CHAPTER 4: DISCUSSION

4.1 Limitations of deterministic mathematical model

Deterministic model is easy to use to simulate dynamics of biological networks. Nevertheless, there are several limitations using deterministic model (Lachor *et al.*, 2011; Wilkinson, 2006). Deterministic model is not accurate when the rate of the reaction is slow. The error increases when concentration of the proteins is low. It does not describe the systems response when the system is bi-stable or multiple stable. Moreover, it does not consider stochastic variation of time responses of the system.

In comparison to deterministic model, the process of stochastic model may evolve in several directions instead of only one direction in the case of deterministic model (Lachor *et al.*, 2011; Wilkinson, 2006). Gillespie algorithm is the most popular algorithm developed for stochastic simulation in chemical reactions (Gillespie, 1976; Lachor *et al.*, 2011). Stochastic model takes in consideration of the random fluctuation of protein molecules along the time in the system. Thus, it shall be applied to simulate the dynamics of FFL in order to compensate the limitations of deterministic model.

4.2 Parameter values used in simulating the dynamics of FFL

The change of parameter values used in simulating the dynamics of FFL generally does not affect the simulation results in terms of the response time and acceleration pattern with the exception of I3-FFL and I4-FFL. The acceleration pattern in I3-FFL and I4-FFL is sensitive to the maximal promoter activity of the second transcription factor. The acceleration pattern in I3-FFL and I4-FFL strongly requires higher values of maximal promoter expression level of the second transcription factor than the first transcription factor.

4.3 Response time and acceleration pattern in peculiar FFL

Coherent FFL shows whether the response of expression of gene *Z* is immediate or delayed. Mangan & Alon, (2003) described that there is opposite response time (delay or no delay) of protein *Z* production/repression in C1-FFL with AND and OR gates in term of the presence or absence of the S_x. As elaborated in Chapter 3, similar to C1-FFL, the response time of protein *Z* production/repression in C2-FFL with AND gate is opposite to the case of C2-FFL with OR gate. However, the response time is similar in the cases of C3-FFL and C4-FFL with AND and OR gates.

The response time of target gene expression in coherent FFL plays important roles in transcriptional regulation. The delay in target gene expression usually is required when

the gene product is less favourable in the biological system such as the arabinose sugar utilisation as secondary carbon source in *E. coli* as mentioned in Section 1.7. The delay in shutdown of target gene expression (e.g. C1-FFL with OR gate) prolongs the viability of the gene product. For example, *fliL* expression in *E. coli* is a type of C1-FFL with SUM input function, Boolean logic input similar to OR gate (Kalir *et al.*, 2005). FliL promotes *E. coli*'s flagella movement in response to environmental change such as availability of nutrient sources and temperature change. When the first transcription factor, FlhDC, is removed, the concentration of FliL remains at functional level for a period of time to allow further movement of the *E. coli*. If harmful element such as toxic compound exists in the environment, C1-FFL with SUM input function helps *E. coli* to further swim away from the site by prolonging the flagella movement.

Incoherent FFL shows different regulatory pattern from coherent FFL. All incoherent FFL show acceleration in production of protein *Z* at either the ON or OFF step. As described by Mangan & Alon, (2003), I1-FFL with AND gates shows accelerating pattern at the ON step. The production of protein *Z* drastically over shoots at certain time point and decreases back to the concentration at steady-state. In particular, the acceleration in production of gene *Z* in I3-FFL and I4-FFL requires strong promoter activity of gene *Y* in comparison to I2-FFL. This indicates that I2-FFL does not require gene *Y* promoter activity for acceleration in production of protein *Y*. In fact, a strong promoter activity of gene *Y* also enables acceleration in production of protein *Y* to acquire a significantly large amount of protein *Z* within a relatively shorter period of time.

The acceleration of target gene expression satisfies the requirement of certain biochemical process. For example, when glucose starvation occurs in *E. coli*, *gal* system in *E. coli* is activated when galactose is available as carbon source (Mangan *et al.*, 2006). I1-FFL promotes the production of enzymatic products of *galE* at a much faster rate in comparison to simple regulation to utilise galactose. This action is performed to ensure *E. coli* is survived by quickly resuming carbon metabolism when galactose is available as sole carbon source.

4.4 Challenge in determining the FFL types

It is often critical to determine the type of FFL of transcription networks in *E. coli* in some cases. For example, some transcription factors such as IHF act as dual regulators which are able to positively and negatively regulate target genes such as *ompF*. If the regulation type for IHF-*ompF* interactions was positive, the FFL type becomes I4-FFL instead of C2-FFL. Thus, the choice of positive or negative regulation is determined by literature review as well as the logical sense whether it contributes benefits to the whole biochemical network.

4.5 Biological functions of FFL

A total of 28 FFL transcription networks of *E. coli* was shown to be involved in various biological functions such as maltose utilisation, arabinose utilisation, anaerobic respiration and antibiotic response (Chao *et al.*, 1997; Csiszovszki *et al.*, 2011; Eichler *et*

al., 1996; Götz & Goebe, 2010; Kahramanoglou *et al.,* 2006; Martin *et al.,* 1996; Park *et al.,* 1997; Raberg *et al.,* 2011; Reidl & Boos, 1991; Schleif, 2010; Tsui & Freundlich, 1991; Unden & Bongaerts, 1997). Nine peculiar FFL were involved in maltose utilisation, anaerobic respiration and osmoregulatory response, respectively (Table 4.1; Cai & Inouye, 2002; Castillo-Keller *et al.,* 2006; Chao *et al.,* 1997; Huang *et al.,* 1990; Park *et al.,* 1997; Reidl & Boos, 1991; Tsui *et al.,* 1988; Tsui & Freundlich, 1991; Unden & Bongaerts, 1997).

The total number of FFL found in this study is restricted to the datasets from RegulonDB and is less than the total number of FFL published by Mangan & Alon, (2003). Thus, more datasets and literature review are needed in order to obtain higher number of total FFL counts.

4.6 Peculiar FFL transcription networks associated with maltose utilisation in *E. coli*

The cyclic AMP receptor protein (CRP) is a global regulator that has been identified to trigger transcriptional activation for 176 operons and repression for 16 operons *in vitro* (Shimada *et al.,* 2011; Zheng *et al.,* 2004). MalI is the repressor of *malXY* which is homologous to *E. coli* LacI, GaIR, or CytR and is repressed by CRP (Reidl *et al.,* 1989). *malX* encodes an enzyme belonging to the phosphotransferase system (PTS) that can recognize glucose and maltose as substrate and is also involved in the glucose and maltose transport system (Boss & Shuman, 1998; Lloyd *et al.,* 2008; Lloyd 2010; Reidl & Boos, 1991; Schlegel *et al.,* 2002). *malY* encodes a pyridoxal phosphate-containing enzyme which however, down-regulates the maltose system by inactivating *malT*, the key component to

activate other *mal* genes except *malI* and *malXY* in maltose utilisation. Thus, *malY* plays important role in preventing the maltose system from overexpression. In combination of *malX* and *malY*, the *malXY* encodes novel enzyme II of PTS.

The interactions among *crp*, *malI* and *malXY* form a type of C4-FFL with AND gate, which the transcription of *malXY* depends on the presence of *crp* and absence of *malI* (Figure 4.1). When cAMP is present, it serves as activation signal to activate CRP into its active form. Before the active form of CRP accumulates to its concentration sufficient to repress MalI production, production of *malXY* is repressed in the presence of *malI* activated by unknown inducer at the beginning and is resumed when active form of CRP has accumulated to a concentration which is able to trigger transcriptional initiation of *malXY* and repress production of MalI, the repressor which causes delay in MalXY production (Schlegel *et al.*, 2002). Together with other maltose system genes, the process of maltose utilisation occurs. When cAMP is absent, it results in immediate shut down of *malXY* transcription and activation of *malI* without delay**.**

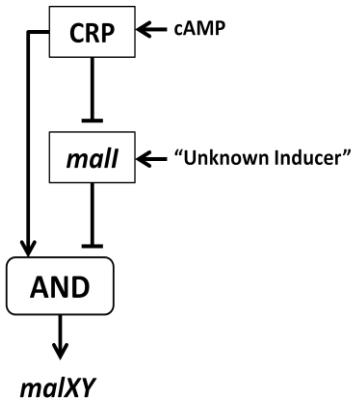

Figure 4.1: The interaction among *crp*, *malI* and *malXY* forms a type of C4-FFL with AND gate.

4.7 Peculiar FFL transcription networks associated with anaerobic respiration in *E. coli*

E. coli is facultative anaerobe which undergoes anaerobic respiration in the absence of oxygen (Jones *et al.*, 2011). In anaerobic respiration of *E. coli*, fumarate and nitrate are used as electron receptor when oxygen is depleted through a global regulator, FNR (Fumarate and Nitrate Reduction) protein (Becker *et al.*, 1996; Dean & Savageau, 2010; Levanon & Bennett, 2005; Salmon *et al.*, 2003; Shaw & Guest, 1982). FNR is responsible for expression of anaerobic respiratory chain and also regulation of the transition between aerobic and anaerobic state.

Arc (Aerobic Respiration Control) is a transcription factor made of arcAB which encodes the regulator for anaerobic respiration, and is positively regulated FNR (Levanon & Bennett, 2005; Perrenoud & Sauer, 2005). ArcA is activated during anaerobic growth.

ArcB is a histidine kinase which phosphorylates itself. The phosphoryl group is then transferred to ArcA under anaerobic state. Phosphorylated ArcA inhibits the production of enzymes participating in aerobic respiration. Simultaneously, It also activates the production of enzymes involved in fermentation.

Table 4.1 shows the list of operons and their products found in peculiar FFL which participate in anaerobic metabolism in *E. coli*. *cyoABCDE* and *cydAB* encode quinol oxidase bo_3 and quinol oxidase *bd*, respectively which catalyse the oxidation of ubiquinol-8 and the reduction of oxygen to water in aerobic respiratory chain (Cotter *et al.*, 1990; Govantes *et al.*, 2000; Unden & Bongaerts, 1997; Wall *et al.*, 1992).

Table 4.1: List of operons in peculiar FFL and their products regulated by FNR and ArcA in anaerobic respiratory system.

Operons	**Products**
cyoABCDE	Quinol oxidase bo_3
sdhCDAB	Succinate dehydrogenase
cydAB	Quinol oxidase *bd*
sucABCD	α-ketoglutarate dehydrogenase (*sucAB*) and succinyl coenzyme A synthetase (*sucCD*)
ndh	NADH dehydeogenase II
nuoABCEFGHIJKLMN	NADH dehydeogenase I
icd	Isocitrate dehydrogenase

α-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase (succinyl-CoA synthetase), are encoded by the *sucAB* and *sucCD* genes, respectively in the tricarboxylic acid (TCA) cycle (Park *et al.*, 1997; Unden & Bongaerts, 1997). α-ketoglutarate dehydrogenase converts α-ketoglutarate into succinyl-CoA and carbon dioxide via

oxidative decarboxylation process. Succinyl-CoA synthetase catalytically separates succinyl-CoA into succinate and CoA. *sdhCDAB* is located near to *sucABCD* which catalyses the oxidation of succinate to fumarate in the TCA cycle and reduces ubiquinone in the membrane.

NADH dehydrogenase I (NdhI), encoded by *nuoABCEFGHIJKLMN*, serves as primary dehydrogenase in the aerobic respiratory chain (Archer & Elliott, 1995; Gyan *et al.*, 2006; Meng *et al.*, 1997). Ndh1, together with NdhII (NADH dehydrogenase II) encoded by *ndh*, form NADH dehydrogenase complex which catalyse oxidation of NADH into NAD^+ in the TCA cycle. Ndh1 and ndh2 are used during fumarate and aerobic/nitrate respiration, respectively. Isocitrate dehydrogenase (ICD) encoded by *icd* catalyses the conversion of isocitrate to α-ketoglutarate, with production of NADH and carbon dioxide in the TCA cycle (Chao *et al.*, 1997).

The operons (except the global regulators, *fnr* and *arcA*) involved in anaerobic respiration were identified to play positive roles in the TCA cycle, part of the aerobic respiration instead of anaerobic respiration. It is not surprising that the target operons fall into the category of peculiar FFL, mostly C3-FFL with exception of the regulation of *cydAB* and *ndh* which belong to I3-FFL. Literature reviews suggested that these operons are regulated by both *fnr* and *arcA* which strongly implies that the Boolean input logic for the peculiar FFL in anaerobic respiration is AND gate.

In the case of C3-FFL with AND gate, four operons: *cyoABCDE*, *icdA*, *nuoABCEFGHIJKLMN* and *sdhCDAB_b0725_sucABCD* are negatively regulated by *fnr* and *arcA* (Archer & Elliott, 1995; Cotter *et al.*, 1990; Park *et al.*, 1997; Wall *et al.*, 1992). In order to switch the metabolism nature from aerobic to anaerobic state, the expression of these operons which encourages activation of aerobic respiratory system is terminated to favour the uses of fumarate and nitrate as receptor in anaerobic respiratory system. When oxygen is depleted, the respiratory system is immediately switched from aerobic state to anaerobic state in order to maintain the energy supply to the *E. coli* cell which is lower than the energy supplied by aerobic respiratory system. When oxygen is supplied, the anaerobic metabolism is still the energy-producing mechanism for *E. coli* cell will not be immediately switched off to allow the aerobic respiration to resume as energy-producing mechanism.

In the case of I3-FFL with AND gate, two operons: *cydAB* and *ndh* are negatively regulated by *fnr* and positively regulated by *arcA* (Cotter *et al.*, 1990; Meng *et al.*, 1997; Wall *et al.*, 1992). When oxygen is depleted, the respiratory system is immediately switched from aerobic state to anaerobic state similar to the regulation of the operons in C3-FFL. However, when oxygen is present, *cydAB* and *ndh* are positively regulated by *arcA* in opposite to the regulation of the other four operons in C3-FFL. When oxygen is present, *arcA* accelerates the expression of the target operons and stimulates the production of the operon products higher than their usual concentration at certain time point.

The reason why *cydAB* and *ndh* are regulated in the type of I3-FFL but not C3-FFL is yet to investigate. Both C3-FFL and I3-FFL differ at the OFF step of the regulation which is the resume of aerobic respiratory system when oxygen is present. Fumarate and nitrate are no longer used as electron receptor. It is probably because the expression of certain operons such as *cyoABCDE* is delayed. The expression of specific operons has to accelerate at the beginning in order to trigger the expression of other essential operons as soon as the oxygen is available for utilisation since oxygen is a better electron receptor to produce higher energy yield than fumarate and nitrate. *cydAB* and *ndh* are strongly associated with *cyoABCDE* and *nuoABCEFGHIJKLMN*, respectively to form clusters of enzyme complex. Thus, the favour of I3-FFL in regulation of *cydAB* and *ndh* may serve as starting enzymes to initial the whole aerobic respiratory process.

Majority of peculiar FFL found in datasets from RegulonDB are involved in anaerobic metabolism in *E. coli*. Interestingly, the operons which are regulated by two global regulators, FNR and ArcA, participate in aerobic respiration rather than anaerobic respiration. Multiple operons are regulated by few global regulators probably because of saving energy or cost in synthesising too many regulators. The transition between aerobic and anaerobic state is simply archived by switching between activation and repression of the regulation. The reason why two of the operons involved in anaerobic respiration: *cydAB* and *ndh* are regulated in the pattern of I3-FFL as opposite of the others is unknown but it is believed that the acceleration of the operon products is required in initialising the whole process.

4.8 Peculiar FFL transcription networks associated with osmoregulatory response in *E. coli*

In ihf-*ompB-ompC and* ihf-*ompB-ompF* networks of osmoregulatory system, *ihf* encodes integration host factor (IHF), a direct participant in certain site-specific recombination events such as the integration of λ phage viral genome into the host genome (Friedman, 1988). The protein products of the *ompB* operon, OmpR and EnvZ, are required for the expression and osmoregulation of the major outer membrane proteins, OmpF and OmpC (Cai & Inouye, 2002; Castillo-Keller *et al.*, 2006).

OmpR and EnvZ are members of a family of bacterial signal transduction proteins. OmpR is a DNA-binding protein which regulates *ompF* and *ompC* transcription (Forst *et al.*, 1989; Nikaido & Vaara, 1985). EnvZ is thought to act as an osmosensor and to affect *ompF* and *ompC* by phosphorylating and dephosphorylating OmpR. Nonphosphorylated OmpR functions as a transcription activator for *ompC* and phosphorylated OmpR functions as a transcription activator for *ompF*. Both phosphorylated and nonphosphorylated OmpR are co-existing and thus both OmpC and OmpF are available simultaneously. OmpC and OmpF construct part of the porin system which produces relatively nonspecific pores or channels that allow the passage of small hydrophilic molecules across the outer membrane. The *ihf-ompB-ompC* and *ihf-ompB-ompF* networks are built favourably in C2-FFL than in C1-FFL. OmpC and OmpF production depend on the presence of *ompB* product or the absence of IHF. Thus, these two FFL are associated with OR gate.

Under normal circumstances, *ompB*, *ompC* and *ompF* in *E. coli* are regulated only in response to the osmotic change in the environment without the invasion of λ phage. *ihf* is involved in λ phage viral infection of *E. coli* and is able to repress the expression of *ompB*, *ompC* and *ompF* (Huang *et al.*, 1990; Tsui *et al.*, 1988; Tsui & Freundlich, 1991). Thus, the repression of *ompB*, *ompC* and *ompF* transcription serves as a defense mechanism against infection which results in shut down of *ompB*, *ompC* and *ompF* transcription and no production of OmpC and OmpF. Therefore, porin is not synthesised on the membrane. Due to the presence of certain concentration of OmpB, the repression of OmpC and OmpF production will not occur unless the concentration of OmpB decreases. Thus, there is dependency of *ompC* and *ompF* expression on OmpB and repression of OmpC and OmpF production is delayed.

The lack of porin further disallows the passage of molecules across the outer membrane and therefore prevents the *E. coli* cell from infection. If the FFL type was C1-FFL, IHF positively regulates porin synthesis and causes detrimental effect to the survival of *E. coli*. When the IHF concentration decreases, there is no delay in resuming the porin synthesis to ensure quick restoration of osmoregulatory system in *E. coli*.

4.9 Validation of results of FFL dynamics by experimental approach

The computational results of dynamics simulation of peculiar FFL were not validated by laboratory experiments in this study. Nevertheless, the dynamics results of two major FFL types: C1-FFL and I1-FFL have been validated by researchers experimentally (Kalir *et al.*, 2005; Mangan *et al.*, 2006; Mangan *et al.*, 2003).

Kalir *et al.* (2005) showed delayed GFP (Green Fluorescent Protein) fluorescent emission of FliL production in the absence of FliA (C1-FFL with OR gate at OFF step). GFP fluorescence remains at steady-state of approximately 0.9 before it decreases. As elaborated in the research finding of Mangan *et al.* (2003), the results of GFP fluorescence of *araBAD* expression in C1-FFL with AND gate showed delay of AraBAD production at the ON step when the activation signal, L-arabinose, is present. Mangan *et al.* (2006) proved that *galE* expression exhibits acceleration pattern by quantitative measurement of GFP fluorescence which is similar to the dynamics of I1-FFL with AND gate at the ON step.

Despite the fact that the dynamics of FFL reflects the real biological phenomenon in terms of the response time and pattern of gene expression, and the concentration of gene product, various parameters such as the maximal promoter expression level and degradation/dilution rate are not able to be determined based on experimental results. Thus, the parameter values need to be pre-defined to simulate the dynamics of FFL.

4.10 Association of FFL with other network motifs

Although FFL network motif is useful in understanding the transcriptional regulation involving three elements/nodes, it is often associated with different types of network motifs in a complex biological network (Prill *et al.*, 2005). Auto-regulation network motif may be found within the FFL (Kalir *et al.*, 2005). Two or more FFL may form a larger and more complex FFL termed multi-output FFL (Kashtan *et al.*, 2004). The influence of such association on dynamics of FFL was not investigated in this study.

4.11 Effect of RNA interference on dynamics of FFL

RNA interference (RNAi) is a process which represses the gene transcription at post-transcriptional event by short non-coding RNA (Shabalina *et al.*, 2008). Two types of short non-coding RNA are produced in eukaryotic cells: miRNA and siRNA. Both miRNA and siRNA inhibit translation by different mechanisms. miRNA (MicroRNA) prevents translation by partially binding to the complementary section of mRNA sequence. siRNA (small interfering RNA) prevents translation by degrading the mRNA sequence. In prokaryotic cells, small RNA which functions like microRNA and siRNA such as stress-induced Hfq-binding small RNA of *E. coli* is also able to induce gene-silencing at translational level (Morita *et al.*, 2006).

RNAi may affect the simulation result of dynamics of the whole FFL dynamics with the dynamics of mRNA taken into consideration. When the mRNA translation is disrupted by RNAi, the scaling factor of protein amount produced per mRNA unit, π, which depends on the concentration of functional mRNA, is also greatly reduced. The rate of change of protein depending on mRNA concentration is altered. Thus, the FFL dynamics interfered by RNAi may not represent the actual FFL transcription network in certain biological system under this condition.

The small non-coding RNA involved in RNAi, especially miRNA, may themselves act as transcriptional repressors in the regulatory network of FFL (Herranz & Cohen, 2010). The FFL type which comprises miRNA as one of the nodes is called miRNA-mediated FFL. The possible role of RNAi molecules as transcriptional repressors in FFL in *E. coli* was not investigated in this study. Nevertheless, miRNA mediated-FFL has been described to be the key factor of the development of squamous carcinoma in the head and neck (Cohen *et al.*, 2009). The three nodes which make up this miRNA-mediated FFL are protein kinase Cα, miR15a (microRNA) and cyclin E. miR15a directly activates cyclin E expression and protein kinase Cα activates and represses cycline E and miR15a, respectively. Cyclin E promotes DNA synthesis and eukaryotic cell growth. miR15a inhibits the expression of cyclin E and prevent uncontrolled growth of tumour cells. When protein kinase Cα is present, it forms C4-FFL together with the other two elements. Protein kinase Cα represses miR15a and promotes cyclin E expression. The consequence of this C4-FFL network is the uncontrolled development of squamous carcinoma. Thus, protein kinase Cα is the key protein causing development of squamous carcinoma.

CHAPTER 5: CONCLUSIONS

CellDesigner 4.2 can be a useful tool for investigating the dynamics of network motifs such as FFL. This is because simulation results from CellDesigner can be used as a starting point to discuss and understand the behavior of transcription networks that contain peculiar FFL. Without this, it is difficult to understand the biological network processes from a quantitative perspective. On the other hand, *in silico* simulation of FFL dynamics can only hypothetically explain a gene regulation phenomenon. Laboratory experiments still need to be performed to validate the simulation results. In this study, *E. coli* was used as the model organism to study the FFL dynamics due to the relatively complete understanding that we have for this organism at the molecular, biochemical and physiological levels. This approach is also useful to study the FFL dynamics of transcription networks of other kinds of organisms with available datasets of transcriptional regulation. Nevertheless, more work needs to done to understand the physiology of the organisms of interest in order to study their biological dynamics.