occupancy of the DNA operator within the promoter by the lac repressor,

$$\text{transcription} \propto \frac{[O]}{[O]_{tot}} \tag{42}$$

In order to model experimental data, we need to compute the occupancy of the DNA operator in terms of the thermodynamic constants (K_{RR^*}, K_{RE}, K_{R^*E}, K_{RO}, and K_{R^*O}) and the total concentration of repressor, effector and operator ($[R]_{tot}$, $[E]_{tot}$, and $[O]_{tot}$).

We begin by defining the constants,

$$K_{RR^*} = \frac{[R^*]}{[R]} \tag{43}$$

$$K_{RE} = \frac{[RE]}{[R][E]} \tag{44}$$

$$K_{[R^*E]} = \frac{[R^*E]}{[R][E]} \tag{45}$$

$$K_{RO} = \frac{[RO]}{[R][O]} \tag{46}$$

$$K_{R^*O} = \frac{[R^*O]}{[R^*][O]} \tag{47}$$

We also need to define the total concentrations of operator in terms of the individual bound and conformational states,

$$\begin{aligned}[O]_{tot} = [O] + [RO] + 2[REO] + [RE_2O] + \\ [R^*O] + 2[R^*EO] + [R^*E_2O]\end{aligned} \tag{48}$$

The factors of 2 in from of the [REO] and [R*EO] states are due to the degeneracy of effector binding to the dimer. The effector could be bound to either the left monomer or right monomer which look identical, hence they are degenerate. To account for both species we need to multiply by two.

We can state the various bound species from Equation (48) in terms of the thermodynamic constants and the unbound concentrations of repressor, operator and effector ([R], [O], and [E], respectively),

$$[RO] = K_{RO}[R][O] \tag{49}$$

$$[REO] = K_{RE}K_{RO}[R][E][O] \tag{50}$$

$$[RE_2O] = K_{RE}^2 K_{RO}[R][E]^2[O] \tag{51}$$

$$[R^*O] = K_{RR^*}K_{R^*O}[R][O] \tag{52}$$

$$[R^*EO] = K_{RR^*}K_{R^*E}K_{R^*O}[R][E][O] \tag{53}$$

$$[R^*E_2O] = K_{RR^*}K_{R^*E}^2 K_{R^*O}[R][E]^2[O] \tag{54}$$

We then note the following equality for simplification,

$$(1+[E]K_{RE})^2 = 1+2[E]K_{RE}+[E]^2K_{RE}^2 \tag{55}$$

Plugging in Equations (49) - (54) and using the equality in Equation (55) into Equation (48),

$$O_{tot}=[O]+[O][R]K_{RO}(1+[E]K_{RE})^2+[O][R]K_{RR^*}K_{R^*O}(1+[E]K_{R^*E})^2 \quad (56)$$

We then rearrange put Equation (56) in the form which is proportional to transcription as in Equation (42),

$$\frac{[O]}{[O]_{tot}}=\frac{1}{1+[R](K_{RO}(1+[E]K_{RE})^2+K_{RR^*}K_{R^*O}(1+[E]K_{R^*E})^2)} \quad (57)$$

We now have the occupancy of the operator in terms of the thermodynamic parameters and the free species of repressor and effector. Unfortunately, we do not know the free species concentrations, only the total concentrations.

The experiment we would like to model however allows for some simplifications of Equation (57). We measured GFP production from a reporter plasmid which has a single operator within the promoter region. This plasmid has a very low copy number and hence the number of DNA operators in the cell is approximately 20. Meanwhile, the repressor is constitutively expressed from a second plasmid with also with a low copy number (~10-20) (Daber et al., 2009). This promoter, the natural promoter of the *lacI* gene in *E. coli*, has been estimated to make approximately 40 lac repressor dimers per promoter (Oehler et al., 1994), hence we have approximately 400-800 dimers per cell. Since a fraction of the operator itself is free, and the affinity of repressor for operator is high, we can assume that free operator concentration is approximately 0 with respect to the total repressor concentration,

$$[R]_{tot} \gg [O] \approx 0 \quad (58)$$

We then can approximate the total repressor concentration in terms of free species and thermodynamic binding constants,

$$[R]_{tot}=[R]+2[RE]+[RE_2]+[R^*]+2[R^*E]+[R^*E_2]+ [RO]+2[REO]+[RE_2O]+[R^*O]+2[R^*EO]+[R^*E_2O] \quad (59)$$

Using the assumption of Equation (58), we can say that all of the operator bound species of the repressor only marginally contribute to the total, which leads to the approximation that,

$$[R]_{tot}\approx[R]+2[RE]+[RE_2]+[R^*]+2[R^*E]+[R^*E_2] \quad (60)$$

$$[R]_{tot}\approx[R]((1+[E]K_{RE})^2+K_{RR^*}(1+[E]K_{R^*E})^2) \quad (61)$$

Solving for the free repressor concentration,

$$[R]\approx\frac{[R]_{tot}}{(1+[E]K_{RE})^2+K_{RR^*}(1+[E]K_{R^*E})^2} \quad (62)$$

We then can substitute Equation (62) into Equation (57)to arrive at,

$$\frac{[O]}{[O]_{tot}}=\frac{1}{1+\frac{[R]_{tot}(K_{RO}(1+[E]K_{RE})^2+K_{RR^*}K_{R^*O}(1+[E]K_{R^*E})^2)}{(1+[E]K_{RE})^2+K_{RR^*}(1+[E]K_{R^*E})^2}} \quad (63)$$

We then define the two following constants,

$$f=\frac{(1+[E]K_{RE})^2}{(1+[E]K_{RE})^2+K_{RR^*}(1+[E]K_{R^*E})^2} \quad (64)$$

$$s = \frac{(1+[E]K_{R^*E})^2}{(1+[E]K_{RE})^2 + K_{RR^*}(1+[E]K_{R^*E})^2} \tag{65}$$

$$r = [R]_{tot} K_{RO} \tag{66}$$

$$r^* = [R]_{tot} K_{RR^*} K_{R^*O} \tag{67}$$

$$\frac{[O]}{[O]_{tot}} = \frac{1}{1+rf+r^*s} \tag{68}$$

The final assumption required is regarding the free effector concentration, [E]. We assume that free effector concentration does not significantly differ from the total effector concentration. The error in this assumption will decrease as total effector concentrations comes to greatly exceed total repressor and total operator concentrations, which occurs in our experimental system at full induction. We then approximate Equations (64) and (65),

$$f \approx \frac{(1+[E]_{tot}K_{RE})^2}{(1+[E]_{tot}K_{RE})^2 + K_{RR^*}(1+[E]_{tot}K_{R^*E})^2} \tag{69}$$

$$s \approx \frac{(1+[E]_{tot}K_{R^*E})^2}{(1+[E]_{tot}K_{RE})^2 + K_{RR^*}(1+[E]_{tot}K_{R^*E})^2} \tag{70}$$

Using these assumptions we have finally stated the occupation of the operator solely in terms of total repressor and effector concentrations and the thermodynamic constants. We will use Equations (68), (69), and (70) to understand our experimental data in this chapter in the frame of the MWC model of allostery.

The phenotypic properties of the lac genetic switch can also be understood in terms of this model. We define three phenotypic properties: leakiness is the expression without effector (basal expression), dynamic range is the difference between maximal induction or co-repression and the leakiness, and $[E]_{50}$ is the concentration of effector needed for half maximal induction or co-repression.

To model leakiness, we set $[E]_{tot} = 0$ and find,

$$f = s = \frac{1}{1 + K_{RR^*}} \tag{71}$$

$$\text{Leakiness}(L) = \frac{[O]}{[O]_{tot}}([E]_{tot} = 0) = \frac{1}{1 + \frac{r + r^*}{1 + K_{RR^*}}} \tag{72}$$

Similarly we can find dynamic range by measuring the operator occupancy at saturating effector concentration ($[E]_{tot} \rightarrow \infty$) and subtracting leakiness. At very large effector concentrations, the terms multiplied by $[E]_{tot}$ are much larger than 1, hence,

$$f([E]_{tot} \rightarrow \infty) \approx f \approx \frac{([E]_{tot} K_{RE})^2}{([E]_{tot} K_{RE})^2 + K_{RR^*}([E]_{tot} K_{R^*E})^2} = \frac{K_{RE}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2} \tag{73}$$

$$s([E]_{tot} \rightarrow \infty) = \frac{K_{R^*E}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2} \tag{74}$$

$$\frac{[O]}{[O]_{tot}}([E]_{tot} \rightarrow \infty) = \frac{1}{1 + \frac{r K_{RE}^2 + r^* K_{R^*E}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2}} \tag{75}$$

$$\text{Dynamic Range}(D) = \frac{[O]}{[O]_{tot}}([E]_{tot} \rightarrow \infty) - L =$$

$$D = \frac{1}{1 + \frac{r K_{RE}^2 + r^* K_{R^*E}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2}} - \frac{1}{1 + \frac{r + r^*}{1 + K_{RR^*}}} \tag{76}$$

Finally, the $[E]_{50}$ is calculated by solving for the effector concentration at half maximal induction which is equal to the leakiness plus half of the dynamic range. We therefore need to solve the following equation for $[E]_{tot}$,

$$L + D/2 = \frac{1}{1 + rf + r^*s} \tag{77}$$

A solution for $[E]_{50}$ exists if we assume that the R* conformation has effectively very low affinity for DNA (relative to r) and we set r* = 0. In that case we have,

$$L + D/2 = \frac{1}{1 + \frac{r(1 + [E]_{50} K_{RE})^2}{(1 + [E]_{50} K_{RE})^2 + K_{RR^*}(1 + [E]_{50} K_{R^*E})^2}} \tag{78}$$

We then define the following constant and use it to re-arrange Equation (78),

$$\frac{1}{\varphi} = \frac{1}{L + \frac{D}{2}} - 1 \tag{79}$$

$$\frac{1}{\varphi} = \frac{r(1+[E]_{50}K_{RE})^2}{(1+[E]_{50}K_{RE})^2 + K_{RR^*}(1+[E]_{50}K_{R^*E})^2} \tag{80}$$

We then multiply the denominators on both sides of the Equation and find,

$$\begin{aligned}&[E]_{50}^2(K_{RE}^2(\varphi r - 1) - K_{RR^*}K_{RE}^2) + \\ &[E]_{50}(2K_{RE}(\varphi r - 1) - 2K_{RR^*}K_{R^*E}) + \varphi r - K_{RR^*} - 1 = 0\end{aligned} \tag{81}$$

If we then define the following constants we can easily solve for $[E]_{50}$,

$$A = K_{RE}^2(\varphi r - 1) - K_{RR^*}K_{RE}^2 \tag{82}$$

$$B = 2K_{RE}(\varphi r - 1) - 2K_{RR^*}K_{R^*E} \tag{83}$$

$$C = \varphi r - K_{RR^*} - 1 \tag{84}$$

$$[E]_{50} = \frac{-B \pm \sqrt{(B^2 - 4AC)}}{2A} \tag{85}$$

Equation (85) of course only holds when r* = 0 and $[E]_{50}$ must be real and greater than zero. Now that we have a solution to the MWC model, we can generate mutants of the lac repressor and test them using the solution.

3.3 Mutations to the DNA Recognition Domain

The DNA recognition domain of the lac repressor has been shown to be positions 1, 2 and 6 of the DNA recognition helix which corresponds to positions tyrosine 17 (Y17), glutamine 18 (Q18), and arginine 21 (R21). The recognition helix is aligned within the

major groove of DNA and makes specific contacts with positions 4, 5, and 6 of the lac DNA operators.

The recognition domain of the lac repressor has been an important target for mutagenesis in order to explore how proteins use helix-turn-helix motifs to specifically bind DNA. First it was shown that specificity could be changed by combining Y17, Q18, and R22 mutations with mutations to the operator at positions 4, 5 and 6 (Sartorius et al., 1989). Next, a mutant library was created where all three positions were mutated to all 20 amino acids creating approximately 8000 mutant repressors. This library was screened against the O1 and Lsym operators with the goal of finding tighter binding mutants (Daber & Lewis, 2009). They found 33 different mutants that repressed transcription through O1, however only with mutations to positions 17 and 18; any mutation to position 22 failed to repress. Interestingly, only one mutant, Q18M, was found to bind O1 tighter than wild type repressor. Finally, they found a relationship between the repression of these mutants and dynamic range; tighter binding could lead to less leakiness of expression but it also decreased the maximal inducibility. This led to the hypothesis that DNA specificity mutants primarily change the lac repressor-DNA operator affinity (K_{RO}).

We sought to test this hypothesis *in vivo* by globally modeling the transcriptional regulation phenotype of dimeric lac repressor with and without mutations to the DNA specificity residues.

Three point mutations were selected for analysis: Y17I which is leakier in regulating O1 than wild type, Q18A which is also leakier, and Q18M which is less leaky in regulating O1. All three point mutations were separately cloned using inverse PCR

mutagenesis into the pLacI plasmid which constitutively expresses dimeric lac repressor (11 C-terminal residues are truncated) from the natural constitutive promoter of the *lacI* gene from *E. coli*.

The wild type repressor and three point mutants were tested using the prokaryotic GFP regulation assay (Figure 3.2).

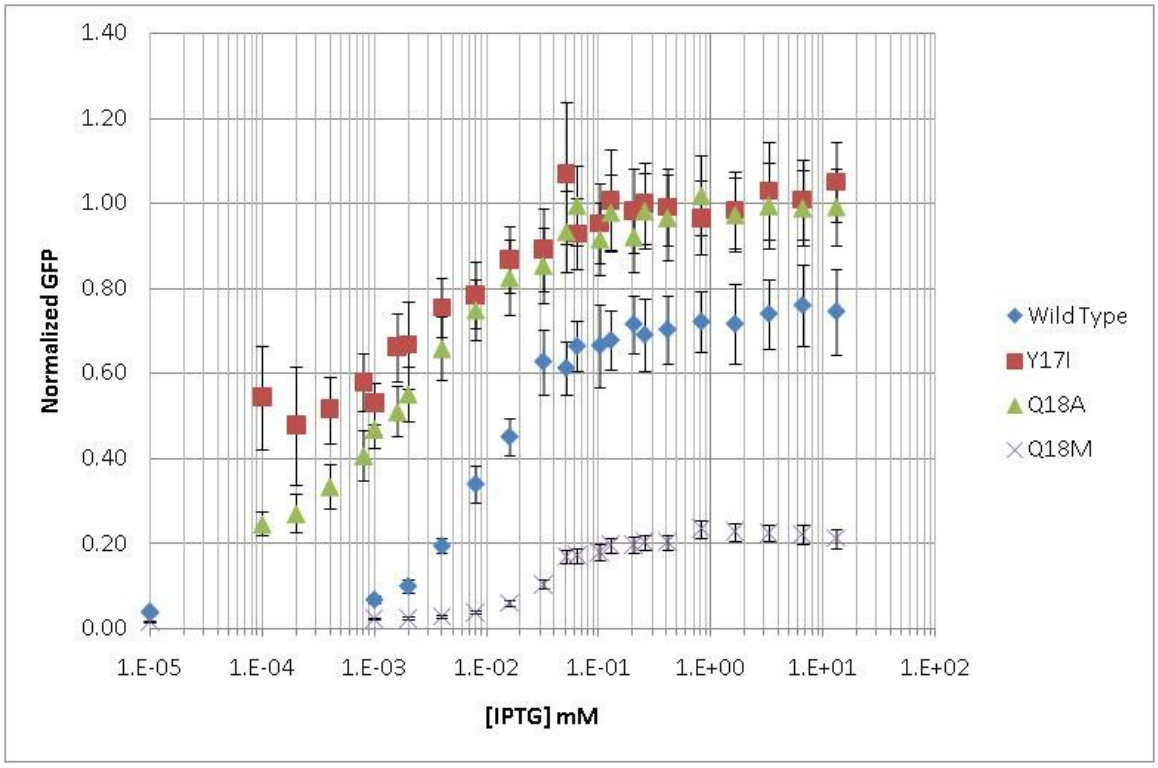


Figure 3.2. **Regulation of GFP by headpiece mutants of the lac repressor and O1.** Positions 17 and 18 both reside within the recognition helix of the N-terminal headpiece of the lac repressor and are known to be important for specificity in DNA binding. Point mutations Y17I and Q18A both are leakier (higher expression without IPTG) than wild type lac repressor whereas Q18M is less leaky. The dynamic range is decreased in all mutants with respect to wild type lac repressor.

Figure 3.2 shows that all three point mutants can still regulate GFP expression using the O1 operator. Y17I is the leakiest mutant (~50% leakiness) and it essentially induces to constitutive expression of the promoter at [IPTG] ≥ 0.1 mM. Q18A is marginally less leaky than Y17I (~25% leakiness) and also reaches constitutive levels of

expression upon induction with [IPTG] ≥ 0.1 mM. Q18M is approximately half as leaky (~2% leakiness) than wild type (~4% leakiness) and its maximal induction is severely limited.

It is of note that the dynamic range of the genetic switch appears to be limited by the upper and lower bounds. We can model this behavior using our MWC model of the lac genetic switch. Figure 3.3 shows simulated lac genetic switch data where only the affinity for the R state of DNA is changed ($r = [R]_{tot}K_{RO}$ is altered). Figure 3.3.B clearly demonstrates the limiting nature of the upper and lower bounds on dynamic range; a median value of r produces a minimally leaky switch that maximally induces. This is precisely the relationship we see in the experimental data of Figure 3.2 where the only mutation is to the specificity region of the N-terminal DNA binding domain. We hypothesize therefore that point mutations to this regions specifically change the affinity of repressor for DNA (K_{RO} and K_{R^*O}). It of course is possible that the point mutations change the total concentration of repressor ($[R]_{tot}$), however given the current experiment we have no means of measuring the repressor concentration. This is a limitation of this study and it will be addressed later in this chapter.

We globally modeled the three headpiece mutants of the lac repressor dimer along with the wild type by assuming the only difference could be in the K_{RO} thermodynamic parameter. We therefore performed regression between simulated curves and experimental curves looking for thermodynamic parameters that fit the data. For the four data sets there are 11 fit parameters (r^{WT}, r^{Y17I}, r^{Q18A}, r^{Q18M}, r^{*WT}, r^{*Y17I}, r^{*Q18A}, r^{*Q18M}, K_{RE}, K_{R^*E}, K_{RR^*}) however it was previously found that for wild type dimeric lac repressor K_{R^*O} is effectively zero (Daber et al., 2009) so we will set all of the r* parameters to zero and fit the remaining 7 parameters.

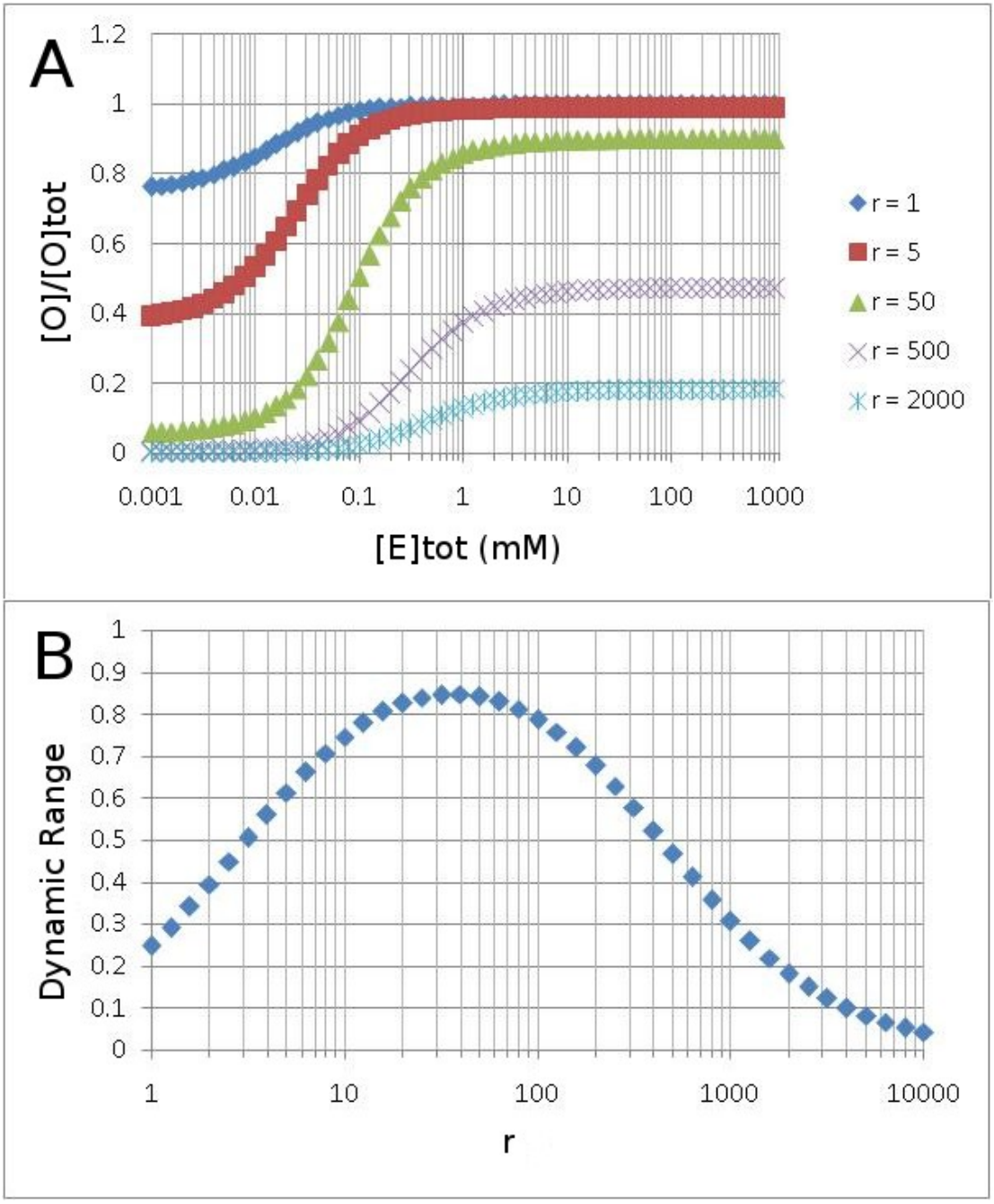


Figure 3.3. **Simulated lac genetic switch data as a function of changing r.** The following thermodynamic parameters were used: r* = 0, K_{RE} = 4 mM^{-1}, K_{R*E} = 60 mM^{-1}, K_{RR*} = 2. **(A)** Simulated induction curves with an effector shows that lower r (lower R_{tot} and/or K_{RO}) makes a leakier switch that more fully induces, whereas higher r leads to tighter and tighter switches that cannot fully turn on. **(B)** The dynamic range shows that a happy medium exists when altering r where the switch is minimally leaky and maximally inducible.

There is another complication to the modeling that also needed to be addressed before thermodynamic parameters could be measured. Figure 3.4 shows the effect of altering K_{RR^*} on the induction of the genetic switch.

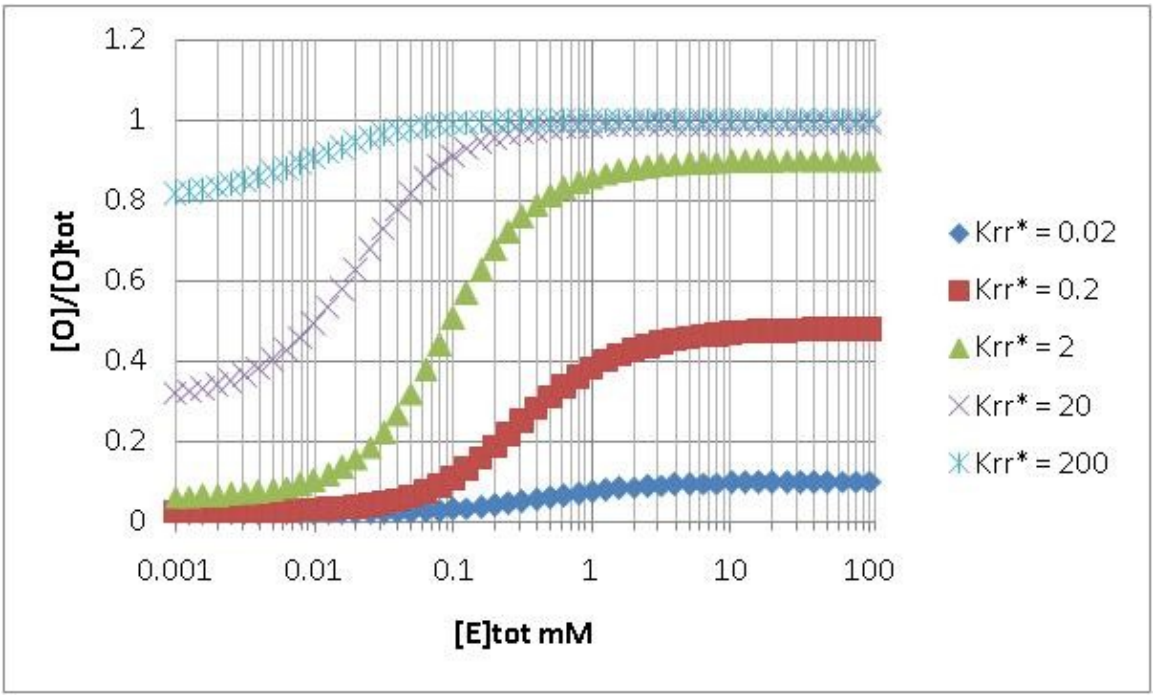


Figure 3.4. **Simulated lac genetic switch data as a function of changing r.** The following thermodynamic parameters were used: r=50, r* = 0, K_{RE} = 4 mM^{-1}, K_{R^*E} = 60 mM^{-1}. Changes to K_{RR^*} mirror changes to r because it effectively makes more or less of the high affinity conformation, R, of the protein. Since the r parameter also contains the repressor concentration, these two parameters become linked.

Changes to K_{RR^*} on the induction curve of the lac genetic switch look very, very similar to changes to K_{RO}. From the model alone, it is unclear whether the headpiece mutants are altering DNA affinity or the conformational equilibrium. From a structural perspective, the location of the point mutations in the recognition domain of the DNA binding domain strongly indicates the former, however it would be better if we could eliminate the possibility that K_{RR^*} is changing. Analyzing the limits of the MWC equilibrium actually provides a mechanism to isolate K_{RR^*} from r, and hence investigate which may be changing for this data. We start by re-defining Equation (75),

$$y_2 = \frac{[O]}{[O]_{tot}}([E]_{tot} \to \infty) = \frac{1}{1 + \frac{r K_{RE}^2 + r^* K_{R^*E}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2}} \tag{86}$$

Looking at the formula for leakiness, Equation (72), we can isolate r in both equations pretty easily (assuming r* = 0),

$$\frac{1}{y_2} - 1 = \frac{r K_{RE}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2} \tag{87}$$

$$\frac{1}{L} - 1 = \frac{r}{1 + K_{RR^*}} \tag{88}$$

The ratio of Equations (87) and (88) then eliminates the r parameter and gives us an experimental measure that isolates K_{RR^*},

$$z = \frac{\frac{1}{y_2} - 1}{\frac{1}{L} - 1} = \frac{K_{RE}^2 + K_{RR^*} K_{RE}^2}{K_{RE}^2 + K_{RR^*} K_{R^*E}^2} \tag{89}$$

Figure 3.3.A and Figure 3.4 show the same resultant phenotype from changing r and K_{RR^*} separately, however Figure 3.5 shows a large difference in the z parameter. Obviously, there is no effect when altering r, as z is independent of r. There is however a significant decrease in z when you increase K_{RR^*}.

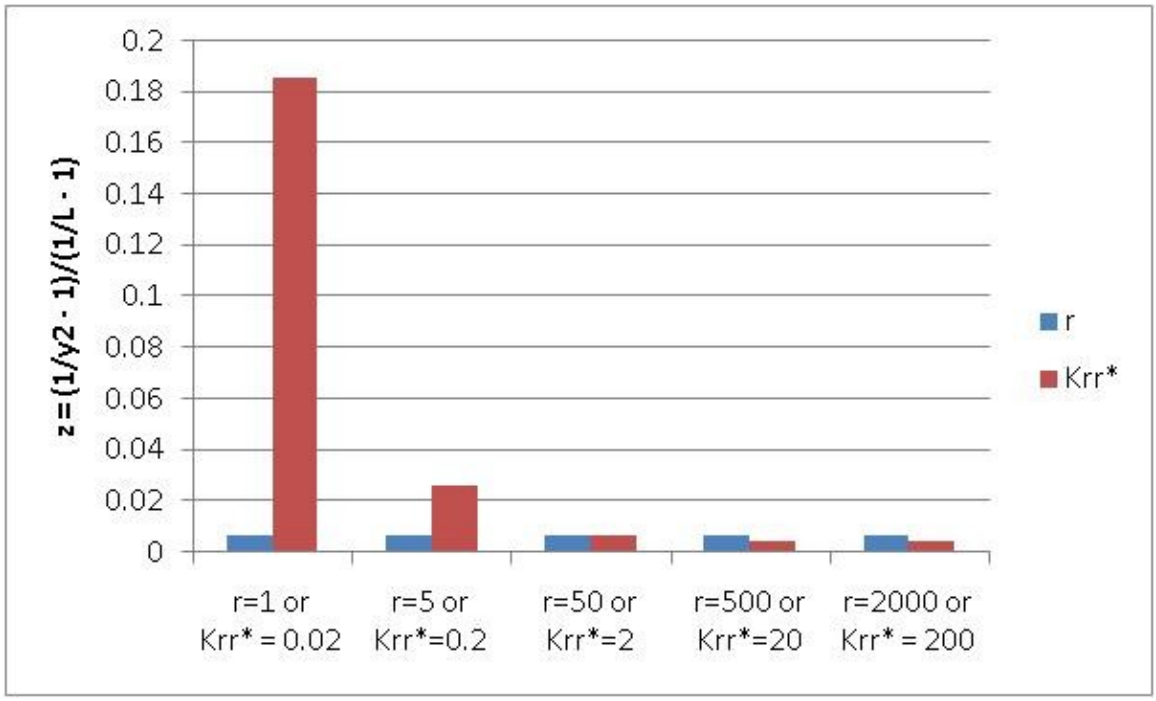


Figure 3.5. **Calculation for the z parameter as a function of either r or Krr*.** Thermodynamic parameter values for the blue bars (where r changes) were: r* = 0, K_{RE} = 4 mM^{-1}, K_{R*E} = 60 mM^{-1}, K_{RR*} = 2. Since z is independent of r, there is no change. The thermodynamic parameter values for the red bars (where K_{RR*} changes) were: r=50, r* = 0, K_{RE} = 4 mM^{-1}, K_{R*E} = 60 mM^{-1}. As K_{RR*} increases (more repressor in the R* conformation) the z parameter decreases.

Figure 3.6 shows the z values for wild type and the three headpiece mutants. Y17I and Q18A are approximately unchanged from wild type, while Q18M has a much higher value. The z parameter calculation is especially sensitive to small changes in small numbers due to the inversion, hence small changes in the leakiness measurement will have large effects on the experimental calculation of z. The question is then: is the difference between wild type and Q18M z due to experimental error in measuring small signals or does Q18M alter the K_{RR*} parameter. We can address this question by fitting the Q18M data set while changing either r or K_{RR*}, calculating z and seeing if the magnitude of the z change is comparable with experiment.

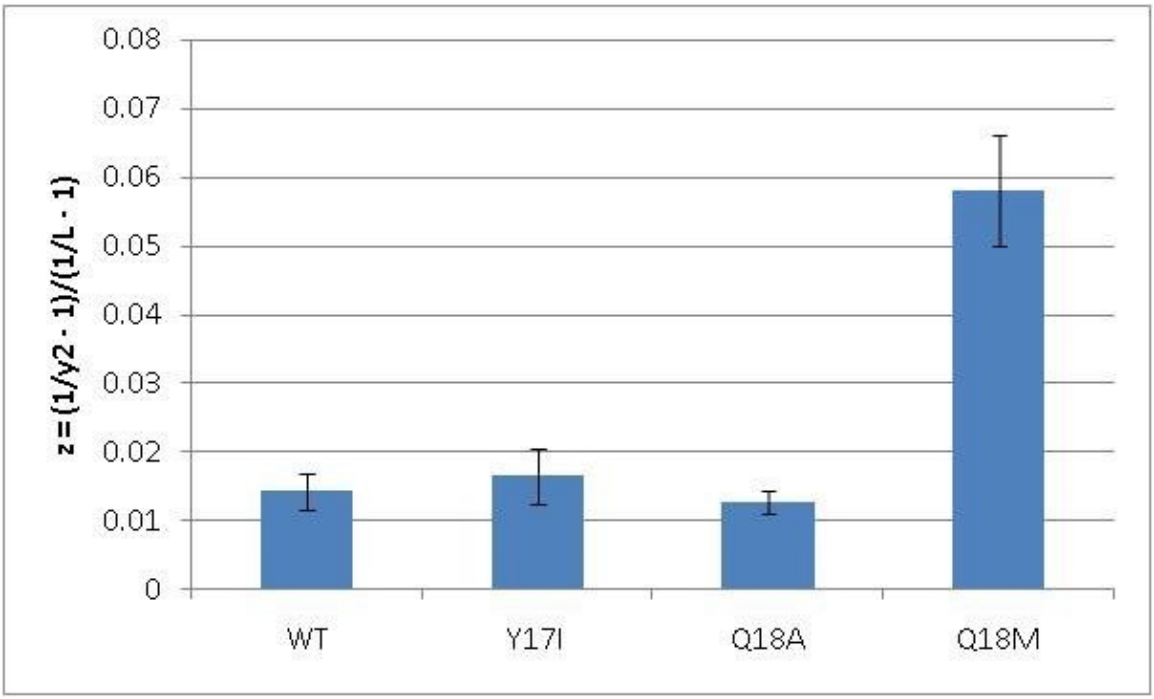


Figure 3.6. **Calculated z values from wild type and the three headpiece mutants of the lac represor.** Y17I and Q18A have approximately the same z value as WT, indicating the altered phenotype is a result of changing the r parameter. Q18M has a higher value which could be due to changes in K_{RR^*}.

The final complication in fitting arises from the non-uniqueness of solution due to the compensatory nature of r and K_{RR^*}. We will bootstrap our value of r based upon the findings of Daber, Sharp, and Lewis who determined that r = 150 +- 50 for a hetero-dimeric form of the lac repressor (Daber et al., 2009). Their finding was based upon fitting z parameters for their experimental data.

We then globally fit the four experimental curves given r^{WT} = 150 fitting for three parameters common to all four experiments (K_{RR^*}, K_{RE}, and K_{R^*E}) and with one independent parameter for each of the headpiece curves (r^{Y17I}, r^{Q18A}, and r^{Q18M}). Table 3.2 shows the fit parameters. We then refit the Q18M looking for altered K_{RR^*} given: r=150, K_{RE} = 60 mM^{-1}, K_{R^*E} = 500 mM^{-1}. This fit is given in Table 3.3.

Construct	r	K_{RR^*}	K_{RE} (mM^{-1})	K_{R^*E} (mM^{-1})	Leakiness	Dynamic Range	$\log_{10}$ $[E]_{50}$ (mM)
Wild Type	150 ± 50	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.044	0.7	-1.94
Y17I	8.6 ± 0.07	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.44	0.5	-2.66
Q18A	18.0 ± 0.2	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.28	0.7	-2.49
Q18M	1404 ± 1	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.005	0.2	-1.56

Table 3.2. Fit values where only the r parameter was altered between the 4 different experimental curves.

Construct	r	K_{RR^*}	K_{RE} (mM^{-1})	K_{R^*E} (mM^{-1})	Leakiness	Dynamic Range	$\log_{10}$ $[E]_{50}$ (mM)
Wild Type	150 ± 50	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.044	0.7	-1.94
Q18M	150 ± 50	0.65 ± 0.1	60 ± 0.2	500 ± 5	0.005	0.2	-1.56

Table 3.3. Fit values where Q18M were re-fit looking for changes in K_{RR^*} instead of changes in r.

We then calculated z and compared to experimentally calculated z to determine if Q18M primarily changes r or K_{RR^*}. Table 3.4 shows the results and the difference between theoretical and experimental z values.

Construct	z Predicted	z Experimental	$\lvert\Delta z\rvert/z$
Wild Type	0.017 ± 0.0003	0.014 ± 0.003	0.15
Y17I	0.017 ± 0.0003	0.016 ± 0.004	0.02
Q18A	0.017 ± 0.0003	0.013 ± 0.002	0.25
Q18M (Table 3.2)	0.017 ± 0.0003	0.058 ± 0.008	2.45
Q18M (Table 3.3)	0.040 ± 0.007	0.058 ± 0.008	0.47

Table 3.4. **Predicted z values versus experimental values of z.** The last column shows the difference relative experimental z where lower values are better.

The normalized difference values in Table 3.4 show that Y17I and Q18A phenotypes are well explained by only changing the r parameter whereas the Q18M mutant is best explained by altering the K_{RR^*} parameter.

It would be expected that the Y17I and Q18A mutations primarily alter K_{RO} due to their location within the DNA recognition helix of the lac repressor, however these mutations could also alter the total concentration of the repressor, $[R]_{tot}$. It is unclear how Q18M could alter the conformational equilibrium parameter but we also do not have a good structural picture of the N-terminal domain of the lac repressor in the effector bound conformation. Crystal structures do not show density (likely indicating mobility of this region) for the N-terminal domain when IPTG is bound in the effector binding pocket. Q18M makes the conformational equilibrium go from greater than 1 (favoring the low operator DNA affinity conformation R*) to less than 1 (favoring the R conformation). This mutation takes a large polar residue (glutamine) and changes it to a still large, more apolar residue (methionine). Perhaps the solvation of residue 18 between the R and R* states plays a significant role in the very close energetic balance between these two states, significant enough to alter the conformational equilibrium. Central to this idea, however, is that the methionine can still make good specific contacts within the major groove of DNA.

Figure 3.7 shows the experimental data with the best fit MWC curves for wild type lac repressor and the three head piece mutants.

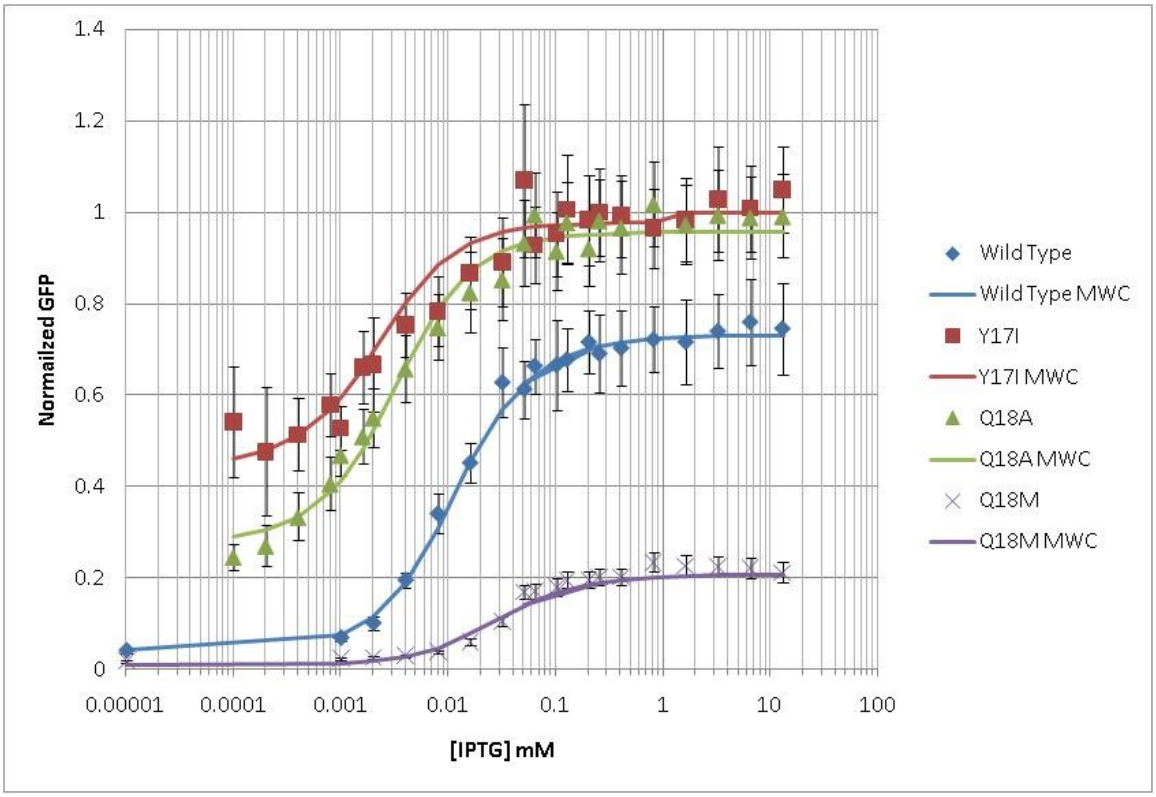


Figure 3.7. **Best fit values for wild type, dimeric lac repressor and three point mutations to the DNA recognition domain.** The MWC model is very accurate at reproducing the experimental curves.

3.4 Mutations to the Effector Binding Pocket

The effector binding pocket of the lac repressor includes a large number of residues that both make specific contacts with the repressor and which are important for making secondary, stabilizing contacts. The residues arginine 197 (R197), asparagine 246 (N246), and aspartic acid 274 (D274) form hydrogen bonds to the O2 and O3 hydroxyls of sugar effector molecules. As these are essential for effector binding, they are poor candidates for mutation if we are trying to change the thermodynamic binding parameters without completely destroying binding. Much like our approach in the DNA binding domain, the residues that are responsible for specificity in effector binding were instead targeted. Effectors are known that are inducers, co-repressors and neutral and

the difference between the molecules is on the substituent groups on the first and sixth carbons of the sugar ring (Barkley, Riggs, Jobe, & Bourgeois, 1975). These substituent groups make contacts with residues 79, 148, 161, 291, 293 and 296. These residues were targeted by making 6 libraries where each site was individually mutated to all 20 amino acids (120 mutants screened) and their *in vivo* GFP regulatory phenotype was recorded (Daber et al., 2011).

Here we analyze a subset of these mutations that have phenotypes which are of particular interest. Again, we would like to globally fit the GFP regulation of these mutants with the goal of linking effector pocket point mutations to changes in the effector binding parameters (K_{RE} and K_{R^*E}). It is also possible that point mutations in the effector binding pocket could alter the conformational equilibrium (K_{RR^*}) (Figure 3.4). The two effector binding parameters can change in two significant ways: first, the magnitudes could change (stronger or weaker binding) and second, the relative magnitudes between the two affinities could change. Here we define the relative magnitude as,

$$X = \frac{K_{R^*E}}{K_{RE}} \tag{90}$$

We can then use the MWC model to predict how these two changes alter the regulatory phenotype. Changes in the magnitudes of the binding affinities cause either a left or right shift in the curve, which is reflected as a change to the half maximal effector concentration, $[E]_{50}$ (Figure 3.8). Lower magnitudes of K_{RE} and K_{R^*E} increase $[E]_{50}$ (more effector is needed to bind) whereas higher affinities decrease $[E]_{50}$ (less effector is needed). Alternatively, changing the ratio between the two effector binding affinities changes the dynamic range of the switch (Figure 3.9). Larger X makes a more inducible

switch because more repressor goes to the R* conformation upon addition of effector. The effector binding affinities, predictably, do not change the leakiness. Any changes in leakiness must be due to either changes in r (likely due to changes in $[R]_{tot}$) or K_{RR^*}.

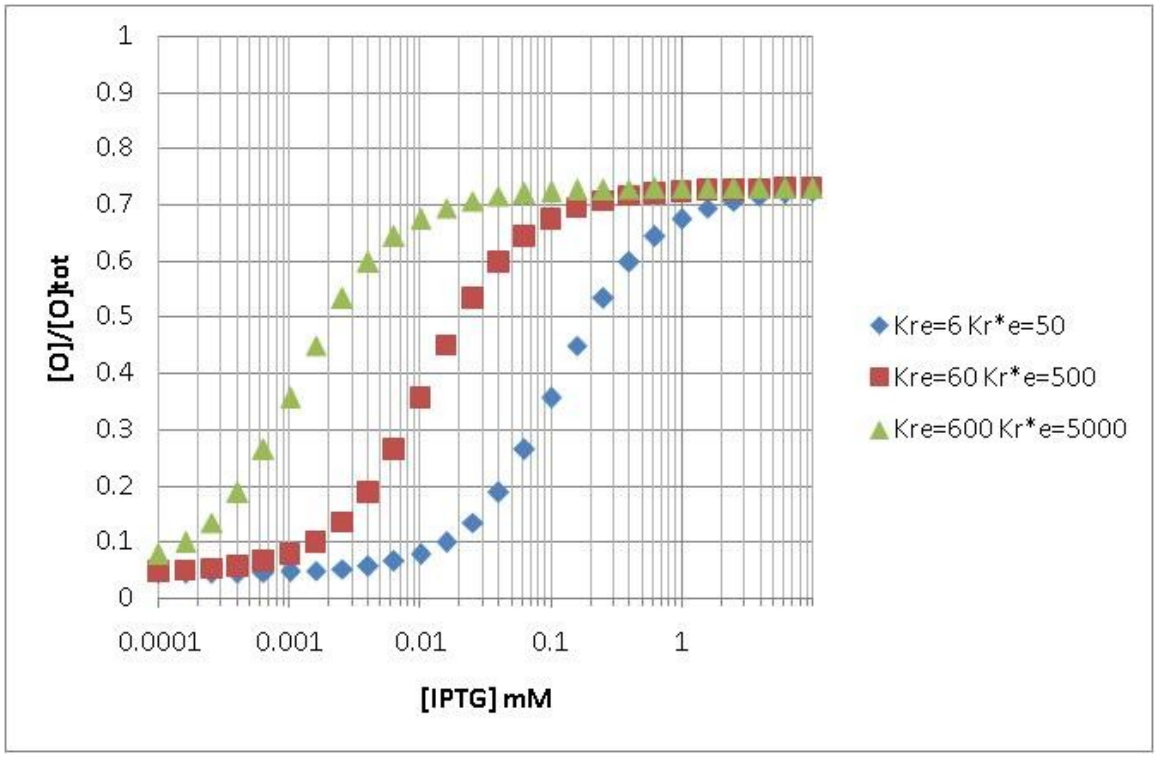


Figure 3.8. **Effect of changing the magnitudes of the effector affinity on MWC gene regulation.** Leakiness and dynamic range are unchanged. There is however a significant effect on the midpoint of maximal induction, $[E]_{50}$. Higher affinity to effector decreases the $[E]_{50}$, and vice versa.

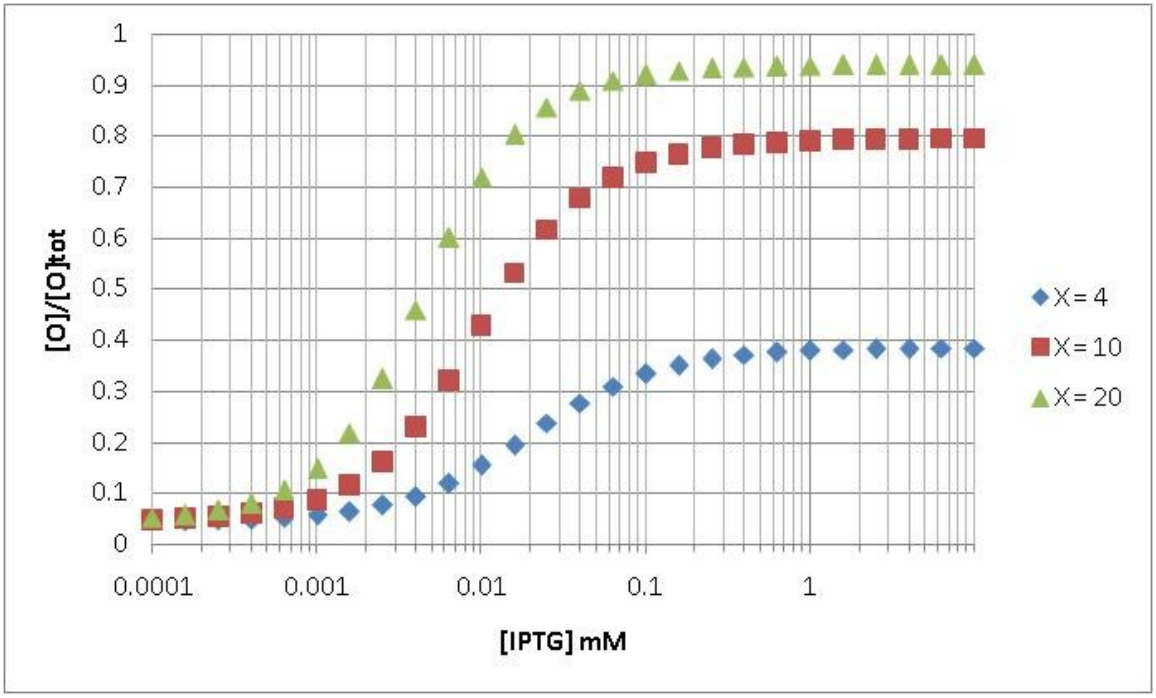


Figure 3.9. **Effect of changing the relative magnitudes of the effector affinities (X) on MWC gene regulation.** Leakiness is unchanged. There is however a significant increase in dynamic range as X increases due to effector more efficiently moving lac repressor to the R* conformation. $[E]_{50}$ is marginally affected by changing the ratio X.

We then want to select effector pocket mutants that change the gene regulatory phenotype in various ways with the goal of globally linking these phenotypic changes to the underlying thermodynamic parameters. Our hypothesis is that we can model every effector pocket mutant phenotype only through changes to K_{RE}, K_{R^*E}, and K_{RR^*}.

Table 3.5 shows the selected point mutants along with an effector that gives rise to a particular phenotype. This list includes many interesting phenotypes: tighter and weaker IPTG binding, more ideal on-off genetic switches, and even inverted switches which go from on to off with a dynamic range comparable to wild type lac repressor.

Three mutations to phenylalanine 161 (F161) were found to have a more ideal on-off behavior (Figure 3.10). All three mutants are less leaky than wild type and induce

to a higher level with IPTG. We hypothesize that this region alters both the conformational equilibrium and the effector binding.

Mutation	Effector	Phenotype
Wild Type	IPTG	Canonical inducing
Wild Type	ONPG	Canonical co-repressing
L148D	ONPG	Inverted genetic switch
L148W	ONPG	Inverted genetic switch
F161N	IPTG	Better on-off switch
F161S	ONPG	Weakly inducing with canonical co-repressor
F161T	IPTG	Better on-off switch
F161W	IPTG	Better on-off switch
Q291I	IPTG	Decreased leakiness
Q291K	IPTG	Weaker IPTG binding
Q291M	IPTG	Stronger IPTG binding
Q291R	IPTG	Non-responsive to IPTG
L296W	IPTG	Less leaky but weaker IPTG binding
L296W	ONPG	Weakly inducing with canonical co-repressor

Table 3.5. Point mutations to the effector binding pocket of the lac repressor and their phenotypic response with specific effectors.

Four point mutations to glutamine 291 (Q291) primarily change maximal expression (higher for Q291M and much, much lower for Q291R) or the mid-point of induction (lower effector needed for Q291M and higher effector needed for both Q291I and Q291K) (Figure 3.11). We hypothesize that this region primarily alters the effector binding affinities in both magnitude and in relative affinity between the two conformational states of the lac repressor.

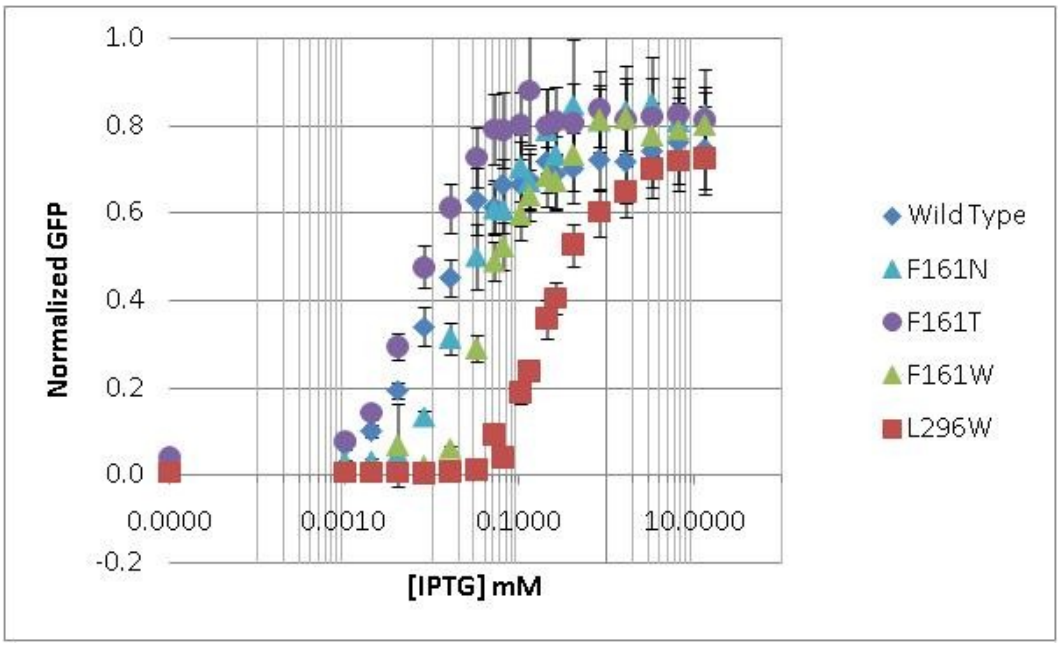


Figure 3.10. **IPTG induction profiles of lac repressor dimer and point mutants in the effector binding domain.** Mutations to F161 appear to create more ideal repressors. L296W is still inducible with IPTG but at a much, much lower affinity than wild type.

Two point mutations to leucine 148 (L148) have the the notable phenotype of being inverted switches with the effector ONPG (Figure 3.12). They exhibit much leakier expression and co-repress very effectively. We hypothesize that these mutations alter the conformational equilibrium while ONPG binding remains similar to wild type lac repressor which normally co-represses with ONPG. Alternately, two point mutations, F161S and L296W, are seen to induce with ONPG. Of particular interest is L296W which also induces with IPTG (Figure 3.10). Somehow, wild type is co-repressable with ONPG and inducible with IPTG, whereas L296W is inducible with both. We hypothesize that these two point mutations alter the relative effector binding affinities of the two conformational states of the lac repressor.

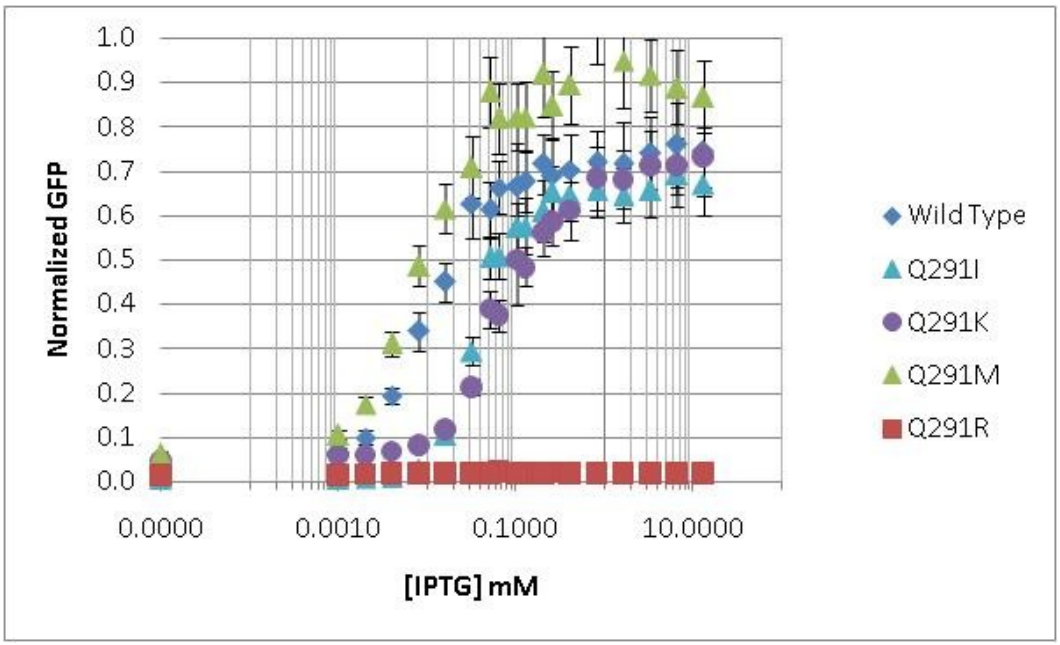


Figure 3.11. **IPTG induction profiles of lac repressor dimer and point mutants in the effector binding domain.** Mutations to Q291 alter the affinity of the lac repressor for effector, both higher and lower and between the two conformations.

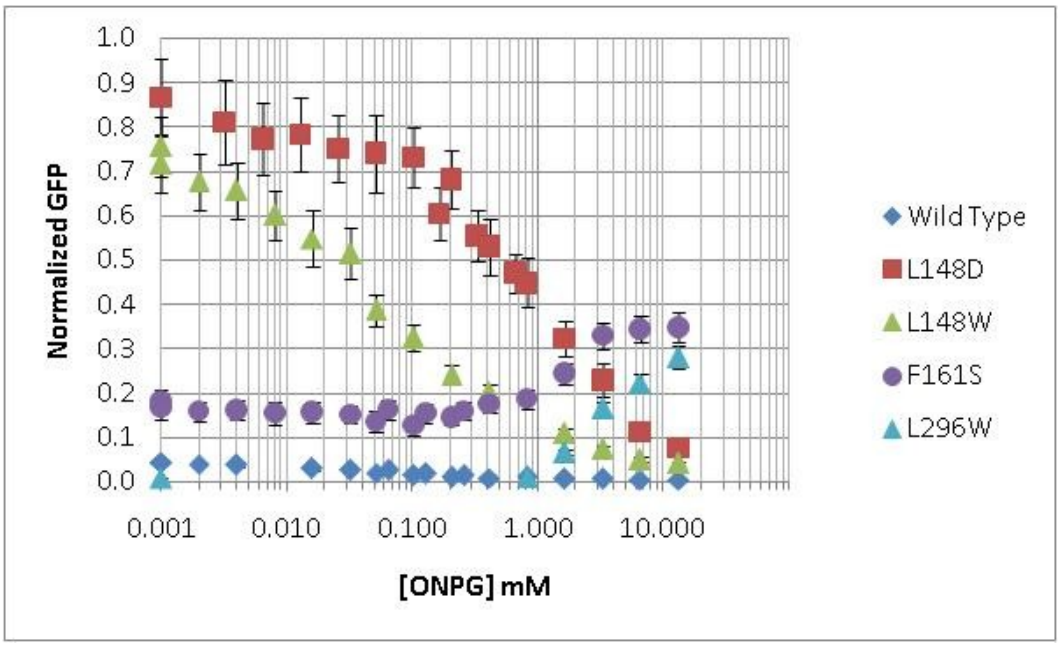


Figure 3.12. **ONPG induction and co-repression profiles of lac repressor dimer and point mutants in the effector binding domain.** Mutations to L148 appear to create inverted repressors. F161S and L296W on the other hand change ONPG from a co-repressor to an inducer.

We globally fit the effector pocket mutants along with wild type dimeric lac repressor induction and co-repression curves. We assumed that total repressor concentration is the same for every mutant and that operator DNA affinity is unchanged, hence we used r^{WT} = 150 for every pocket mutant. The same thermodynamic parameters from the headpiece fitting were used for wild type with IPTG. Table 3.6 gives the final fit parameters.

Construct	Effector	r	K_{RR^*}	K_{RE} (mM^{-1})	K_{R^*E} (mM^{-1})	Leakiness	Dynamic Range	Log_{10} $[E]_{50}$ (mM)
Wild Type	IPTG	150 ± 50	5.8 ± 0.07	60 ± 0.2	500 ± 5	0.044	0.7	-1.94
Wild Type	ONPG	150 ± 50	5.8 ± 0.07	13 ± 1.5	1.3 ± 0.9	0.044	0.7	-1.94
L148D	ONPG	150 ± 50	525 ± 7	2.4 ± 0.2	0.46 ± 0.05	0.78	-0.7	-0.012
L148W	ONPG	150 ± 50	325 ± 6	18 ± 0.7	3.2 ± 0.2	0.69	-0.6	-1.13
F161N	IPTG	150 ± 50	0.98 ± 0.34	23 ± 0.8	670 ± 127	0.013	0.8	-1.64
F161S	ONPG	150 ± 50	26.5 ± 0.01	0.25 ± 0.1	0.54 ± 0.02	0.16	0.3	0.57
F161T	IPTG	150 ± 50	2.7 ± 1.4	82 ± 6.5	1415 ± 293	0.024	0.8	-2.21
F161W	IPTG	150 ± 50	0.47 ± 0.23	18 ± 9	615 ±307	0.01	0.8	-1.42
Q291I	IPTG	150 ± 50	0.07 ± 0.02	26 ± 2.8	1889 ±294	0.007	0.7	-1.46
Q291K	IPTG	150 ± 50	5.9 ± 0.04	11 ± 0.3	86 ± 5	0.044	0.7	-1.18
Q291M	IPTG	150 ± 50	8.9 ± 0.6	44 ± 3.1	515 ± 51	0.062	0.8	-2.07
Q291R	IPTG	150 ± 50	1.33 ± 0.03	251 ± 19	296 ± 25	0.015	0.0	-2.36
L296W	IPTG	150 ± 50	0.04 ± 0.01	3.8 ± 0.1	409 ± 38	0.007	0.73	-0.66
L296W	ONPG	150 ± 50	0.04 ± 0.01	0.4 ± 0.01	17.5 ± 1.7	0.007	0.36	0.65

Table 3.6. Globally fit thermodynamic parameters from fitting wild type dimeric lac repressor and effector pocket mutants to the MWC model of allostery.

Effector pocket point mutants alter leakiness through changes in conformational equilibrium

The assumption that effector pocket mutations do not alter DNA binding affinity means the conformational equilibrium must change if leakiness differs from wild type. This set of mutants shows a wide variety of altered leakiness. L148D and L148W are much, much leakier with ~85% and ~75% basal expression, respectively. This requires nearly 100 fold and 55 fold increases in K_{RR^*} to push more of the repressor in the low operator DNA affinity conformation, R*. Conversely Q291I and L296W are extremely non-leaky with ~1% and ~0.5% basal expression, respectively. These mutants push the conformational equilibrium to ~80 and ~150 fold less than wild type. Equation (8) links conformational equilibrium and energy changes which we can re-state for the lac repressor conformational equilibrium,

$$K_{RR^*} = \exp\left(\frac{-\Delta G}{k_B T}\right) \tag{91}$$

Experiments were carried out at 37C, so we can measure the energy difference between the two states,

$$k_B T = 1.987 e^{-3} \frac{kcal}{mol\,K} 310\,K = 0.616 \frac{kcal}{mol} \tag{92}$$

$$\Delta G = -k_B T \ln\left(K_{RR^*}\right) \tag{93}$$

Table 3.7 gives energy differences for all of the effector pocket mutants and also double difference measurements from wild type where,

$$\Delta\Delta G = \Delta G^{mut} - \Delta G^{WT} \quad (94)$$

Construct	Krr*	ΔG	ΔΔG
WT	5.8 ± 0.07	-1.08 ± 0.01	
L148D	525 ± 7	-3.86 ± 0.01	-2.78 ± 0.01
L148W	325 ± 6	-3.56 ± 0.01	-2.48 ± 0.01
F161N	0.98 ± 0.34	-0.01 ± 0.21	1.1 ± 19
F161S	26.5 ± 0.01	-2.02 ± 0.002	-0.94 ± 0.0001
F161T	2.7 ± 1.4	-0.61 ± 0.32	0.47 ± 0.25
F161W	0.47 ± 0.23	0.47 ± 0.30	1.55 ± 1
Q291I	0.07 ± 0.02	1.64 ± 0.18	2.72 ± 0.29
Q291K	5.9 ± 0.04	-1.09 ± 0.01	-0.01 ± $4x10^{-5}$
Q291M	8.9 ± 0.6	-1.35 ± 0.04	-0.26 ± 0.01
Q291R	1.33 ± 0.03	-0.18 ± 0.01	0.91 ± 0.07
L296W	0.04 ± 0.01	1.98 ± 0.15	3.07 ± 0.24

Table 3.7. Calculated differences in energy between the two structural conformations (ΔG) in kcal/mol based upon the conformational equilibrium parameter, K_{RR^*}. Double energy differences are also given which is the difference from the wild type (WT) value (ΔΔG) in kcal/mol.

Wild type lac repressor has an energy difference between its two conformational space of only about 1 kcal/mol, which is small. It has however been noted that for simple two state folding proteins the entire folding energy is only on the order of 1-10 kcal/mol due to enthalpic and entropic concerns (Muñoz & Eaton, 1999). The energy difference between R and R* of 1 kcal/mol is therefore very significant. This difference is made more prominent in the aforementioned point mutations: L148D and L148W which stabilize the R* conformation by 2-3 kcal/mol relative to R and Q291I and L296W which

stabilize the R conformation by 2-3 kcal/mol relative to R*. Three of these mutations are apolar to polar residues and the third is polar to apolar so none is conservative and each could potentially play a significant role in the stability of each conformation. Interestingly, the only mutation that does not appear to alter the conformational equilibrium is Q291K which are both polar and of similar size but lysine is positively charged. This may be a close to neutral mutation with respect to conformation but Q291K has a drastically lower affinity for IPTG.

In general, the energy scale of 0-3 kcal/mol due to point mutation is not unreasonable given a range of 1-10 kcal/mol folding energy. It is plausibile that these mutations are indeed altering the conformational equilibrium.

Effector pocket point mutants alter $[E]_{50}$ through a subtle interplay between effector affinity and conformational equilibrium

Several of the mutants we tested have altered affinity for their effectors. This altered affinity takes two forms: changes in amount of effector needed to induce or co-repress ($[E]_{50}$) and changes in maximal expression which alters the dynamic range.

First, we look at mutants that have altered $[E]_{50}$. Only two mutants had higher $[E]_{50}$ for IPTG than wild type repressor: F161T and Q291M. Interestingly, these two mutants show the flexibility of the allosteric approach in altering phenotype. Normally to improve binding a more perfect pocket must be developed. F161T has essentially the same IPTG affinity for the R state as wild type (82 ± 6.5 mM^{-1} and 60 ± 0.2 mM^{-1}, respectively) but has a 3-fold higher affinity in the R* state (1415 ± 293 mM^{-1} and 500 ± 5 mM^{-1}, respectively). F161T does indeed form a better pocket for only one of the lac repressor conformations. This however is not the only way to improve $[E]_{50}$ in an

allosteric system. Q291M has approximately the same IPTG affinity in the R* state as wild type (515 ± 51 mM^{-1} and 500 ± 5 mM^{-1}, respectively) but actually impairs binding of the R conformation to IPTG by nearly a third (44 ± 6.5 mM^{-1} and 60 ± 0.2 mM^{-1}, respectively). Nature does not always need to make a more perfect pocket, it could simply break the right part while leaving the rest in tact. Central to both mutations however is conformational equilibrium greater than 1 which means the lac repressor mutants still favor the R* state. Both mutants have increased the inducibility factor X (Equation (90)) and the protein prefers the low operator DNA affinity state. The net effect is less IPTG is needed to do the same job: induce the genetic switch.

Most of the mutants had lowered $[E]_{50}$. The mutation F161W is a good counter-example to Q291M. Much like Q291M, the R affinity for IPTG is reduced versus wild type (18 ± 9 mM^{-1} and 60 ± 0.2 mM^{-1}, respectively) while maintaining R* affinity relative to wild type (615 ± 307 mM^{-1} and 500 ± 5 mM^{-1}, respectively). On the surface, F161W is an even better example with an even higher inducibility factor X, however the $[E]_{50}$ is much lower than both Q291M and wild type. This is the result of the point mutation switching the conformational equilibrium towards the R state (0.47 ± 0.23). Repressor is too favorable to the R state and despite the better X it still takes more IPTG to induce. F161N follows the same route as F161W; higher X cannot compensate for a lower conformational equilibrium so $[E]_{50}$ is lower. Q291I provides the same counter-example for F161T. It drastically improves R* binding of IPTG nearly four-fold but it comes at the cost of a conformational equilibrium vastly in favor of the R conformation (0.07 ± 0.02).

Q291K maintains a positive conformational equilibrium, however the affinity for effector in the R* state is decreased over 5-fold. L296W simply is worse across the

board; both IPTG affinities are lower and conformational equilibrium greatly favors the R state. Both mutants have a lower $[E]_{50}$ than wild type.

The final mutant of interest is Q291R. This mutant very weakly induces and does so because the inducibility factor X is nearly 1. Q291R strengthens the ability to bind ITPG in the R conformation while weakening binding to R* until they are approximately the same. The repressor binds effector but it fails to push the effective equilibrium between the two conformations; they both just bind the same amount of IPTG and DNA occupancy is unaffected.

Effector pocket point mutants alter the dynamic range by changing the relative effector affinities (X) and leakiness

Dynamic range is of course tied intimately to leakiness which has been established to be a function of conformational equilibrium. We will instead focus on how effector pocket mutants can maximize dynamic range by maximizing the relative effector affinity, X.

Q291R of course is the best example of a flattened dynamic range. X has been brought nearly to unity resulting in the collapse of the dynamic range of the switch.

Alternately, several mutants excel at separating the two conformations affinities for effector. L296W has the highest X for IPTG with a greater than 100-fold preference for the R* state. Despite the greatly decreased leakiness of this mutant it can induce the nearly wild type levels. The cost of course is paid in $[E]_{50}$ as discussed previously. Q291I follows nearly an identical path but doesn't have quite as poor an $[E]_{50}$.

F161N, F161T, F161W and Q291M all have dynamic ranges that exceed wild type lac repressor and do so by having larger X. Only F161T achieves the necessary balance of effector binding and conformational equilibrium to be less leaky, have a larger dynamic range and do so with less IPTG (lower $[E]_{50}$) than wild type. F161T is a simply a more ideal on-off genetic switch than wild type dimeric lac repressor.

In the case of co-repression, a decreased X is desirable. L148D and L148W both bind ONPG weaker than wild type lac repressor however they maintain approximately the same co-repressing factor X. Given the large increase in the conformational equilibrium of both L148D and L148W such that the repressor vastly favors the R* state, the very low ratio X suddenly becomes much more relevant.

Leakiness is now very large (~80% and ~70%) and so the dynamic range for co-repression with ONPG goes from practically non-existant for wild type to negative 60-70% for the two mutants.

3.5 Relative Lac Repressor Concentrations in Prokaryotes

One key assumption was made throughout this chapter. Namely, that repressor concentration does not change with point mutation. It proved non-trivial to reproducibly measure the very small amount of lac repressor produced by the cells and a repressor tagging strategy was developed to enhance the signal. The fluorescent protein mCherry was fused to the C-terminus of the lac repressor to allow for direct measurement of the lac repressor in living cells.

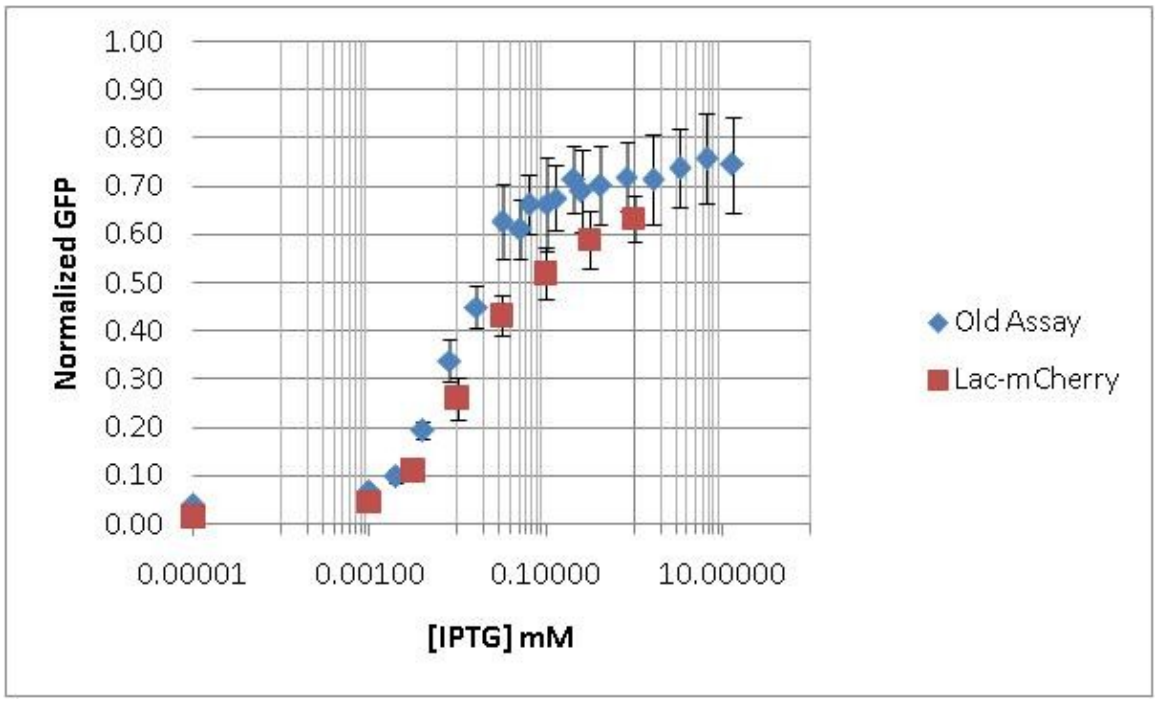


Figure 3.13. **Verifying that Lac-mCherry functions *in vivo*.** The assay conditions have changed (EPB229 cells in MOPS minimal media) and a C-terminal tag was added, however the induction profile with IPTG is relatively unchanged. In particular, Lac-mCherry is slightly less leaky and does not induce quite as high as previously. It is close enough however to compare between mutants to see if overall levels of repressor have changed upon point mutation.

The lac repressor-mCherry fusion protein (hereafter called Lac-mCherry) was cloned into the pLacI plasmid with an 11-base pair linker to make a construct that has been previously shown to be functional (Lau et al., 2004). This plasmid was first used in the previous assay with DH5α cells double transformed with reporter and repressor plasmids grown in LB media. Unfortunately, no mCherry signal could be measured as the media contributed considerable autofluorescence in that excitation/emission range. We found MOPS minimal media did not have autofluorescence in the mCherry region. DH5α cells however did not grow in MOPS minimal media supplemented with glucose presumably due to significant amount of alterations to the bacteria, many of which impair infectivity and growth. We did find that a strain, EPB229, developed from MG1655 cells

with the natural lac operon deleted, was capable of growing in the MOPS minimal media with glucose. We therefore altered the assay to EPB229 cells, double transformed with Lac-mCherry repressor and a reporter plasmid, grown in MOPS minimal media with glucose.

We first verified that the IPTG induction of wild type lac repressor was the same as Lac-mCherry. Figure 3.13 shows Lac-mCherry does function correctly in *E. coli* but it has altered leakiness and maximal expression compared to wild type dimeric lac repressor in the original assay. The addition of the weakly dimerizing mCherry could potentially change the conformational equilibrium and even the overall level of repressor in the cell. The fact is however that Lac-mCherry is functional and can be used to measure differences in expression upon point mutation.

We next made point mutations to Lac-mCherry and measured both GFP regulation and mCherry signal *in vivo*. We saw no significant deviations from expected GFP regulatory phenotype due to the addition of the mCherry tag. Figure 3.14 shows the relative signal from mCherry normalized to wild type Lac-mCherry. All of the tested mutants have expression slightly higher than wild type repressor, however all the changes are less than two-fold. As the phenotypic changes were much more significant than this, we can safely say that our initial assumption was correct: $[R]_{tot}$ is relatively unchanged by point mutation. The fitting and conclusions regarding the r parameter are then valid.

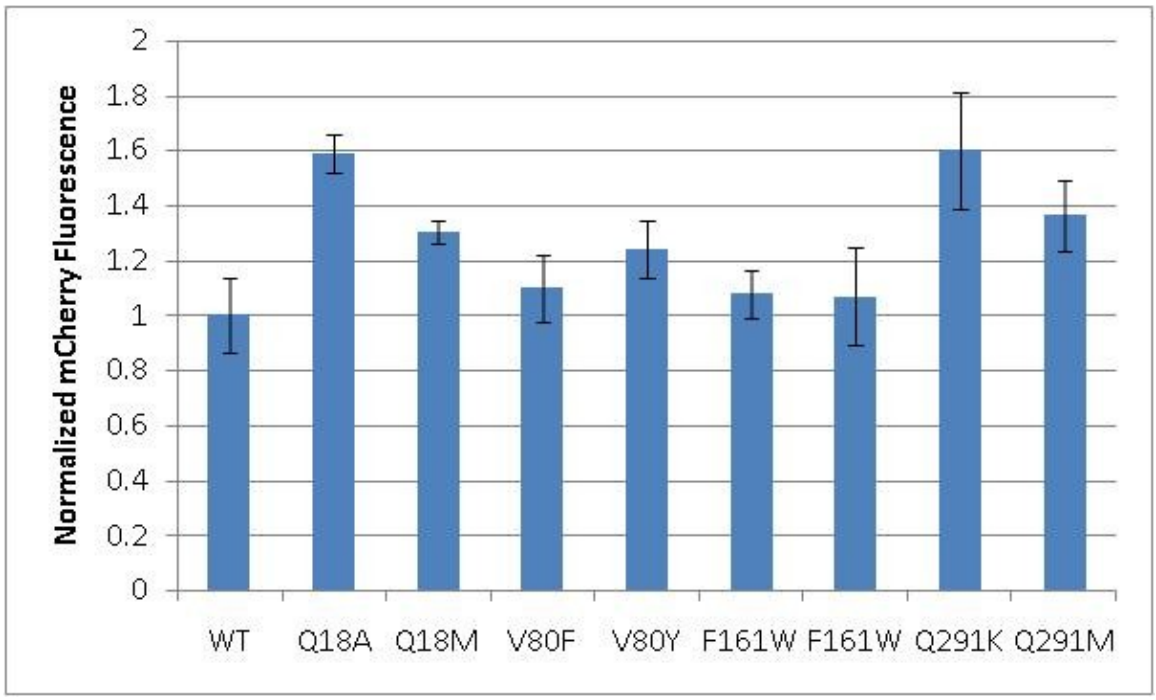


Figure 3.14. **Relative expression of Lac-mCherry and point mutants.** All of the mutants tested had expressions that varied less than two-fold from wild type repressor. This verifies the initial assumption that $[R]_{tot}$ is unchanged upon point mutation.

3.5 Discussion

We sought to link point mutations in structurally interesting regions of the lac repressor to specific changes in thermodynamic parameters. We were able to definitely link headpiece mutants to repressor-operator affinity and effector pocket mutants to repressor-effector affinities, however both stories are complicated by changes to the conformational equilibrium between the R and R* state.

We found mutations to the specificity domain of the DNA binding region can have phenotypes explained by changes to the r parameter of our MWC model solution. Furthermore, we found that the *in vivo* concentration of lac repressor does not change enough upon point mutation, hence changes to r must be reflected in changes to the affinity of repressor for operator DNA (K_{RO}). One point mutation, Q18M, is potentially

better explained by altering the conformational equilibrium, K_{RR^*}. We also cannot rule out this mutation changing both K_{RR^*} and K_{RO} with the net effect of a less leaky repressor. Any experimental measurement of Q18M – operator DNA affinity will be an effective affinity that is a combination of the entire MWC equilibrium and hence it may still not be possible to discern changes between K_{RO} and K_{RR^*}. Our approach where we use normalized expression at saturating effector to separate the two thermodynamic parameters may be the only way to accurately separate these two parameters. It would not be unreasonable however that the conformational equilibrium is capable of being altered by this point mutation as the balance between the two states appears to be incredibly malleable as evidenced by the effector pocket mutant data.

Mutants in the effector pocket were incredibly effective in creating phenotypically interesting genetic switches through single point mutations. Nearly every point mutant changed the leakiness of the switch. Repressor-operator affinity should be unchanged and the overall concentration of the lac repressor does not greatly change, hence the r parameter is the same same for effector pocket mutants. Leakiness changes must then be reflected in changes of the conformational equilibrium, K_{RR^*}. The conformational equilibrium was able to change in both directions (in favor of either the R or R* conformation) by up to 3 kcal/mol of energy upon single point mutations which provides an astounding ability to increase leakiness up to 70% or lowers it nearly 7-fold. Furthermore, effector pocket point mutations were able to both increase and decrease the magnitudes of the repressor-effector affinities for both conformations. More importantly, the affinity changes did not always mirror each other, as in both don't increase together or decrease together. This means the ratio of the two affinities, X, can increase or decrease allowing for bold changes in the dynamic range of the switch;

higher dynamic range than wild type and even flipped dynamic range creating induction from co-repression. Of note was the precise interplay between the three thermodynamic parameters K_{RE}, K_{R^*E}, and K_{RR^*}. Only one mutant, F161T, had all three parameters move in the correct directions to make a switch with improvements in all three phenotypic parameters necessary for making a more ideal on-off genetic switch.

We showed a link between structural regions of the lac repressor and the thermodynamic binding parameters associated with these regions. Furthermore, we showed that the conformational equilibrium of the dimeric lac repressor is balanced on a razors edge and is easily tipped one way or the other through point mutation. The interplay between binding affinity and conformational equilibrium leads to a wide array of genetic switches that are potentially relevant given the right gene that needs regulation.

Chapter 4:
Linking *In Vivo* to *In Vitro*

4.1 Introduction

In Chapter 3 we sought to link the *in vivo* functioning of the lac genetic switch to the thermodynamic parameters of the MWC model by specifically perturbing the switch in key structural locations and looking for corresponding changes in the model parameters. This approach worked very well, linking DNA affinity to the DNA binding domain and effector affinity to the effector pocket and finally showing that the conformational equilibrium is malleable.

A study by Poelwijk, et al. looked for unique phenotypes through random mutagenesis of the lac repressor (Poelwijk, de Vos, & Tans, 2011). Mutants were identified which exhibited an inverted repression behavior; a phenotype also found by in Chapter 3 by mutating the effector binding domain. Interestingly, Poelwijk's mutations were in regions physically distinct from either the DNA or effector binding domains. One potential explanation is that the mutations destabilize the folded form of the repressor, altering the conformational landscape. Mutagenesis of the repressor can result in more than just predictable changes of thermodynamic binding constants.

Central to all of these studies is the use of *in vivo* data to understand the behavior of genetic switches. It has been pointed out that a lack of corroborating *in vitro* evidence prevents the identification of other processes which may significantly play into the equilibrium, such as non-specific DNA binding or effector uptake (Tungtur, Skinner, Zhan, Swint-Kruse, & Beckett, 2011). They attempted to measure the thermodynamic binding of a LacI/GalR hybrid repressor both *in vitro* and *in vivo*. Notably, a DNA pull

down assay was used to quantify the *in vivo* concentration of their hybrid repressor. Unfortunately, they were unable to rectify a greater than 25-fold difference between their two data sets. This indicates that they are missing a significant contributor to the genetic switch by only analyzing *in vivo* data.

Here we sought to overcome the limitations of past studies three ways: 1.) measure the *in vivo* concentration of the lac repressor, 2.) measure the *in vitro* transcription of purified lac genetic switch, and 3.) use an assumption free solution to the MWC equilibrium to model both *in vitro* and *in vivo* data.

We were able to measure lac repressor concentration *in vivo* and use *in vitro* transcription to assess the purified lac genetic switch. Furthermore, we found excellent agreement between *in vitro* and *in vivo* data when molecular crowding was taken into consideration. We do however find that the repressor-DNA affinity is much lower than has previously been measured. Additional concerns, such as effector uptake and non-specific DNA binding do not appear to play significant roles.

4.2 Measuring the *In Vivo* Concentration of the Lac Repressor

We sought a method where we could simultaneously measure repressor concentration and transcriptional regulation and thus chose to fluorescently tag the repressor. The fluorescent protein mCherry was chosen due to minimal auto-fluorescence from MOPS minimal media and minimal spectral overlap with our reporter gene YFP. Furthermore, a dimeric Lac-mCherry fusion construct is known to be functional *in vivo* (Lau et al., 2004). The goal is to measure raw mCherry fluorescence and OD_{600} in growing *E. coli* cells and convert those measurements to an intracellular concentration of lac repressor (Figure 4.1).

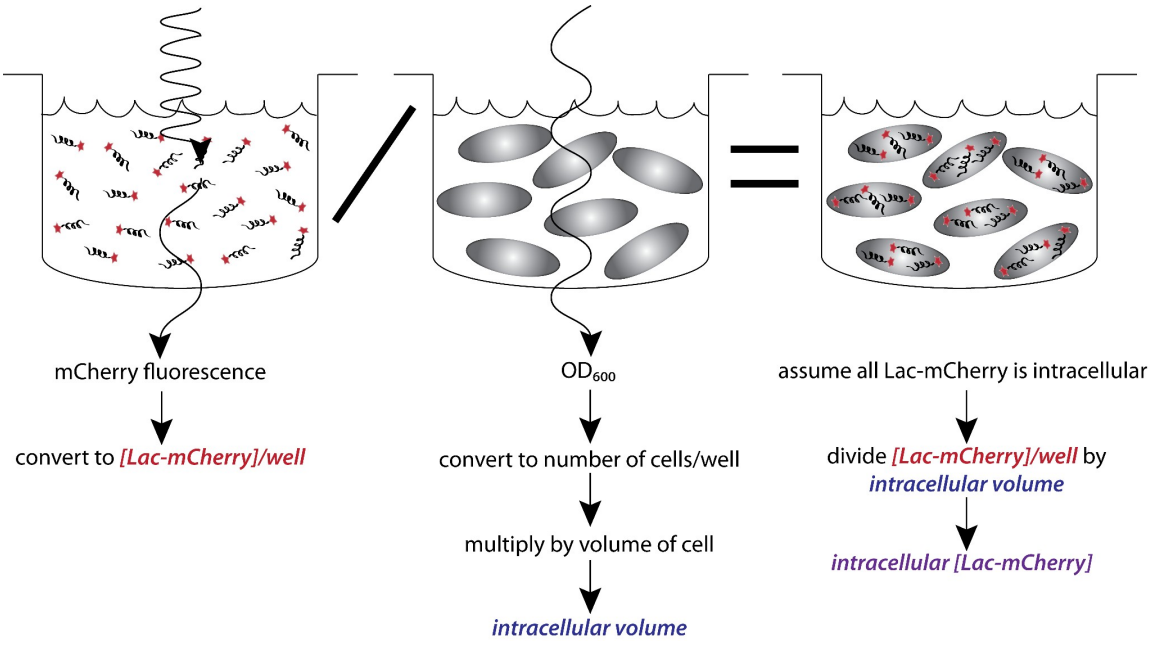


Figure 4.1. Measuring intracellular Lac-mCherry concentration. Raw mCherry fluorescence and OD_{600} are measured on a plate reader. Calibration curves for both were established given our experimental setup (cell line, plasmids, media, amount of sample loaded, plates and plate reader). Raw fluorescent signal is converted to concentration of Lac-mCherry per well. Raw OD_{600} signal is converted to the fraction of well volume that is intracellular. Dividing Lac-mCherry well concentration by intracellular volume fraction effectively concentrates the Lac-mCherry to be intracellular. These two measurements, combined with the appropriate calibrations, allow a quick and accurate measurement of intracellular Lac-mCherry concentration.

A linear relationship was established for OD_{600} and cell count. We estimate the volume of *E. coli* growing in glucose supplemented minimal media to be $1x10^{-15}$ L (Tungtur et al., 2011). We then measured OD_{600}, calculated the number of cells and multiplied by volume of the cell to calculate the fraction of the well that is intracellular. A linear relationship was also established for purified Lac-mCherry fluorescence and concentration of Lac-mCherry.

We assume all of the Lac-mCherry is intracellular; therefore we divided the Lac-mCherry concentration by the fraction of volume that is intracellular. Using this method, we can quickly and accurately measure *in vivo* Lac-mCherry concentrations.

Intracellular Lac-mCherry concentration in EPB229 cells

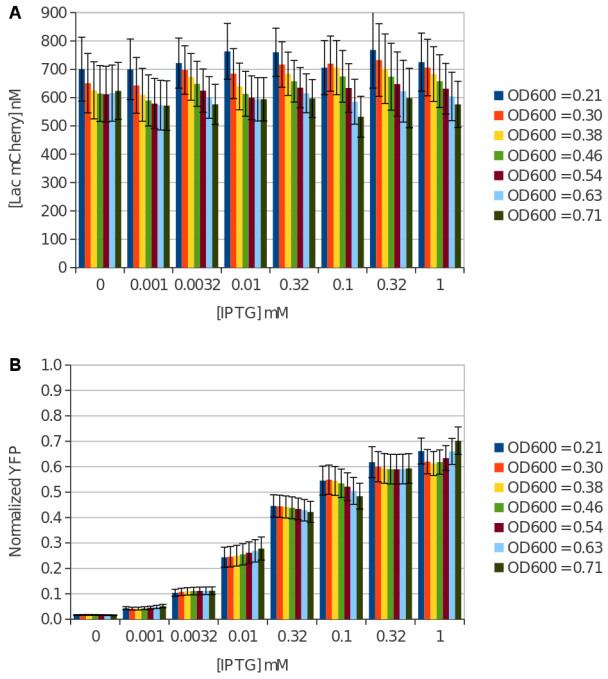


Figure 4.2. *In vivo* Lac-mCherry and YFP regulation show no growth dependence. **(A)** Lac-mCherry was calculated for growing *E. coli* cells and found to have minimal OD_{600} dependence. As expected for a constitutively expressed gene, there is no change in Lac-mCherry concentration with increasing IPTG concentration. **(B)** Normalized YFP simultaneously measured and again no OD_{600} dependence was found throughout the exponential growth phase. In stark contrast the Lac-mCherry, a distinct induction profile is measured for YFP.

EPB229 cells expressing Lac-mCherry and the reporter plasmid were grown in varying concentrations of the inducer IPTG. Intracellular concentration of Lac-mCherry was calculated from mCherry fluorescence and OD_{600} and found to be 664 ± 90 nM at

mid-log growth phase (OD_{600} = 0.6) (Figure 4.2.A). As expected for a constitutively expressed gene, minimal variation was seen with IPTG and cell growth (Figure 4.2.B).

We then converted to molecules per cell,

$$6.6e^{-7}M * 1e^{-15}\frac{L}{cell} * 6.022e^{23}\frac{molecules}{mole} = 397\frac{molecules}{cell} \quad (95)$$

We have previously estimated the copy number of our plasmid to be ~10-20 plasmids/cell (Daber et al., 2009). This corresponds to approximately 20-40 Lac-mCherry dimers per plasmid which agrees well with previous estimates of ~40 Lac repressor dimers per plasmid for our promoter (Oehler et al., 1994).

4.3 Measuring the *In Vivo* Regulation of YFP

In addition to mCherry fluorescence and OD_{600} measurements, YFP fluorescence was measured in cells as a function of IPTG. Unregulated expression was established by measuring OD_{600} and YFP in cells co-transformed with O1 YFP reporter and a plasmid which does not contain any repressor (pABD34). These positive control cells were grown in tandem with cells containing both reporter and repressor and grown in a variety of IPTG concentrations.

Positive controls showed no IPTG dependence as expected, so data from every sample was combined to determine an overall positive control polynomial fit. YFP fluorescence is seen to increase as cells grow as would be expected due to the increased number of cells per µL. We remove this bias and normalize regulated YFP expression by dividing by the positive control fit curve.

Normalized YFP expression was then measured as a function of OD_{600} and IPTG (Figure 4.4, blue squares). Almost no OD_{600} dependence can be noted in the induction profile (Figure 4.2.B). The YFP signal is repressed without IPTG and is approximately 1-5% of unregulated expression. Upon induction with saturating IPTG we see a robust YFP increase to approximately 60% of the unregulated expression.

4.4 Measuring the *In Vitro* Regulation of mRNA

While the *in vivo* experiment measures translation product (fluorescing YFP) we know the lac repressor actually regulates mRNA production. Previously, our lab has determined a linear relationship between mRNA and fluorescence protein signal allowing us to use fluorescence as a proxy for mRNA regulation *in vivo*. The situation *in vitro* is reversed; it is much easier to measure mRNA production.

We used the Maxiscript T7 *in vitro* transcription kit (Ambion) which produces mRNA from linearized DNA with a T7 promoter. We then measured incorporation of radioactive labeled CTP into mRNA. The T7 promoter was modified to add an O1 operator DNA site and we were able to modulate Lac-mCherry and IPTG concentrations. A positive control of constitutive mRNA production is established by not adding any Lac-mCherry.

We first established that radioactively labeled mRNA was linearly observable by constitutively producing mRNA and loading various dilutions onto polyacrylamide gels and established that mRNA concentration was linearly related to the concentration of mRNA loaded on the gel. Positive controls were included for every experiment and were used for normalization.

The additional benefit of *in vitro* transcription is the flexibility in dosing not only IPTG, but also Lac-mCherry. We exploited this flexibility by first titrating in Lac-mCherry without IPTG present and with saturating IPTG (1mM) (Figure 4.3.A). As expected, increasing concentration of Lac-mCherry decreases mRNA production. Furthermore, addition of IPTG returns mRNA signal to near constitutive levels.

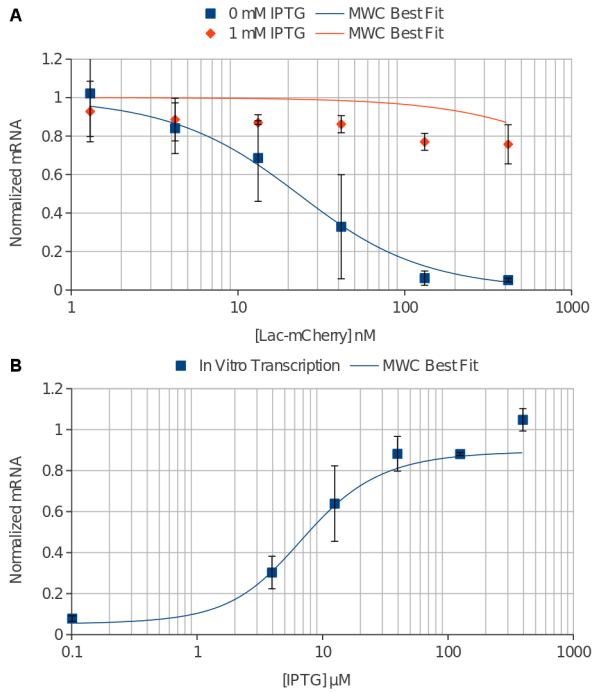


Figure 4.3. ***In vitro* transcription controlled by the lac repressor is accurately fit by the MWC model. (A)** Lac-mCherry was added at varying concentrations with 11nM O1 DNA and mRNA was quantitated (blue squares). The repression was relieved upon addition of 1mM IPTG (orange diamonds). The data was globally fit by the MWC model and an accurate solution was found for the Lac-mCherry titration (solid blue line). Unfortunately, the model predicts higher induction than was measured experimentally (solid orange line). **(B)** IPTG was added at varying concentrations with 333nM Lac-mCherry and 11nM O1 DNA and again mRNA was quantitated (blue squares). A robust induction profile was measured showing induction up to approximately 90% of constitutive expression. The global fit also accurately fits the IPTG titration data (solid blue line).

We then titrated IPTG at a fixed Lac-mCherry concentration (Figure 4.3.B). The induction of mRNA is seen to very closely resemble that of the *in vivo* data, but it is

noticeably leakier. Maximal repression was about 5% and maximal induction was approximately 90%.

4.5 Modeling Using MWC Thermodynamic Equilibrium

Finally, we sought to simultaneously model the *in vivo* and *in vitro* data using the Monod, Wyman, and Changeux (MWC) model of thermodynamic equilibrium. Previously, we have relied upon approximate solutions of the lac genetic switch equilibrium to model *in vivo* induction profiles. This solution assumes that the total repressor concentration greatly exceeds operator concentration ($R_{tot} >> O_{tot}$). This condition does not hold for our *in vitro* experiment where we titrated in Lac-mCherry nor would it necessarily be true in all *in vivo* systems. Therefore, we sought a solution to the equilibrium that held for every potential input.

We modeled transcription using the Monod, Wyman and Changeux (MWC) thermodynamic equilibrium. Normalized expression is then modeled as free operator concentration, [O], divided by total operator concentration, $[O]_{tot}$.

We then want to take known quantities and measured normalized regulation ($[O]/[O]_{tot}$) and measure the equilibrium thermodynamic parameters of our model. We know the total concentrations of repressor ($[R]_{tot}$), operator DNA ($[O]_{tot}$), and effector ($[E]_{tot}$). We also will act as if we know all of the thermodynamic constants. We want to make no assumptions so that both *in vivo* and *in vitro* data can be directly compared.

In Section 3.2 we started the problem by defining all of the thermodynamic parameters in Equations (43) - (47) and formulas for total operator in Equation (48) and total repressor in Equation (59). We have also already defined some of the bound

species in Equations (49) - (54). We therefore need to define the missing formula for total effector concentration,

$$[E]_{tot} = [E] + 2[RE] + 2[RE_2] + 2[R^*E] + 2[R^*E_2] + 2[ROE] + 2[ROE_2] + 2[R^*OE] + 2[R^*OE_2] \quad (96)$$

We therefore begin by writing the equation for total repressor solely in terms of free species and thermodynamic constants,

$$[R]_{tot} = [R] + 2[R][E]K_{RE} + [R][E]^2K_{RE}^2 + [R]K_{RR^*} + 2[R][E]K_{RR^*}K_{R^*E} + [R][E]^2K_{RR^*}K_{R^*E}^2 + [R][O]K_{RO} + 2[R][O][E]K_{RO}K_{RE} + [R][O][E]^2K_{RO}K_{RE}^2 + [R][O]K_{RR^*}K_{R^*O} + 2[R][O][E]K_{RR^*}K_{R^*O}K_{RE} + [R][O][E]^2K_{RR^*}K_{R^*O}K_{RE}^2 \quad (97)$$

We then make the following definitions,

$$\alpha_1 = 1 + K_{RR^*} \quad (98)$$

$$\beta_1 = 2K_{RE} + 2K_{RR^*}K_{R^*E} \quad (99)$$

$$\gamma_1 = K_{RE}^2 + K_{RR^*}K_{R^*E}^2 \quad (100)$$

$$\gamma_2 = 2K_{RO}K_{RE} \quad (101)$$

$$\delta_1 = K_{RO}K_{RE}^2 \quad (102)$$

$$\beta_2 = K_{RR^*}K_{R^*O} \quad (103)$$

$$\gamma_3 = 2K_{RR^*}K_{R^*O}K_{R^*E} \tag{104}$$

$$\delta_2 = K_{RR^*}K_{R^*O}K_{R^*E}^2 \tag{105}$$

Substituting into Equation (97),

$$[R] = \frac{[R]_{tot}}{\alpha_1 + [E]\beta_1 + [E]^2\gamma_1 + [O](K_{RO} + [E]\gamma_2 + [E]^2\delta_1 + \beta_2 + [E]\gamma_3 + [E]^2\delta_2)} \tag{106}$$

The equation has been organized such that polynomials in [E] are apparent. As long as we only add and multiply polynomials, they can trivially be treated as symbolic functions for further simplification. We define the following polynomials,

$$B_1 = \alpha_1 + [E]\beta_1 + [E]^2\gamma_1 \tag{107}$$

$$B_2 = K_{RO} + \beta_2 + [E](\gamma_2 + \gamma_3) + [E]^2(\delta_1 + \delta_2) \tag{108}$$

Now substituting back into Equation (109),

$$[R] = \frac{[R]_{tot}}{B_1 + [O]B_2} \tag{109}$$

We next want to follow the same path for [E] and [O]. Inspection of Equations (48), (59), and (106) show that we have already done the most complicated case. We can then quickly arrive at,

$$[O] = \frac{[O]_{tot}}{1 + [R]B_2} \tag{110}$$

The effector equation is similar, but it has a few extra coefficients of two within its equations. We define two more polynomials,

$$A_1 = \beta_1 + [E]2\gamma_1 \tag{111}$$

$$A_2 = \gamma_2 + \gamma_3 + 2[E](\delta_1 + \delta_2) \tag{112}$$

Substituting into Equation (106),

$$[E]_{tot} = [E] + [R][E]A_1 + [R][E][O]A_3 \tag{113}$$

We can then eliminate [O] by substituting Equation (110) into Equations (107) and (113). Since we can only multiply and add polynomials, we multiply the denominator of Equation (110) on both sides.

Substituting into Equation (107),

$$[R]_{tot} + [R]B_2[R]_{tot} = [R]B_1 + [R]^2 B_1 B_2 + [R]B_2[O]_{tot} \tag{114}$$

We then define the following polynomials,

$$\varphi_1 = B_1 B_2 \tag{115}$$

$$\varphi_2 = B_1 + B_2([O]_{tot} - [R]_{tot}) \tag{116}$$

Substituting into Equation (114),

$$[R]^2 \varphi_1 + [R]\varphi_2 = [R]_{tot} \tag{117}$$

The substitution of Equation (110) into Equation (113) requires the following definitions,

$$\psi_1 = [E] A_1 B_2 \tag{118}$$

$$\psi_2 = [E](B_2 + A_1 + A_2[O]_{tot}) - B_2[E]_{tot} \tag{119}$$

We then arrive at,

$$[R]^2 \psi_1 + [R]\psi_2 = [E]_{tot} - [E] \tag{120}$$

In principal we can get this down to a single equation, but in order to do so the final polynomial becomes of a much higher order which prevents accurate computational solutions.

The strategy is then to guess at the free effector concentration to calculate Equations (115)-(116) and (120)-(121). Equations (117) and (120) can then be solved for [R] by looking for the roots to the equation. When the correct free effector concentration is found the roots of Equation (117) and Equation (120) will converge. By minimizing the difference between the roots a solution can be reached. All other concentrations are then trivial to calculate once [R] and [E] are known. Custom Matlab (Mathworks) software was written to numerically solve the MWC equilibria (Matlab File Exchange ID #40602).

The accuracy of the solution is easily checked by using the bound and free species concentrations to calculate the total species concentrations and thermodynamic parameters. Calculated values should agree with input values.

Using the assumption-free solution to measure thermodynamic parameters

Experimentally we know the total concentrations ($[R]_{tot}$, $[E]_{tot}$, $[O]_{tot}$) and normalized transcription/expression ($[O]/[O]_{tot}$). We want to measure the thermodynamic constants (K_{RR^*}, K_{RE}, K_{R^*E}, K_{RO}, K_{R^*O}). This leaves 5 independent constants in the MWC model to fit to the experimental data. The large number of independent constants results in a myriad of non-unique solutions to the equations. This complication was limited by the following algorithm.

First, since it is widely reported to be effectively zero, K_{R^*O} was set to be very, very small ($1x10^{-10}$ nM^{-1}). This leaves four independent parameters.

Next, it had been observed from previous studies that the ratio of K_{R^*E} to K_{RE} is well defined when the concentration of repressor greatly exceeds that of operator. Under this assumption, a simpler solution of the MWC equilibrium exists as explored in Section 3.2. We isolated a subset of the *in vitro* data where this condition was true and used a non-linear least squares fitting algorithm to measure the ratio $X = K_{R^*E}/K_{RE}$ as a function of conformational equilibrium. The ratio was seen to asymptote at approximately 13.75. This value is then used to reduce the number of independent constants to 3 (K_{RR^*}, K_{RE}, and K_{RO}).

We then simultaneously fit the *in vitro* data to obtain the best fit thermodynamic parameters using a non-linear least squares algorithm in Matlab (Table 4.1). The model accurately fits both the IPTG (Figure 4.3A) and the lac repressor doping (Figure 4.3B) *in vitro* transcription experiments. The fit values agree well with values obtained in the

literature with the exception of repressor-DNA affinity. The repressor-DNA affinity was measured to be 0.4 ± 0.2 nM^{-1}. This is significantly weaker than the 100-3333 nM^{-1} that has been measure previously (Sharp, 2011). It does agree well with an estimated value of 1 nM^{-1} for lac repressor-DNA affinity that prevails under conditions within the *E.* coli cell (Muller-Hill, 1996). The thermodynamic equilibrium value (6.3 ± 3.3) does not significantly differ from that measured previously by our group. The repressor-IPTG affinity ($7.6x10^{-4}$ ± $2.5x10^{-4}$ nM^{-1} for the higher affinity conformation) was found to be slightly higher than previously published values ($2.3x10^{-4}$ nM^{-1}) but it is generally within agreement. The ratio of affinities for the two conformations (13.7) was in good agreement with previously measured values.

Using the in vitro *thermodynamic parameters to predict* in vivo *genetic regulation*

The raison d'être for *in vitro* measurements is to inform what is occurring *in vivo.* One of the central difficulties in using *in vitro* measurements is the lack of a well enough defined *in vivo* system to directly compare it with. Furthermore, a model is required which can accurately function in both circumstances and provide useful predictions. We then seek to fully define our *in vivo* experiment to model it with the *in vitro* determined thermodynamic parameters.

We estimate the copy number of our operator reporter plasmid to be ~ 20 copies per cell (Daber et al., 2009). This then gives us,

$$[O]_{tot} = \frac{20\,\text{molecules}}{6.02\,e^{23}\,\frac{\text{molecules}}{\text{mole}}} * \frac{1}{1\,e^{-15}\,L} * 1\,e^{9}\,\frac{nM}{M} = 33\,nM \qquad (121)$$

The strain of *E. coli* used has the lac genetic switch deleted from the genome; therefore lac permease is also deleted. It is then assumed that IPTG enters the cell through passive diffusion and has the same concentration as the media.

	This study	Daber, Sharp, and Lewis†	Chapter 3	Sharp, Set 1 ‡	Sharp, Set 2 ‡	Sharp, Set 3 ‡	Müller-Hill §
K_{RR^*} = $[R^*]/[R]$	6.3 ± 3.4	2 ± 0.5	5.8 ± 0.07				
K_{RO} (nM^{-1})	0.42 ± 0.21			3330	100	1510	1
K_{RE} (nM^{-1})	$5.6x10^{-5}$ ± $1.8x10^{-5}$		$6x10^{-5}$ ± $2x10^{-7}$				
K_{R^*E} (nM^{-1})	$7.6x10^{-4}$ ± $2.5x10^{-4}$		$5x10^{-4}$ ± $5x10^{-6}$	$2.3x10^{-4}$	$2.3x10^{-4}$	$2.3x10^{-4}$	
K_{R^*O} (nM^{-1})	$1.0x10^{-10}$						
R_{tot} (nM) [with 40% crowding]	664 ± 90 [1660 + 225]						
r = K_{RO} * R_{tot} [with 40% crowding]	278 [697]	150 ± 50	150 ± 50				
X = K_{R^*E} / K_{RE}	13.7 ± 0.13	15 ± 3	8.28				

Table 4.1. Fit values from the MWC models compared with literature values. All fit parameters agree with the exception of repressor-operator DNA affinity (K_{RO}). †(Daber et al., 2009) ‡(Sharp, 2011) §(Müller-Hill, 1996).

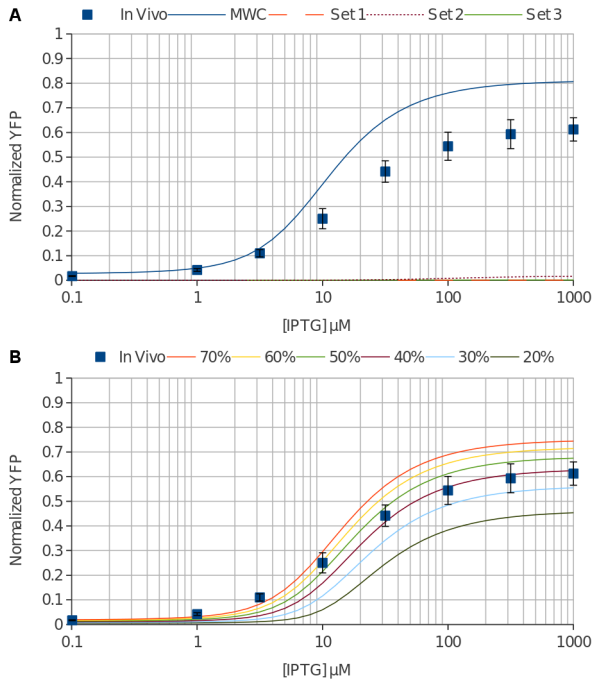


Figure 4.4. ***In vivo* regulation by the lac repressor is accurately predicted with molecular crowding. (A)** YFP under control of the lac repressor was measured in *E. coli* cells at varying concentrations of IPTG (blue squares). We used the measured intracellular concentration of the lac repressor (660nM) and the fit values from *in vitro* transcription to predict the *in vivo* induction curve with the MWC model (solid blue line). The model agrees at low IPTG concentrations but shows a significant over-prediction at full induction. Our repressor-DNA affinity was much lower than previously published values, so we also modeled three curated data sets from Sharp [10] (dashed orange, dotted purple, and solid green lines). All three predict greatly over-repressed YFP expression and do not fit the *in vivo* data. **(B)** Molecular crowding is known to play a significant role in cells. We modeled this by estimating the available volume in percentage for our repressor and calculated an effective repressor concentration. We modeled several percentages and 40% available volume (solid purple line) accurately reproduces the *in vivo* regulation from the *in vitro* transcription derived thermodynamic constants. This corresponds to an effective repressor concentration of 1.6µM.

Figure 4.4.A shows the simulated *in vivo* data along with experimentally determined values. Clearly, the model predicts accurate leakiness but higher induction than is observed. There is a known molecular crowding effect in living cells due to the density of molecules. Figure 4.4.B shows the effect of including molecular crowding on the predicted *in vivo* induction curve. The model shows excellent agreement with

experiment at a molecular crowding of 40% which estimates effective *in vivo* repressor concentration to be 1.6 μM. Furthermore, this value agrees well with estimates of 20%-40% available space *in vivo* (Kubitschek & Friske, 1986).

Since there is a notable deviation in repressor-DNA affinity with previous measurements, the same analysis was carried out for the three curated data sets from Sharp (Sharp, 2011). Using the values from the literature, we find that they do not in any case come close to replicating our *in vivo* data (Figure 4.4.A). The DNA affinities are much too high for the measured DNA and repressor concentrations. At these affinities the switch is essentially completely off and cannot be induced with any concentration of IPTG. Crowding only enhances the deviation from experiment as it further increases the concentration of repressor.

Simulating native in vivo *lac genetic switch phenotype*

The thermodynamic constants from our *in vitro* data better represents our *in vivo* model system. The question then is: which set of thermodynamic parameters could effectively regulate the native lac genetic switch?

Essentially we have rebuilt the lac operon with the *lacZ*, *lacY* and *lacA* polycistronic message replaced by the reporter gene YFP and the dimeric lac repressor constitutively expressed by its native promoter. We have a higher copy number of both the reporter and repressor plasmids (~20 copies per cell) which increases both the operator and repressor concentrations above that normally found in the cell. A secondary deviation is the removal of the tetramerization domain and multiple operator DNA sites (O2 and O3 additionally exist on the genome) which simplifies our analysis. The cooperativity of the native tetrameric lac repressor is known to decrease leakiness

approximately 10-fold, so we might expect a dimeric lac repressor with one operator (O1) to have some leakiness in its repression (Oehler et al., 1994).

As previously mentioned, *in vivo* lac repressor dimer concentration was measured to be ~40 dimers per cell, which gives,

$$[R]_{tot} = \frac{40\,\text{molecules}}{6.02\,e^{23}\,\frac{\text{molecules}}{\text{mole}}} * \frac{1}{1\,e^{-15}\,L} * 1\,e^{9}\,\frac{nM}{M} = 66\,nM \tag{122}$$

And we know there is one operator per cell,

$$[O]_{tot} = \frac{1\,\text{molecules}}{6.02\,e^{23}\,\frac{\text{molecules}}{\text{mole}}} * \frac{1}{1\,e^{-15}\,L} * 1\,e^{9}\,\frac{nM}{M} = 1.7\,nM \tag{123}$$

Using these values, along with the experimentally determined binding constants derived from this study and those curated by Sharp, we can simulate dimeric lac repressor induction curves at native conditions. Figure 4.5.A shows that the values determined in this study predict a leaky repressor that is maximally inducible. The much higher DNA affinities of the curated data sets all produce over-repressed curves that do not show good induction.

The over-repression is even more prominent as cell crowding is considered. Using the value of 40%, which gives R_{tot} = 66 nM / 0.4 = 165 nM, we find that the over-repression of the high affinity DNA sets all produce curves that weakly induce or do not induce at all (Figure 4.5.B). The predicted curve using our thermodynamic parameters again provides reasonable induction (~10% leakiness up to ~95% maximal induction). While this level of leakiness would be intolerably high for efficient regulation of the lac

operon, the restoration of the tetramerization domain would significantly decrease the leakiness while minimally impairing inducibility.

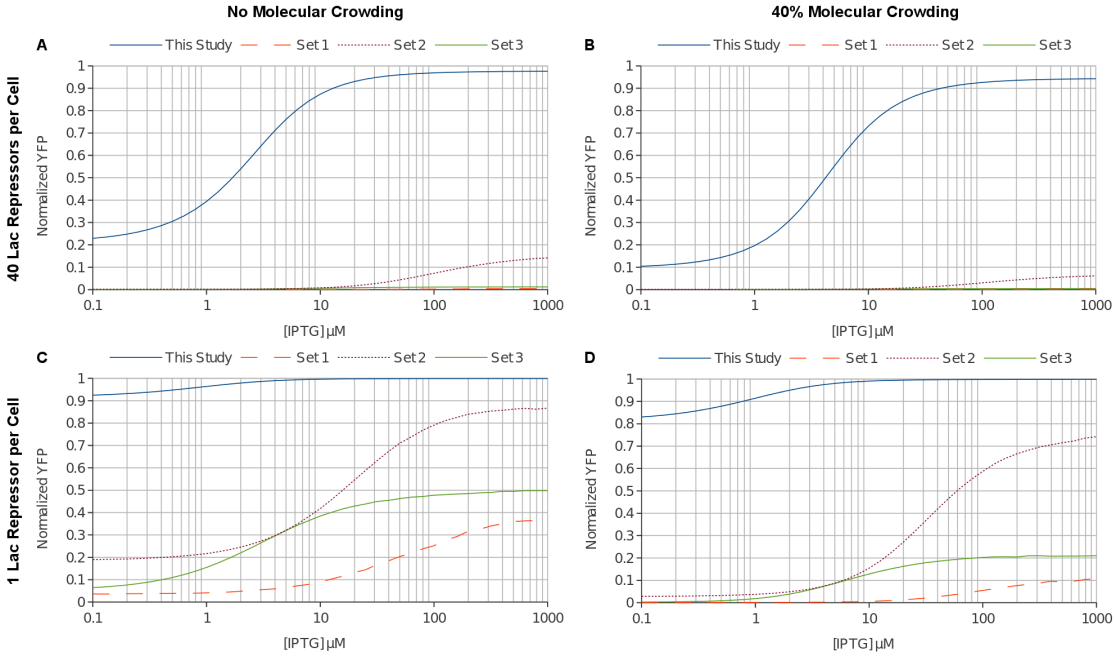


Figure 4.5. **Simulating a simplified lac operon from *in vitro* derived thermodynamic constants.** The correct repressor-DNA affinities must be able to provide robust switching under conditions naturally experienced by *E. coli.* With this in mind, we modeled a dimeric lac repressor regulating a gene with a single operator sequence. **(A)** The natural lac promoter makes ~66nM of lac repressor dimer and one operator is at ~ 1.7nM in the cell. We modeled these conditions for the thermodynamic parameters from this study and for the three curated data sets of Sharp. The predicted curve from this study shows a reasonable repression and induction profile (solid blue line). Only Set 2 from Sharp is weakly inducible (dotted purple line). **(B)** Including molecular crowding (40% available volume) enhances the situation. The curated data sets do not make useful switches. Alternately, the predicted induction curve from *in vitro* transcription derived constants shows a leaky switch that induces very well (solid blue line). **(C)** We next sought to model the minimal possible repressor to find a condition where the curated data sets produce reasonable induction curves. 1 molecule of dimer per cell (~1.7nM) does show good induction profiles for set 2 (dotted purple line) and set 3 (solid green line). Set 1 still shows a switch that can marginally be induced and would likely not be useful (dashed orange line). **(D)** Molecular crowding effects again enhance the repressor concentration and only set 2 could reasonably regulate a gene (dotted purple line). The values from this study (solid blue line) predict a very leaky switch. Although the second curated set could effectively regulate the gene at this concentration, in reality a single dimer and single operator DNA binding would be dominated by stochastic events creating an inherently unstable switch.

If we consider the lowest possible concentration of lac repressor (1 molecule/cell; R_{tot} = 1.7 nM; with 40% crowding R_{tot} = 4.25 nM) we find that second curated data set does produce reasonable induction curves, even if 40% crowding is taken into consideration (Figure 4.5.C and Figure 4.5.D). Unfortunately, in this regime the binding would be highly stochastic and hence noisy, which would not produce stable repression. Furthermore, this level of repressor expression does not agree with published values. While it is technically possible for these affinities to be accurate, it is highly improbable. The first and third data sets would require less than 1 molecule of dimeric lac repressor per cell to be functionally useful according to our model.

Given the wide range of repressor-operator DNA affinities (100nM – 3333nM) it can be reasonably concluded that these values must contain significant artifacts from the experimental techniques. Techniques such as gel shift assays, where molecular “caging” effects are known to be significant, and nitrocellulose filter binding assays, where the binding is removed from the solution phase, were used to create the curated data sets. Our measurement of repressor-DNA binding affinity did require an indirect measurement, namely transcription; but it did occur in the solution phase. We attribute the difference in values to differences in experimental setup.

4.6 Discussion

We have reproduced the transcriptional regulation of the lac repressor dimer *in vitro* and shown that it accurately reproduces the *in vivo* repression of YFP under control of the lac repressor. Accurate modeling of the *in vivo* data required an estimate of 40% cellular crowding in the cell, which agrees with previous estimates. Non-specific DNA

binding and IPTG uptake did not appear to have any significant effect. Crowding could be tested *in vitro* through crowding agents such as BSA or PEG. Alternative explanations are potentially possible such as fluctuations in the size of the *E. coli*. What is essentially important is that the concentration of lac repressor in the cell greatly affects the maximal induction given our thermodynamic parameters. The curve is extremely sensitive in that region to changes in repressor concentration. So only an approximately two-fold increase in repressor concentration is sufficient to replicate the *in vivo* data. Whether the lac repressor concentration is increased by molecular crowding or by decreased *E. coli* volume would have to be tested by further experiments.

The measured thermodynamic binding parameters match well for IPTG binding and conformational equilibrium, except there is significantly lower repressor/operator DNA affinity measured (by approximately 3-4 orders of magnitude). This discrepancy was modeled and it was demonstrated that the affinity measured in this study is capable of reproducing not only the *in vivo* data from this study, but also can predict reasonable induction curves at concentrations of repressor and DNA that are naturally seen by *E. coli*. We therefore conclude that lac repressor DNA affinity is significantly weaker than previous *in vitro* measures and more in line with the estimates for repressor-DNA affinity at *in vivo* conditions where we do find good agreement with previously published values.

Finally, this study highlights the difficulty in using *in vitro* data generated from experimental techniques that are divorced from conditions closer to that of the cell. Experimental artifacts may greatly overshadow actual values, which should come as no surprise in the case of lac repressor binding to operator DNA where the published binding constant has changed 33-fold as experimental techniques have changed. The difficulty in *in vitro* measurements is well known in the field as is evidenced by the large

consideration given to differences in buffer conditions (Ha, Capp, Hohenwalter, Baskerville, & Record, 1992), DNA length (Khoury, Lee, Lillis, & Lu, 1990), and even hydrostatic pressure (Royer, Chakerian, & Matthews, 1990). Techniques such as gel filtration or nitrocellulose filter binding assays are excellent at differentiating binding strength between point mutants; they are limited in comparison with *in* vivo results. Using experimental setups which more closely mimic the *in vivo* system can significantly improve the ability of the predictive capabilities of *in vitro* experiments. They do come with the caveat that the data interpretation is not as straightforward as simple binding experiments.

Chapter 5:
An Autogenously Regulated Lac Repressor for AAV-mediated Eukaryotic Gene Therapy

5.1 Introduction

Bacterial gene regulation is interesting due to its simplicity. The lac genetic switch consists of a single protein that binds to a known short operator DNA sequence all regulated by a small sugar molecule. Eukaryotic gene regulatory systems are not so simple (Roeder, 2003). Could we use the much simpler bacterial systems to regulate specific genes of interest in eukaryotes?

The first attempt to translate any bacterial gene regulation to eukaryotes was the *lexA* protein. *LexA* was shown to block transcription when its operator was placed within a promoter (Brent & Ptashne, 1984). The lac repressor was then tested in a mouse L cell line for the ability to regulate a transgene (Hu & Davidson, 1987). Prior to this paper it was unknown if the protein could fold, dimerize, tetramerize, migrate to the nucleus, or even bind DNA which might be in a chromatin bound state. Hu and Davidson found that not only does the lac repressor block transcription, it can do so with an operator located in three different regions within the promoter: between the start codon and the transcription start site, between the transcription start site and the TATA box, and upstream of the TATA box. Not only does the lac repressor bind, it is capable of stopping transcription from a wide variety of locations. Furthermore, they showed that repression was relieved upon addition of IPTG. They did, however, note that only 10% of the repressor was localized to the nucleus. Addition of a nuclear localization signal localized all of the repressor to the nucleus but with the odd caveat that operator DNA binding was

reduced 6-fold (Hu & Davidson, 1991). The lac genetic switch was shown to transiently and stably regulate a Simian Virus promoter, SV40, in cell culture (Brown et al., 1987; Figge, Wright, Collins, Roberts, & Livingston, 1988). A quantitative assay was developed that showed stable HEK293 cell lines expressing constitutive lac repressor and regulated *lacZ* were capable of 5-fold induction with IPTG from an Epstein-Barr virus derived promoter (Biard, James, Cordier, & Sarasin, 1992; Liu, Feliciano, & Stambrook, 1989). The lac repressor was shown to work in tandem with the Tet repressor to create a regulatory cascade capable of responding to IPTG, tetracycline, or both (Aubrecht, Manivasakam, & Schiestl, 1996). Plant cells were also shown to be capable of regulating gene expression using the lac genetic switch (Ulmasov, Capone, & Folk, 1997).

Up till this point, the lac genetic switch was shown to be functional in both transient and stable eukaryotic cell lines with the operator in varying positions in various promoters. The next logical step was to test the lac repressor in a transgenic mouse model (Scrable & Stambrook, 1997). This first required re-encoding the *lacI* gene to use more optimal mammalian codons (*synlacI*) (Zhang, Zubay, & Goldman, 1991). Scrable and Stambrook found that the *E. coli lacI* gene was transcriptionally silenced whereas the mammalian optimized *synlacI* gene was ubiquitously transcribed. The full lac genetic switch was then shown to functional in the mouse (Cronin et al., 2001). Mice were developed with the tyrosinase gene under control of the lac operator and it was shown that coat color of the mice could reversibly be changed upon feeding of IPTG. Interestingly, they found that the *synlacI* gene previously described did not form measurable protein product nor was it functionally capable of regulating gene expression. The found a cryptic splice site within the *synlacI* gene that, when reverted to

the *lacI* sequence, creates functional lac repressor in cells. This mouse model was later used to investigate the timing of tyrosinase expression in embryogenesis to study occulocutaneous albinism (Cronin, Ryan, Talley, & Scrable, 2003). The lac genetic switch functionality was also shown to persist through embryonic stem cell differentiation (Caron et al., 2005) and it was shown to be functional in living plants (Moore, Gälweiler, Grosskopf, Schell, & Palme, 1998).

Clearly the lac repressor is capable of functioning in gene regulation in eukaryotes. It is not, however, optimized for that purpose. Furthermore, the use of the lac genetic switch on the animal level has required making transgenic animals where the genome has directly been modified. These approaches are not useful in human gene therapy for both practical and safety reasons. One leading method of gene therapy delivery is the use of Adeno-Associated Virus (AAV). AAV virus can be targeted to specific tissues and the delivered genes exist on an episomally separate from the genome (Penaud-Budloo et al., 2008), and the delivered genes can be stably expressed over long periods of time. One serious drawback is that the size of the viral capsid creates an upper limit on the length of the gene that can be delivered to the cell; only so much DNA can be packed within the AAV viral capsid.

With this in mind, we sought to develop a lac genetic switch that would provide robust gene regulation in AAV-mediated gene therapy. This required making a genetic switch that is as small as possible and optimizing the minimal switch for gene regulation in Eukaryotes.

5.2 Development of an Autogenously Regulated Lac Repressor

The lac repressor is constitutively expressed in *E. coli* from its own promoter. Therefore, the lac genetic switch is usually applied by inserting a promoter driving lac repressor and a second promoter driving the gene of interest under control of a lac operator DNA sequence. Constitutively expressed repressor is the exception, not the rule, in the LacI/GalR family (Swint-Kruse & Matthews, 2009). Many repressors are autogenously regulated; they control their own expression.

The question then is: what will happen if we make an autogenously regulated lac repressor? It is first instructive to consider the behavior of a properly tuned constitutively expressed lac repressor on the lac genetic switch. Looking back on the modeling of Chapter 3, we saw that the effect of changing repressor concentration was the same as changing the repressor-operator DNA affinity, K_{RO}. Figure 3.3.A shows that higher repressor concentration (higher r) will make a less leaky switch that is not as inducible. Lower concentration of repressor will make a leakier switch that easily maxes out induction. Figure 3.3.B shows that an ideal lac repressor concentration exists where the dynamic range is maximal and this spot coincides with a non-leaky switch. The wild type lac repressor promoter has evolved to constitutively express the lac repressor at just this correct range. If we want to utilize the constitutively expressed lac repressor genetic switch in a different organism we would need to optimize the promoter and/or the repressor-operator DNA affinity to get the maximal dynamic range. In making stable cell lines, one alternative approach is to incorporate a random number of repressor promoters and reporter promoters into the genome and then select for optimal dynamic range. This is just not possible for gene therapy; you cannot select for good cells in a patient.

An autogenously regulated switch on the other hand auto-tunes the repressor concentration at equilibrium. As repressor is made it begins to shut off the promoter. At equilibrium, the promoter will make just enough repressor to maintain its off state. This means however that the autogenously regulated is inherently leaky; it *must* let enough transcription occur to keep the repressor concentration high enough to shut off the switch. A constitutively expressed switch has no necessary leakiness. There are benefits to autoregulation however; the effector response is linearized (Nevozhay, Adams, Murphy, Josic, & Balázsi, 2009), the response to effector is faster (Rosenfeld, Elowitz, & Alon, 2002), and the genetic switches are more stable (Becskei & Serrano, 2000). All of these benefits, combined with a smaller overall transgene, led us to develop an autogenously regulated lac genetic switch with the goal of minimizing the inherent leakiness of the switch.

An autogenously regulated lac genetic switch in E. coli

We first made an autogenously regulated lac genetic switch by placing the *lacI* gene fused to mCherry after the natural lac promoter. A reporter gene, YFP, was included after an 18 base pair spacer (Figure 5.1). Prokaryotes are capable of expressing a poly-cistronic message where multiple genes are expressed off of a single promoter (which is how the *lacZ*, *lacY*, and *lacA* genes are naturally expressed), therefore we expected this single plasmid to make both a lac repressor-mCherry fusion protein and the YFP reporter gene.

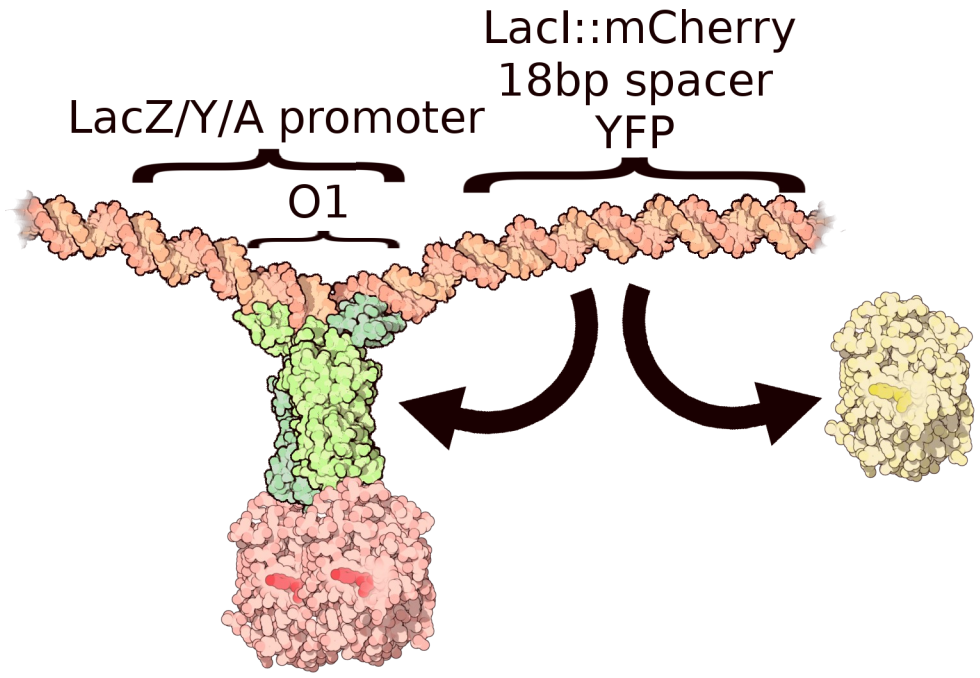


Figure 5.1. **Plasmid organization for prokaryotic autogenously regulated lac genetic switch.** The native *lacZ/Y/A* promoter under control of the O1 natural lac operator controls two genes: lac repressor fused to mCherry and YFP. Lac-mCherry binds to O1 leading to autoregulation and the YFP acts as a cytoplasmic readout of the switch.

The autogenously regulated prokaryotic switch was tested in EPB229 cells growing in minimal media supplemented with glucose. Both mCherry and YFP were measured in living cells in the same manner as the *in vivo* experiments of Chapter 4. Figure 5.2 shows that the autogenously regulated lac genetic switch successfully turns off YFP expression and induces to a lower value than the constitutive lac genetic switch. Essentially, the autoregulation diminishes the overall dynamic range of the prokaryotic switch, which is expected because the promoter that drives the lac repressor in *E. coli* is well tuned to have a maximal dynamic range switch. Figure 5.3 shows that the autoregulation also regulates the lac repressor. The induction is much less than what is seen for YFP. One thing to note is that the calculated lac repressor concentration from

the autogenously regulated switch is actually much higher than was measured from the constitutive plasmid in Chapter 4.

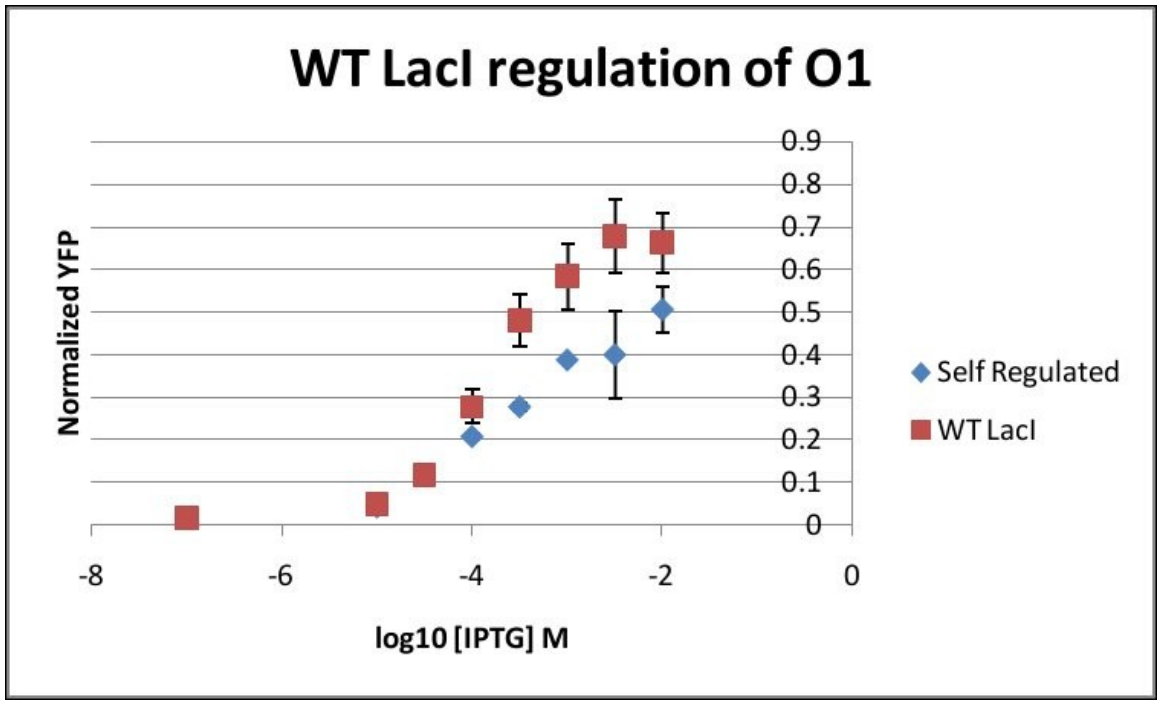


Figure 5.2. Autogenously regulated lac genetic switch versus constitutively regulated. The autogenously regulated (Self Regulated) switch has a similar leakiness to the constitutively regulated (WT LacI) but it cannot induce to as high of levels.

The reporter operator is much stronger than the natural *lacI* promoter that makes constitutive lac repressor. The autogenously regulated stronger promoter actually makes more repressor than the constitutive weak promoter. That is why leakiness is not greater in the autogenously regulated as expected. On the flip side, there is so much repressor that the switch cannot turn fully on. Clearly the strength of the promoter that is being autogenously regulated plays a significant role in the functional output of the genetic switch.

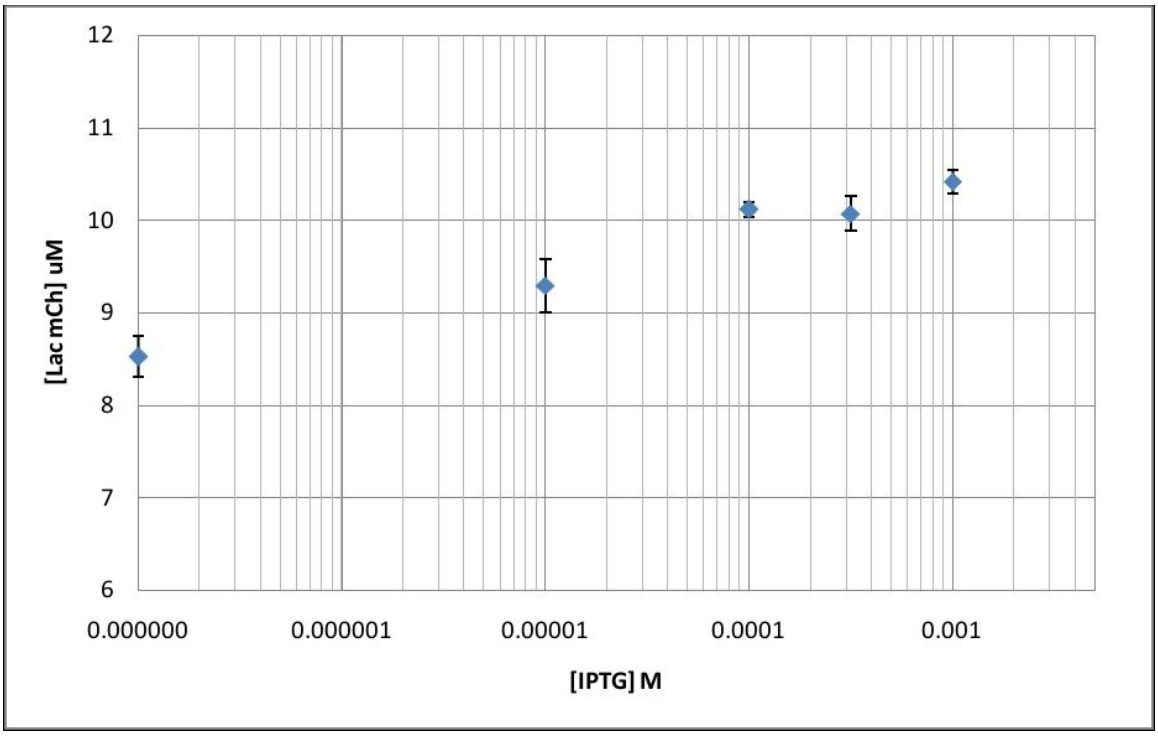


Figure 5.3. **Autoregulation of the lac repressor as measured by mCherry signal.** Lac repressor levels are seen to rise upon IPTG induction as would be expected for a autogenously regulated repressor. The measured levels of repressor are higher than measured from the constitutive plasmid.

It is unlikely that placing an autogenously regulated switch on most promoter will result in a gene regulation profile that fits the needs of gene therapy. We would like to use the known mutants of the lac repressor to be able to improve upon the genetic switch. The question is: do lac repressor mutants still give rise to the same phenotypes in an autogenously regulated switch? We tested both the DNA binding domain mutants (Figure 5.4) and effector pocket mutants (Figure 5.5) from Chapter 4 on the autogenously regulated lac genetic switch in prokaryotes. The two DNA binding domain mutants clearly follow the same phenotypic changes seen in the constitutive switch therefore we can change leakiness and dynamic range by mutating the DNA binding domain. The effector pocket mutants also exhibited changes in leakiness, dynamic

range, and $[E]_{50}$ as previously found. Therefore we can mutate the effector pocket to try to optimize any of the three phenotypic properties.

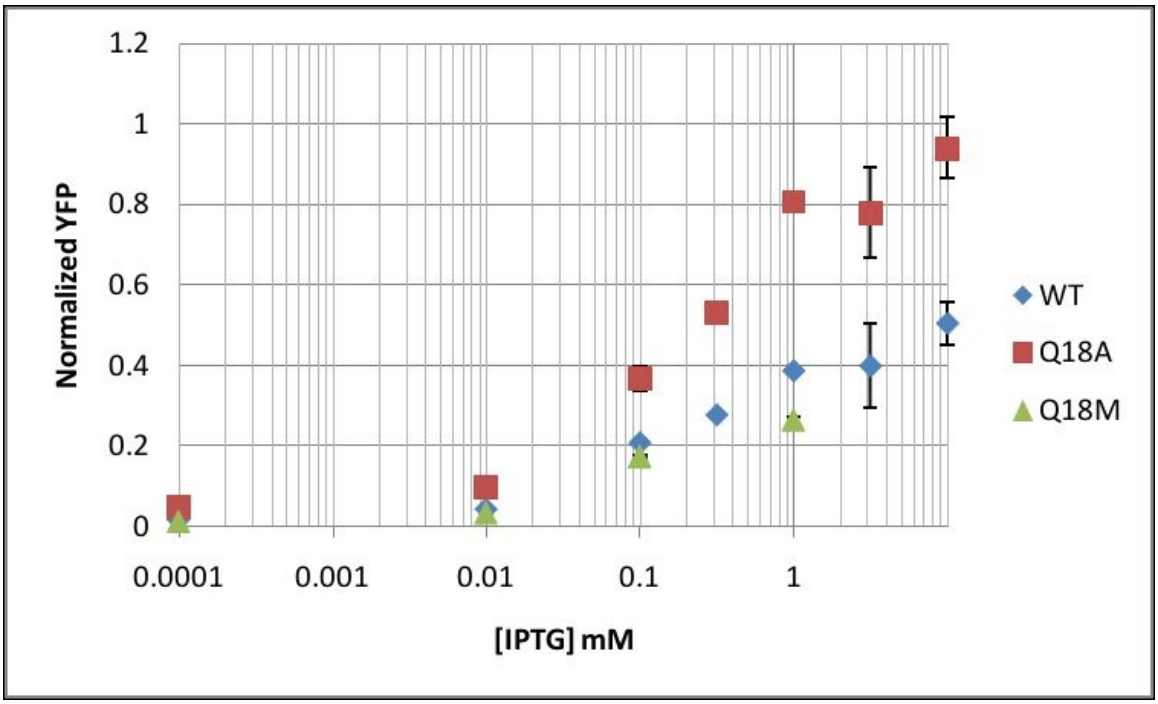


Figure 5.4. DNA binding domain mutants in the prokaryotic autogenously regulated switch. Q18A is still leakier and Q18M is still less leaky than wild type, just as observed in the constitutive switch. The tradeoff between leakiness and dynamic range is still present.

The demonstration of the autogenously regulated lac genetic switch in *E. coli* essentially established that, as expected, the lac repressor is functional autogenously regulated and that the phenotype is malleable through point mutation of either the DNA binding domain or the effector binding pocket. Furthermore, the strength of the promoter significantly effects the levels of protein that are made.

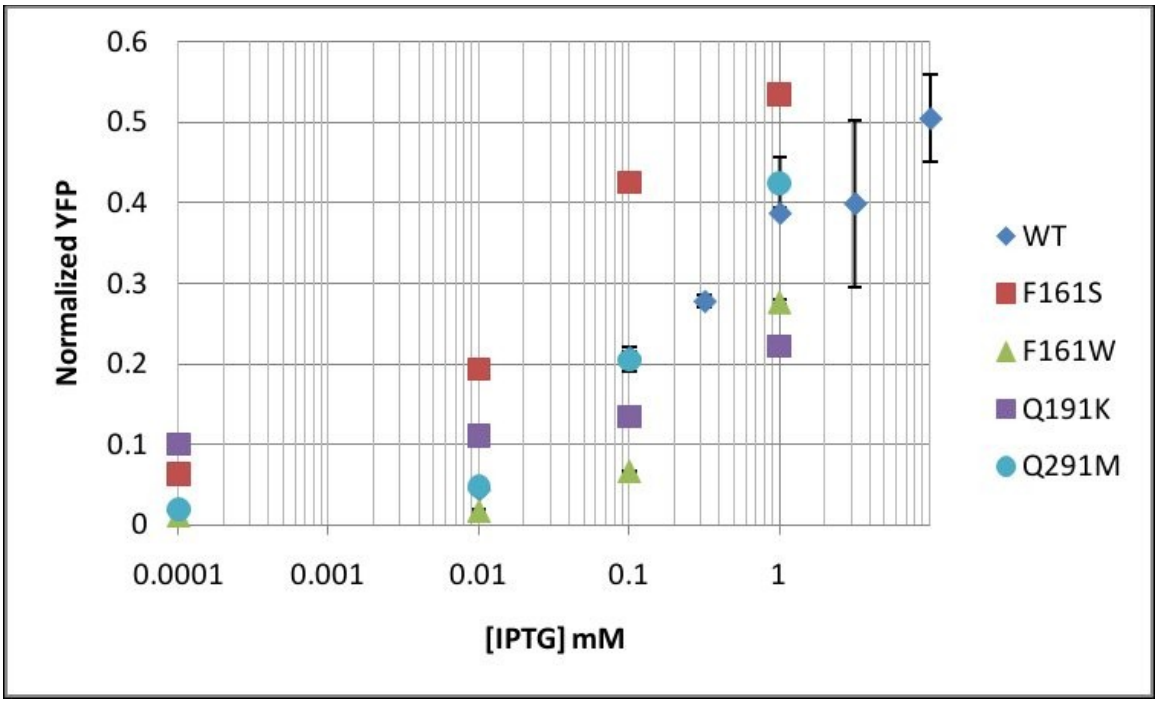


Figure 5.5. Effector pocket mutants in the prokaryotic autogenously regulated switch. Four point mutants in the effector binding pocket are seen to alter the leakiness, dynamic range and $[E]_{50}$ of the autogenously regulated switch. Clearly this region of the repressor is still an ideal target for mutagenesis to alter the phenotype of the of the switch.

5.3 The Autogenously Regulated Genetic Switch in HEK293T Cells

We next sought to test the autogenously regulated lac repressor in transiently transfected HEK293T cells to test its ability to regulate a reporter gene. We first developed a plasmid that has the minimal CMVI promoter with the Lsym lac operator between the TATA box and the transcription start site (Figure 5.6). Eukaryotes do not translate poly-cistronic mRNA messages so an alternative strategy is needed to express the two genes. We first used an Internal Ribosomal Entry Sequence (IRES) that allows the ribosome to translate a second gene within the mRNA and therefore make an autogenously regulated switch. We used YFP as the reporter gene and the lac repressor sequence was the codon optimized sequence corrected for a splice site, which

we term *EuLac*, which was then fused the mCherry *(Cronin et al., 2001)*. EuLac::mCherry fusion has the canonical nuclear localization sequence (NLS) on the C-terminus to localize the repressor to the nucleus.

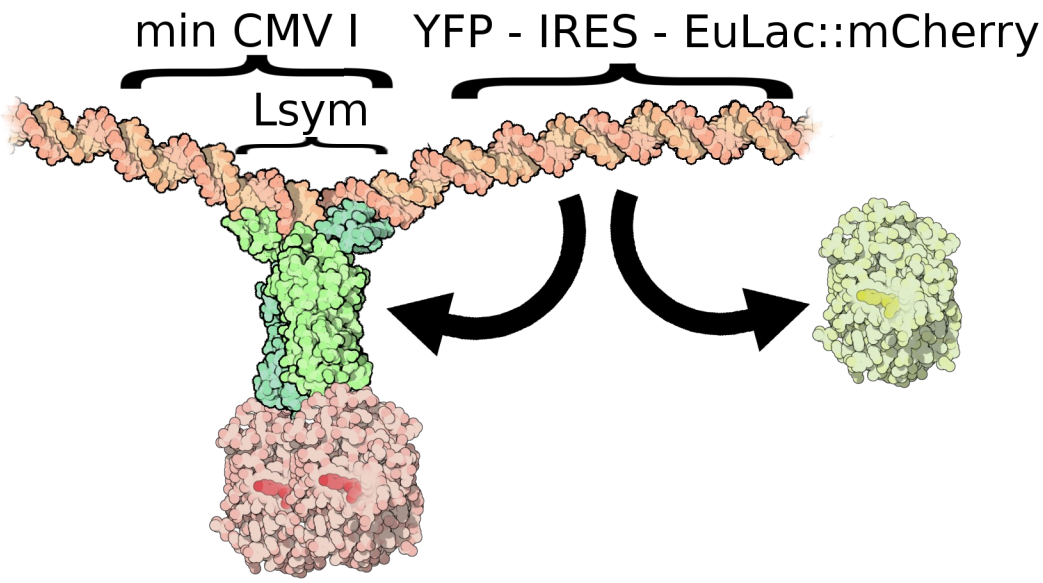


Figure 5.6. The pIRES-EuLacmCh-YFP plasmid. This plasmid expresses an autogenously regulated dimeric lac repressor mCherry fusion construct which can bind one Lsym operator site and has YFP as a readout.

This autogenously regulated IRES plasmid was transiently transfected into HEK293T cells and media was supplemented either with or without 2.5 mM IPTG. Figure 5.7 shows that both the YFP and the lac repressor are induced upon addition of IPTG. YFP is dispersed throughout the cell while the mCherry signal is localized to the nucleus (labeled blue by Hoechst nuclear stain). The autogenously regulated IRES switch clearly is functional in transient transfections of HEK293T cells. This construct is not however ideal for applications in AAV-mediated gene therapy. The mCherry tag is unnecessary now that we know the NLS correctly localizes lac repressor to the nucleus. Second, the IRES is more than 300 base pairs. A smaller sequence is known to be able to express two genes in eukaryotes from a single promoter in the 2A sequence. The 2A

sequence encodes for a protein that causes a slippage as it is being translated through the ribosome. The result is that, with a certain frequency, two proteins are made rather than one large fusion protein. One benefit of the 2A sequence is that the two proteins are made in equimolar amounts which is not necessarily true for IRES sequences.

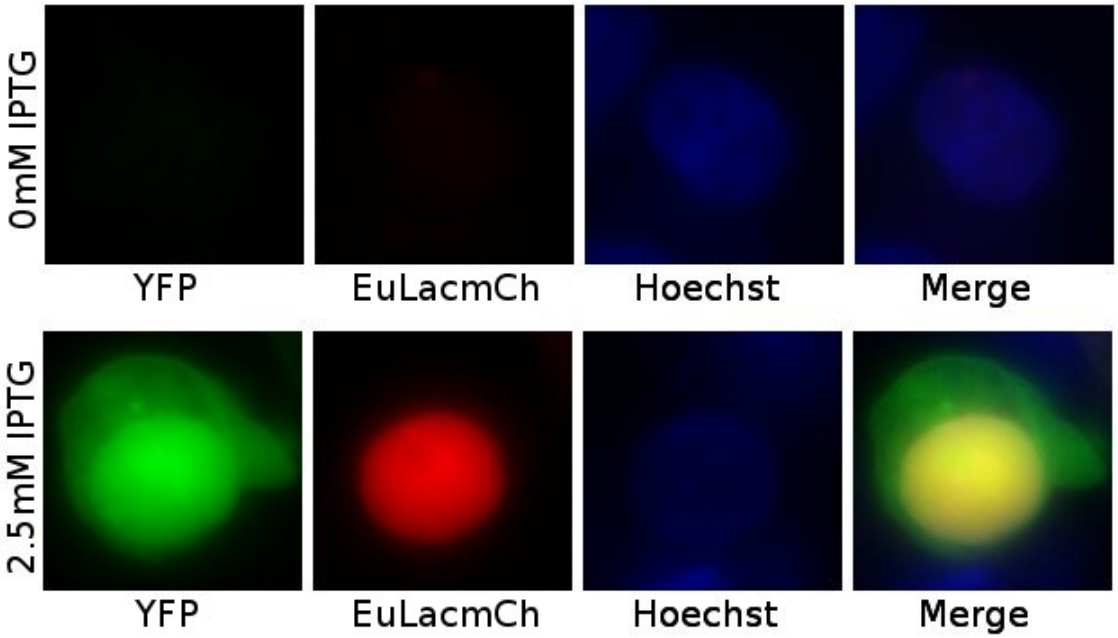


Figure 5.7. Autogenously regulated IRES lac genetic switch in HEK293T cells. YFP and lac repressor (mCherry) both are induced upon addition of IPTG. YFP is dispersed throughout the cell whereas the mCherry signal is localized to the nucleus (labeled by Hoechst stain).

We next developed a minimal length cassette with the goal of using it in an AAV viral delivery vector. This cassette again uses the same minimal CMV I promoter with a Lsym operator DNA sequence. The gene order has been switched so the *EuLac* gene is first, followed by the 2A sequence, followed by GFP as the readout (Figure 5.8). Extraneous DNA was trimmed and unique restriction enzyme cleavage sites were designed around every gene in the cassette for easier downstream cloning. The same plasmid was made with luciferase as the readout gene.

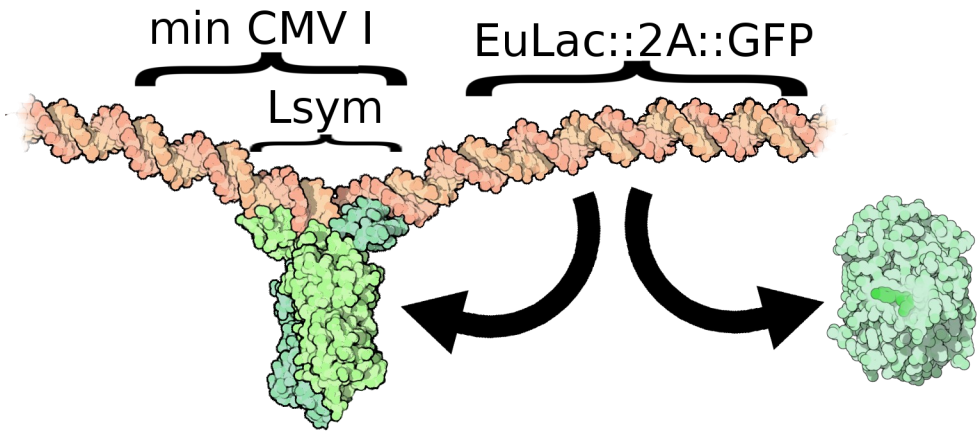
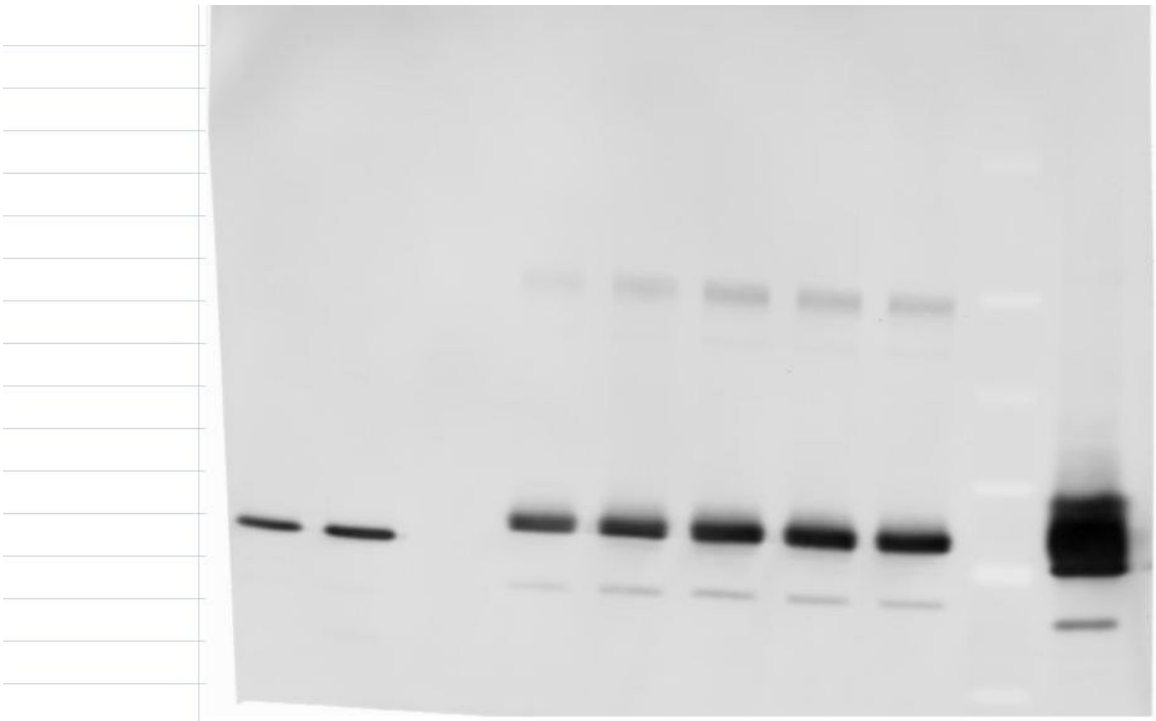


Figure 5.8. The pSW-GFP plasmid. This plasmid expresses an autogenously regulated dimeric lac repressor construct which can bind one Lsym operator site and has GFP as a readout. This cassette easily fits into the AAV viral capsid. The same plasmid expressing luciferase rather than GFP is called pSW-Luc.

We first need to verify the efficiency of the 2A sequence to split the two proteins. HEK293T cells were transiently transfected with either: 1. the original autogenous IRES plasmid, 2. the new autogenous 2A plasmid, or 3. a constitutively expressed GFP under control of the chicken β-actin (CBA) promoter. Cells were grown in varying concentrations of IPTG and whole cell lysate was harvested for western. We performed a western with α-GFP antibodies (Figure 5.9). The original plasmid only has a single band for YFP (which is also detected with α-GFP) which induces ~ 2-fold with IPTG. The autogenous 2A plasmid has three primary bands: the highest band corresponds to Lac::2A::GFP fusion protein, the middle band is truncated 2A fused to GFP, and the third band is likely a degradation product. Comparing the highest band to the middle band we estimate ~90% breakage of the fusion protein to separated lac repressor and GFP. We also see induction of GFP with increasing IPTG, again approximately 2-fold (Figure 5.10).



Construct	1	1		2	2	2	2	2		CBA
[IPTG] mM	0	2.5		0	0.01	0.06	0.63	2.5		0

Figure 5.9. α-GFP western of HEK293T cells. Construct 1 is the autogenously regulated IRES plasmid from Figure 5.6. Construct 2 is the autogenously regulated 2A plasmid from Figure 5.8. CBA is the CBA promoter constitutively expressing GFP. We see induction with IPTG for both construct 1 and construct 2, both with maximal induction of approximately 2-fold. For the 2A cassette there is clearly a higher, unbroken fusion protein. We see approximately 90% splitting of the fusion protein into the two desired proteins, lac repressor and YFP.

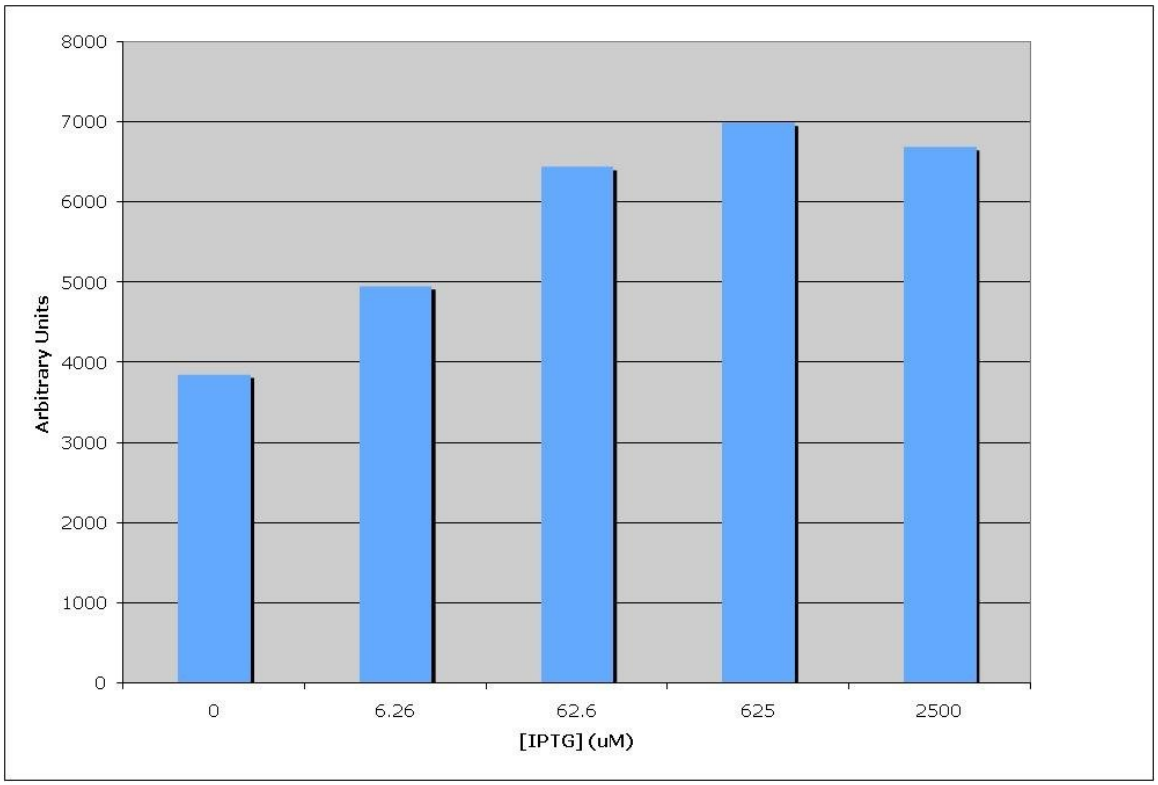


Figure 5.10. Quantification of autogenously regulated 2A cassette from western. We see clear induction at only 6.26 μM IPTG and a maximal 2-fold induction.

5.4 The Autogenously Regulated Genetic Switch in Different Cell Types

A secondary question is whether the switch functions correctly in different cell types. Ideally the switch would function well in every cell type as gene therapy has many potential targets: liver, muscle, eye, brain, etc. We therefore set out testing the switch to measure its functionality in two additional cell culture lines: COS-7 (**C**V-1 (simian) in **O**rigin and carrying the **S**V40 genetic material) which are fibroblast-like and derived from monkey kidney tissue, and ARPE-19 (**R**etinal **P**igment **E**pithelial) which are human derived RPE cells.

We transiently transfected the autogenously regulated 2A plasmid from Figure 5.8 into COS-7 cells. Figure 5.11 shows the quantified induction curve which again shows a 2-fold induction with IPTG.

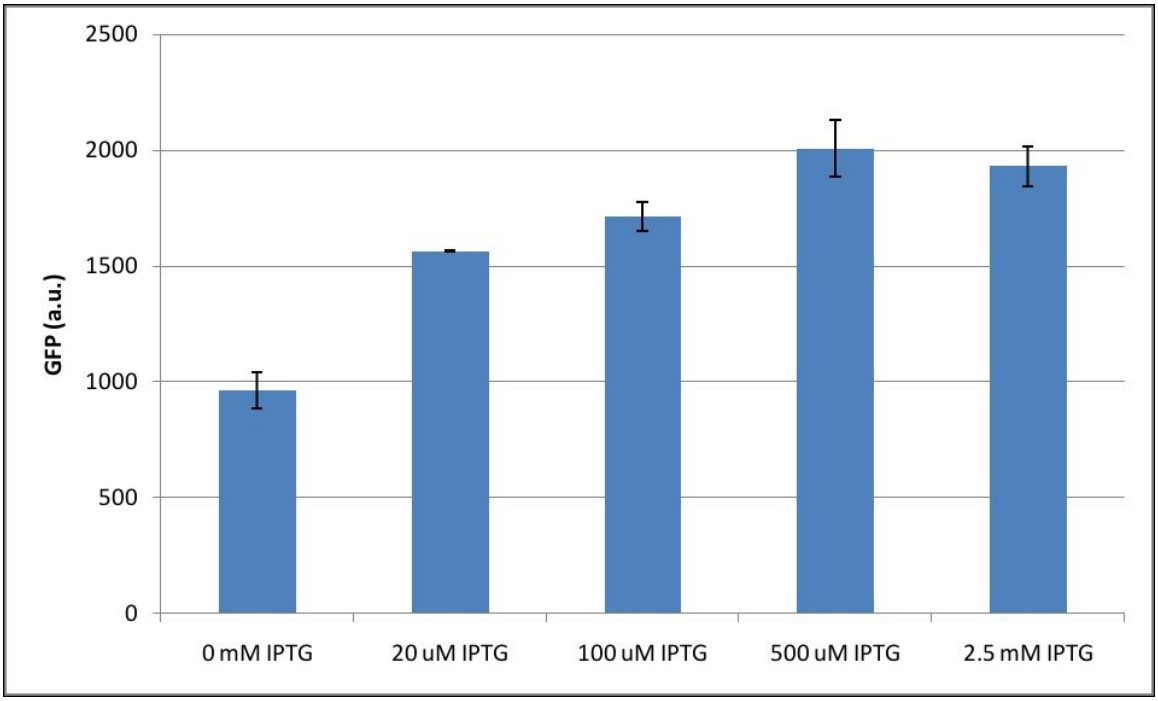


Figure 5.11. Induction profile of autogenously regulated 2A cassette in COS-7 cells.

We next wanted to quantitatively test the 2A autogenously regulated cassette in ARPE-19 cells. We made a large lac truncation from residues 18-290 to create a positive control cassette where the readout gene is unregulated. We then transiently transfected HEK293T cells with the autogenously regulated 2A cassette and separately transiently transfected HEK293T cells with the positive control plasmid. Figure 5.12 shows the normalized expression of the 2A autogenously regulated cassette. The switch is not very efficient in its current form; it only represses approximately half the constitutive expression and induces to 80%. Clearly, we have a lot of room to improve the switch.

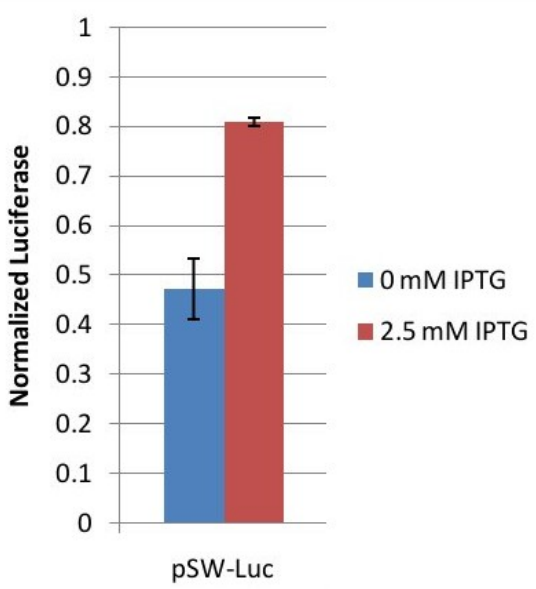


Figure 5.12. Normalized regulation of the autogenously regulated 2A cassette in ARPE-19 cells. The cassette poorly represses, reducing expression only down to ~ 45%. This switch approximately has two-fold induction. This switch would not be useful in AAV-mediate gene therapy.

The switch functions in three different cell lines and has approximately the same induction ratio (~2-fold) in all three cell lines. There appears to be no cell specific bias, however if the switch is to be used in a new setting it would strongly be recommended to test its functionality in cell culture first.

5.5 Using Tetrameric Lac Repressor to Improve the Autogenously Regulated Genetic Switch

The lac repressor construct we used has the 11 C-terminal residues of the lac repressor truncated to create a dimeric repressor. While we have used a dimeric repressor for clarity of understanding and studying allostery in Chapters 3 and 4, here we just want a more effective switch. With that goal in mind we sought to restore the tetramerization domain and multiple operator DNA sequences to restore the allosteric of lac repressor to two DNA operators.

First, we cloned a second Lsym operator DNA sequence 92 base pairs downstream of the first Lsym site (Figure 5.13). We did not necessarily expect this downstream site to be able to directly regulate transcription. We transiently transfected dimeric lac repressor with one Lsym operator (pSW-Luc) and dimeric lac repressor with two Lsym opeartors (pSW-Luc Lsym x2) into ARPE-19 cells and assayed for luciferase with and without IPTG (Figure 5.14). The addition of a second operator decreased both leakiness and maximal expression.

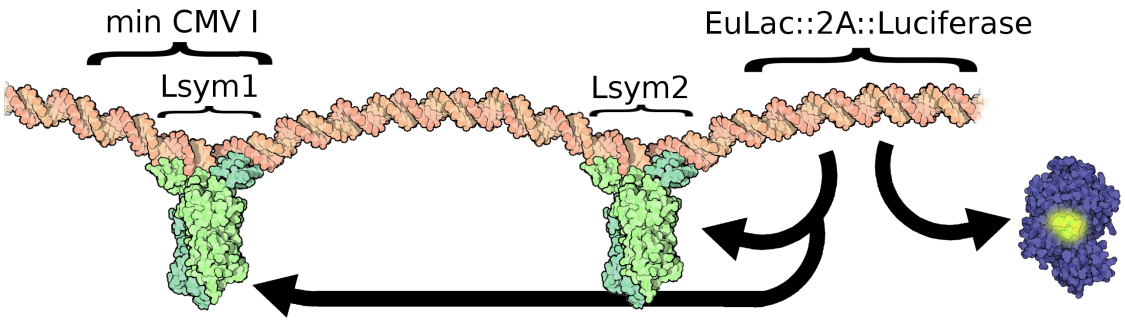


Figure 5.13. The pSW-Luc Lsym x2 plasmid. This plasmid expresses an autogenously regulated dimeric lac repressor construct which can bind two Lsym operator sites and has luciferase as a readout.

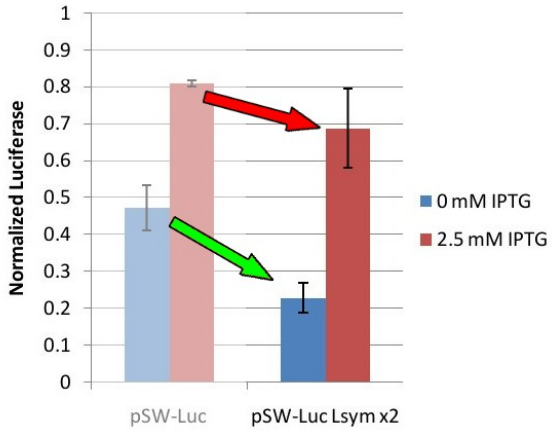


Figure 5.14. Effect of adding of a second Lsym operator with dimeric lac repressor. The pSW-Luc construct expresses dimeric lac repressor and has a single Lsym operator DNA. The pSW-Luc Lsym x2 construct has a second Lsym operator DNA site 92 base pairs downstream. The addition of this second operator decreases leakiness (green arrow) but also decreases maximal expression (red arrow).

Next, the 11 C-terminal residues were re-inserted into the *EuLac* gene to make what we will call *EuLacTet* to make the pSW2-Luc/GFP plasmids (Figure 5.15). This gene still has the 11 amino acid linker and NLS sequence following the full lac repressor itself, not to mention the partial 2A protein appended to the C-terminus. We first wanted to see if the *EuLacTet* protein could form tetramers after 2A breakage. We cloned the *EuLacTet* with a 6x Histidine tag (His-tag) into our standard pBAD expression vector and followed the standard lac repressor purification protocol. We also purified the original construct without the 11 C-terminal insertion which should be dimeric. These putative dimeric and tetrameric proteins were run on a native electrophoresis gel where oligomeric state will be preserved and stained for protein. The native gel confirmed that the tetrameric protein ran at twice the molecular weight of the dimeric protein (Figure 5.16).

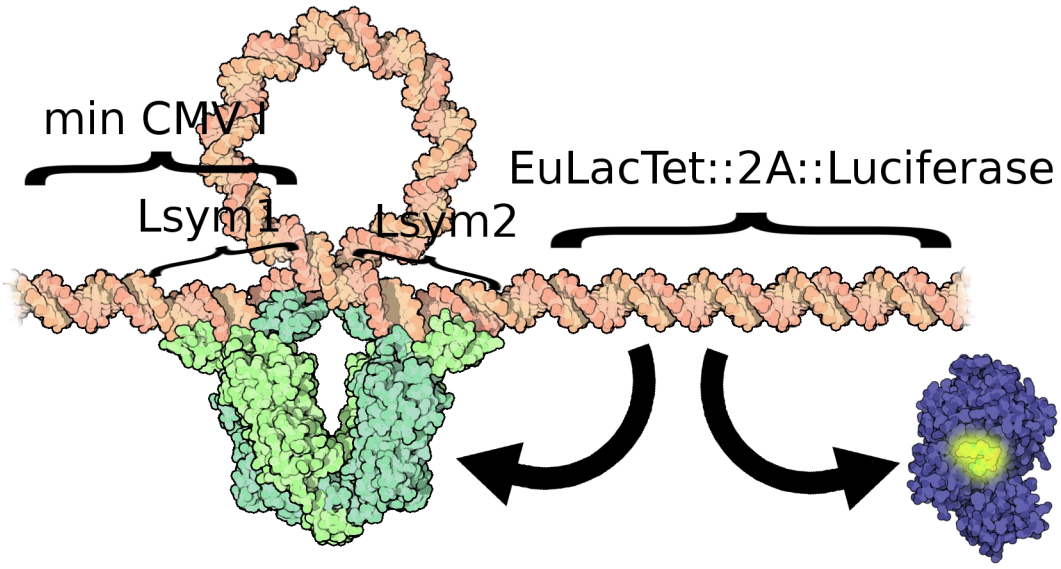


Figure 5.15. The pSW2-Luc plasmid. This plasmid expresses an autogenously regulated tetrameric lac repressor construct which can bind two Lsym operator sites and has luciferase as a readout. Two DNA operators are known to allosterically bind tetrameric lac repressor to increase the operator binding affinity and decrease repression. One model of this behavior is through DNA loops (pictured) wherein the spacing between the two operator DNA sites is critical.

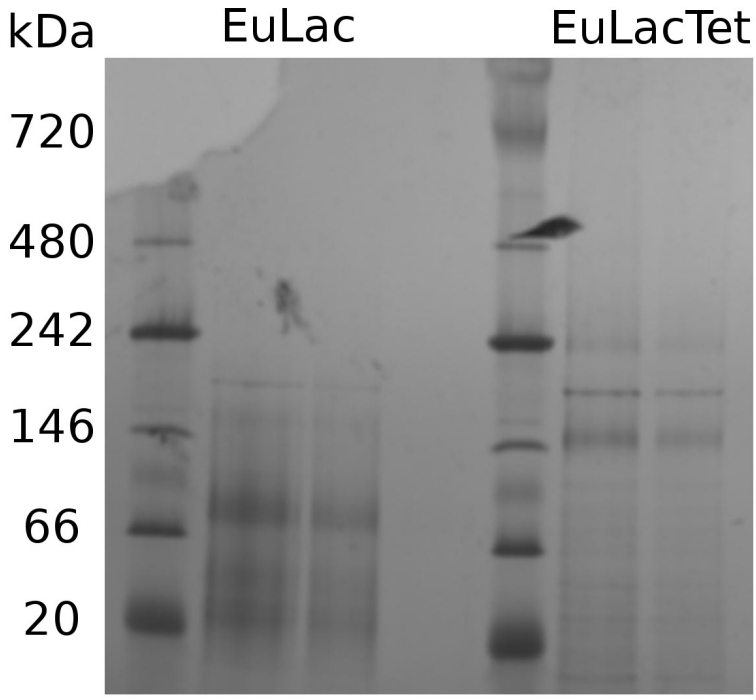


Figure 5.16. Native gel electrophoresis to discern oligomerization state. The EuLac construct which lacks 11 C-terminal residues known to form a coiled-coil domain has two primary bands near ~84 kDa and ~42 kDa which correspond to dimeric and monomeric species, respectively. The EuLacTet construct, which has the 11 C-terminal residues restored, forms a single tight band near the predicted ~174 kDa which corresponds with a tetrameric species. The highest band in both wells is a contamination from the purification. Both constructs are oligomerizing as designed.

We transiently transfected tetrameric lac repressor with two Lsym operators (pSW2-Luc) into ARPE-19 cells and assayed for luciferase with and without IPTG (Figure 5.17). The addition of the tetramerization domain decreased leakiness and increased maximal expression over the dimeric construct. Overall however, the potential cooperative effect of tetrameric lac repressor binding to two operators was less than has typically been seen in *E. coli*. This construct has an approximately 3-3.5-fold induction with IPTG and the leakiness is decreased to ~25% of constitutive expression. These are both significant gains from the starting pSW-Luc construct, however it does not appear

that it is necessarily due to the allosteric advantages of binding two operators with tetrameric lac repressor.

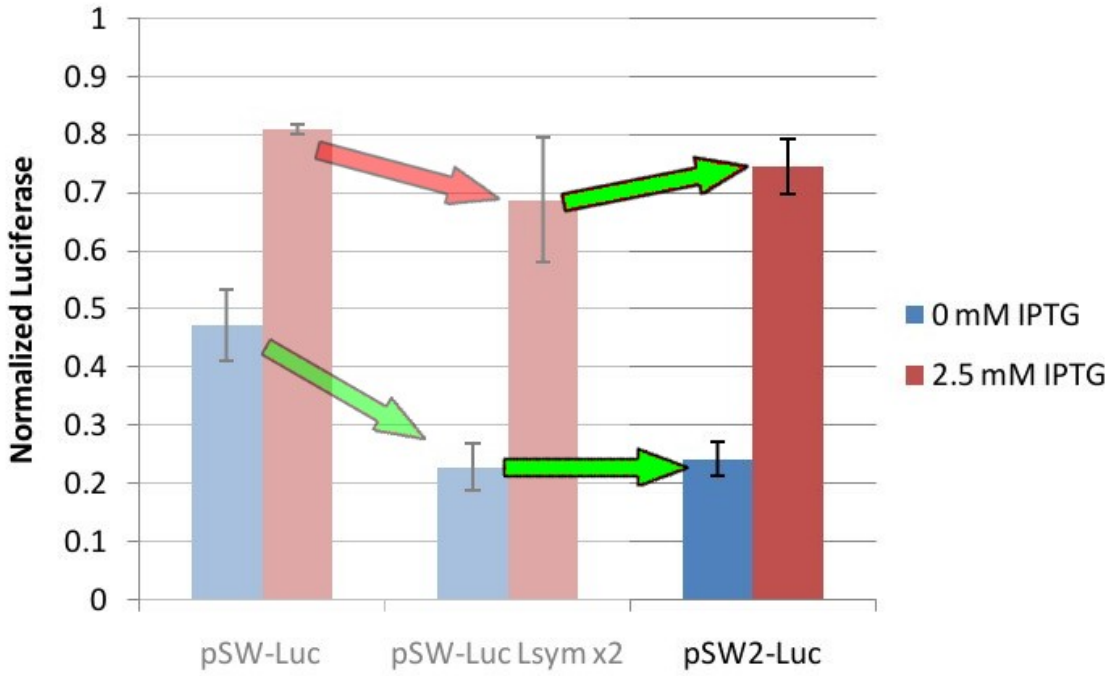


Figure 5.17. Effect of adding of a second Lsym operator with tetrameric lac repressor. The pSW2-Luc construct expresses tetrameric lac repressor and has two Lsym operator DNA sequences. The decreased leakiness is maintained (lower right green arrow) and the maximal expression is marginally improved (upper right green arrow). Overall a significant improvement from the starting pSW-Luc construct has been obtained.

To measure if the tetramerization is playing a significant role we sought to repeat a classic experiment from *E. coli*. It is known that the spacing between the two operator DNA sites critically effects the ability of lac repressor to regulate gene expression resulting in a periodicity of optimal spacing (Müller et al., 1996). Optimal spacing was found at 70.5, 81.5 and 92.5 base pairs with significant drop offs in repression by altering the spacing even 1 or 2 base pairs. With this in mind we inserted or removed one or two base pairs between the two operator sites and also changed to spacing entirely to 70 base pairs. We tested these constructs in transient transfections of ARPE-19 cells with

and without IPTG (Figure 5.18). There was no significant difference in measured induction ratios between these constructs. Either the cooperative effect is minimal in cell culture or further construct optimization is needed.

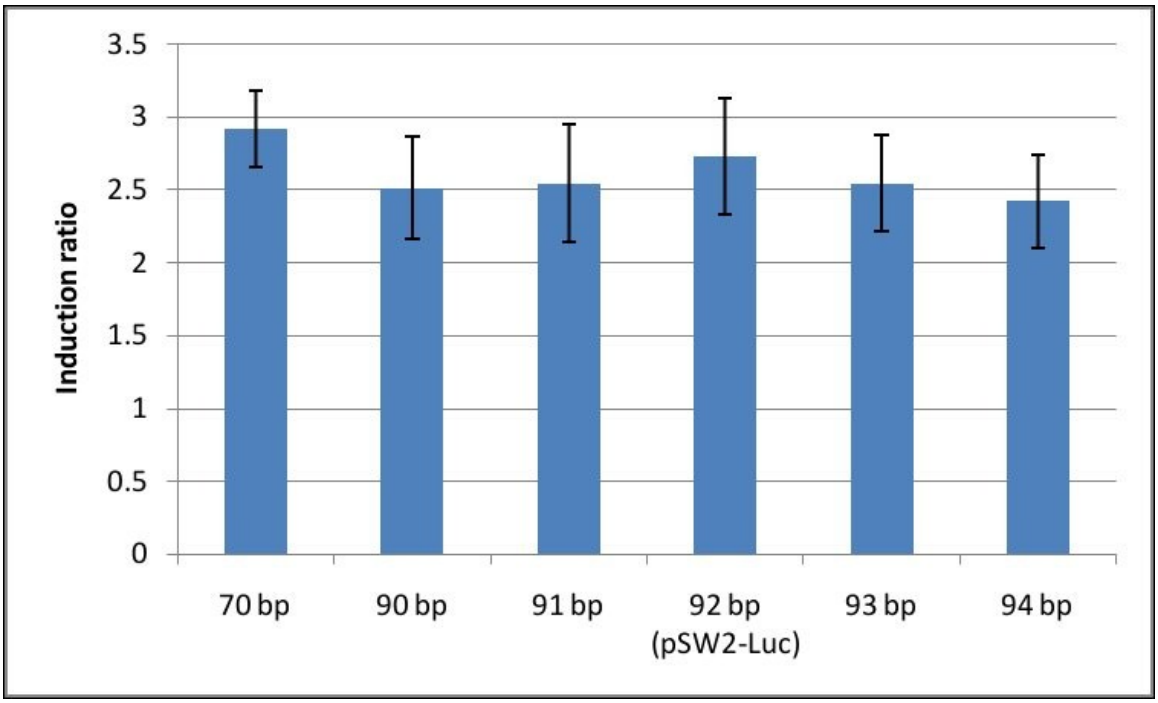


Figure 5.18. Alternate Lsym operator DNA base pair spacings. It has been observed in *E. coli* that optimal repression by tetrameric lac repressor occurs when the two operator sequences are properly spaced. We found no significant change in induction ratio with alternative spacings.

We did reduce leakiness and increase dynamic range by using tetrameric lac repressor and two operator DNA sites. It is unclear however if the system is working as has been previously reported in bacteria.

5.6 Using Lac Repressor Mutations to Improve the Autogenously Regulated Genetic Switch

We have ample evidence that point mutations of the lac repressor can decrease leakiness, increase dynamic range, and change the affinity for effector molecules such

as IPTG. We selected two mutants from Chapter 3 that should improve our eukaryotic autogenously regulated switch: Q18M which decreases leakiness and dynamic range and F161W which decreases leakiness and improves dynamic range.

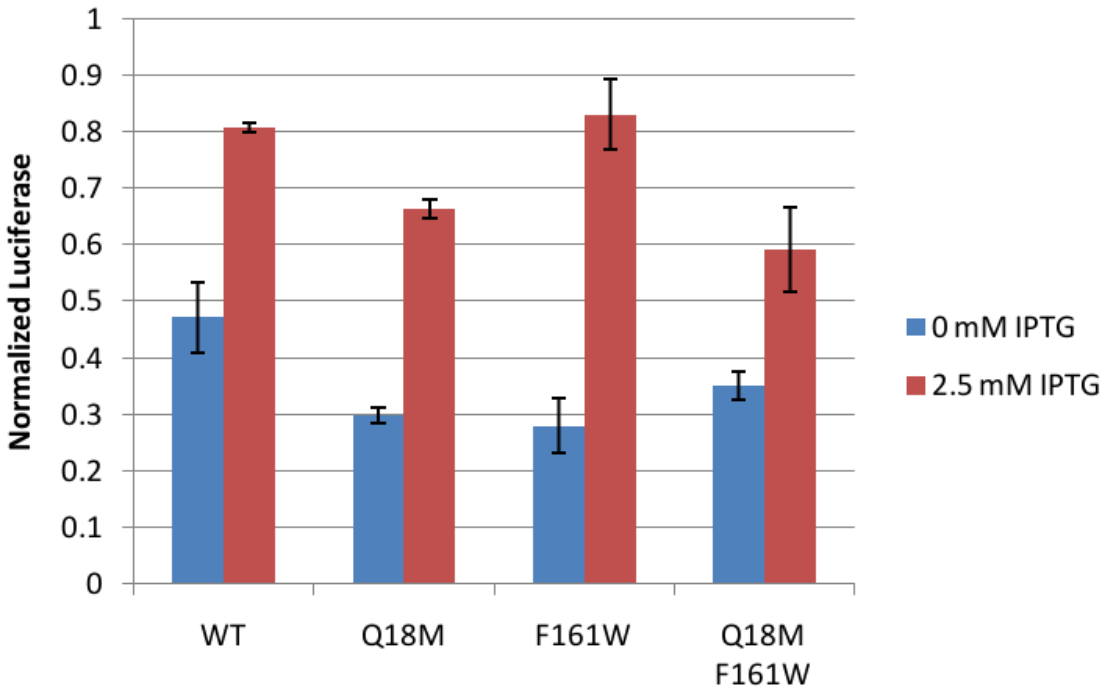


Figure 5.19. Point mutations of dimeric lac repressor with one Lsym operator (pSW-Luc). Q18M decreases leakiness from wild type (WT) lac repressor but it also decreases the maximal induction. This is the classic tradeoff between leakiness and dynamic range that was studied in Chapters 3 and 4. F161W decreases leakiness below that even of Q18M and maintains the same maximal induction. The combination mutant was less leaky than WT but leakier than either point mutation alone, and it had the worst dynamic range.

We tested Q18M and F161W and the combination mutant Q18M F161W in the pSW plasmid in ARPE-19 cells with and without IPTG. We found that both the Q18M mutation and the F161W mutation functioned as has been previously found in *E. coli* in the non-autogenous genetic switch (Figure 5.19). Q18M exhibited the classic tradeoff between decreased leakiness and decreased dynamic range that was found in Chapter 3 when only the repressor-operator DNA affinity, K_{RO}, is increased. F161W had both

decreased leakiness and increased dynamic range, just as was found in bacteria, due to its simultaneous changes to repressor-effector affinities and conformational equilibrium. We were unsure if the mutations would have additive effects; the Q18M F161W double point mutant shows that the two mutations do not work in concert. We hypothesized in Chapter 3 that Q18M does not primarily alter K_{RO} but instead the conformational equilibrium K_{RR^*}, which is also altered by F161W. If that is the case it is unsurprising that the two mutants do not work together.

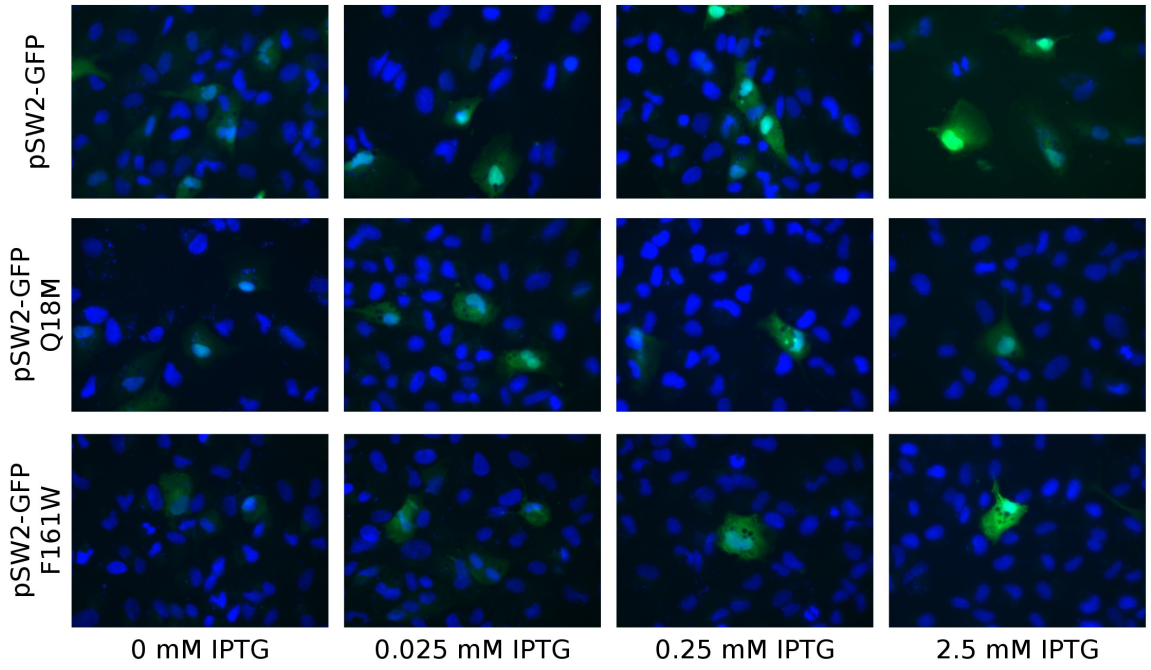


Figure 5.20. Microscopy of point mutants of pSW2-GFP. GFP microscopy shows that there is a clear induction of GFP with increasing IPTG concentration. The intensity of the Q18M point mutant is significantly less both at no IPTG (basal leakiness) and maximally induced with 2.5 mM IPTG. This is more consistent with what we saw with the dimeric cassette pSW-Luc. F161W looks very similar to wild type repressor. Of note is the unexpected increased GFP signal in the nucleus (DAPI stained, blue).

We next tested the two single mutants with the tetrameric cassette. The luciferase assay does not have a control for number of cells and therefore is not ideal for absolute comparison between different plasmids. It does function very well for

establishing intra-plasmid behavior (i.e. measuring IPTG induction ratios). With that in mind, we decided to measure the Q18M and F161W mutants in the tetrameric cassette regulating GFP (pSW2-GFP). We transiently transfected ARPE-19 cells and grew the cells in varying concentrations of IPTG. Cells were fixed and GFP was measured by fluorescent microscopy (Figure 5.20). Q18M was seen to be less leaky than both wild type and F161W and similarly has a decreased maximal expression. F161W did not appear to have the improved leakiness and dynamic range that was measured for the dimeric cassette.

Point mutations to the lac repressor appear to be potentially useful in improving the phenotypic qualities of the autogenous genetic switch, however their application is not straightforward due to the more complicated nature of the switch. Both Q18M and F161W functioned similarly in the dimeric autogenous switch with a single Lsym operator, but only Q18M worked as designed in the tetrameric autogenous switch with two Lsym operators. This approach can certainly be used to further improve the switch but point mutations will have to be verified to be functional in the each autogenous switch design.

5.7 Comparing with Tet-On Technology

The gold standard for prokaryotic derived eukaryotic gene regulation is the tet repressor derived tet-On and tet-Off systems. To see if our levels of leakiness and dynamic range are comparable with these systems we first set out to make a single plasmid tet-On system. We did this by isolating the tet-On activator and its promoter from the pTet-On plasmid and inserting it into the pTRE-GFP plasmid. This plasmid will constitutively express Tet-On activator and control the expression of GFP through the Tet

Regulatory Element (TRE) via the effector doxycycline; we called this plasmid pTRE-On-GFP.

ARPE-19 cells were transiently transfected with pTRE-On-GFP side by side with the pSW2-GFP and mutants (Figure 5.21). The pTRE-On plasmid clearly induces GFP upon addition of doxycycline. The inherent leakiness and dynamic range of the pTRE-On plasmid is very similar to the wild type tetrameric pSW2 plasmid. It seems to induce to a slightly higher level. The Q18M tetramer is less leaky than the pTRE-On which is encouraging if minimal leakiness is required.

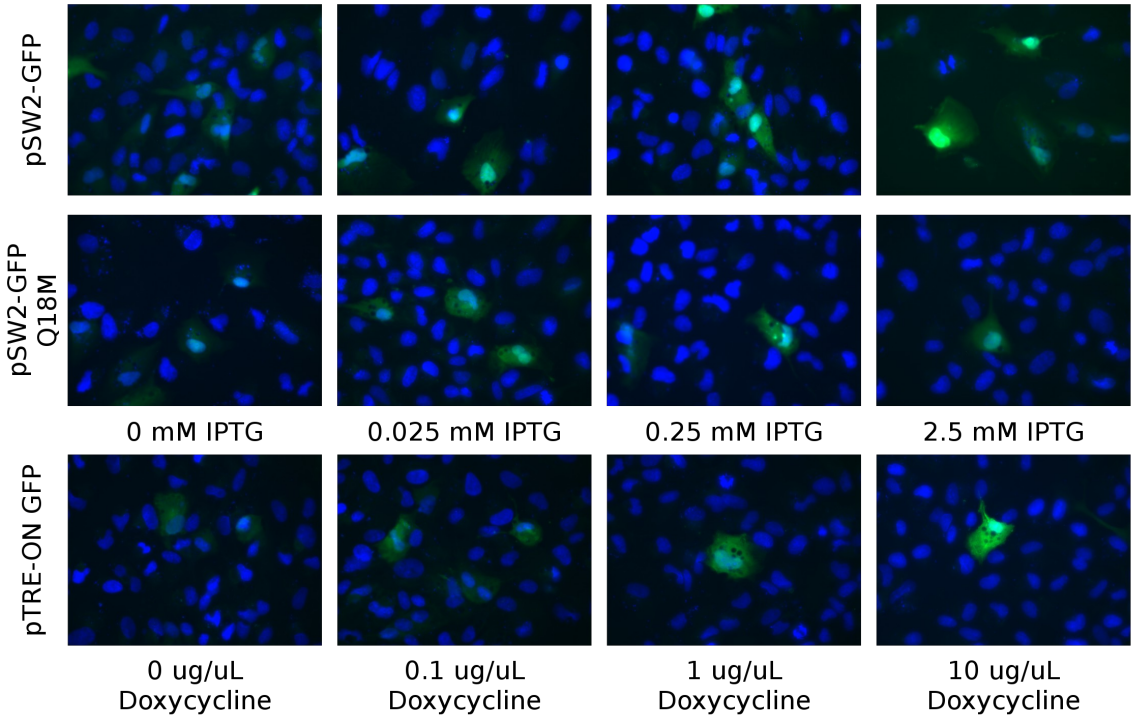


Figure 5.21. Microscopy of point mutants of pSW2-GFP and pTRE-On-GFP. Our autogenously regulated switch performs very similarly to the tet-On technology. Q18M is actually less leaky than tet-On despite the limits of autogenous regulation. The dynamic range of the tet-On switch however appear to be superior. Note the lac of nuclear localization of GFP in the TRE-On system.

One thing that was revealed by the microscopy was an unexpected localization of GFP to the nucleus. This was much more prevalent at the highest IPTG

concentrations of the pSW2 plasmids (and hence highest GFP concentrations). This could be due to un-broken Lac repressor::2A::GFP fusion proteins being localized to the nucleus through the NLS signal on the repressor. This certainly will have a confounding effect on the analysis. It is unknown if for some reason the 2A is breaking less efficiently in this experiment (although no IPTG dependence was noted in the western). It is also unknown if the repressor::2A::GFP protein is dimeric or tetrameric which could be impeding the ability of tetrameric lac repressor to allosterically bind the two Lsym operators. Finally, the half-life of the fusion protein in the nucleus versus the half life of GFP in the cytoplasm will certainly have some effect on the quantification of these experiments.

The TRE-On plasmids do not have this concentration of GFP in the nucleus. There is no 2A site or NLS associated with the GFP on the TRE-On plasmid in anyway. This strongly hints that the high nuclear GFP signal in the lac repressor expreriments is due to the 2A and/or NLS signals.

5.8 Testing the Autogenously Regulated Genetic Switch in Mice

The end goal of this work is to develop a lac repressor based genetic switch that is usable in AAV-mediate gene therapy. To move towards that goal we wanted to test our autogenously regulated cassette in a mouse model.

We inserted the cassette from the pSW2 plasmid which has two Lsym operators and tetrameric lac (Figure 5.15) into a plasmid which will make serotype 8-AAV virus. Virus was made at a virus core facility and confirmed to function in HEK293T cells.

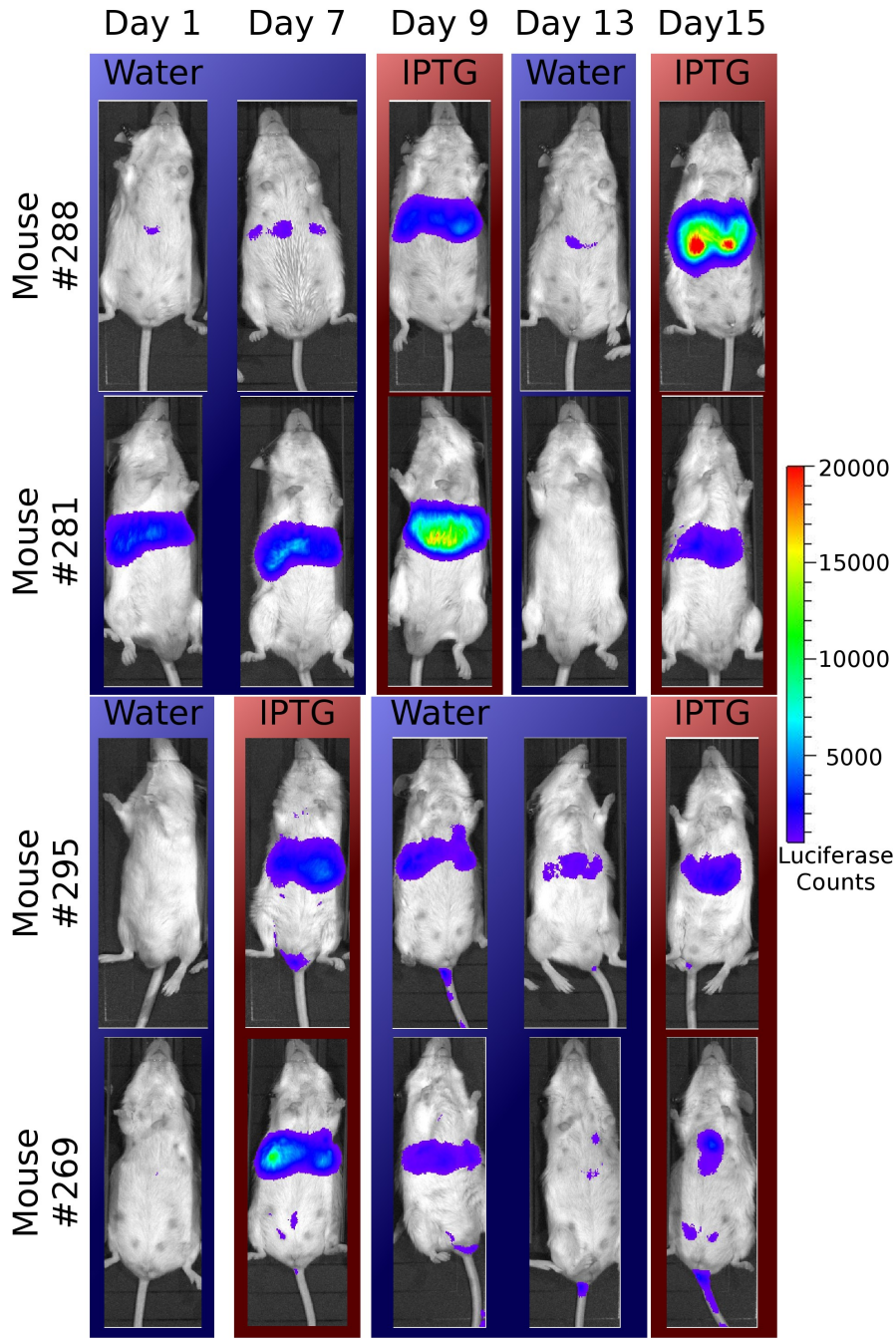


Figure 5.22. Autogenously lac genetic switch controlling luciferase in mice. Mice were live images for luciferase signal after being either on normal water (blue boxes) or after being gavaged with 1M IPTG (red boxes). There is significant mouse to mouse variation, however we always see potent induction after IPTG gavage. Luciferase is seen to drop within 2 days of water feeding and re-induce. The autogenous lac genetic switch is a reversible regulator of a transgene in mice.

We performed tail-vein injections of 12 nude mice with SW2-Luc (same cassette as pSW2-Luc (Figure 5.15)) virus (Multiplicity of Infection (MOI): 1e11) to target the liver. The 12 mice were split into two cohorts of 6 to differentiate between time dependent changes in luciferase signal and IPTG dependent changes in luciferase signal. Non-injected mice were also included as a control and showed no measurable luciferase signal.

Four mice died in the study, either due to complications with gavage or unknown circumstances. Of the remaining mice, two from each cohort appeared to have luciferase correctly targeted to their livers in the imaging. Many mice had significant luciferase signal in muscle tissue in the tail and surrounding region and were excluded from further analysis. Mice were fed sequential diets of either water or gavaged (force feeding) 1M IPTG. We previously confirmed that gavage of 1M ITPG provides concentrations of IPTG in excess of 100 μM in brain, retina, vitreous, muscle, liver, kidney, serum and urine. Mice were given intraperitoneal (IP) injections of luciferin and scanned on a live animal imaging scanner throughout the experiment (Figure 5.22).

All four animals have low but measurable levels of luciferase one month after the tail-vein injection. There is noticeable animal to animal variation with mouse #281 being particularly leaky. The mice were not age and gender matched and they were all given the same amount of viral vector regardless of weight which could account for some of the variation.

The first cohort (Mice #288 and #281) were initially kept on a normal water diet throughout the first week (Figure 5.23). There is a ~2-fold increase in luciferase signal throughout this time. The second cohort (Mice #295 and #269) were gavaged two times

a day for three days; there is a potent increase (>20-fold increase) in luciferase signal in these two mice (Figure 5.24).

The first cohort was then gavaged with IPTG and saw a ~4-fold increase in luciferase. Meanwhile the second cohort was put back on regular water and saw a ~3-fold decrease in signal after 2 days and ~8-fold decrease after 6 days. Both cohorts were then re-gavaged with IPTG to re-induce and there was a wide range of luciferase signal increase from ~3-fold to ~30-fold. This re-induction confirms that the mice still have AAV infected cells capable of regulating luciferase signal through the lac autogenous genetic switch.

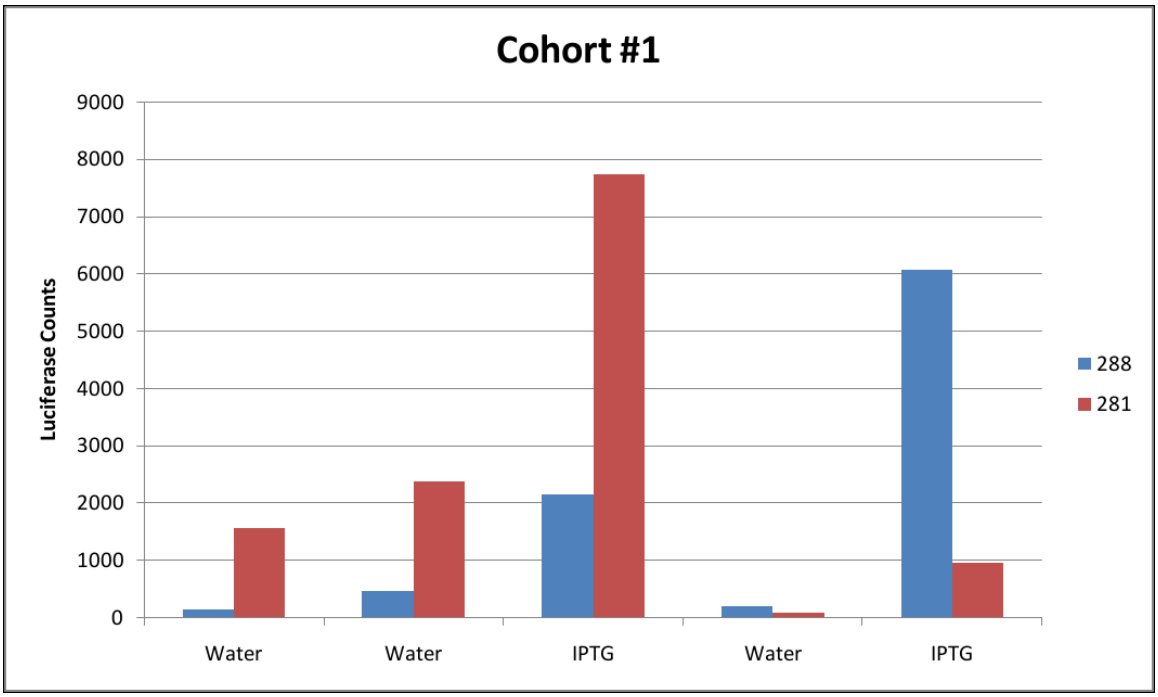


Figure 5.23. Quantification of luciferase counts for cohort 1 mice. The overall signal level between the two mice is quite different but the same trend exists: luciferase signal increases upon IPTG gavage and is low on a water diet.

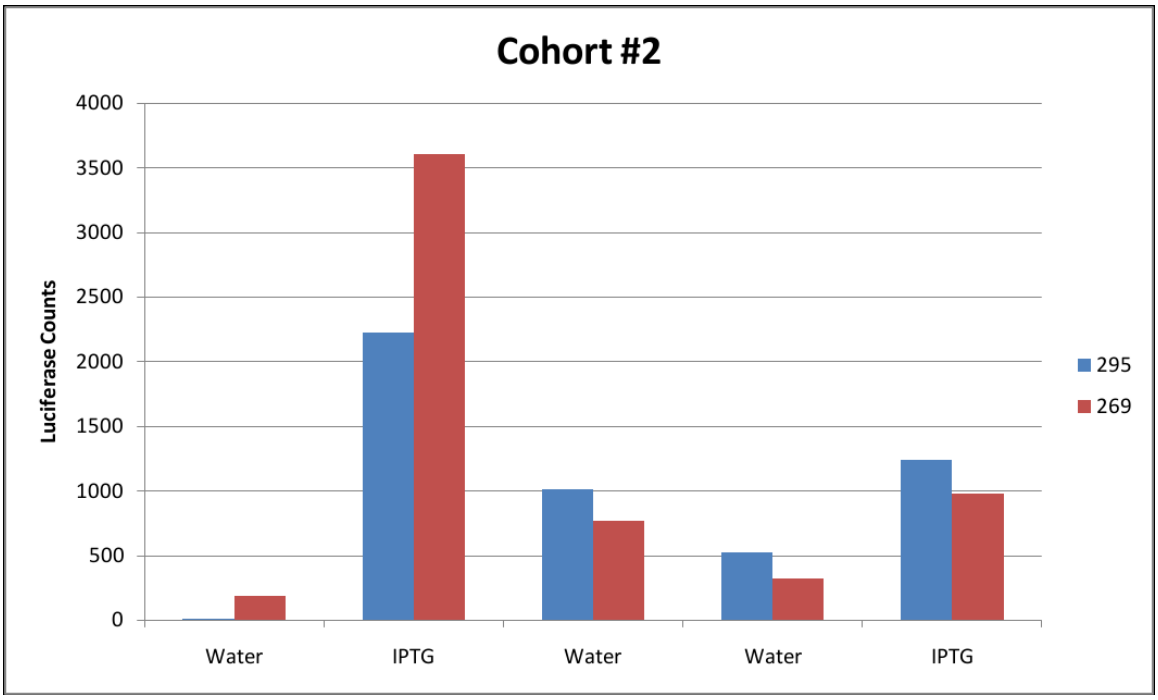


Figure 5.24. Quantification of luciferase counts for cohort 2 mice. Again the expected trend is observed: luciferase signal increases upon IPTG gavage and is low on a water diet.

Since both cohorts were induced with IPTG and repressed upon removal of IPTG during water feeding, we averaged across cohorts (Figure 5.25). The mouse to mouse variability of luciferase counts is very large which could be due to several factors: we did not age and sex match the mice and we injected the same amount of viral vector regardless of mouse weight.

We then used the averages to measure first induction ratio, repression ratio with removal of IPTG (water feeding), and second induction ratio (Figure 5.26). Again there is a very large variability but the trend is striking; there is an ~8-fold first induction, ~14-fold repression, and ~8-fold second induction.

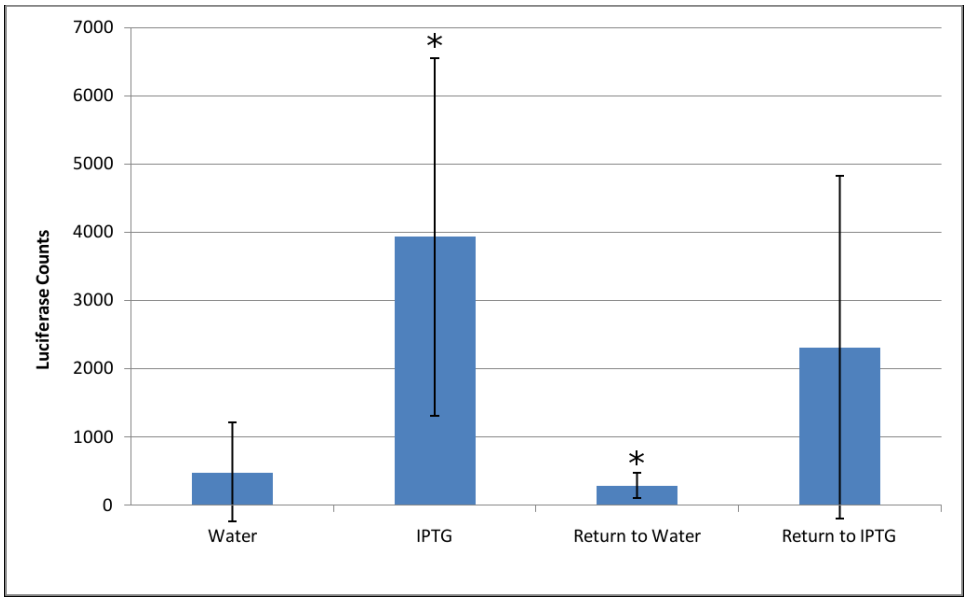


Figure 5.25. Average luciferase counts for mice throughout the experiment. The luciferase signal significantly increases with the first gavage ($p < 0.05$) compared with the initial water readings and significantly decreases upon returning to water ($p < 0.05$). The luciferase signal increases a second time upon return to IPTG gavage, however the change is not significant due to the large variation between mice.

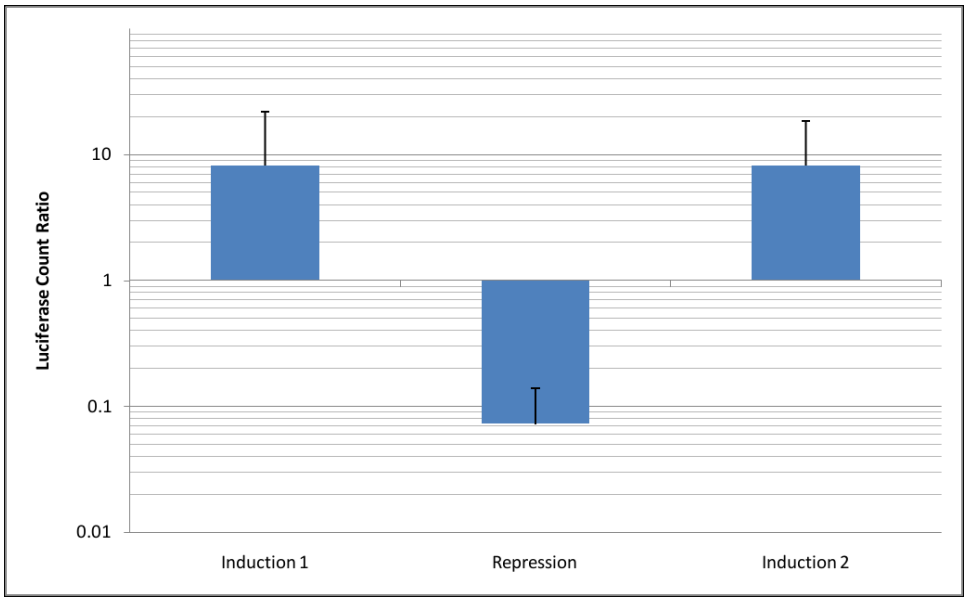


Figure 5.26. Luciferase ratios upon switching from water to IPTG (Induction) or IPTG to water (Repression). A ratio of 1 would be no change upon switching to or from IPTG. Both the first and second induction had large luciferase signal increases with ~8-fold increases. The repression was stronger with a ~14-fold decrease in signal upon removal of IPTG.

The repression ratio is notably larger than the first induction ratio. Since the lac repressor is autogenously regulated in this switch, we are inducing not only luciferase but also lac repressor; more lac repressor will be present in the nucleus after induction. Upon removal of IPTG from the diet this excess repressor can and will bind the two operator DNA sites and repress transcription. Since the lac repressor concentration will exceed the equilibrium concentration before induction it is not surprising that the switch would go even more “off”.

5.9 Discussion

The goal of this chapter was to develop a lac genetic switch that was functional in AAV-mediated gene therapy. We showed that an autogenously regulated lac repressor, which is small enough to pack into an AAV viral capsid along with a transgene, is functional in both prokaryotes and eukaryotes. Furthermore, the eukaryotic switch can function with either the IRES or 2A sequence to mediate the expression of both the lac repressor and the transgene from a single promoter. We showed that the eukaryotic autogenous switch is functional in HEK293T, COS-7, and ARPE-19 cell culture as well as live mice. The switch appears to be broadly applicable to different cell types.

The first generation of autogenous switches were not particularly efficient gene regulators; only repressing about half of the constitutive level and having a 2-fold induction ratio. We showed that the lac repressor point mutations from Chapters 3 and 4 also change the gene regulatory phenotype in the autogenously regulated lac genetic switch in prokaryotes. We then showed that these same point mutations can be used to decrease the leakiness of the switch or to increase the dynamic range. Finally, we found

that multiple DNA operator sites and a tetrameric lac repressor could both decrease the leakiness of the switch and increase the dynamic range.

The autogenous switch was shown to function from transductions of AAV virus-encoding transgenes in cell culture and in living mice. The living mice were stably expressing the AAV-encoded luciferase nearly 1 month post infection and were capable of reversibly regulating the expression of luciferase in their liver.

A curious finding is that the induction ratios measured in mice actually surpass those from cell culture. The nature of our transient transfection experiments could partially be to blame. The time window in which the cells are dense enough for experiment and not too dense to have contact inhibition effects is small. The cells may not have had enough to time reach an equilibrium before we were required to measure their transgene signal. Regardless, it is very encouraging that the switch appears to function better in living organisms which of course is much more relevant to our end goal of developing a regulator for AAV-mediated gene therapy.

Chapter 6
Conclusions

The lac repressor hasn't changed, but its measured properties have (Sharp, 2011). This is due to the wide variety of techniques, buffers, constructs, etc. that have been applied over the years in order to best measure the property of interest given the best technique of the day. This thesis sought to bring together this often discordant information to link protein structure to genetics to *in vivo* phenotype to *in vitro* measurements all through the use of the MWC model of allostery.

We have found that point mutations of certain amino acids can specifically alter repressor binding affinities to effectors and that these changes agree with the structural location of those amino acids. Two residues in the specificity region of the DNA binding domain (Y17 and Q18) make specific contacts with positions 5 and 6 of the DNA operator sequences. Point mutations to these two residues (Y17I and Q18A) result in gene regulatory phenotypes that are well explained by reductions to the repressor-operator DNA affinity, K_{RO}. All other thermodynamic parameters are unaffected; mutants bind effector with the same affinity and have the same conformational equilibrium. One residue in the specificity region of the effector binding domain (Q291) makes specific contacts with the substituent group of thiogalactosides. A point mutation to this residue (Q291K) results in a gene regulatory phenotype that is well explained by reductions to the repressor-effector affinities, K_{RE} and K_{R^*E}. All other thermodynamic parameters are unaffected; this mutant binds operator DNA with the same affinity and has the same conformational equilibrium.

These three point mutations are a wonderful confirmation of the MWC model of allostery for the lac genetic switch. Many mathematical functions can recreate the curve shape of proteins binding to ligands, however they are just that: functions. They have no basis in biophysical reality other than they happen to look like reality. One could throw a ball and see that the trajectory looks a lot like a parabola; the fact that the physics of throwing the ball predict a parabolic trajectory give the mathematical function meaning. Now we can throw a ball, fit it to a parabola, and use that fit to measure the initial velocity of the throw. Analogously, we see the normalized expression from the lac genetic switch as a function of effector, we fit it to the curve shape predicted by the MWC model, and use that to fit the inherent thermodynamic parameters of the lac repressor. The point mutations Y17I, Q18A, and Q291K show that we can use the structure of the lac repressor to specifically perturb individual binding affinities. The MWC model of the lac repressor splits the binding affinities into individual properties that could be independently manipulated to create unique gene regulatory phenotypes; these three point mutations exactly fall into that model. Structure agrees with genetics agrees with *in vivo* phenotype given the MWC model of allostery.

Now of course we studied much more than these three point mutants. The lion share of point mutations were seen to perturb the conformational equilibrium between the two primary structural conformations of the lac repressor, R and R*. Furthermore, we saw that the conformational equilibrium could be manipulated from a variety of locations in the protein: Q18, L148, F161, Q291, and L296. Using the Boltzmann we saw that the wild type lac repressor only has an energetic difference of about 1 kcal/mol between the R and R* state (the R* state has lower energy and is favored). It is not surprising therefore that even a single point mutation could alter the conformational

equilibrium. We measured conformational equilibrium parameters of lac repressor point mutations that changed the energy between the two states as much as 3 kcal/mol (in favor of the R* state for L148D and L148W; in favor of the R state for Q291I and L296W). The Boltzmann distribution is sensitive to the exponential of the energy; therefore even relatively small energy differences have profound changes on the populations of the two conformations. For these point mutations we see the conformational equilibrium change nearly 100-fold which has profound effects on the phenotype of the lac genetic switch.

The observation that the conformational equilibrium is changed upon point mutation may be the strongest test of the MWC model of allostery. The entire basis of the MWC model is that the net function of a population of proteins can be altered by shifting the population to a different conformation with altered properties. Multiple structural conformations with different properties is inherent in this model. Consider instead a model of the lac repressor with only one conformation, R, which can bind operator DNA with an affinity K_{RO} and effector with an affinity K_{RE}. We then mutate the effector binding pocket (F161W for example) and see that the genetic switch is drastically less leaky: how is this possible? The point mutation is nowhere near the DNA binding domain so K_{RO} should be unaffected. The lac repressor still induces upon ITPG, better even than wild type (Figure 3.10), so the lac repressor itself must be folding correctly and functioning. The only other explanation is that there is more repressor, however we measured the total repressor concentration and found no change from wild type *in vivo* (Figure 3.14). Additionally, more repressor would impede maximal expression. A single conformation model simply cannot explain the F161W mutant given the data on hand. The MWC model of the lac repressor perfectly explains the *in vivo*

phenotype. The conformational equilibrium between the two states is tipped 1.5 kcal/mol in favor of the R state reducing the leakiness. Meanwhile the mutation is in the effector binding pocket and increases the affinity for IPTG for the R* state relative to the affinity for the R state; adding IPTG more greatly populates the R* state than wild type leading to greater induction. The structural location of the F161W mutant combined with the *in vivo* phenotype can only be explained by an MWC model with at least two structural conformations.

Of note are four point mutations from the literature: A110T, A110K, L148F, and S151P (Müller-Hartmann & Müller-Hill, 1996; Swint-kruse et al., 2003). None of these mutations are within the DNA binding domain of the lac repressor yet all four were found to alter the net repressor-DNA affinity; all four altered the leakiness of the switch. Unlike the targeted residues from this study, A110 and S151 are not part of the effector binding pocket; both are within the N-terminal region of the core domain. The structure clearly indicates that neither residue contributes to DNA binding or effector binding directly, yet point mutations to both regions alter both parameters simultaneously. Again, the conformational equilibrium parameter comes to the rescue. Figure 3.4 shows that, given thermodynamic binding values similar to wild type lac repressor, increasing the conformational equilibrium (favoring the R* state) creates a leakier genetic switch which responds to less effector; *effective* DNA affinity is lowered and *effective* IPTG affinity is increased. This is exactly what is observed for A110T and L148F. Conversely, if the conformational equilibrium is decreased (favoring the R state) then the genetic switch is less leaky and needs more IPTG to induce; *effective* DNA affinity is increased and *effective* IPTG affinity is decreased. This is exactly what is observed for A110K and S151P. The structural location of these residues is perfectly consistent with the ability to

alter the stability of the repressor conformations and hence change conformational equilibrium upon point mutation. A conformational equilibrium between the two states that is close to one is ideal as it is most easily changed by single point mutations. Wild type lac repressor has a conformational equilibrium of approximately 6-fold in favor of the R* state which corresponds to only a 1 kcal/mol energy difference between the two states. The conformational equilibrium is easy to perturb in the lac repressor due to the small energy difference between its two conformations.

The other effector pocket mutants showed that the interplay between the conformational equilibrium, the magnitudes of the repressor-effector affinities, and the ratio of the two repressor-effector affinities define the resultant phenotype. The DNA binding domain mutants showed that conformational equilibrium and repressor-DNA affinity could be altered but they had the same resultant phenotype. The effector binding pocket turned out to be much more interesting in terms of potential phenotypes upon point mutation. So why did point mutations to the DNA binding domain give such a one dimensional response? It is because the output (transcription and translation of a gene) is linked to the occupancy of the operator DNA by repressor. RNA polymerase does not care if IPTG is bound to the repressor but it does care if the IPTG bound repressor drastically lowers the occupancy of DNA. You can alter the phenotype of the lac genetic switch through mutating the DNA binding domain but to get more ideal lac genetic switches (lower leakiness, higher dynamic range, lower $[E]_{50}$) you need to target the allosteric effector site, namely the effector binding pocket.

Many novel lac repressor genetic switches have been identified with properties that could potentially be impactful. One obvious future direction would be the use of

tighter binding lac repressors in the regulation of toxic proteins being recombinantly expressed.

Utilizing genetics of the lac repressor to link structure to *in vivo* phenotypes with the MWC model was a powerful technique. The next obvious "missing link" in the lac repressor literature was the connection between *in vitro* data derived from purified compenents and *in vivo* gene regulation taking place in living bacteria. We can measure binding affinities via fits of the *in vivo* data to the MWC model but if our measured affinities do not agree with the isolated proteins, DNA and effectors we are claiming to model we are potentially missing key components of the system. Furthermore, it is not a foregone conclusion that *in vivo* results will agree with *in vitro* predictions. One group found a 25-fold discrepancy for a hybrid repressor between *in vivo* and *in vitro* data (Tungtur et al., 2011). For the lac repressor itself, one study found very good agreement between published *in vitro* values and *in vivo* values (Sharp, 2011). There was however a major limitation to the data of the lac repressor study; the concentration of the lac repressor itself was not taken into account. We did measure the lac repressor concentration to be around 650 nM *in vivo* which agrees with previously reported values for our promoter and the estimated copy number. We would expect wild *E. coli* with a single *lacI* promoter expressing lac repressor to have ~10-20 fold less repressor (the copy number of our plasmid) which corresponds to 30-65 nM lac repressor *in vivo*. Using the values from Table 4.1 and the assumption free solution to the MWC model from Chapter 4 we found that the thermodynamic binding values curated by Sharp could not regulate gene expression because repressor-DNA affinities were too high (Figure 4.5). This is not an artifact of the more complicated model solution. If we simply take the repressor-operator DNA affinity values and concentration of the repressor we can

easily calculate the r parameter ($r = K_{RO}\,[R]_{tot}$): Chapter 4, r = 278; Sharp set 1, r = 2.2e6; Sharp set 2, r = 6.6e4; Sharp set 3, r = 1e6. Figure 6.1 shows that the three curated sets from the Sharp paper all are way over-repressed. We can, as before, model what is occuring in natural *E. coli* by estimating $[R]_{tot}$ = 30 nM which reduces the r values 20-fold and should lessen the drastic over-repression. Figure 6.2 shows that only Set 2 could induce with IPTG given the thermodynamic parameters.

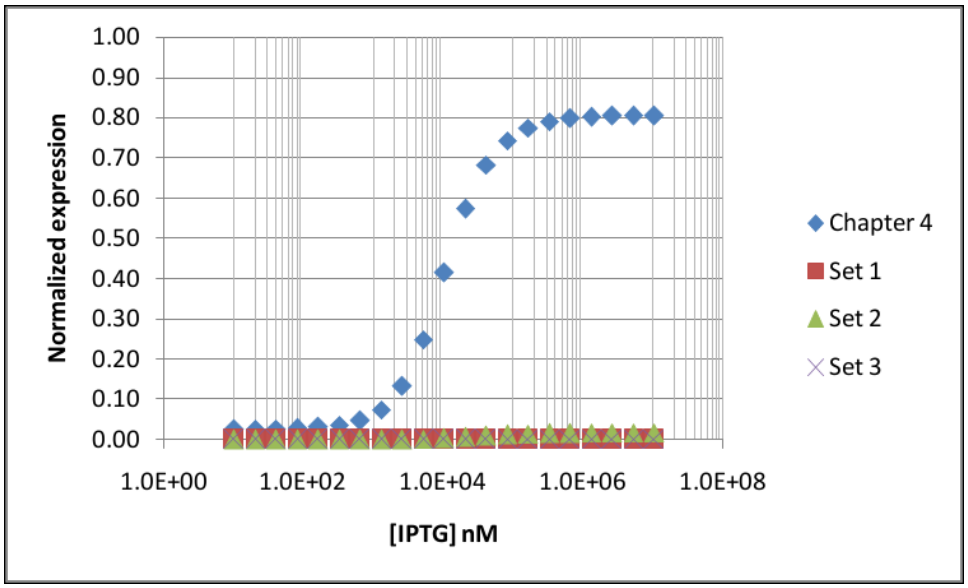


Figure 6.1. Simulated induction curves from various *in vitro* derived values at a repressor concentration of 650 nM. Chapter 4 are the values measured from the *in vitro* transcription experiment. Sets 1, 2 and 3 are derived from the published values in (Sharp, 2011). The *in vitro* values from the literature are drastically over-repressed.

The *in vitro* measured values of Chapter 4 agree with previous *in vitro* measurements for repressor effector binding and conformational equilibrium. The repressor-operator DNA affinity was several orders of magnitude weaker than *in vitro* results but agreed with a published value for affinity at conditions within the *E. coli* cell. Regardless, only the *in vitro* values of Chapter 4 are capable of predicting lac genetic switches that would function at the measured concentration of lac repressor *in vivo*.

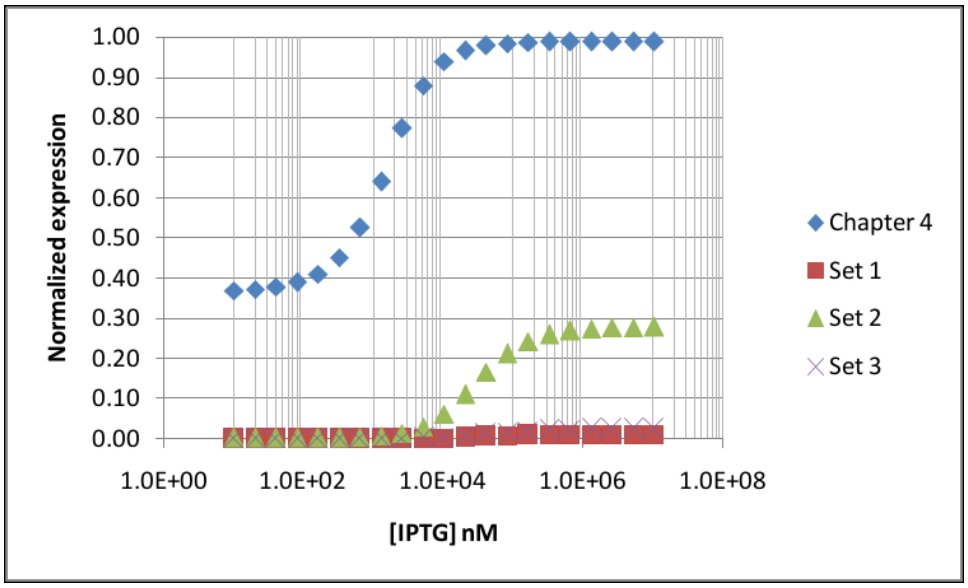


Figure 6.2. Simulated induction curves from various *in vitro* derived values at a repressor concentration of 30 nM. Chapter 4 are the values measured from the *in vitro* transcription experiment. Sets 1, 2 and 3 are derived from the published values in (Sharp, 2011). Chapter 4 predicts a leaky repressor, however the values are derived from dimeric lac repressor with a single operator DNA, which is known to be leakier than the wild type operon. Set 2 predicts a very tight repressor that is capable of some induction. Sets 1 and 3 are still drastically over-repressed.

Chapter 4 shows the inherent difficulty in rectifying *in vivo* and *in vitro* data. We specifically set out with this goal and therefore were able to use the same lac repressor construct and same DNA operator sequence in both our *in vivo* and *in vitro* experiments. When attempting to compare values with literature values the setups rarely agree. Furthermore, we used an *in vitro* technique that attempts to closely mimic the *in vivo* assay. The *in vitro* transcription experiment takes place at the same temperature, in the solution phase, in buffer conditions where RNA polymerase is active and hence reminiscent of *in vivo* buffer, involves RNA polymerase to probe the operator DNA occupancy (although our *in vivo* and *in vitro* experiments do not use the same promoters or RNA polymerases), and uses long DNA strands. The comparison is not perfect but it is closer than most techniques such as gel shift or nitrocellulose binding. Next, we had an *in vivo* system where we could accurately estimate the total concentration of operator

(due to a tightly controlled copy number of its plasmid) and effector (due to the removal of thiogalactoside-permease from the *E. coli* limiting IPTG access to passive diffusion through the cell membrane). Then we were able to measure the repressor concentration throughout the experiment through the mCherry tag on the lac repressor and even able to calibrate our experiment to get an absolute *in vivo* concentration that agreed with previously published values. Finally, we had to develop a solution to the MWC model that would work in all of our experimental conditions. This required coming up with an assumption free solution so total operator, total effector and total repressor are directly represented in the solution. This was important in the fitting of the *in vitro* transcription data where we were able to titrate in lac repressor to more accurately measure repressor-operator affinity. This experiment violated the assumption of our original solution to the MWC model that $[R]_{tot} >> [O]_{tot}$ but it was easily modeled with the assumption free solution.

Despite our best efforts, Figure 4.4.A shows that the *in vitro* thermodynamic parameters predict a leakier genetic switch than we measured *in vivo*. There is approximately a 2-fold difference between our measured repressor-operator DNA affinity and that which would be needed to decrease the leakiness of the switch to agree with *in vivo*. This difference could potentially be rectified by including the well known phenomenon of cell crowding. We modeled the increased effective concentration of the lac repressor by dividing our measured concentration by effective cell volume. When the effective cell volume is 40% of the total *E. coli* volume the *in vitro* data does an excellent job predicting the *in vivo* phenotype. The value of 40% available volume agrees with previously published values for *E. coli* (Kubitschek & Friske, 1986). *In vivo* phenotype can be linked to *in vitro* binding parameters using *in vitro* experiments that attempt to

mimic the *in vivo* setting. It is non-trivial to compare *in vivo* to *in vitro* data and it requires knowing not just the binding constants of the various species but also the concentrations of the species themselves. Trying to compare without knowing the concentration could potentially lead to erroneous agreement.

While it is gratifying to link structure, genetics, *in vivo* data, and *in vitro* data through a single MWC model, the entire enterprise is enriched by a real world application. With that in mind we set out to apply what we know about the lac repressor genetic switch to AAV-mediate gene therapy.

There was one primary limitation in AAV-mediated delivery of a lac genetic switch: delivering two promoters and two genes to a cell through AAV capsids is not possible due to genome size limitations. Only 5 kilobases of DNA can practically be packaged within an AAV capsid (Wu, Yang, & Colosi, 2010). We therefore re-wired the lac repressor to be autogenously regulated; the lac repressor controlled its own expression. We demonstrated that an autogenously regulated lac genetic switch functioned in both *E. coli* and three different eukaryotic cell types in cell culture (HEK293T, Cos-7, and ARPE-19). We found that our first generation of eukaryotic autogenous lac switches were not ideal; we measured 45% leakiness and only 2-fold induction. Just about every method we thought of to improve the switch worked. Addition of a second operator DNA site downstream of the promoter decreased leakiness. Restoration of the tetramerization domain of the lac repressor marginally increased dynamic range while maintaining the decreased leakiness. The Q18M mutation in the DNA binding domain deceased leakiness at the expense of maximal expression; Q18M functions has the same phenotype in *E. coli*. The F161W mutation decreased leakiness and increased dynamic range in the dimeric lac repressor but may

not function the same for the tetrameric lac repressor. Other groups have shown that operator DNA can repress transcription in many locations relative to the promoter (Hu & Davidson, 1987) and that the lac repressor functions with different promoters than the minimal CMV promoter we studied (Brown et al., 1987; Caron et al., 2005; Liu et al., 1989; Ulmasov et al., 1997). We are therefore extremely confident that many more improvements are possible by systematic exploration of lac repressor mutation and promoter/operator DNA combinations. Finally, different gene therapy applications may be optimized by different leakiness and maximal expression values; some transgenes need to be very, very off while others could tolerate leakiness if it meant higher expression upon induction. Tailored autogenous lac repressor genetic switches for specific gene therapy applications could be an incredibly important future direction of this work.

It is currently not known if the tetrameric lac repressor is functioning as desired in the eukaryotic autogenous switch. We have demonstrated that it is a tetramer however the expected gain in dynamic range from cooperativity was not measured. Several important experiments could shed light on the issue. First, if the 2A sequence is not splitting the two genes efficiently enough and causing a dimeric lac repressor-2A-YFP fusion protein to localize to the nucleus as the microscopy suggests, then the 2A needs to be replaced with an IRES sequence which will make two separate genes. All of the lac repressor will necessarily be tetrameric as there can be no C-terminal fusion from this construct. This would easily be tested with microscopy and the cooperativity could be tested with the assay varying the spacing between the first and second Lsym operators. Secondly, the ability of the second operator site to repress is a confounding factor. The second operator could be placed far upstream of the CMV promoter after the

ITR region. This operator could be tested alone to confirm that it does not repress. Once that has been established it could be tested in combination with the first Lsym site to see if there is an improvement. This improvement would have to be attributed to cooperativity of binding by the tetrameric lac repressor. Similarly, this construct could be tested with the dimeric lac repressor to confirm that tetramerization is necessary.

Finally, we were able to test the behavior of our autogenous lac genetic switch on the organism level. We showed that the tetrameric autogenous lac genetic switch was capable of reversible regulation of luciferase in the livers of living mice. Surprisingly, the induction ratios in mice surpassed those measured in cell culture resulting in approximately 8-fold induction upon IPTG gavage and 14-fold repression after mice were returned to normal water drinking. This study will need to be confirmed in an age and sex matched mouse study and the functionality of the switch in additional tissues such as the retina, where the field of gene therapy has enjoyed success, are important future directions.

This thesis studied the lac repressor in purified *in vitro* experiments, in living *E. coli*, in living eukaryotic cells, and finally in living mice. We have found good agreement when comparing *in vitro* to *in vivo E. coli*. We have found good agreement between structurally relevant mutations of the lac repressor and *in vivo E. coli* and even in cell culture of eukaryotic cells. Finally, autogenous switches which were shown to function in cell culture were also shown to function in the liver of a mouse. The functionality of the switch on all of these scales are well explained by the Monod, Wyman, and Changeux model of protein allostery.

Appendices

A.1. An Assumption Free Model of the Autogenous Lac Genetic Switch

We can use the solution for the MWC model from Chapter 4 to model the autogenous lac genetic switch. First, we define the concentration of the lac repressor for a given unregulated promoter to be $[R]_{tot}$. We measured $[R]_{tot}$ for our experimental system in Chapter 4; for ~10-20 copies of the *lacI* promoter in EPB225 cells $[R]_{tot} \approx 650$ nM. It is important to remember that $[R]_{tot}$ will be a function of promoter strength and copy number of the promoters in the cell. The equilibium value of repressor in the cell for the autogenously regulated (AR) system is then defined as,

$$[R]_{tot}^{AR} = \frac{[O]}{[O]_{tot}}[R]_{tot} \tag{124}$$

Of course we know that the autogenously regulated repressor concentration is bound by reality, such that,

$$0\,\text{mM} \leq [R]_{tot}^{AR} \leq [R]_{tot} \tag{125}$$

We then will start with Equation (109) which we will re-define here,

$$[R] = \frac{[R]_{tot}^{AR}}{B_1 + [O]B_2} \tag{126}$$

We will then insert Equation (124) into Equation (126) and isolate the free operator concentration [O],

$$[R]=\frac{\frac{[O]}{O_{tot}}[R]_{tot}}{B_1+[O]B_2} \tag{127}$$

$$[R]B_1+[R][O]B_2=\frac{[R]_{tot}}{[O]_{tot}}[O] \tag{128}$$

$$[R]B_1=(\frac{[R]_{tot}}{[O]_{tot}}-[R]B_2)[O] \tag{129}$$

Now we can insert the formula for free operator from Chapter 4 (Equation (110)) into Equation (129),

$$[R]B_1=(\frac{[R]_{tot}}{[O]_{tot}}-[R]B_2)\frac{[O]_{tot}}{1+[R]B_2} \tag{130}$$

Now we will re-arrange Equation (130) as we did in Chapter 4,

$$[R]B_1+[R]^2B_1B_2=[R]_{tot}-[R][O]_{tot}B_2 \tag{131}$$

$$[R]^2B_1B_2+[R](B_1+[O]_{tot}B_2)=[R]_{tot} \tag{132}$$

$$\varphi_2^{AR}=B_1+B_2O_{tot} \tag{133}$$

$$[R]^2\varphi_1+[R]\varphi_2^{AR}=[R]_{tot} \tag{134}$$

It is now abundantly clear that Equation (134) will take the place of Equation (117). Just as before we have two Equations (134) and (120) which are polynomials of

[R] and [E]. Again, as before, we guess at values of [E], calculate [R] from each equation, and minimize the difference.

Bibliography

Aubrecht, J., Manivasakam, P., & Schiestl, R. H. (1996). Controlled gene expression in mammalian cells via a regulatory cascade involving the tetracycline transactivator and lac repressor. *Gene*, *172*(2), 227–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8682308

Barkley, M. D., Riggs, A. D., Jobe, A., & Bourgeois, S. (1975). Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry1*, *14*(8), 1700–1712. Retrieved from http://proxy.library.upenn.edu:2082/pubmed?otool=upennlib&term=barkley+riggs+1975

Barry, J. K., & Matthews, K. S. (1997). Ligand-induced conformational changes in lactose repressor: a fluorescence study of single tryptophan mutants. *Biochemistry*, *36*(50), 15632–42. doi:10.1021/bi971685r

Barry, J. K., & Matthews, K. S. (1999). Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry*, *38*(20), 6520–8. doi:10.1021/bi9900727

Becskei, A., & Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, *405*(6786), 590–3. doi:10.1038/35014651

Belin, D. (2003). Why Are Suppressors of Amber Mutations So Frequent Among Escherichia coli K12 Strains ?: A Plausible Explanation for a Long-Lasting Puzzle. *Genetics*, *165*(October), 455–456.

Bell, C. E., Barry, J. K., Matthews, K. S., & Lewis, M. (2001). Structure of a variant of lac repressor with increased thermostability and decreased affinity for operator. *Journal of Molecular Biology*, *313*(1), 99–109. doi:10.1006/jmbi.2001.5041

Beltrao, P., Bork, P., Krogan, N. J., & van Noort, V. (2013). Evolution and functional cross-talk of protein post-translational modifications. *Molecular Systems Biology*, *9*(1), 714. doi:10.1002/msb.201304521

Biard, D. S., James, M. R., Cordier, A., & Sarasin, A. (1992). Regulation of the Escherichia coli lac operon expressed in human cells. *Biochimica et Biophysica Acta*, *1130*(1), 68–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1311956

Brent, R., & Ptashne, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. *Nature*, *312*(5995), 612–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6390216

Brown, M., Figge, J., Hansen, U., Wright, C., Jeang, K. T., Khoury, G., … Roberts, T. M. (1987). lac repressor can regulate expression from a hybrid SV40 early promoter containing a lac operator in animal cells. *Cell*, *49*(5), 603–12. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3034429

Buttin, G., Cohen, G. N., Monod, J., & Rickenberg, H. V. (1956). [Galactoside-permease of Escherichia coli]. *Annales de l'Institut Pasteur*, *91*(6), 829–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13395026

Caron, L., Prot, M., Rouleau, M., Rolando, M., Bost, F., & Binétruy, B. (2005). The Lac repressor provides a reversible gene expression system in undifferentiated and differentiated embryonic stem cell. *Cellular and Molecular Life Sciences : CMLS*, *62*(14), 1605–12. doi:10.1007/s00018-005-5123-2

Chakerian, A. E., & Matthews, K. S. (1991). Characterization of mutations in oligomerization domain of Lac repressor protein. *The Journal of Biological Chemistry*, *266*(33), 22206–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1939243

Chakerian, A. E., Tesmer, V. M., Manly, S. P., Brackett, J. K., Lynch, M. J., Hoh, J. T., & Matthews, K. S. (1991). Evidence for leucine zipper motif in lactose repressor protein. *The Journal of Biological Chemistry*, *266*(3), 1371–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1988425

Chang, W., Olson, J. S., & Matthews, K. S. (1993). Lysine 84 Is at the Subunit Interface of lac Repressor Protein. *The Journal of Biological Chemistry*, *268*(23), 17613–17622.

Chen, J., & Matthews, K. S. (1992). Deletion of Lactose Repressor Carboxyl-terminal Domain Affects Tetramer Formation. *The Journal of Biological Chemistry*, *267*(20), 13843–13850.

Chen, J., Surendran, R., Lee, J. C., & Matthews, K. S. (1994). Construction of a Dimeric Repressor: Dissection of Subunit Interfaces in Lac Repressor. *Biochemistry*, *33*, 1234–1241.

Cohn, M. (1957). Contributions of studies on the beta-galactosidase of Escherichia coli to our understanding of enzyme synthesis. *Bacteriological Reviews*, *21*(3), 140–68. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=180896&tool=pmcentrez&rendertype=abstract

Cronin, C. A., Gluba, W., & Scrable, H. (2001). The lac operator-repressor system is functional in the mouse. *Genes & Development*, *15*(12), 1506–1517. doi:10.1101/gad.892001

Cronin, C. A., Ryan, A. B., Talley, E. M., & Scrable, H. (2003). Tyrosinase expression during neuroblast divisions affects later pathfinding by retinal ganglion cells. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(37), 11692–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14684871

Culard, F., & Maurizot, J. C. (1981). Lac repressor - lac operator interaction. Circular dichroism study. *Nucleic Acids Research*, *9*(19), 5175–5184.

Daber, R., & Lewis, M. (2009). Towards evolving a better repressor. *Protein Engineering, Design & Selection : PEDS*, *22*(11), 673–83. doi:10.1093/protein/gzp051

Daber, R., Sharp, K. A., & Lewis, M. (2009). One is not enough. *Journal of Molecular Biology*, *392*(5), 1133–1144. doi:10.1016/j.jmb.2009.07.050

Daber, R., Sochor, M. A., & Lewis, M. (2011). Thermodynamic analysis of mutant lac repressors. *Journal of Molecular Biology*, *409*(1), 76–87. doi:10.1016/j.jmb.2011.03.057

Daber, R., Stayrook, S., Rosenberg, A., & Lewis, M. (2007). Structural analysis of lac repressor bound to allosteric effectors. *Journal of Molecular Biology*, *370*(4), 609–19. doi:10.1016/j.jmb.2007.04.028

Daly, T. J., Olson, J. S., & Matthews, K. S. (1986). Formation of mixed disulfide adducts at cysteine-281 of the lactose repressor protein affects operator and inducer binding parameters. *Biochemistry*, *25*(19), 5468–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3535878

Dong, F., Spott, S., Zimmermann, O., Kisters-Woike, B., Müller-Hill, B., & Barker, A. (1999). Dimerisation mutants of Lac repressor. I. A monomeric mutant, L251A, that binds Lac operator DNA as a dimer. *Journal of Molecular Biology*, *290*(3), 653–66. doi:10.1006/jmbi.1999.2902

Edelstein, S. J., & Le Novère, N. (2013). Cooperativity of allosteric receptors. *Journal of Molecular Biology*, *425*(9), 1424–32. doi:10.1016/j.jmb.2013.03.011

Englander, S. W., Mayne, L., & Krishna, M. M. G. (2012). Protein folding and misfolding: mechanism and principles. *Quarterly Reviews of Biophysics*, *40*(4), 287–326. doi:10.1017/S0033583508004654.Protein

Fairman, R., Chao, H. G., Mueller, L., Lavoie, T. B., Shen, L., Novotny, J., & Matsueda, G. R. (1995). Characterization of a new four-chain coiled-coil: influence of chain length on stability. *Protein Science : A Publication of the Protein Society*, *4*(8), 1457–69. doi:10.1002/pro.5560040803

Falcon, C. M., & Matthews, K. S. (1999). Glycine insertion in the hinge region of lactose repressor protein alters DNA binding. *The Journal of Biological Chemistry*, *274*(43), 30849–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10521477

Falcon, C. M., & Matthews, K. S. (2001). Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. *Biochemistry*, *40*(51), 15650–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11747440

Falcon, C. M., Swint-Kruse, L., & Matthews, K. S. (1997). Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. *The Journal of Biological Chemistry*, *272*(43), 26818–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9341111

Farabaugh, P. J. (1978). Sequence of the lacI gene. *Nature*, *274*(24), 765–769.

Figge, J., Wright, C., Collins, C. J., Roberts, T. M., & Livingston, D. M. (1988). Stringent regulation of stably integrated chloramphenicol acetyl transferase genes by E. coli lac repressor in monkey cells. *Cell*, *52*(5), 713–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2830990

Files, J. G., & Weber, K. (1976). Digestion of lac Repressor by Trypsin. *The Journal of Biological Chemistry*, *251*(11), 3386–3391.

Flashner, Y., & Gralla, J. D. (1988). Dual mechanism of repression at a distance in the lac operon. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(23), 8968–72. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=282636&tool=pmcentrez&rendertype=abstract

Flynn, T. C., Swint-kruse, L., Kong, Y., Booth, C., Matthews, K. S., & Ma, J. (2003). Allosteric transition pathways in the lactose repressor protein core domains : Asymmetric motions in a homodimer. *Protein Science*, *12*, 2523–2541. doi:10.1110/ps.03188303.central

Fukami-Kobayashi, K. (2003). Parallel Evolution of Ligand Specificity Between LacI/GalR Family Repressors and Periplasmic Sugar-Binding Proteins. *Molecular Biology and Evolution*, *20*(2), 267–277. doi:10.1093/molbev/msg038

Geisler, N., & Weber, K. (1977). Isolation of amino-terminal fragment of lactose repressor necessary for DNA binding. *Biochemistry*, *16*(5), 938–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/321012

Gilbert, W., & Maxam, A. (1973). The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3581–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=427284&tool=pmcentrez&rendertype=abstract

Gilbert, W., & Müller-Hill, B. (1966). Isolation of the Lac Repressor. *Proceedings of the National Academy of Sciences*, *56*, 1891–1898.

Gilbert, W., & Müller-Hill, B. (1967). The lac operator is DNA. *Proceedings of the National Academy of Sciences*, *58*, 2415–2421.

Ha, J. H., Capp, M. W., Hohenwalter, M. D., Baskerville, M., & Record, M. T. (1992). Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the “glutamate effect” on protein-DNA interactions. *Journal of Molecular Biology*, *228*(1), 252–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1447786

Hsieh, W. T., & Matthews, K. S. (1981). Tetranitromethane modification of the tyrosine residues of the lactose repressor. *The Journal of Biological Chemistry*, *256*(10), 4856–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7014561

Hu, M. C., & Davidson, N. (1987). The inducible lac operator-repressor system is functional in mammalian cells. *Cell*, *48*(4), 555–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3028641

Hu, M. C., & Davidson, N. (1991). Targeting the Escherichia coli lac repressor to the mammalian cell nucleus. *Gene*, *99*(2), 141–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2022328

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*(3), 318–356. doi:10.1016/S0022-2836(61)80072-7

Jobe, A., & Bourgeois, S. (1972). lac Repressor-operator interaction. VI. The natural inducer of the lac operon. *Journal of Molecular Biology*, *69*(3), 397–408. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4562709

Jobe, A., & Bourgeois, S. (1973). Lac Repressor-Operator Interaction VIII. Lactose is an Anti-inducer of the lac Operon. *Journal of Molecular Biology*, *75*, 303–313.

Kaptein, R., Lamerichs, R. M. J. N., Boelens, R., & Rullmann, J. A. C. (1990). Two-dimensional NMR study of a protein-DNA complex. *Biochemical Pharmacology*, *40*(1), 89–96. doi:10.1016/0006-2952(90)90183-L

Kaptein, R., Zuiderweg, E. R., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *Journal of Molecular Biology*, *182*(1), 179–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3889346

Khoury, a M., Lee, H. J., Lillis, M., & Lu, P. (1990). Lac repressor-operator interaction: DNA length dependence. *Biochimica et Biophysica Acta*, *1087*(1), 55–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2205296

Kim, J., Zweib, C., Wu, C., & Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene*, *85*, 15–23.

Kleina, L. G., & Miller, J. H. (1990). Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *Journal of Molecular Biology*, *212*(2), 295–318. doi:10.1016/0022-2836(90)90126-7

Kramer, H., Niemoller, M., Amouyal, M., Revet, B., Wilcken-Bergmann, B. Von, & Müller-Hill, B. (1987). lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *The EMBO Journal*, *6*(5), 1481–1491.

Kubitschek, H. E., & Friske, J. A. (1986). Determination of bacterial cell volume with the Coulter Counter. *Journal of Bacteriology*, *168*(3), 1466–7. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=213663&tool=pmcentrez&rendertype=abstract

Lau, I. F., Filipe, S. R., Søballe, B., Økstad, O.-A., Barre, F.-X., & Sherratt, D. J. (2004). Spatial and temporal organization of replicating Escherichia coli chromosomes. *Molecular Microbiology*, *49*(3), 731–743. doi:10.1046/j.1365-2958.2003.03640.x

Lewis, M. (2005). The lac repressor. *Comptes Rendus Biologies*, *328*(6), 521–48. doi:10.1016/j.crvi.2005.04.004

Lewis, M. (2013). Allostery and the lac Operon. *Journal of Molecular Biology*, *425*(13), 2309–16. doi:10.1016/j.jmb.2013.03.003

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Helen, C., Schumacher, M. A., … Lu, P. (1996). Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science*, *271*, 1247–1254.

Liu, H. S., Feliciano, E. S., & Stambrook, P. J. (1989). Cytochemical observation of regulated bacterial beta-galactosidase gene expression in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(24), 9951–5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=298620&tool=pmcentrez&rendertype=abstract

Loth, K., Gnida, M., Romanuka, J., Kaptein, R., & Boelens, R. (2013). Sliding and target location of DNA-binding proteins: an NMR view of the lac repressor system. *Journal of Biomolecular NMR*, *56*(1), 41–9. doi:10.1007/s10858-013-9723-0

Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., & Miller, J. H. (1994). Genetic Studies of the lac Repressor. XIV. Analysis of 4000 Altered Escherichia coli lac Repressors Reveals Essential and Non-essential Residues, as well as “Spacers” which do not Require a Specific Sequence. *Journal of Molecular Biology*, *240*, 421–433.

Mckay, D. B., Pickover, C. A., & Steitz, T. A. (1982). Escherichia coli luc Repressor is Elongated with its Operator DNA Binding Domains Located at Both Ends. *Journal of Molecular Biologyj*, *156*, 175–183.

Milk, L., Daber, R., & Lewis, M. (2010). Functional rules for lac repressor-operator associations and implications for protein-DNA interactions. *Protein Science : A Publication of the Protein Society*, *19*(6), 1162–72. doi:10.1002/pro.389

Miller, J. H. (1984). Genetic Studies of the lac Repressor XII. Amino Acid Replacements in the DNA Binding Domain of the Escherichia coli lac Repressor. *Journal of Molecular Biology*, *180*, 205–212.

Miller, J. H., & Schmeissner, U. (1979). Genetic studies of the lac repressor. X. Analysis of missense mutations in the lacI gene. *Journal of Molecular Biology*, *131*(2), 223–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/114666

Monod, J. (1942). *Recherches sur la croissance des cultures bacteriennes*.

Monod, J., Wyman, J., & Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, *12*(1), 88–118. doi:10.1016/S0022-2836(65)80285-6

Moore, I., Gälweiler, L., Grosskopf, D., Schell, J., & Palme, K. (1998). A transcription activation system for regulated gene expression in transgenic plants. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(1), 376–81. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18229&tool=pmcentrez&rendertype=abstract

Müller, J., Oehler, S., & Müller-Hill, B. (1996). Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *Journal of Molecular Biology*, *257*(1), 21–9. doi:10.1006/jmbi.1996.0143

Müller-Hartmann, H., & Müller-Hill, B. (1996). The side-chain of the amino acid residue in position 110 of the Lac repressor influences its allosteric equilibrium. *Journal of Molecular Biology*, *257*(3), 473–8. doi:10.1006/jmbi.1996.0176

Muller-Hill, B. (1996). *The Lac Operon: A Short History Of A Genetic Paradigm* (p. 135).

Müller-Hill, B. (1996). *The Lac Operon: A Short History Of A Genetic Paradigm* (p. 135).

Muñoz, V., & Eaton, W. A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(20), 11311–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18030&tool=pmcentrez&rendertype=abstract

Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., & Balázsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5123–8. doi:10.1073/pnas.0809901106

Nichols, J. C., & Matthews, K. S. (1997). Combinatorial Mutations of lac Repressor. *The Journal of Biological Chemistry*, *272*(30), 18550–18557. doi:10.1074/jbc.272.30.18550

Nussinov, R., & Tsai, C.-J. (2013). Allostery in disease and in drug discovery. *Cell*, *153*(2), 293–305. doi:10.1016/j.cell.2013.03.034

Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., & Müller-Hill, B. (1994). Quality and position of the three lac operators of E. coli define efficiency of repression. *The EMBO Journal*, *13*(14), 3348–3355.

Oehler, S., Eismann, E. R., Kramer, H., & Müller-Hill, B. (1990). The three operators of the lac cooperate in repression. *The EMBO Journal*, *9*(4), 973–979.

Ogata, R. T., & Gilbert, W. (1978). An amino-terminal fragment of lac repressor binds specifically to lac operator. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(12), 5851–4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=393073&tool=pmcentrez&rendertype=abstract

Pace, H. C., Lu, P., & Lewis, M. (1990). lac repressor: Crystallization of intact tetramer and its complexes with inducer and operator DNA. *Proceedings of the National Academy of Sciences*, *87*(March), 1870–1873.

Penaud-Budloo, M., Le Guiner, C., Nowrouzi, A., Toromanoff, A., Chérel, Y., Chenuaud, P., … Snyder, R. O. (2008). Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *Journal of Virology*, *82*(16), 7875–85. doi:10.1128/JVI.00649-08

Pfahl, M. (1981). Mapping of I Gene mutations which lead to repressors with increased affinity for lac operator. *Journal of Molecular Biology*, *147*(1), 175–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7265237

Pfahl, M., Gulde, V., & Bourgeois, S. (1979). “Second” and “third operator” of the lac operon: an investigation of their role in the regulatory mechanism. *Journal of Molecular Biology*, *127*(3), 339–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/430569

Poelwijk, F. J., de Vos, M. G. J., & Tans, S. J. (2011). Tradeoffs and optimality in the evolution of gene regulation. *Cell*, *146*(3), 462–70. doi:10.1016/j.cell.2011.06.035

Roeder, R. G. (2003). Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. *Nature Medicine*, *9*(10), 1239–44. doi:10.1038/nm938

Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology*, *323*(5), 785–793. doi:10.1016/S0022-2836(02)00994-4

Royer, C. a, Chakerian, a E., & Matthews, K. S. (1990). Macromolecular binding equilibria in the lac repressor system: studies using high-pressure fluorescence spectroscopy. *Biochemistry*, *29*(20), 4959–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2194564

Sadler, J. R., Sasmor, H., & Betzt, J. L. (1983). A perfectly symmetric lac operator binds the lac repressor very tightly. *Proceedings of the National Academy of Sciences*, *80*, 6785–6789.

Sartorius, J., Lehming, N., Kisters, B., von Wilcken-Bergmann, B., & Müller-Hill, B. (1989). lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. *The EMBO Journal*, *8*(4), 1265–1270.

Sasse-dwight, S., & Gralla, J. D. (1988). Probing Co-operative DNA-binding in Vivo. *Journal of Molecular Biology*, *202*, 107–119.

Schmitz, A., & Galas, D. J. (1979). The interaction of RNA polymerase and lac repressor with the lac control region. *Nucleic Acids Research*, *6*(1).

Schmitz, A., Schmeissner, U., Miller, J. H., & Lu, P. (1976). Mutations Affecting the Quaternary Structure of the lac Repressor. *The Journal of Biological Chemistry*, *251*(11), 3359–3366.

Scrable, H., & Stambrook, P. J. (1997). Activation of the lac Repressor in the Transgenic Mouse. *Genetics*, *147*(1), 297–304.

Sharp, K. A. (2011). Allostery in the lac operon: population selection or induced dissociation? *Biophysical Chemistry*, *159*(1), 66–72. doi:10.1016/j.bpc.2011.05.007

Slijper, M., Bonvin, A. M., Boelens, R., & Kaptein, R. (1996). Refined structure of lac repressor headpiece (1-56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the lac operator. *Journal of Molecular Biology*, *259*(4), 761–73. doi:10.1006/jmbi.1996.0356

Spott, S., Dong, F., Kisters-Woike, B., & Müller-Hill, B. (2000). Dimerisation mutants of Lac repressor. II. A single amino acid substitution, D278L, changes the specificity of dimerisation. *Journal of Molecular Biology*, *296*(2), 673–84. doi:10.1006/jmbi.1999.3469

Spotts, R., Chakerian, A. E., & Matthews, K. S. (1991). Arginine 197 of lac Repressor Contributes Significant Energy to Inducer Binding. *The Journal of Biological Chemistry*, *266*(34), 22998–23002.

Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J. H., Kisters-Woike, B., & Müller-Hill, B. (1996). 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. *Journal of Molecular Biology*, *261*(4), 509–23. doi:10.1006/jmbi.1996.0479

Swint-Kruse, L., Elam, C. R., Lin, J. W., Wycuff, D. R., & Matthews, K. S. (2001). Plasticity of quaternary structure: twenty-two ways to form a LacI dimer. *Protein Science : A Publication of the Protein Society*, *10*(2), 262–76. doi:10.1110/ps.35801

Swint-Kruse, L., & Matthews, K. S. (2009). Allostery in the LacI/GalR family: variations on a theme. *Current Opinion in Microbiology*, *12*(2), 129–37. doi:10.1016/j.mib.2009.01.009

Swint-kruse, L., Zhan, H., Fairbanks, B. M., Maheshwari, A., & Matthews, K. S. (2003). Perturbation from a Distance : Mutations that Alter LacI Function through Long-Range Effects. *Biochemistry*, *42*, 14004–14016.

Tungtur, S., Skinner, H., Zhan, H., Swint-Kruse, L., & Beckett, D. (2011). In vivo tests of thermodynamic models of transcription repressor function. *Biophysical Chemistry*, *159*(1), 142–51. doi:10.1016/j.bpc.2011.06.005

Ulmasov, B., Capone, J., & Folk, W. (1997). Regulated expression of plant tRNA genes by the prokaryotic tet and lac repressors. *Plant Molecular Biology*, *35*(4), 417–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9349265

Weickert, M. J., & Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *The Journal of Biological Chemistry*, *267*(22), 15869–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1639817

Wu, Z., Yang, H., & Colosi, P. (2010). Effect of genome size on AAV vector packaging. *Molecular Therapy : The Journal of the American Society of Gene Therapy*, *18*(1), 80–6. doi:10.1038/mt.2009.255

Zabin, I., Kepes, A., & Monod, J. (1962). Thiogalactoside Transacetylase. *Journal of Biological Chemistry*, *237*, 253–257.

Zhan, H., Camargo, M., & Matthews, K. S. (2010). Positions 94-98 of the lactose repressor N-subdomain monomer-monomer interface are critical for allosteric communication. *Biochemistry*, *49*(39), 8636–45. doi:10.1021/bi101106x

Zhang, S. P., Zubay, G., & Goldman, E. (1991). Low-usage codons in Escherichia coli, yeast, fruit fly and primates. *Gene*, *105*(1), 61–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1937008