Diverse Two-Dimensional Input Functions Control Bacterial Sugar Genes

Shai Kaplan,1,2,3 Anat Bren,1,3,* Alon Zaslaver,1 Erez Dekel,1 and Uri Alon1
1Department of Molecular Cell Biology
2Department of Biological Chemistry
Weizmann Institute of Science, Rehovot 76100, Israel
3These authors contributed equally to this work.
*Correspondence: anat.bren@weizmann.ac.il
DOI 10.1016/j.molcel.2008.01.021

SUMMARY

Cells respond to signals by regulating gene expression. The relation between the level of input signals and the transcription rate of the gene is called the gene’s input function. Because most genes are regulated by more than one signal, the input functions are usually multidimensional. To understand cellular responses, it is essential to know the shapes of these functions. Here, we map the two-dimensional input functions of 19 sugar-utilization genes at high resolution in living E. coli cells. We find diverse, intricately shaped input functions, despite the similarity in the regulatory circuitry of these genes. Surprisingly, some of the input functions are nonmonotonic, peaking at intermediate signal levels. Furthermore, most of the input functions show separation of variables, in the sense that they can be described as the product of simple functions that depend on a single input. This first broad survey of two-dimensional input functions can be extended to map the logic of gene regulation in other systems.

INTRODUCTION

E. coli responds to a wide array of signals, such as nutrients and stresses, by adjusting the expression level of its genes (Alon, 2006; Neidhardt, 1996; Ptashne and Gann, 2002). Often, the proper response depends on combinations of multiple different signals (Yuh et al., 1998; Shen-Orr et al., 2002; Thieffry et al., 1998). Hence, the bacterium needs to process these signals in order to compute the appropriate response of each gene. The relation between the level of the input signals and the transcription rate of the gene is called the gene input function. Because most genes are regulated by more than one input, the input functions are usually multidimensional. Despite the importance of these functions for the understanding of transcription regulation, high-resolution measurements of input-output relationships with two or more input signals are rare.

To carry out a broad survey at high resolution of two-dimensional input functions, we employed the sugar utilization genes of E. coli. These genes allow the bacterium to use specific sugars as carbon and energy sources. Each sugar system includes transporters that pump the sugar into the cell and enzymes that break it down. Each system also includes a transcription factor that senses the presence of the sugar and accordingly regulates gene expression. In addition, most of the sugar systems are regulated by a master transcription factor called CRP, which senses the starvation state of the cell. This master regulator is activated by cAMP, a small molecule produced in the cell upon glucose starvation (Neidhardt, 1996).

The overall organization of this regulatory system is in the form of a network motif called dense overlapping regulons (Shen-Orr et al., 2002) (DOR, Figure 1), a design that also occurs in other gene systems (Alon, 2006, 2007; Lee et al., 2002). The sugar DOR transcription circuit acts to compute the expression of each sugar gene in response to two inputs: the cognate sugar and cAMP. The system includes several feed-forward-loop circuits (Mangan and Alon, 2003; Mangan et al., 2003, 2006; Shen-Orr et al., 2002), in which CRP regulates the system-specific transcription factor, and both regulate their target genes (Figure 1). Furthermore, several of the transcription factors show autoregulation loops (Becskei and Serrano, 2000; Maeda and Sano, 2006; Rosenfeld et al., 2002) (Figure 1). The input function of each gene is thus determined by a combination of the effects at the promoter level (arrangement and sequence of binding sites) and at the upstream regulatory circuitry.

Sugar genes within this regulatory network are generally expected to be regulated in a qualitatively similar manner: their expression should increase with availability of the sugar, and with glucose starvation (Neidhardt, 1996; Richet, 2000; Schleif, 2003; Semsey et al., 2006; Thattai and Shraiman, 2003). Thus, these input functions are often described as Boolean AND gates: sugar AND glucose starvation. A case study of detailed mapping of the lactose system input function showed, however, an intricately shaped input function, intermediate between an AND and an OR gate (Setty et al., 2003).

Here we used high-throughput measurements in living cells to map the input functions of 19 sugar genes. We found that when mapped at high resolution the shapes of these input functions are intricate and surprisingly diverse. Despite this diversity, most of the input functions can be well described by a product of relatively simple functions, each of a single input variable.

RESULTS

Mapping Two-Dimensional Input Functions in the E. coli Sugar System

To map the input function of each gene, we used fluorescent reporter strains that provide accurate readout of the transcriptional
activity in living cells (Kalir et al., 2001; Mangan et al., 2003, 2006; Zaslaver et al., 2004). In each reporter strain, rapidly folding green fluorescent protein (GFP) is expressed from a low copy plasmid under the control of the promoter region of the gene of interest (Zaslaver et al., 2006). Thus, bacteria turn fluorescent green in proportion to the rate of transcription from the promoter. A comprehensive library of reporter strains is available for *E. coli’s* promoters (Zaslaver et al., 2006). Here, promoters from five different sugar systems—arabinose, maltose, rhamnose, galactose, and fucose—were studied. Cells were grown in 96-well plates, providing 96 combinations of the two input signals (eight levels of cAMP and 12 levels of the cognate sugar). A robotic station set up the plates, incubated them, and periodically moved them to a multiwell fluorimeter that allowed measurement of GFP fluorescence and cell density at a resolution of 8 min over 20 hr of growth. The input function was defined as the promoter activity (rate of GFP accumulation, δGFP/δt/OD, see the Experimental Procedures) averaged over one cell cycle in the exponential phase of growth at each of the 96 input combinations (Setty et al., 2003). Repeat experiments on different days showed about a 10% mean relative error in the input function.

To control the input signals, the experiment was conducted in a chemically defined medium containing glucose. Growth on glucose helps to avoid the effect of sugars on the growth rate of the cells. Glucose enables all strains to grow with similar growth rates (0.65 ± 0.03 doubling/hr at exponential phase) regardless of the level of the second input sugar. A second benefit of glucose is that it reduces the endogenous production of cAMP to very low levels, allowing the level of cAMP to be controlled by its external concentration in the medium. A synthetically designed promoter containing only a CRP binding site indicated that the activity of CRP is modulated in a graded fashion by external cAMP (see Figure S1A available online). Furthermore, an unregulated promoter displays uniform expression that is not dependent on either of the two inputs (Figure S1B).

One potential side effect of glucose is known as “inducer exclusion,” in which glucose inhibits some sugar transporters (Bettenbrock et al., 2006; Neidhardt, 1996). Most of the sugar inducers in this study, however, are not subject to inducer exclusion, as we find by comparing the response to the inducing sugar in the presence and absence of glucose (data not shown). The only exception is galactose: we find that for galactose, inducer exclusion shifts the sugar concentration needed for a given response to higher levels by a factor of about 8 while preserving the shape of the input function. The present experimental design therefore seems to allow the mapping of the input functions of sugar genes—the specific relationship between their input signals and the output promoter activity.

**Diversity Found in the Shapes of Input Function**

We mapped the input functions of 19 *E. coli* sugar genes. We find that the input functions are diverse, with different shapes for different genes. Furthermore, the input functions are more intrinsically shaped than Boolean AND or OR gates (Figure 2). In some systems (arabinose, rhamnose, and fucose), the input function displayed a triangular activation region, where expression is high above a diagonal threshold defined by the two inputs. In these systems, the inducers “help each other,” such that the levels of an inducer needed to activate the promoter to a given level decrease with the level of the second inducer. In other systems, such as the maltose system, the input function showed a rectangular activation region, in which the input function is high when either signal crosses a threshold. In these systems, the inducers did not affect each other’s efficacy. Triangular or rectangular input function shapes were found in about 60% of the present functions. Several other input functions had a different shape that depended mostly on one of the two inputs (galS, mglBAC, and malT, Figures 2G, 2I, and 2K).

Unexpectedly, some of the input functions were found to be nonmonotonic, peaking at intermediate levels of the signals. For example, the promoters for the galactose enzyme operon galETK (Figure 2F) and for the transporter galP (Figure 2H) both peak at intermediate cAMP levels. The peak in the galP input function is about 3-fold higher than the level at saturating cAMP (Figure S3). The fucR promoter, an internal promoter in the fucP operon, shows a function that has two peaks, one at

![Figure 1. Regulatory Circuitry of E. coli Sugar Genes Is in the Form of a DOR Network Motif](image-url)
low fucose levels and the other at intermediate cAMP levels (Figure 2R). It displays reduced expression levels when both signals are low and when both are high. This function resembles a graded version of an “exclusive-OR” (XOR) function (Buchler et al., 2003; Hermsen et al., 2006).

We find that promoters in the same system can show different input functions, despite the fact that they are controlled by the same signals and regulators. This is seen by comparing input functions along columns in Figure 2. These differences can, in certain cases, be related to the biological function of the gene products. One case in which different biological roles corresponded to different input functions can be seen when comparing transcription factors to enzymes and transporters (comparing rows in Figure 2). Transcription factor promoters tend to be less sensitive to the levels of the sugar inducer than other genes in the same system. In one case, \( \text{malT} \), the input function is almost totally independent of the sugar (Figure 2K). Other transcription factors are activated at lower levels of the sugar than the other genes in the system (\( \text{araC} \), \( \text{galS} \), and \( \text{fucR} \); Figures 2B, 2G, and 2R). This may reflect the constitutive need for some basal expression level of these transcription factors, which is required in order to sense the sugar and regulate the other genes in the system.

An additional case where different biological roles corresponded to different input functions can be seen when comparing...
comparing the arabinose transporters. These show similarly shaped input functions, but with quantitative differences. The high-affinity transporter *araFGH* is activated to halfway its maximal level at sugar levels that are about 4-fold lower than that required for halfway activation of the low-affinity but high-throughput transporter *araE* (Figure S4). At sugar levels where *araE* expression is high, *araFGH* promoter activity begins to decline (Figures 2C and 2D; Figure S4), signifying a gradual switch from high-affinity to low-affinity transporter expression with increasing sugar levels. These small but significant differences in input functions would be difficult to observe using standard low-resolution methods.

Separation of Variables in the Input Functions

We next attempted to find a mathematical form that can quantitatively describe the input functions. We found that most of the input functions show a feature called “separation of variables”: they can be described, to a good approximation, as the product of two functions, each of which depends only on one input. Thus, \( f(x, y) = h_1(x)h_2(y) \). In other words, separation of variables means that inducers have multiplicative effects.

The monotonic input functions in this study are well described by a product of two Hill curves, \( f(x, y) = h_1(x)h_2(y) \), where \( h_1(x) = x^{n_1} / (K_1^{n_1} + x^{n_1}) \) and \( h_2(y) = y^{n_2} / (K_2^{n_2} + y^{n_2}) \) (Figures 3A and 3B). The Hill curve for each inducer is evaluated at saturating levels of the other inducer. The parameters \( n_1 \) and \( n_2 \) describe the steepness of the input function with respect to each input. The parameters \( K_1 \) and \( K_2 \) are effective concentrations of each input at which the input function rises halfway to its maximum value. Note that these parameters reflect not only the contribution of transcription factor binding to the promoter, but also the effects of upstream circuitry. Input functions with a triangular-shaped activation region, such as the arabinose system, are described by low Hill coefficients (Figures 3A and 3B). Rectangular-shaped input functions such as in the maltose system correspond to higher Hill coefficients (Figures 3B and 4). Such high Hill coefficients may be partly due to multiple binding sites in the promoters of the maltose genes that allow cooperative effects of MalT (Figure 4 and Richet, 2000). We find that the effective Hill coefficients \( n_1 \) and \( n_2 \) for genes from a given system tend to be similar, whereas the thresholds \( K_1 \) and \( K_2 \) are more variable (Figure S4). All 15 monotonic input functions in this study showed separation of variables to a good approximation (mean fit error of...
The action of CRP, due to this separation of variables, is akin to a linear amplifier that increases the response by a factor that depends only on cAMP and not on the sugar level.

Two of the nonmonotonic input functions (galP and galE) also showed separation of variables to a reasonable approximation (Figures 3C and 4). Here, the function h$_2$ is nonmonotonic. One input function, of the fucR promoter, was not well represented by separation of variables, due to its two-peaked structure.

Separation of variables, if found in other systems as well, can help to reduce the complexity of measuring input functions. The functions h$_1$ and h$_2$ can be measured by varying one input while holding the other input constant. This feature can greatly reduce the number of measurements needed to map a multidimensional input function. For example, to map a function of N = 4 inputs at a resolution of m = 8 concentration levels per input generally requires m$^N = 4096$ measurements, but only requires m·N = 32 measurements if the function has separation of variables. A reasonable strategy might be to measure h$_1$ and h$_2$ for each input separately, and then to sample several combinations of inputs to test whether separation of variables holds in each case.

**DISCUSSION**

The present study provides the first broad survey of two-dimensional input functions at high resolution. Instead of Boolean gates, we find diverse and intricate shapes, including nonmonotonic functions that peak at intermediate levels of the signals. Many of the functions showed triangular- or rectangular-shaped activation regions that were best described using a product of two continuous functions. This highlights the need to measure the input function of each gene in order to fully understand the computations performed by the cell. It is also the first glimpse of the range of computations that can be performed by a DOR network motif (Shen-Orr et al., 2002; Alon, 2007), a motif that is commonly found in transcription networks.

An interesting question raised by this study is why some input functions are nonmonotonic. The nonmonotonic shape of the galP and galE input functions might be related to the dual use of galactose as both a carbon source and a component of the cell wall (Neidhardt, 1996; Semsey et al., 2006). Such duality is
Molecular Cell
Diverse Input Functions Control Bacterial Genes

not shared by the other sugars in this study and may require more intricate computation for the gal sugar genes. The two-peak nature of the fucR input function might arise from the fact that it is an internal promoter: it may provide fine-tuning of FucR expression augmenting the upstream fucP promoter that also drives this gene.

The observed diversity of input functions raises at least two questions for future research. The first question is, how does each input function arise based on (1) the structure of the promoter (Bintu et al., 2005) and (2) the structure of the upstream circuitry such as feed-forward loops, auto-regulation loops, and other network motifs? This can be addressed by using the present assay to measure the effects of perturbations in the system on the detailed shape of the input function (Mayo et al., 2006) or by synthetic approaches of engineered circuits and promoters (Guet et al., 2002; Cox et al., 2007; Anderson et al., 2007; Guido et al., 2006; Sprinzak and Elowitz, 2005). Second, one may ask why is a particular shape selected for a particular gene? Addressing this requires understanding the relation between input functions and the distribution of input signals in the environment in which the organism evolved. This can be approached by frameworks such as cost-benefit evolutionary analysis (Dekel et al., 2005; Kalisky et al., 2007; Alon, 2006; Tanase-Nicola and Wolde, 2007). The present study can in principle be extended to map input functions in other gene systems, with the goal of understanding the design principles of gene control in cells.

EXPERIMENTAL PROCEDURES

Reporter Strains
Most reporter strains in this study are from the fluorescent reporter library developed in our lab and detailed in Zaslaver et al. (2006). Additional reporters for this study were prepared as described (Zaslaver et al., 2006) (malI [genomic coordinates 3550024-3550756], malP [3550024-3550756, comp], malk [4243912-4244387, comp], fucP [2931634-2932304], and fucA [2931634-2932304, comp]).

Growth Conditions and Measurements
Reporter strains (five different reporter strains from a given system + promoterless control strain pUA66 in each experiment) were grown overnight in M9 minimal medium containing 0.4% glucose, 0.2% casamino acids, and 50 μg/ml kanamycin at 37°C. Using a robotic liquid handler (FreedomEvo, Tecan), 96-well plates were prepared with 150 μl of M9 minimal medium containing 0.2% glucose, 0.05% casamino acids, 25 μg/ml kanamycin, and 96 different combinations of the system inducers (eight levels of serially diluted cAMP and 12 levels of serially diluted system-specific sugar). The wells were inoculated with the reporter strain at a 1:600 dilution from the overnight culture. This high dilution factor allowed a prolonged exponential phase. Wells were then covered with 100 μl of mineral oil (Sigma) to prevent evaporation, a step that we previously found not to significantly affect aeration or growth (Ronen et al., 2002; Zaslaver et al., 2004), and transferred into an automated incubator. Cells were grown in an incubator with shaking (6 Hz) at 30°C for about 20 hr. Every 8 min the plate was transferred by the robotic arm into a multwell fluorimeter (Infinite F200, Tecan) that reads the OD (600 nm) and GFP (535 nm). Relative plasmid copy number was assayed by real-time quantitative PCR (Lee et al., 2006) for the plasmid-encoded kanamycin resistance gene, and found not to measurably vary across growth conditions.

Data Analysis
Promoter activity for each well was calculated from the OD and GFP measurements after subtracting the OD and GFP backgrounds as described (1). GFP background was obtained for each well from the promoterless control strain. Promoter activity was calculated by computing the rate of accumulation of GFP per unit time divided by the OD (dGFP/dt/OD) (Ronen et al., 2002). The promoter activity was averaged over a window of 80 min (~1 cell cycle at exponential growth). Over this window, promoter activity was constant to a good approximation (variations less than 20%). Each input function was normalized to its maximal level (maximal promoter activity for each reporter is shown in Figure 4) and smoothed with a median filter (medfilt2 of Matlab 7.0). Growth rate (dOD/dt/OD/h) in units of doubling/hr at this time window was very similar (0.65 ± 0.03 doublings/hr) for all conditions.

Separation of Variables
The functions h1 and h2 correspond to the response to sugar and cAMP, respectively, holding the other input at saturating levels. Hill curves (full lines) are best fits to data. For monotonic input functions, we used Hill functions for h1 and h2, computing best-fit Hill parameters K and n using Matlab 7.0 (Figure 4 shows values with 95% confidence interval). For nonmonotonic input functions galP and galG, a Hill function with a basal level parameter h1(x) = b + (a/x/K)n / (1 + x/K)n was used because these promoters have a significant basal activity. A nonmonotonic function was used for h2, h2(y) = (b + y/K)n / (1 + y/K)n to capture the decline at high cAMP levels. The best fit parameters were for galP h1(b = 0.4, n = 2.3, K = 13, a = 0.65), h2(b = 0.4, n = 4, K = 1.7, K’ = 5.8) and for galG h1(b = 0.25, n = 1.4, K = 33, a = 1.5), h2(b = 0, n = 2.6, K = 1, K’ = 5.5). The fucR function was not modeled in the present study. The separation-of-variables model was obtained by multiplying the values of h1 and h2, f(x,y) = h1(x) h2(y), where x and y are the sugar and cAMP concentrations, respectively. The root-mean-square error (rmse) between f(x,y) and the measured input function was calculated (Figure 4).

SUPPLEMENTAL DATA
Supplemental Data include Supplemental Experimental Procedures, Supplemental References, and five figures and can be found with this article online at http://www.molecule.org/cgi/content/full/29/6/786/DC1/

ACKNOWLEDGMENTS
We thank the US National Institutes of Health, Israel Science Foundation, Mina and Everard Muller Foundation, Human Frontier Science Program, and the Kahn Family Foundation for support. S.K. is supported by Yeshaya Horowitz association through the Center for Complexity Science.

Received: September 24, 2007
Revised: November 21, 2007
Accepted: January 23, 2008
Published: March 27, 2008

REFERENCES
Diverse Input Functions Control Bacterial Genes


