

PECULIAR FEED-FORWARD LOOP NETWORK MOTIFS IN *ESCHERICHIA*
COLI

KOK ENG PIEW

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

PECULIAR FEED-FORWARD LOOP NETWORK MOTIFS IN
ESCHERICHIA COLI

KOK ENG PIEW

(SGJ 100004)

SUBMITTED TO

INSTITUTE OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

IN PARTIAL FULFILMENT

OF THE REQUIREMENT FOR

THE DEGREE OF MASTER OF BIOINFORMATICS

2012

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: KOK ENG PIEW

(I.C/Passport No: 831130-14-5353)

Registration/Matric No: SGJ100004

Name of Degree: MASTER OF BIOINFORMATICS

Title ("this Work"):

PECULIAR FEED-FORWARD LOOP NETWORK MOTIFS IN *ESCHERICHIA COLI*

Field of Study: BIOINFORMATICS

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any Copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date:

Subscribed and solemnly declared before,

Witness's Signature

Date:

Name:

Designation:

ABSTRACT

A feed-forward loop (FFL) is one type of transcription network motifs studied in systems biology. So far, eight types of FFL based on their regulatory patterns have been identified. The majority of FFL found in transcription networks of biological systems have been identified to belong to coherent and incoherent type 1 FFL. Nevertheless, six FFL types (coherent and incoherent type 2, 3 and 4 FFL) are present in low frequencies in *Escherichia coli*. The persistence and not elimination of these “peculiar” FFL through natural selection is perplexing. One way to understanding this conundrum is to study the dynamics of these FFL and empirically study their relative abundance using public databases. To this end, data mining from RegulonDB and simulation using CellDesigner 4.2 were carried out. Coherent FFL shows delay in target gene expression upon the activation or repression by the activated forms of first and second transcription factors. Incoherent FFL shows acceleration of target gene expression that results in over production of target gene product. In particular, the acceleration of target gene expression in incoherent type 3 and 4 FFL are highly dependent on the promoter binding activity of the second transcription factor. We identified 84 of 1702 transcription factor-operon interactions in RegulonDB to be involved in FFL transcription networks in *E. coli*. A total of 28 FFL were identified, nine of them found to participate in biochemical processes such as maltose utilization, arabinose utilization and anaerobic respiration. The efficacy of peculiar FFL was discussed in the context of these processes. We find that *in silico* simulation of FFL dynamics using CellDesigner 4.2 provide explicit results that are useful for guiding biological interpretation of important biological processes in *E. coli*. It is also a practical

means of generating useful hypotheses in gene regulatory networks for experimental validation in the laboratory.

ABSTRAK

Gelung suap hadapan (GSH) merupakan salah satu jenis motif rangkaian transkripsi yang dikaji dalam sistem biologi. Secara keseluruhannya, terdapat lapan jenis FFL berdasarkan corak pengawalaturan yang telah dikenal pasti. Kebanyakan GSH yang dikenal pasti dalam rangkaian transkripsi sistem biologi ialah jenis koheren dan jenis 1 tak-koheren. Walaubagaimanapun, terdapat enam jenis GSH (jenis 2 hingga 4 koheren dan tak-koheren) wujud dalam frekuensi yang rendah dalam *Escherichia coli*. Kewujudan GSH "luar biasa" ini dan kegagalan penyingkiran mereka melalui pemilihan semula jadi menjadi kemusykilan. Satu cara untuk memahami teka-teki ini adalah untuk mengkaji dinamik GSH ini dan mengkaji secara empirikal kelimpahan relatif mereka dengan menggunakan pangkalan data awam. Untuk tujuan ini, perlombongan data dari RegulonDB dan simulasi menggunakan CellDesigner 4.2 telah dijalankan. GSH yang koheren menunjukkan kelewatan dalam ekspresi gen sasaran apabila pengaktifan atau penindasan oleh faktor transkripsi pertama dan kedua yang diaktifkan. GSH tidak koheren menunjukkan pecutan ekspresi gen sasaran yang menyebabkan lebih pengeluaran produk gen sasaran. Khususnya, pecutan ekspresi gen sasaran dalam GSH tak-koheren jenis 3 and 4 adalah sangat bergantung kepada aktiviti pengikatan promoter faktor transkripsi kedua. Kami telah mengenal pasti 84 daripada 1702 interaksi faktor transkripsi-operon di RegulonDB yang terlibat dalam rangkaian transkripsi GSH di dalam *E. coli*. Sejumlah sebanyak 28 GSH telah dikenal pasti, sembilan daripada mereka didapati mengambil bahagian dalam proses

biokimia seperti penggunaan maltosa, penggunaan arabinosa dan respirasi anarobik. Keberkesanan GSH luar biasa telah dibincangkan dalam konteks proses ini. Kami mendapati bahawa simulasi *in silico* dinamik GSH dengan menggunakan CellDesigner 4.2 memberi keputusan yang jelas yang berguna untuk membantu panafsiran proses biologi penting dalam *E. coli*. Ia juga adalah satu cara praktikal bagi menjana hipotesis yang berguna dalam rangkaian pengawalaturan gen untuk pengesahan eksperimen di makmal.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my advisor, Dr. Khang Tsung Fei, for his excellent guidance throughout this project. Working on this project has been a great learning experience for me and I am grateful to him for giving me this opportunity.

I am also thankful to my parents, elder sister, elder brother, friends and colleagues for all the love and support. I would never have been able to finish this project without the motivation from my friends, help from my colleagues, and moral support from my family.

TABLE OF CONTENT

| <u>No.</u> | <u>Topic</u> | <u>Page</u> |
|-------------------|--|--------------------|
| I. | ABSTRACT | iii |
| II. | ACKNOWLEDGEMENT | vi |
| III. | TABLE OF CONTENT | vii |
| IV. | LIST OF FIGURES AND TABLES | ix |
| V. | LIST OF ACRONYMS | xii |
| | | |
| 1.0 | INTRODUCTION | |
| 1.1 | Biological networks in systems biology | 1 |
| 1.2 | Transcription network model | 2 |
| 1.3 | Network motifs – Individual building block of biological network complex | 4 |
| 1.4 | Types of transcription network motifs | 4 |
| 1.5 | Feed-forward loop (FFL) | 6 |
| 1.6 | Kinetics and dynamics of feed-forward loop | 9 |
| 1.7 | Importance of FFL in biological systems | 13 |
| 1.8 | Peculiar FFL types: The FFL types other than C1-FFL and I1-FFL | 16 |
| 1.9 | Objectives of study | 17 |
| 2.0 | METHODOLOGY | |
| 2.1 | Mathematical model for dynamics simulation | 18 |
| 2.2 | Simulating the dynamics of FFL | 22 |
| 2.3 | Extraction of <i>Escherichia coli</i> transcription network dataset from RegulonDB | 28 |

| | |
|---|----|
| 3.0 RESULTS | |
| 3.1 Dynamics of C2-FFL with AND/OR gate | 31 |
| 3.2 Dynamics of C3-FFL with AND/OR gate | 34 |
| 3.3 Dynamics of C4-FFL with AND/OR gate | 36 |
| 3.4 Dynamics of I2-FFL with AND gate | 39 |
| 3.5 Dynamics of I3-FFL with AND gate | 42 |
| 3.6 Dynamics of I4-FFL with AND gate | 45 |
| 3.7 Peculiar FFL transcription networks in <i>E. coli</i> | 48 |
| 4.0 DISCUSSION | |
| 4.1 Limitations of deterministic mathematical model | 49 |
| 4.2 Parameter values used in simulating the dynamics of FFL | 49 |
| 4.3 Response time and acceleration pattern in peculiar FFL | 50 |
| 4.4 Challenge in determining the FFL types | 52 |
| 4.5 Biological functions of FFL | 52 |
| 4.6 Peculiar FFL transcription networks associated with maltose utilisation in <i>E. coli</i> | 53 |
| 4.7 Peculiar FFL transcription networks associated with anaerobic respiration in <i>E. coli</i> | 55 |
| 4.8 Peculiar FFL transcription networks associated with osmoregulatory response in <i>E. coli</i> | 60 |
| 4.9 Validation of results of FFL dynamics by experimental approach | 62 |
| 4.10 Association of FFL with other network motifs | 63 |
| 4.11 Effect of RNA interference on dynamics of FFL | 63 |
| 5.0 CONCLUSIONS | 65 |
| REFERENCES | 66 |
| APPENDICES | 73 |

LIST OF FIGURES AND TABLES

| <u>Figure</u> | <u>Title</u> | <u>Page</u> |
|----------------------|--|--------------------|
| 1.1 | Transcriptional activation of gene Y by transcription factor X . | 2 |
| 1.2 | Three mechanisms of transcriptional repression of gene Y by transcriptional repressor X : (1) Re-organising the chromatin by recruiting chromatin-remodelling proteins, (2) interfering the binding of transcription factors to promoter sequences, and (3) modifying the structure of RNA polymerase. | 3 |
| 1.3 | Types of simple regulation of gene Y by transcription factor X . (a) Positive autoregulation (NAR), and (b) negative autoregulation (NAR). | 5 |
| 1.4 | Single-input modules network motifs involved in regulation of multiple target genes $Y_1, Y_2, Y_3, \dots Y_n$ by transcription factor X . | 5 |
| 1.5 | Single-input modules network motifs involved in regulation of multiple target genes $Y_1, Y_2, Y_3, \dots Y_n$ by a set of transcription factors $X_1, X_2, X_3,$ and X_4 . | 6 |
| 1.6 | Feed-forward loop with transcriptional regulation by activation and/or repression. X, Y and Z are transcriptional factors. X activated by signal S_x regulates both Y and Z while Y activated by signal S_y regulates only Z . | 7 |
| 1.7 | Eight types of FFL. (a) Coherent FFL, and (b) Incoherent FFL. | 8 |
| 1.8 | FFL with Boolean input gates. (a) FFL with AND gate, and (b) FFL with OR gate. | 9 |
| 1.9 | Dynamics of C1-FFL with AND gate. K_{xy} and K_{yz} represent activation threshold for regulation of gene Y by transcription factor X and regulation of gene Z by transcription factor Y , respectively. | 11 |
| 1.10 | Dynamics of I1-FFL with AND gate. K_{xy} represents activation threshold and K_{yz} represents repression threshold for regulation of gene Z transcription by transcription factor X and transcriptional repressor Y , respectively. | 13 |
| 1.11 | <i>ara</i> system (C1-FFL) and <i>lac</i> system (simple regulation) regulated by CRP activated by its activation signal, cAMP. | 14 |
| 1.12 | Distribution of FFL counts in transcription networks of <i>E. coli</i> and <i>S. cerevisiae</i> . | 16 |

| | | |
|-------------|---|-----------|
| 2.1 | Properties overview of CellDesigner version 4.2. | 23 |
| 2.2 | A simple biological network constructed with two three molecules and two transition directions. | 23 |
| 2.3 | Kinetics for a network of simple regulation, beta with a parameter value of 5. | 24 |
| 2.4 | C2-FFL transcription network model constructed with CellDesigner version 4.2. | 26 |
| 2.5 | Dynamics simulation panel of CellDesigner version 4.2. | 27 |
| 2.6 | Downloadable datasets for TF-operon interactions. | 28 |
| 2.7 | Negative regulation of <i>malI</i> gene by CRP. | 29 |
| 2.8 | Negative regulation of <i>malXY</i> operon by MalI. | 29 |
| 2.9 | Positive regulation of <i>malXY</i> operon by CRP. | 29 |
| 2.10 | Information of <i>malXY</i> operon. | 30 |
| 3.1 | Dynamics of C2-FFL with AND gate at ON and OFF step. | 32 |
| 3.2 | Dynamics of C2-FFL with OR gate at ON and OFF step. | 33 |
| 3.3 | Dynamics of C3-FFL with AND gate at ON and OFF step. | 35 |
| 3.4 | Dynamics of C3-FFL with OR gate at ON and OFF step. | 36 |
| 3.5 | Dynamics of C4-FFL with AND gate at ON and OFF step. | 38 |
| 3.6 | Dynamics of C4-FFL with OR gate at ON and OFF step. | 39 |
| 3.7 | Dynamics of I2-FFL with AND gate at ON and OFF step. | 41 |
| 3.8 | Dynamics of I3-FFL with AND gate at ON and OFF step. | 43 |
| 3.9 | The effect of gene <i>Y</i> promoter activity on acceleration of protein <i>Z</i> production in I3-FFL with AND gate at OFF step. | 44 |
| 3.10 | Dynamics of I4-FFL with AND gate at ON and OFF step. | 46 |
| 3.11 | The effect of gene <i>Y</i> promoter activity on acceleration of protein <i>Z</i> production in I4-FFL with AND gate at ON step. | 47 |
| 4.1 | The interaction among <i>crp</i> , <i>malI</i> and <i>malXY</i> forms a type of C4-FFL with AND gate. | 54 |

| <u>Table</u> | <u>Title</u> | <u>Page</u> |
|---------------------|---|--------------------|
| 3.1 | List of peculiar FFL associated with various biological functions extracted from TF-operon interactions, RegulonDB. | 48 |
| 4.1 | List of operons in peculiar FFL and their products regulated by FNR and ArcA in anaerobic respiratory system. | 56 |

LIST OF ACRONYMS

| <u>No</u> | <u>Acronym</u> | <u>Abbreviation</u> |
|------------------|-----------------------|-------------------------------------|
| 1 | Arc | Aerobic Respiration Control |
| 2 | C1-FFL | Coherent Type 1 Feed-Forward Loop |
| 3 | C2-FFL | Coherent Type 2 Feed-Forward Loop |
| 4 | C3-FFL | Coherent Type 3 Feed-Forward Loop |
| 5 | C4-FFL | Coherent Type 4 Feed-Forward Loop |
| 6 | cAMP | Cyclic Adenosine Monophosphate |
| 7 | CRP | Cyclic AMP Receptor Protein |
| 8 | DNA | Deoxyribonucleic Acid |
| 9 | DOR | Dense Overlapping Regulons |
| 10 | FFL | Feed-Forward Loop |
| 11 | FNR | Fumarate and Nitrate Reduction |
| 12 | GFP | Green Fluorescent Protein |
| 13 | GTP | Guanosine Triphosphate |
| 14 | I1-FFL | Incoherent Type 1 Feed-Forward Loop |
| 15 | I2-FFL | Incoherent Type 2 Feed-Forward Loop |
| 16 | I3-FFL | Incoherent Type 3 Feed-Forward Loop |
| 17 | I4-FFL | Incoherent Type 4 Feed-Forward Loop |
| 18 | ICD | Isocitrate dehydrogenase |
| 19 | IHF | Integration Host Factor |
| 20 | miRNA | MicroRNA |
| 21 | NADH | Nicotinamide Adenine Dinucleotide |

| | | |
|-----------|-------|---------------------------------|
| 22 | NAR | Negative Autoregulation |
| 23 | NdhI | NADH dehydrogenase I |
| 24 | NdhII | NADH dehydrogenase II |
| 25 | Omp | Outer Membrane Porin |
| 26 | PAR | Positive Autoregulation |
| 27 | PTS | Phosphotransferase System |
| 28 | RNA | Ribonucleic Acid |
| 29 | RNAi | RNA Interference |
| 30 | SBML | Systems Biology Markup Language |
| 31 | SIM | Single-Input Modules |
| 32 | siRNA | Small Interfering RNA |
| 33 | TCA | Tricarboxylic Acid |
| 34 | XML | Extensible Markup Language |
| 35 | URL | Uniform Resource Locator |