doi:10.1016/j.jmb.2009.06.039 J. Mol. Biol. (2009) 391, 661–670

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

A Novel Molecular Switch

Robert Daber and Mitchell Lewis*

Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 37th and Hamilton Walk, Philadelphia, PA 19104-6059, USA

Received 3 March 2009; received in revised form 5 June 2009; accepted 16 June 2009
Available online 21 June 2009

Transcriptional regulation is a fundamental process for regulating the flux of all metabolic pathways. For the last several decades, the lac operon has served as a valuable model for studying transcription. More recently, the switch that controls the operon has also been successfully adapted to function in mammalian cells. Here we describe how, using directed evolution, we have created a novel switch that recognizes an asymmetric operator sequence. The new switch has a repressor with altered headpiece domains for operator recognition and a redesigned dimer interface to create a heterodimeric repressor. Quite unexpectedly, the heterodimeric switch functions better than the natural system. It can repress more tightly than the naturally occurring switch of the lac operon; it is less leaky and can be induced more efficiently. Ultimately, these novel repressors could be evolved to recognize eukaryotic promoters and used to regulate gene expression in mammalian systems.

Edited by J. Karn

Keywords: directed evolution; repression; heterodimer; chimeric operator

Introduction

The viability of a cell often depends on its ability to successfully regulate a variety of cellular processes. In virtually all organisms, critical cellular events are controlled by modulating gene expression. Jacob and Monod were the first to suggest that a molecular switch could enhance or repress gene expression and, as a consequence, regulate cellular processes.[1] To understand how molecular switches function, two classical systems have been extensively analyzed: the epigenetic switch from the bacteriophage lambda, which controls phage life cycle and development, and the switch that regulates lactose metabolism in *Escherichia coli*. In both instances, the switch is a two-component system that is composed of a repressor protein and an operator DNA site. When the repressor binds to the operator, a complex that physically blocks transcription by an RNA polymerase is established. The switch from bacteriophage lambda functions as a two-state device; transcription is either on or off. The molecular switch that regulates the lac operon is slightly more complex, since it behaves more like a potentiometer or dimmer switch. Inducer molecules potentiate the stability of the repressor–operator complex and, as a consequence, the level of transcription is modulated.

The switch that regulates the lac operon has been well characterized (for a recent review, see Wilson *et al.*[2]). The repressor is a 360-amino-acid protein that has a modular structure. It contains an NH_2-terminal or "headpiece" domain (~60 residues) that binds specifically to operator DNA and a COOH-terminal "core" domain that binds inducers. The monomeric repressor self-associates into a dimer of dimers, where a dimeric repressor molecule recognizes the operator using the classical helix–turn–helix (HTH) motif. The HTH provides the scaffolding for specific side chains that recognize the bases in the major groove of the operator. The operator of the *lac* operon, a short stretch of DNA (~17 base pairs), is composed of two nearly identical half-sites that are located between the end of the *lacI* gene and the beginning of the *lacZ* gene.[3] The COOH-terminal core domain consists of two subdomains: an N-terminal domain (NTD) and a C-terminal domain (CTD). The inducer molecule binds to a pocket located at the interface of these two domains. When bound to the repressor, the NTDs adopt a conformation that is different from the conformation when bound to DNA. In contrast, the CTDs are structurally invariant and are responsible for stabilizing the dimeric structure of the repressor.[4]

Although regulatory mechanisms are somewhat more complex in eukaryotes than in prokaryotes,

*Corresponding author. E-mail address: lewis@mail.med.upenn.edu.

Abbreviations used: HTH, helix–turn–helix; NTD, N-terminal domain; CTD, C-terminal domain; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein.

the molecular switch of the lac operon has been adapted to control gene expression in mammalian systems. Hu and Davidson were among the first to exploit the switch of the lac operon to control reporter gene expression reversibly in mammalian cells.[5] Building upon these results, Figge *et al.* demonstrated that when the lac repressor was inserted into a mammalian cell it was able to gain access to the mammalian chromosome and regulate a stably integrated reporter gene.[6] In addition, Itzhaki *et al.* incorporated the lac operator into a mammalian promoter by homologous recombination and demonstrated that this bacterial switch could regulate the expression of endogenous loci.[7] Even more dramatically, Scrable's laboratory adapted the lac regulatory system to control gene expression in the mouse; demonstrating this bacterial switch can be used to regulate complex events in higher organisms.[8]

Inspired by the observation that the switch of the lac operon can be used to regulate gene expression in a variety of systems, we embarked on a project to create a novel repressor that could recognize altered operator sequences that were distinct from and not recognized by the wild-type repressor. Although proteins containing zinc finger domains have been previously generated to recognize preselected target sequences,[9,10] the lac repressor is equipped with the ability to repress as well as induce transcription repeatedly based on environmental conditions. Therefore, we hypothesized that the bacterial switch that regulates the lac operon could be genetically altered to regulate endogenous gene expression in mammalian systems by directly recognizing and repressing a mammalian promoter. Unlike classical operator sites, these target sequences are not symmetric, making repression by a homodimeric lac repressor impossible. To circumvent this obstacle, we created a heterodimeric repressor where each monomer recognizes a distinct operator half-site. The unique heterodimeric repressors have two different DNA binding domains as well as a dimer interface that cannot homodimerize. Interestingly, these novel repressors function better than the wild-type repressor; they bind more tightly to the chimeric operator than the wild-type repressor binds to the natural operator, thereby reducing the basal level of transcription and have an increased dynamic range such that the difference between the on and off (induced and repressed) states is more pronounced.

Results

Creating a symmetric switch with altered specificity

In bacteria, many repressors homologous to the lactose repressor regulate transcription of inducible genes.[11] These proteins, referred to as the LacI/GalR family, have similar structures and regulate transcription in an analogous fashion. All of the LacI/GalR family members have a headpiece domain that contains an HTH motif, which recognizes its operator site. The sequence conservation in the HTH region suggests that these proteins bind to similar operators. Discrimination between the different operators is most likely due to a few nonconserved residues on the HTH motif. In addition to the HTH, all members of this family use a hinge helix for binding to the operator. Both the purine and the lac repressors recognize the central portion of the operator by placing a pair of helices in the minor groove of the operator.[12,13] Members of the LacI/GalR family also have a CTD that is responsible for both effector binding and oligomerization. These domains have significant sequence homology, suggesting that the architecture of the CTD is also well conserved even though members of the LacI/GalR superfamily respond to a variety of effector molecules. Transcription regulated by these repressors is modulated by allosteric effectors, including galactose, fructose, maltose, ribulose, hypoxanthine, xanthine, and a spectrum of β-galactosides. These molecules bind to the core of the repressors and act either as inducers or as co-repressors. Each metabolite functions either by disrupting or by stabilizing the repressor–operator complex. Even though the effector binding sites have similar architecture, the scaffolding is draped with different amino acid side chains to create the unique specificity. Not only are the repressor molecules in the LacI/GalR superfamily conserved, but the operator sequences also possess noticeable homology. In fact, all of the operator sequences that are recognized by this superfamily are pseudo-palindromic. The primary differences in the operators are localized to the peripheral regions, while the central base pairs are highly conserved. For example, the operators of the lac and gal operons differ only at positions 2 and 4 in the operator sequence, yet these differences are sufficient for specificity.[14]

Intrigued by the high level of conservation and the subtle differences that lead to distinct DNA binding patterns, Muller–Hill's laboratory attempted to dissect the "code" defining protein–DNA recognition within the family of proteins that contain the HTH domain.[14–16] By systematically altering each base in the operator and each residue in the recognition helix, they demonstrated that the first two residues on the recognition helix specifically recognize base pairs 4 and 5 of the lac operator, and that residue 6 recognizes the sixth base of the operator. Structural studies have confirmed that the most critical residues for specific operator recognition are Y17, Q18, and R22.[13,17] These three amino acids protrude from the recognition helix of the repressor and form specific interactions with the bases in the major groove by contacting bases 4, 5, and 6 of the operator. A closer look at the recognition helix of LacI/GalR family members suggests that the recognition helix of the gal repressor differs from that of the lac repressor at two of the three positions; valine replaces tyrosine and an alanine replaces glutamine. It was hypothesized and shown that a mutant lac repressor containing two amino acid substitutions,