CHAPTER 2: METHODOLOGY

The dynamics of various peculiar FFL was investigated using deterministic model. Simulation studies were done using CellDesigner version 4.2 (Funahashi *et al.*, 2008; Kitano *et al.*, 2005). FFL datasets from *E. coli* were downloaded from RegulonDB database version 7.4 (Gama-Castro *et al.*, 2011; Collado-Vides *et al*., 2011) to empirically determine the frequency of peculiar FFL. Details of these procedures are described in the following subsections.

2.1 Mathematical model for dynamics simulation

Consider the regulation of gene Y by transcription factor X. The rate of change of concentration of gene Y product, protein Y over unit time (t) is a function of the concentration of active form of X, X^*; it is given by

$$\frac{dY}{dt} = f(X^*), \tag{2.1}$$

where $f(X^*)$ is taken to be the Hill function, which is monotone and S-shaped. It is increasing when the transcription factor acts as an activator, and decreasing when it acts as a repressor. The Hill input function for an activator is given by

$$f(X^*) = \frac{\beta}{1 + K^{*n}/X^n}, \tag{2.2}$$

where β is the maximal expression level of the promoter, K is the activation threshold, and n is a constant for controlling the steepness of the curve; larger n leads to input functions that are more step-like. For regulation by repression, the Hill input function for a repressor is given by:

$$f(X^*) = \frac{\beta}{1 + X^{*n}/K^n}, \tag{2.3}$$

where K in this context is the repression threshold.

The step function or indicator function is another useful input function for studying the dynamics of transcription networks. Consider the indicator function for activation (X activates Y):

$$f(X^*) = \beta I(X^* > K_{xy}). \tag{2.4}$$

This can be obtained as the limiting case of equation (2.2) as n approaches infinity. Similarly, the indicator function for repression (X represses Y),

$$f(X^*) = \beta I(X^* < K_{xy}) \tag{2.5}$$

is obtained by taking the limit of equation (2.3) as n tends to infinity. In the case of FFL with AND gate, gene Z is regulated by both X^* and Y^* while in FFL with OR gate, regulation of gene Z is dependent on either X^* or Y^*. With the characteristics of Boolean input gate, the step function for activation of gene Z in FFL with AND gate is given by

$$f(X^*, Y^*) = \beta I(X^* > K_{xz}) I(Y^* > K_{yz}). \tag{2.6}$$

The step function for activation of gene Z in FFL with OR gate is given by

$$f(X^*, Y^*) = \beta \max\{I(X^* > K_{xz}), I(Y^* > K_{yz})\}. \quad (2.7)$$

In FFL with AND gate, gene Z is switched off when either X^* or Y^* falls below activation threshold. The step function for repression of gene Z in FFL with AND gate is given by

$$f(X^*, Y^*) = \beta\{I(X^* < K_{xy}), I(Y^* < K_{yz})\}. \quad (2.8)$$

In FFL with OR gate, gene Z is switched off when both X^* and Y^* fall below activation threshold. The step function for repression of gene Z in FFL with OR gate is given by

$$f(X^*, Y^*) = \beta I(X^* < K_{xz}) I(Y^* < K_{yz}) \quad (2.9)$$

The advantage of using the indicator function instead of the Hill function is that it greatly simplifies the qualitative understanding, the solution and analysis of the dynamic equation of FFL. For example, consider the activation of gene Y by transcription factor X. When the concentration of X^* is below K_{xy}, gene Y is switched off to zero expression level and no transcription of gene Y occurs. When the concentration of X^* is above K_{xy}, gene Y is switched on to maximal expression level and initiates the transcription of gene Y. The link between equations (2.2) with (2.4) and (2.3) with (2.5) can be exploited by setting n to be large (n = 30) when performing simulation using CellDesigner (see Section 2.2).

In practice, protein is synthesised and degraded, so a realistic model must incorporate both processes. Consider a simple regulation of gene Y by transcription factor X. The product of gene Y is subject to degradation by enzymes and is also subject to dilution due to the expansion of the cellular environment. Thus, the degradation/dilution rate of the product is proportional to its concentration αY, where α is the degradation/dilution rate. The complete differential equation for describing the dynamics of Y protein is given by

$$\frac{dY}{dt} = \beta_y - \alpha Y \tag{2.10}$$

This is a first-order linear differential equation that can be easily solved using basic calculus (Foerster, 2005). Its solution is given by

$$Y = \frac{\beta}{\alpha} + \left(Y_o - \frac{\beta}{\alpha} \right) e^{-\alpha t}, \tag{2.11}$$

where Y_o in this context is the initial concentration of gene Y product. β/α is the concentration of gene Y product at steady-state. Using equation (2.8), the rate of change of Y for FFL with both AND and OR gate depends only on X, and is given by

$$\frac{dY}{dt} = \beta_y I(X^* > K_{xy}) - \alpha_y Y. \tag{2.12}$$

For the rate of change of Z with AND gate, we have

$$\frac{dZ}{dt} = \beta_z I(X^* > K_{xy}) I(Y^* > K_{yz}) - \alpha_z Z. \tag{2.13}$$

For the rate of change of Z with OR gate, we have

$$\frac{dZ}{dt} = \beta_z \max\{I(X^* > K_{xz}), I(Y^* > K_{yz})\} - \alpha_z Z. \tag{2.14}$$

To provide a more realistic simulation of the whole FFL dynamics in a real transcription network, the dynamics of mRNA needs to be considered as well. Similar to the dynamics of protein, mRNA is also produced at maximal expression level when the transcription is switched on, and is also subject to degradation and dilution. The rate of change of mRNA concentration of gene Y, Y_m, is given by

$$\frac{dY_m}{dt} = \beta_m - \alpha_m Y_m. \tag{2.15}$$

The production of protein Y requires the existence of its mRNA, Y_m. Thus, the change of protein Y is dependent on the concentration of $Y_{m.}$ The rate of change of protein Y depending on Y_m is given by

$$\frac{dY}{dt} = \pi Y_m - \alpha Y, \tag{2.16}$$

where π is a scaling factor for the amount of protein amount produced per unit $Y_{m.,}$ and Y_m can be easily found using equation (2.11).

2.2 Simulating the dynamics of FFL

The dynamics of FFL of interests were simulated *in silico* using a systems biology software called CellDesigner (version 4.2; Funahashi *et al.*, 2008; Kitano *et al.*, 2005). It uses SBML (Systems Biology Markup Language; Hucka *et al*. 2003) as the format to represent biological network. The most attractive feature of this software is the ease in which biological networks can be constructed graphically in an intuitive way and the incorporation of simulation tools to solve the dynamics equations with different parameter settings.

Figure 2.1 shows the properties overview of this software. The main toolbar at the top panel provides access to main functions such as exporting the file in various formats and importing models from different sources of databases. Below the main toolbar, the nodes of a network can be given an explicit context, such as generic proteins, genes, ion channels, etc. The edge connecting two nodes can be given a direction and assigned an

interaction such as transition, translation, inhibition and catalysis (Figure 2.2). The user can change the initial conditions of the network such as the starting concentration of various protein and mRNA molecules.

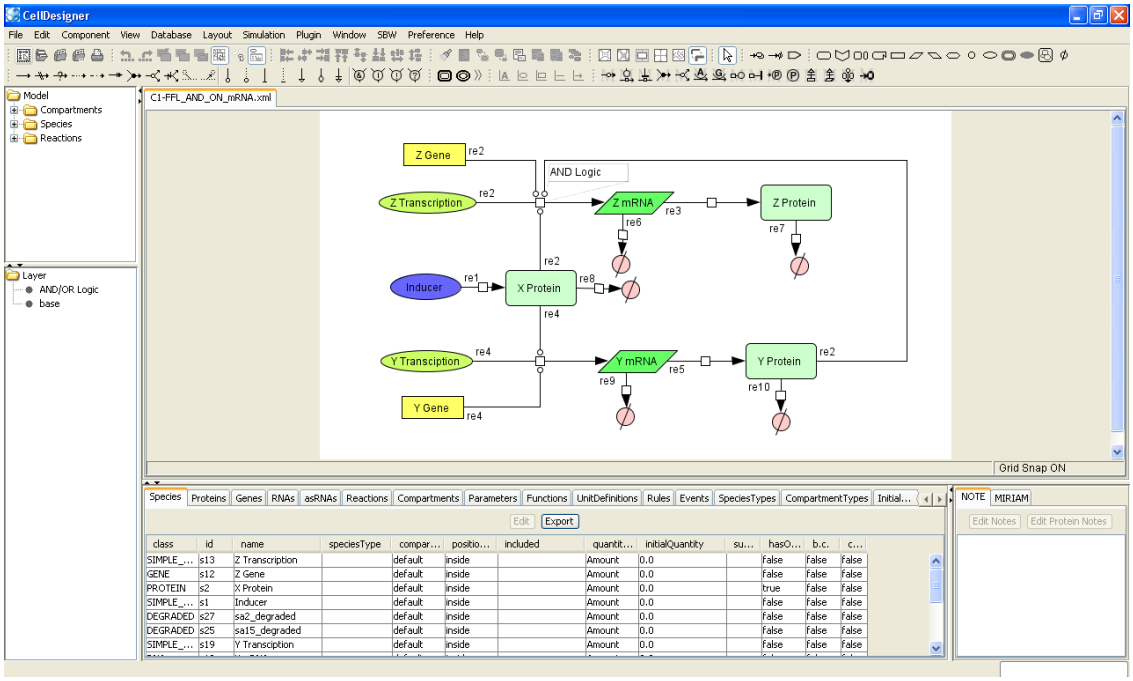


Figure 2.1: Properties overview of CellDesigner version 4.2.

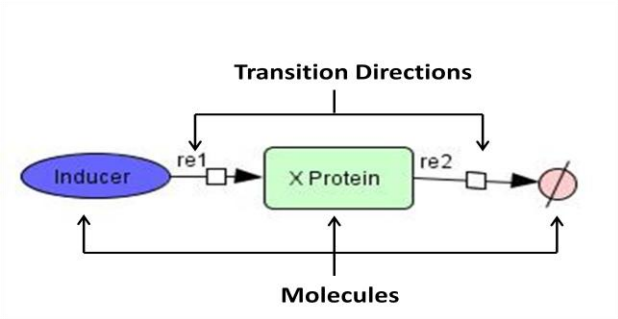


Figure 2.2: A simple biological network constructed with two molecules and two transition directions.

For each of the directed edge, a kinetics equation may be inserted with a set of parameters. The square located on an interaction edge allows the specification of the relevant dynamics equation. For example, the directed edge from inducer to *X* protein simple regulation network is simply inserted with a dynamic equation with "*beta*" equals to 5 (Figure 2.2). The other regulatory directions in a larger network may be inserted with different kinetics equation according to given mathematical models and implement the simulation to obtain the network dynamics. In this study, the dynamics of mRNA was also considered to construct the whole FFL network. It is advantageous because incorporation of mRNA into the FFL network is able to simulate the biological network in a more realistic way than standard analyses which often omit the dynamics of mRNA.

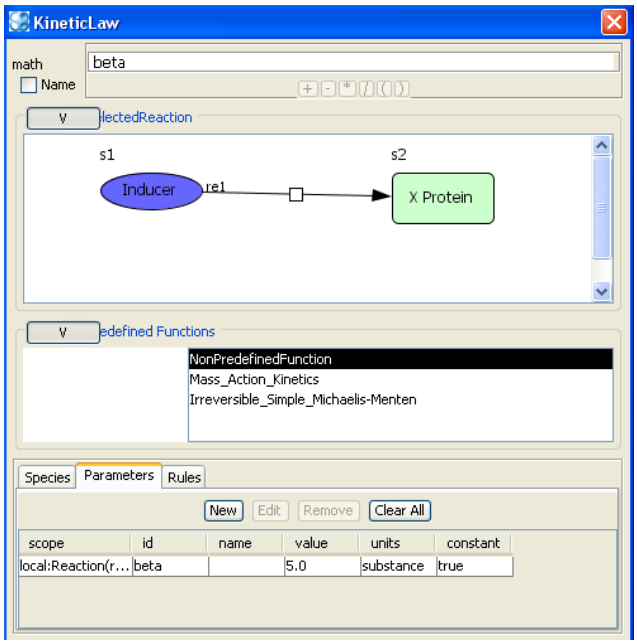


Figure 2.3: Kinetics for a network of simple regulation, *beta* with a parameter value of 5.

Six transcription network models belong to peculiar FFL types. The first group consisted of coherent FFL: C2-FFL, C3-FFL and C4-FFL with AND and OR gate. The second group consisted of incoherent FFL: I2-FFL, I3-FFL, and I4-FFL with only AND gate. The dynamics of incoherent FFL with OR gate was not simulated in this study because of degenerating patterns in production of protein *Z*. In incoherent FFL, the indirect path has the opposite overall sign against the direct path. In the case of I2-FFL and I3-FFL, activation or repression of protein *Z* production through either direct or indirect path permanently results in no production of protein *Z*. In the case of I4-FFL, activation or repression of protein *Z* production through either direct or indirect path permanently results in concentration of protein *Z* at steady-state.

Figure 2.4 shows an example of FFL transcription network model (C2-FFL) constructed using CellDesigner (version 4.2). The dynamics of these peculiar FFL transcription network models were simulated using appropriate mathematical equations in Section 2.1. Briefly, the model in Figure 2.4 is a CellDesigner representation of the following model: *X* protein is activated by its inducer and it represses (line with a bar at the end) the transcription of *Y* and *Z* genes. Following transcription, *Y* and *Z* mRNAs are translated into *Y* and *Z* proteins. The *Y* protein also acts as a transcription factor to activate the transcription of *Z* gene (line with a circle at the end). All mRNAs and proteins are subject to degradation and dilution (pink circle with a slash).

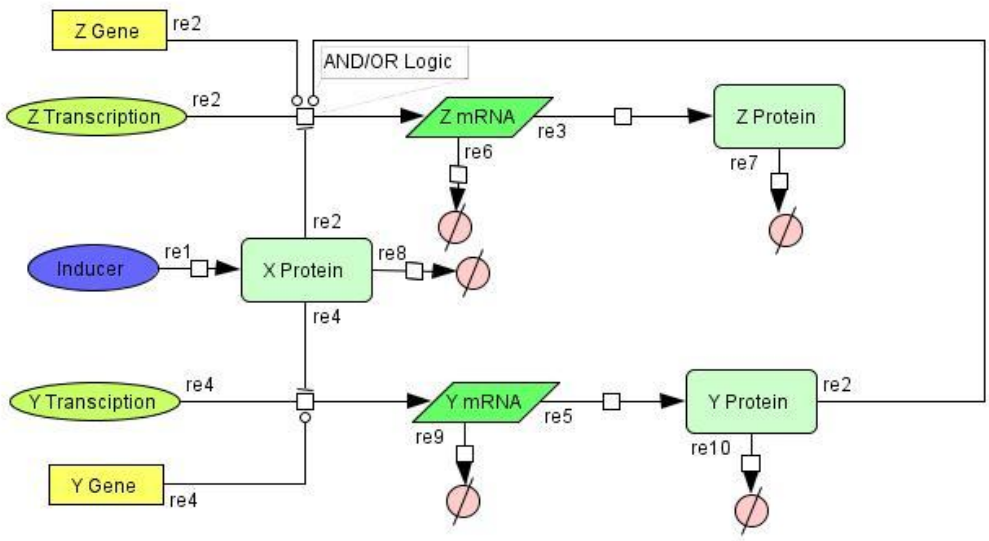


Figure 2.4: C2-FFL transcription network model constructed with CellDesigner version 4.2.

In order to systematically compare the simulation results, all parameters were standardised with a set of default values. The values of maximal promoter activity (β), activation/repression coefficient (K), index for steepness of the Hill function (n), degradation/dilution rate (α) and amount of protein produced per unit mRNA (π) were set to 5, 3, 30, 1 and 5, respectively. In certain simulations such as those involving the X protein in FFL at the OFF step, a starting concentration was given to the molecules to observe the pattern of production decrease. Once the values for all parameters had been set and the starting concentrations were initialised, the simulation process was begun.

Figure 2.5 shows the results of simulating the dynamics of C2-FFL in CellDesigner version 4.2. The simulation time was set to end at 100 units of time. Number of points to construct dynamics curve was set to 5000 points to obtain a smoother curve. Other setting

such as error tolerance and the equation solver remained default for these simulations. After the settings were adjusted, the simulation was run and the raw data of dynamics simulation for 5000 points was obtained as shown in the right panel. These raw data were then used to produce the results plots using R version 2.14.1 (R Development Core Team, 2011). The dynamics results of peculiar FFL obtained *in silico* using deterministic mathematical model were applied to study the naturally-existing peculiar FFL transcription network using *E. coli* as model organism.

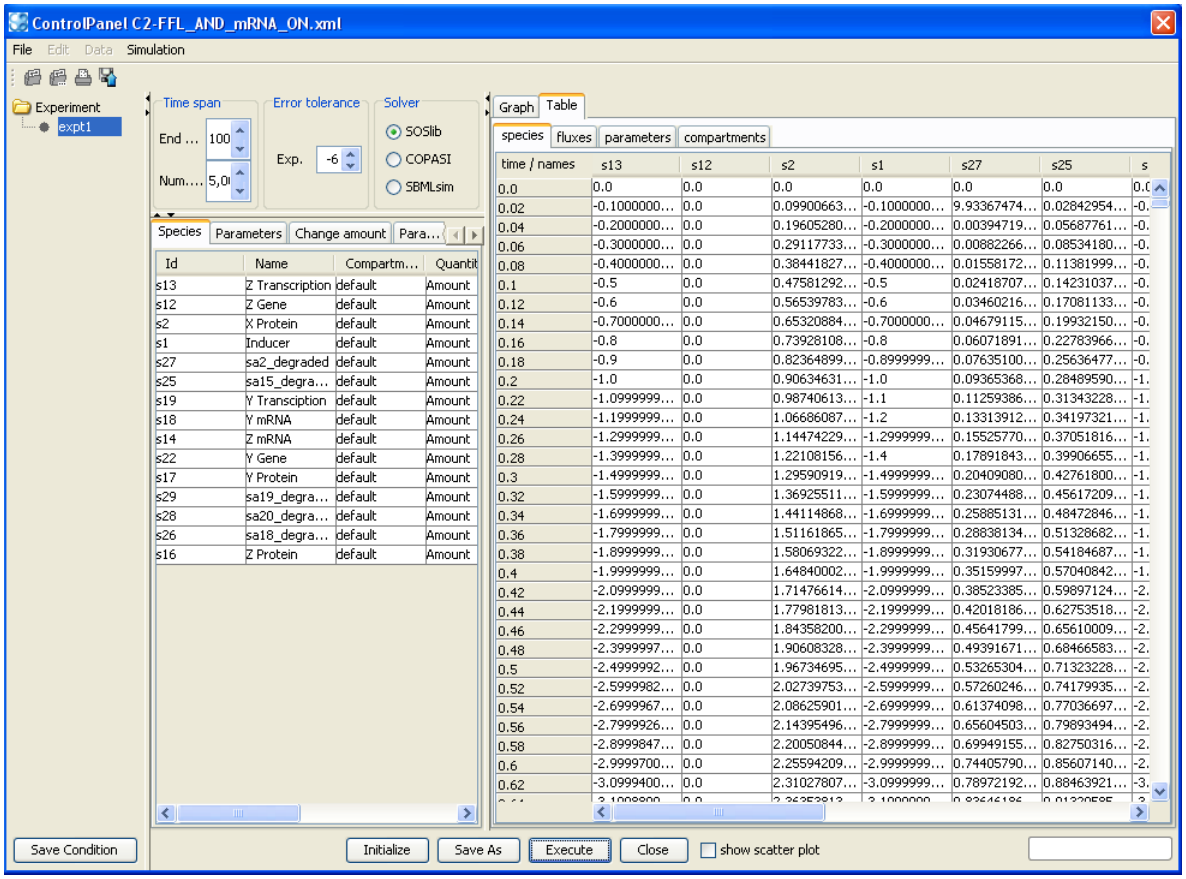


Figure 2.5: Screenshot of simulation outcome in CellDesigner version 4.2.

2.3 Extraction of *Escherichia coli* transcription network dataset from RegulonDB

Transcription network datasets were downloaded from RegulonDB version 7.4 (Gama-Castro *et al.*, 2011; Salgado *et al.*, 2012). RegulonDB is a database which contains regulatory network, gene and operon information for *E. coli* K12 (Huerta *et al.*, 2011; Salgado *et al.*, 2012). This database provides useful transcription network information, such as the gene involved and their interactions with transcription factor. In particular, "TF-operon interactions" under the category of Regulatory Network Interactions was selected so that the dataset of transcription networks could be downloaded.

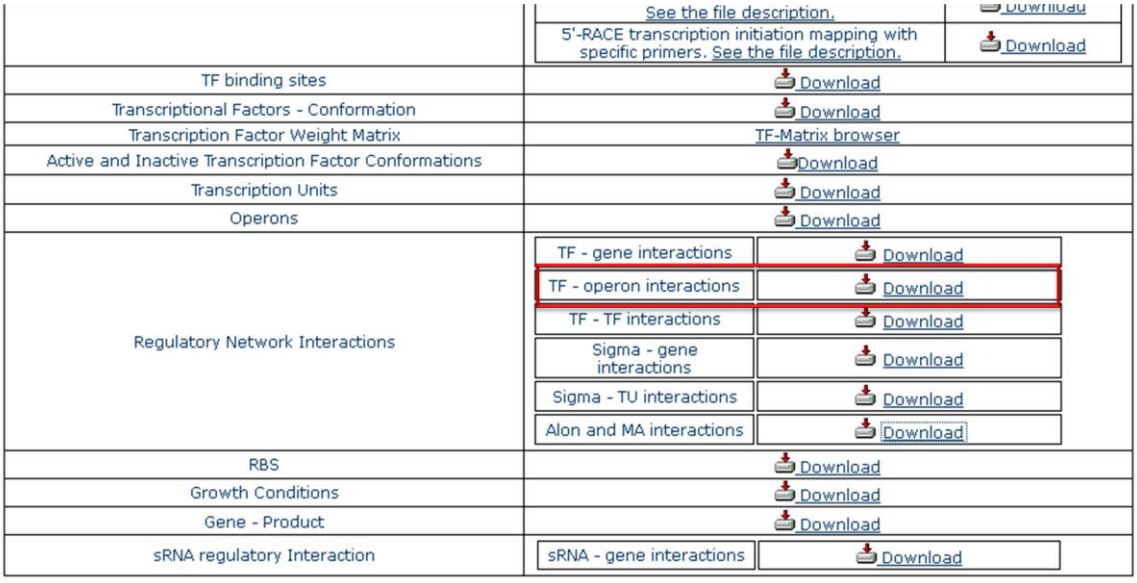


Figure 2.6: Screenshot of downloadable datasets for TF-operon interactions.

After downloading the data, all the interactions were carefully examined to identify the regulatory network interactions belonging to peculiar FFL types. For example, while scanning through the list of "TF-operon interactions", *malXY* operon is positively regulated

by CRP (Figure 2.7) and negatively regulated by MalI (Figure 2.8). Together with a negative regulation of *malI* gene by CRP (Figure 2.9), the network involving CRP, MalI and *malXY* operon was identified to be a type of C4-FFL. With the combination of both interactions, it can be verified that CRP, the product of *malI* gene and the product of *malXY* operon map to the *X*, *Y* and *Z* proteins in C4-FFL, respectively. The information of genes involved in peculiar FFL such as *malXY* operon was also obtained from this database (Figure 2.10).

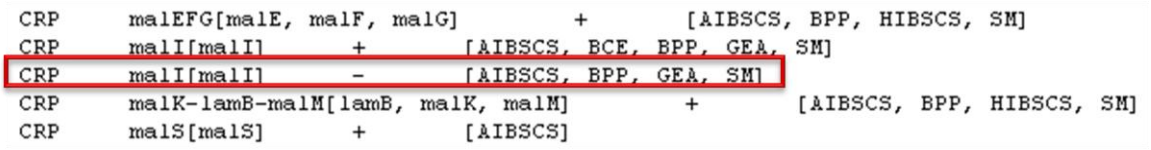


Figure 2.7: Negative regulation of *malI* gene by CRP.

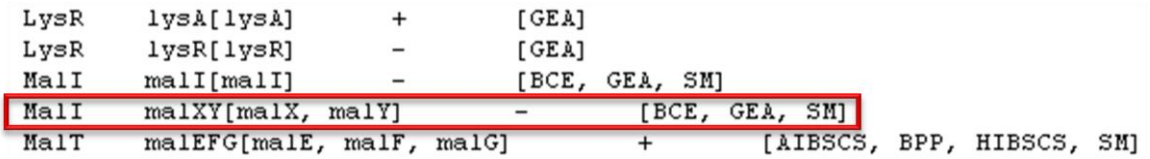


Figure 2.8: Negative regulation of *malXY* operon by MalI.

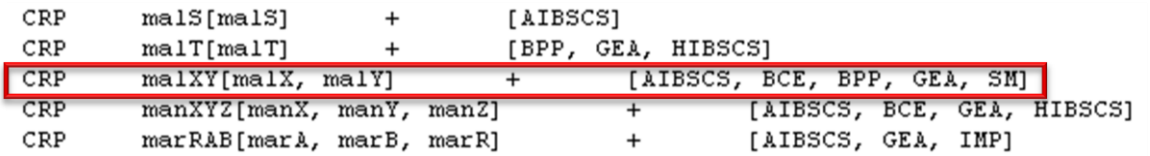
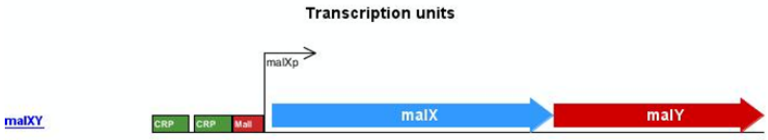


Figure 2.9: Positive regulation of *malXY* operon by CRP.



Object color & Code

OPERON:	Info
Name:	malXY
TRANSCRIPTION UNIT:	Info
Name:	malXY
Synonym(s):	OP00176
Gene(s):	*malX*, *malY* **Genome Browser M3D Gene expression**
Note(s):	The *malY* and *malX* genes are defined as part of the same operon according to their chromosomal positions, their induction pattern, and the polar effect on *malY* caused by an insertion in *malX*, but it is not clear whether *malY* is the last gene in this operon. In addition, the 5' portion of the *malY* gene contains stretches of sequences that would allow the formation of two stable stem-and-loop structures. The possible significance of these structures is unknown.
Evidence(s):	[BTE1] Boundaries of transcription experimentally identified [PM] Polar mutation
Reference(s):	[1] Otsuka J., et al., 1996 [2] Reidl J., et al., 1991
PROMOTER:	

Figure 2.10: Information of *malXY* operon.

Using the available FFL datasets downloaded from RegulonDB version 7.4 and results of simulation of the dynamics of these FFL, the necessity of retaining peculiar FFL for the cellular processes in *E. coli* was investigated.