Peer-reviewed version available at *Int. J. Mol. Sci.* **2018**, *19*, 2335; doi:10.3390/ijms19082335

Large-scale Functional Analysis of CRP-mediated Feed-Forward Loops

Chi-Dung Yang[1, 2, 5§], Hsi-Yuan Huang[1, 3§], Sirjana Shrestha[1, 5], Yen-Hua Chen[4], Hsien-Da Huang[1, 5*] and Ching-Ping Tseng[5*]

[1]Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu 300, Taiwan

[2]Institute of Population Health Sciences, National Health Research Institutes, Miaoli 350, Taiwan

[3]Department of Laboratory Medicine, China Medical University Hospital, Taichung 420, Taiwan

[4]Department of Microbiology and Immunology, Weill Cornell Medicine, Cornell University, New York 10021, USA

[5]Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 300, Taiwan

[§]Chi-Dung Yang and Hsi-Yuan Huang should be considered joint first author.

[*]To whom correspondence should be addressed.

Abstract

The feed-forward loop (FFL) is an important and basic network motif to understand specific biological functions. Cyclic-AMP (cAMP) receptor protein (CRP), a transcription factor (TF), mediates catabolite repression and regulates more than 400 genes in response to changes in intracellular concentrations of cAMP in *Escherichia coli*. CRP participates in some FFLs like *araBAD* and *araFGH* operons and adapt to fluctuating environmental nutrients thus enhancing the survivability of *E. coli*. Although computational simulations have been used to explore the potential functionality of FFLs, a comprehensive study of the functions of all structural types based on *in vivo* data is lacking. Also, the regulatory role of CRP-mediated feed-forward loops (CRP-FFLs) remain unclear to date. Using EcoCyc and RegulonDB, we identified 393 CRP-FFLs in the *E. coli.* Dose-response genomic microarray of *E. coli* revealed dynamic gene expression of each target gene of CRP-FFLs in response to a range of cAMP dosages. All eight types of FFLs were present in CRP regulon with various expression patterns of each CRP-FFL, that were further divided into five functional groups. Microarray and reported regulatory relationships identified 202 CRP-FFLs which were directly regulated by CRP in these eight types of FFLs. Interestingly, 30% (147/482) of genes were directly regulated by CRP and CRP-regulated TFs, indicating that these CRP-regulated genes were also regulated by other CRP-regulated TFs responding to environmental

signals through CRP-FFLs. Furthermore, we applied gene ontology annotation to reveal the biological functions of CRP-FFLs.

Keywords: feed-forward loop (FFL); cAMP receptor protein (CRP); transcription factor (TF)

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

Introduction

To survive, microbes must accustom to certain environmental changes, produced by physiological signals such as nutrients, salts, and molecular species, and physicochemical states such as temperature, redox potential, and osmolality [1]. To react to those environmental signals, *Escherichia coli* regulates gene expressions using many different gene regulators, such as transcription factors (TFs) [2].

Recently databases like EcoCyc 22.0 [3] and RegulonDB 9.4 [4, 5] have reported 201 TFs in *E. coli*. Linking the regulation information between TFs and their targets forms a complicated gene regulation network (GRN). Many computational analyses have attempted to assess the importance of TFs, which include information including the number of TF target genes, and interactions between other TFs [6-8]. Complex transcription regulation networks (TRN) are investigated by analyzing small blocks of networks called "network motifs". Feed-forward loop (FFL) is an important network motif in *E. coli* gene regulation that contains two TFs (*X* and *Y*) and one or more genes. In an FFL, TF *X* regulates TF *Y*. *X* and *Y* regulate target gene *Z* and jointly regulates its transcription rate. Depending on the definition, eight types of FFLs are formed. An FFL is "coherent" if the direct effect (positive or negative) of *X* on *Z* is the same as the indirect effect through the second TF *Y*. On the contrary, if the direct effect of *X* on gene *Z* is opposite from the indirect effect through *Y*, the FFL is "incoherent". Coherent and incoherent FFLs each include four types which are as follows: coherent type 1 (Coh1), coherent type 2 (Coh2), coherent type 3 (Coh3), coherent type 4 (Coh4), incoherent type 1 (InCoh1), incoherent type 2 (InCoh2), incoherent type 3 (InCoh3) and incoherent type 4 (InCoh4) [9, 10]. In *E. coli*, coherent FFLs serve as a sign-sensitive delay element in an *ara* system [11], and incoherent FFLs accelerate the response time of a *gal* system [12]. In addition, there are two other simple topological generalizations of FFL obtained by replicating the appropriate nodes, called multi-Y and multi-output FFLs [13]. In the case of multi-Y FFLs, the target gene *Z* is regulated by single TF *X* and multiple *Y*. In the case of multi-output FFLs, the multiple target genes *Z* are simultaneously regulated by a single TF *X* and *Y*.

The basic theory of FFL has been applied in different fields. For example, FFL can be easily adjusted to produce a single-pulse response and further regulate blood sugar in diabetic patients [14]. In human disease, the latest research has found that the interaction of FFL members, miRNA, TF, and gene interaction may promote effective treatment for ischemic stroke [15]. FFL also plays an important role in the maintenance of

homeostasis in biological response [16]. However, in the field of microbiology, although biological databases have documented the regulatory relationship between transcription factors and genes, there is still lack of integrated experiments and analysis. For the first time, we presented the comprehensive FFLs study. We chose the *E. coli* as a research model and achieved a complete analysis of CRP mediated FFLs.

In *E. coli*, CRP has been reported to be a global TF which regulates over 400 genes [17, 18]. In CRP-mediated feed-forward loops (CRP-FFLs), CRP and a CRP-regulated TF are denoted as *X* and *Y*, respectively. Although Alon *et al.* reported the kinetic features of CRP-FFLs [12] in *E. coli*, due to the limitation of a small sample size to support statistical significance and information on gene interaction, only Coh1, Coh2, and Coh4 FFLs were reported. Evidence indicated that there are many potential FFLs in the CRP regulon, but data for the interaction between gene regulations does not provide a further understanding of the biological functions of CRP-FFLs in *E. coli.* In addition, databases like EcoCyc and RegulonDB revealed that many genes are involved in multi-Y FFLs, such as GadA, a glutamate decarboxylase enzyme coded by the *gadA* gene which confers resistance to acid conditions and is regulated by CRP and CRP-regulated TFs, GadX and FIS. With CRP, GadX and FIS can respectively co-regulate *gadA* gene expression and form Coh2 and InCoh2 FFLs. In the following section, we described single-Y and multi-Y FFLs involved in CRP-FFLs as CRP-mediated single-Y FFL (CRP-sFFL) and CRP-mediated multi-Y FFL (CRP-mFFL) respectively. We defined a CRP-sFFL as an FFL that contains CRP and one CRP-regulated TF, and a CRP-mFFL as a motif that involves CRP and two or more CRP-regulated TFs to regulate one target gene.

To clarify the functional CRP-FFLs in *E. coli*, cAMP dose-response analysis was performed by microarray, thus determining the capacity of CRP to regulate target genes expression involved in CRP-FFLs at various concentrations (0–10 mM) of cAMP. Accurately controlling the activity of CRP in cAMP dose-response experiments requires that intracellular concentration of cAMP could be maintained. Therefore, we applied the *cyaA* and *tolC* double mutant strain that has lost the ability to produce and exhaust cAMP to measure the relative expression level of a target gene in the presence or absence of cAMP [19].

This study aims to establish a comprehensive understanding of the biological functions of CRP-FFLs by combining known gene regulation information and dose-response microarray experiments. The CRP-FFLs can be further divided into CRP-sFFLs and CRP-mFFLs which exist in various expression patterns. We combined

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

target genes with similar profiles for gene ontology (GO) enrichment analysis to further explore the functional categories of these genes. In brief, we provided the first comprehensive study using CRP-FFLs to understand how *E. coli* adapts to environmental changes under different concentrations of cAMP (**Figure 1**).

Results

CRP-mediated feed-forward loops in *Escherichia coli*

We collected 3958 interactions between TFs and genes from EcoCyc and RegulonDB. Among these interactions, CRP regulates 482 genes which contain 51 TFs. From these, we identified 393 CRP-FFLs consisting of 292 genes and 43 TFs. To obtain and clarify the functional CRP-FFLs in *E. coli*, cAMP dose-response experiments were performed by microarray as follows. We first filtered out the genes with expression levels of all cAMP added samples related to zero cAMP control that are less than the absolute value of 0.5 from microarray data. Microarray results are summarized in a heat map in **Figure 2** and **Additional File 1**. Finally, we obtained 202 CRP-FFLs consisting of 147 genes and 37 TFs to determine the optimal number of clusters for each type of CRP-FFLs. The statistics of data in eight types of CRP-FFLs is shown in **Table 1**. The distribution of CRP-FFLs in *E. coli* based on statistics of Mangan *et al*. (10), EcoCyc and RegulonDB, and CRP-FFLs in this study is compared in **Figure 3** and **Additional File-1**. Some previously-reported CRP-regulated genes did not exhibit significant differential expressions in our experimental conditions, probably because these genes require other transcription factors expressed in some specific environmental conditions [17, 18]. The data were used to construct CRP-FFLs which were then subjected to further analysis.

CRP-FFLs characterized by distinct cAMP dose-responses

After the CRP-FFLs were found, we subjected the target genes of all available CRP-FFLs to hierarchical clustering based on the cAMP dose-response microarray data. We divided CRP-FFLs into two groups; CRP-sFFLs and CRP-mFFLs, in order to investigate the relationship between them. Records from EcoCyc and RegulonDB along with the microarray data were used to distinguish 101 CRP-sFFLs and 101 CRP-mFFLs among 202 CRP-FFLs. **Figure 4A** presents the expression profile of target genes regulated by CRP-sFFLs, whereas **Figure 4B** presents the CRP-mFFLs profile of target genes regulated by more than one CRP-regulated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

TF. We investigated whether each of the CRP-FFLs could be assigned to a specific cluster. Furthermore, according to their distinct cAMP dose-response profiles, the CRP-FFLs were divided into five groups (from SG1 to SG5) in CRP-sFFLs and six groups (from MG1 to MG6) in CRP-mFFLs.

For example, both group SG2 in CRP-sFFLs and group MG6 in CRP-mFFLs can form similar expression profiles as cAMP increases (**Figures 4A** and **4B**). The cAMP dose-responses are nearly identical and are characterized by significant levels of CRP-FFLs until cAMP reaches a high concentration. These results indicate that the dose-response of each type of CRP-FFL may be related to their topology and transcriptional regulation with CRP and CRP-regulated TFs.

Gene ontology enrichment analysis of FFLs in CRP regulatory network

Since CRP-sFFLs and CRP-mFFLs share similar cAMP-dose response profiles, we combined the genes of similar profiles for gene ontology enrichment analysis to further explore the functional categories of these genes. The combinations of CRP-sFFL groups and CRP-mFFL groups are shown in **Table 2** and **Additional File 2**. According to the results of gene ontology enrichment analysis, we classified the target genes of CRP-sFFLs and CRP-mFFLs into five functional groups (FG1 to FG5).

Carbohydrates are the main carbon source for ATP synthesis, and they also serve as the signals from the environment. CRP is a well-known global regulator of carbohydrate metabolism [18]. To determine which CRP-FFLs were used by *E. coli* for different carbon sources, we surveyed the FFLs involved in functions related to sugar metabolism. We found that most of the CRP-mediated Coh1 and InCoh1 FFLs were involved in the regulation of sugars. Alon *et al.* (16) mentioned that the concentration of glucose was low in the mammalian colon, therefore, we expected that the concentration of cAMP would be sufficiently high to form the cAMP-CRP complex and regulate CRP-regulon in intestinal bacteria. In agreement with the findings of Alon, our cAMP-dose response data indicated that the gene expression of the most target genes which were involved in Coh1 and InCoh1 FFLs in carbon utilization pathways increased significantly under high concentrations of cAMP. The results also showed that groups SG2 and MG6 were dominated by CRP-FFLs with the transporter function and carbon source utilization. In contrast, the CRP-FFLs involved in the iron and pH homeostasis preferred to use all CRP-FFL types except for the Coh3 FFL. Other features are summarized in **Table 3** and **Additional File**

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

2. Many CRP-FFLs were involved in more than one functional groups and further study of the biological functions of specific FFLs in different biological activity is worthwhile.

Discussion

Integration of different FFLs types diversifies the output response

One of the key questions regarding the function of network motifs is whether a simple three gene FFL is sufficient and/or necessary to perform a specific biological function. Moreover, most current studies in bioinformatics depend solely on preexisting gene integration data and are thus limited by the number of FFLs which have been identified [20]. Here, we focused on a master transcription factor to improve the CRP regulon by combining cAMP dose-response microarray data.

Although the proposed approach has enriched our knowledge about the structure-function relationships in CRP-mediated gene interaction networks, there are uncertainties created by several factors such as the presence of unknown transcriptional regulations, and a transcription factor sometimes serving as a dual regulator [2]. Therefore, a structure such as the described CRP-sFFL may indeed include more complex regulatory networks, such as CRP-mFFLs or feedback regulations. To minimize bias resulting from the lack of comprehensive gene interaction data, we compared the clustering results of the cAMP-dose response profiles for CRP-sFFLs and CRP-mFFLs and unexpectedly found the profiles to be quite similar, with the clustering result shown in **Figure 4**. In fact, four out of five expression profile clusters derived from CRP-sFFLs were also observed in CRP-mFFLs. In this case, there are two possible explanations. On the one hand, due to the limited gene interaction data between CRP-regulated transcription factors and target genes, the so-called "CRP-sFFLs" might be indeed part of CRP-mFFLs. On the other hand, the profile of CRP-mFFLs might be dominated by a master CRP-sFFL, whose promoter had a higher affinity or multiple CRP binding sites, and thus presents a profile similar to a specific CRP-sFFL. If the later turns out to be the case, we expect that the frequency of FFLs in the cAMP dose-response profile of CRP-sFFLs would be similar to that of CRP-mFFLs in a similar profile.

On the basis of the hierarchical clustering results, the expression profiles of the target genes in CRP-sFFLs and CRP-mFFLs can be subdivided into five and six groups respectively. While comparing the hierarchical clustering results of CRP-sFFLs and CRP-mFFLs, we found that the group SG1 profiles are only observed in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

CRP-sFFLs, and such a group is dominated by Coh1 and InCoh1 FFLs. In contrast, other four groups; SG2, SG3, SG4, and SG5 share similar profiles with the CRP-mFFLs. Given that CRP acts as an activator in Coh1 and InCoh1 FFLs and the average correlation coefficient (~0.99) of profiles being high in the SG1 group, we hypothesized that this profile pattern represents one of the basic profiles of Coh1 and InCoh1 FFLs in CRP-sFFLs.

However, our results also demonstrated that one FFL could be used in more than one dose-response profile. The genes in groups SG2 and MG6 were all regulated by Coh1, Coh4, InCoh1, and InCoh4 FFLs, regardless of whether they were regulated by CRP-sFFLs or CRP-mFFLs. All FFL combinations that gave rise to a dose-response profile similar to that of the group genes were composed of a combination of Coh1, Coh4, InCoh1, and InCoh4 FFLs. It is worth noting that CRP acts as an activator in these FFLs, suggesting that CRP alone was sufficient to increase the expression of the target genes, whereas the transcriptional regulation of other TFs was required to modulate the expression of target genes when the cAMP concentration was above 0.3 mM.

The expression profiles in the SG3 group can be observed in most CRP-sFFLs, except for Coh3 and InCoh4 FFLs. In this cluster, the expression of target genes decreased as the concentration of cAMP increased. The decreasing of such genes fits into a linear form in log scale. In general, such a group is a counterexample of the SG2 group, because CRP acts as a repressor in Coh2, InCoh2, and InCoh3 FFLs. In agreement with the above observation of combining FFLs, a composition of Coh2 with InCoh2 FFLs in the MG4 group can produce a profile similar to that in the SG3 group. Nevertheless, some FFLs such as Coh1, InCoh1, and InCoh4, and their combination in the MG2 group also formed a dose-response profile like that in the SG3 group. Although a CRP, acting as an activator, regulates these FFLs, it is possible that the activity of *Y* is higher than that of CRP. If this *Y* is an activator, when the DNA binding activity of the CRP increases along with the concentration of cAMP, the occupation of *Y* on the promoter of the target gene is gradually replaced by CRP, resulting in a decrease of *Z* gene expression. In contrast, for Coh4 and InCoh1 FFLs, where *Y* is a repressor, the repression effect of *Y* may outweigh the strength of CRP.

In addition, with increasing or decreasing concentrations of cAMP, some FFLs showed a bell-shaped dose-response curve in the SG4 (**Figure 4**) group and the expression of the target genes in SG4 group peaked when

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

the concentration of cAMP ranged from 0.1 to 0.3 mM. Theoretically, to achieve a bell-shaped response curve, the target gene should be regulated by two arms of regulation with opposite effects [21]. Therefore, we can expect that some incoherent FFLs can produce such a profile. Indeed, an InCoh1 FFL in CRP-sFFLs can generate a bell-shaped profile. Interestingly, the combination of Coh2 and Coh3 FFLs in the MG1 group or InCoh1 and InCoh4 FFLs in the MG5 group can generate similar profiles, probably because such combinations include transcription factors with opposite regulation effects.

In contrast to the bell curve, some combinations of FFLs were characterized by a good curve distribution, which is an inverse bell curve. In CRP-sFFLs, InCoh2 FFLs in the SG5 group exhibited a good curve dose-response profile, similar to the combination of Coh2 with InCoh2 FFLs in the MG3 group. Our results demonstrated that each FFL can produce some specific dose-response profiles, which can be generated by combining specific FFLs. These results not only provide clues for the biological functions of FFLs but also highlight the principles of simplified complex gene regulatory networks with CRP-mFFLs. **Table 3** provides detailed information for CRP-sFFLs and CRP-mFFLs.

Functional analysis of FFLs in CRP-regulated networks

Analysis of cAMP-dose response profiles derived from the target genes in either CRP-sFFLs or CRP-mFFLs revealed that they shared four similar response profiles. To explore the biological significance of these motifs, we further conducted a gene-ontology analysis. Strikingly, genes within the same clusters were involved in the same biological pathways. Indeed, the more similar the biological processes in which the genes participated, the more similar their dose-response profiles under various concentrations of cAMP. The gene ontology analysis revealed five major functional groups each involving a group of genes demonstrating similar cAMP responses and network motifs.

Unlike the other CRP-sFFL profile clusters, the profile of the SG1 group was only found in target genes in CRP-sFFLs, suggesting that such profiles may only reflect the function of Coh1 and InCoh1 FFLs. Therefore, we can confirm that, at least for those genes in the SG1 group, CRP-sFFL is sufficient to generate a specific profile. Gene ontology analysis clearly showed that the SG1 group was predominated by genes involved in central metabolism processes, including the tricarboxylic acid cycle, cellular respiration, and the generation of energy

Peer-reviewed version available at *Int. J. Mol. Sci.* **2018**, *19*, 2335; doi:10.3390/ijms19082335

by the oxidation of organic compounds. The expression of these genes increased along with the concentration of cAMP. The increased expression begins to form trace amount of cAMP (0.01 mM) and reached a stationary phase as the concentration of cAMP reached 1 mM. This result is in accordance with a previously reported mechanism in which the expression of TCA cycle-related genes are activated as the cell consumes environmental glucose compounds to facilitate the TCA cycle [22-24].

Except for the SG1 group, the remaining CRP-sFFLs profiles were also found in the dose-responses of CRP-mFFLs. For example, target genes in the SG2 group of CRP-sFFLs and the MG6 group of CRP-mFFLs were both characterized by a sigmoid curve. Interestingly, their profiles represented a switch-like behavior to accelerate the expression of target genes only if the concentration of cAMP was above 0.3 mM. In other words, the regulation allows target genes to reach their maximum expression levels until the cAMP concentration exceeds a certain threshold.

This behavior is consistent with carbolite repression, a well-known mechanism in *E. coli* [25, 26]. When *E. coli* is grown in a medium with mixed carbon sources, it first consumes glucose but represses the expression of the transporters of other carbon sources. The expression of the transporters for those alternative carbohydrate compounds are activated after glucose concentrations have dropped below a certain level. This mechanism is important for energy conservation.

In contrast to the above-mentioned clusters, there were three clusters in which the expression of target genes showed an inverse correlation with the concentration of cAMP. In the SG3 group of CRP-sFFLs, and the MG2 and MG4 groups of CRP-mFFLs, the expressions of target genes were repressed by the cAMP-CRP complex resulting in the decrease of their expressions as the cAMP level rose. These gene profiles were found to be involved in stress response, especially iron and pH homeostasis.

Importantly, according to the database records, CRP activated the genes in the MG2 group, thus this group consisted of Coh1 and InCoh1 FFLs. However, the microarray data showed an opposite trend where the expression profile was inversely correlated with the concentration of cAMP. For example, the Fec operon composed of FecABC genes that encodes the ferric citrate transporter was reported to be activated by the cAMP [27], whereas our results showed that the operon is indeed repressed by the cAMP. We speculated that this is due to experimental limitations in the previous studies which did not use mutant strains to allow the precise

control of cAMP levels. Further experiments, such as promoter activity assays and the northern blot, are required to redefine the regulatory effect of the cAMP on these genes.

Due to the availability of a small number of identified genes within the SG4 group with MG1 and MG5, and within the SG5 group with MG3, we could not retrieve any significant GO terms. The relevant biological functions and their directly associated genes are provided in **Figure 5** and **Table 2**. The genes in the SG4 group of CRP-sFFLs and the MG1 and MG5 groups of CRP-mFFLs demonstrated similar bell-shaped profiles. Interestingly, the expression levels of the MG1 and MG5 groups were respectively degraded and elevated when the concentration of cAMP exceeded 1 mM as compared to their basic expression levels. This is probably due to the high degree of intracellular cAMP activating the other unknown regulation in the MG5 group. The genes in the FG4 functional group showed biological function in the L-glutamate biosynthetic process, electro transport chain, and iron transport. One of the key processes in L-glutamate biosynthesis requires the activity of glutamate synthase, which requires NADPH. Such an enzyme has four substrates (L-glutamine, 2-oxoglutarate (α-ketoglutarate), NADPH, and H+) and generates the two products: L-glutamate and NADP+. This reaction helps bacteria produce L-glutamate using NADPH as the acceptor and iron as a cofactor. The remaining FG5 functional group was involved in the acid resistance system and nitrite reductase, as shown in **Table 2**.

Figure 5 illustrates the relationships between gene expression profiles, the structure of regulatory motifs and the function of CRP-regulated genes. It is worth noting that, for the presented biological functions, most of the functional groups investigated in this study surrounded and pointed towards the TCA cycle. For example, the major biological function of the SG2 and MG6 groups is controlling the carbohydrate transporter related genes. The transport of carbon sources, especially glucose, is an important field of research. When *E. coli* consumes glucose, the accumulated intracellular cAMP activates CRP, thus enhancing the expression of glucose transport and TCA cycle-related genes. Such an outcome would help *E. coli* to increase glucose absorption from the environment. This combination of different CRP-FFLs allows *E. coli* to expend more glucose to promote the production efficiency of ATP in the TCA cycle. Our analysis raises applications of FFLs associated with CRP-FFLs worthy of further study.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

Application of FFLs

The accumulation of our knowledge on each FFL contributes to an understanding of their sophisticated designs. Combining previous research on the kinetic feature of distinct FFLs [12] with our findings on the dose-response curve reveals that each FFL presents a specific control element which could potentially be applied in artificial circuit designs. Unlike electronic circuits, the topology and interactions between two genes can be easily changed by introducing a repressor or cooperator, or by shutting down the expression of the regulator. Thus, the expression patterns of cellular genes indicate that cells are used to cope with current growth conditions. A more detailed investigation on how different FFLs work together will enable us to simplify the design of artificial biological circuits as well as to define how an organism makes a specific decision in response to intracellular and environmental signals. Such studies will advance a new field that uses network motifs to understand the complexity of biological functions, such as memory, asymmetric cell deviation, and complex diseases.

Materials and Methods

The data analysis flow for the CRP-FFLs is shown in **Figure 1**. It comprised of two steps: (i) construction of CRP-FFLs from public resources and gene expressions in cAMP dose-responses and (ii) hierarchical clustering analysis of gene expressions to reveal hidden patterns in both CRP-sFFLs and CRP-mFFLs.

Construction of CRP-FFLs

To construct CRP-FFLs, all transcription factors and target genes regulated by CRP were collected from two public resources, namely EcoCyc and RegulonDB. Experimental results support the existence of regulatory interactions through both strong evidence (e.g., binding of purified proteins, DNaseI footprinting, and site mutation) and weak evidence (binding of cellular extracts and transcriptional fusions). Assuming that these CRP-regulated transcription factors regulate the target genes, some transcription factors and target genes that couldn't form CRP-FFLs were removed. In defining the eight types of CRP-FFLs, the dual or unknown regulatory effect of the TF on the regulated gene were not considered. Finally, CRP-FFLs using cAMP dose-response microarray were identified and discussed.

Identification of CRP-regulated genes using cAMP dose-response microarray

All strains used in this study were *E. coli* K12 BW25113 derivatives generated from the Keio collection system, which was kindly provided by the National Institute of Genetics of Japan [28]. For cAMP dose-response experiments, cells were grown in a minimum medium supplied with 0.4% glucose until the log phase, and then spun down and suspended in a fresh minimum medium with 0.4% glycerol and different concentrations of cAMP (0 mM, 0.01 mM, 0.03 mM, 0.1 mM, 0.3 mM, 1 mM, 3 mM and 10 mM) [19]. Cells were incubated with the cAMP medium for 15 minutes and then harvested for RNA isolation. To examine expression profiles of target genes involved in CRP-FFLs, the transcriptome expression of cells incubated with different concentrations cAMP using the GeneChip® *E. coli* Genome 2.0 Array were compared, enabling examination of gene expression profiles of *E. coli* under various conditions to better understand the biological pathways involved. Differences in microarray expression revealed a large number of genes under the control of cAMP during the exponential phase.

Analysis of microarray data

For each cAMP dosage, microarray experiments were performed in duplicate. Bioconductor software written in the R statistical programming language (http://www.bioconductor.org) was used to analyze the gene expression microarray data. Microarray CEL files were background corrected, normalized and expression value was calculated using the Robust Multi-chip Average (RMA) algorithm [29], resulting in log2 expression values. Each average expression value of treatments in the log 2 ratio was compared with the cells grown in the minimum medium without cAMP.

Hierarchical clustering and functional analysis

The hierarchical clustering analysis in the target gene expressions of CRP-FFLs from the cAMP dose-response microarray data was adopted. The insignificantly expressed genes with an absolute log 2-fold-change below 0.5 compared with no cAMP control were filtered out. Hierarchical clustering analysis based on the Pearson's correlation coefficient for similarity measurements and the average-linkage method was performed to group

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

the genes affected by the metabolic change. Finally, interesting genes were classified into five functional groups using the Term Enrichment tool [16, 30]. The maximum *p*-value 0.01 is considered to determine whether any GO terms annotate a specified list of genes at a frequency greater than that would be expected by chance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CPT and HDH conceived and supervised the project. CDY carried out all experimental concepts and wrote the manuscript. HYH and YHC participated in the design and computational analyses. SS edited and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was funded by the Ministry of Science Technology, Taiwan, under the contract number MOST 106-2627-M-009-002, MOST 106-2319-B-400-001, and MOST 106-2633-B-009-001. This work was particularly supported by "Aiming for the Top University Program" of the National Chiao Tung University and Ministry of Education, Taiwan. This work was particularly supported by Ministry of Health and Welfare, Taiwan, under the contract number MOHW106-TDU-B-212-144005. This paper (work) is particularly supported by the Ministry of Education through the SPROUT Project- Center For Intelligent Drug Systems and Smart Bio-devices (IDS2B) of National Chiao Tung University, Taiwan.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

References

1. Freyre-Gonzalez, J. A. a. T.-Q., L. G., *Analyzing Regulatory Networks in Bacteria*. 2010.
2. van Hijum, S. A.; Medema, M. H.; Kuipers, O. P., Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiology and molecular biology reviews : MMBR* **2009,** 73, (3), 481-509, Table of Contents.
3. Keseler, I. M.; Mackie, A.; Santos-Zavaleta, A.; Billington, R.; Bonavides-Martinez, C.; Caspi, R.; Fulcher, C.; Gama-Castro, S.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Muniz-Rascado, L.; Ong, Q.; Paley, S.; Peralta-Gil, M.; Subhraveti, P.; Velazquez-Ramirez, D. A.; Weaver, D.; Collado-Vides, J.; Paulsen, I.; Karp, P. D., The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res* **2017,** 45, (D1), D543-D550.
4. Gama-Castro, S.; Salgado, H.; Santos-Zavaleta, A.; Ledezma-Tejeida, D.; Muniz-Rascado, L.; Garcia-Sotelo, J. S.; Alquicira-Hernandez, K.; Martinez-Flores, I.; Pannier, L.; Castro-Mondragon, J. A.; Medina-Rivera, A.; Solano-Lira, H.; Bonavides-Martinez, C.; Perez-Rueda, E.; Alquicira-Hernandez, S.; Porron-Sotelo, L.; Lopez-Fuentes, A.; Hernandez-Koutoucheva, A.; Del Moral-Chavez, V.; Rinaldi, F.; Collado-Vides, J., RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res* **2016,** 44, (D1), D133-43.
5. Mendez-Cruz, C. F.; Gama-Castro, S.; Mejia-Almonte, C.; Castillo-Villalba, M. P.; Muniz-Rascado, L. J.; Collado-Vides, J., First steps in automatic summarization of transcription factor properties for RegulonDB: classification of sentences about structural domains and regulated processes. *Database (Oxford)* **2017,** 2017.
6. Madan Babu, M.; Teichmann, S. A., Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic acids research* **2003,** 31, (4), 1234-44.
7. Janga, S. C.; Salgado, H.; Martinez-Antonio, A., Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic acids research* **2009,** 37, (11), 3680-8.
8. Ma, H. W.; Kumar, B.; Ditges, U.; Gunzer, F.; Buer, J.; Zeng, A. P., An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. *Nucleic acids research* **2004,** 32, (22), 6643-9.
9. Beisel, C. L.; Storz, G., The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in Escherichia coli. *Molecular cell* **2011,** 41, (3), 286-97.
10. Mangan, S.; Alon, U., Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** 100, (21), 11980-5.
11. Mangan, S.; Zaslaver, A.; Alon, U., The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of molecular biology* **2003,** 334, (2), 197-204.
12. Alon, U., Network motifs: theory and experimental approaches. *Nature reviews. Genetics* **2007,** 8, (6), 450-61.
13. Kashtan, N.; Itzkovitz, S.; Milo, R.; Alon, U., Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **2004,** 20, (11), 1746-58.
14. Urrios, A.; Gonzalez-Flo, E.; Canadell, D.; de Nadal, E.; Macia, J.; Posas, F., Plug-and-Play Multicellular Circuits with Time-Dependent Dynamic Responses. *ACS synthetic biology* **2018,** 7, (4), 1095-1104.

15. Nampoothiri, S. S.; Fayaz, S. M.; Rajanikant, G. K., A Novel Five-Node Feed-Forward Loop Unravels miRNA-Gene-TF Regulatory Relationships in Ischemic Stroke. *Mol Neurobiol* **2018**.
16. Antoneli, F.; Golubitsky, M.; Stewart, I., Homeostasis in a feed forward loop gene regulatory motif. *J Theor Biol* **2018,** 445, 103-109.
17. Shimada, T.; Fujita, N.; Yamamoto, K.; Ishihama, A., Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PloS one* **2011,** 6, (6), e20081.
18. Zheng, D.; Constantinidou, C.; Hobman, J. L.; Minchin, S. D., Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic acids research* **2004,** 32, (19), 5874-93.
19. Hantke, K.; Winkler, K.; Schultz, J. E., Escherichia coli exports cyclic AMP via TolC. *J Bacteriol* **2011,** 193, (5), 1086-9.
20. Hecker, M.; Lambeck, S.; Toepfer, S.; van Someren, E.; Guthke, R., Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems* **2009,** 96, (1), 86-103.
21. Buetti-Dinh, A.; Ungricht, R.; Kelemen, J. Z.; Shetty, C.; Ratna, P.; Becskei, A., Control and signal processing by transcriptional interference. *Mol Syst Biol* **2009,** 5, 300.
22. Chen, Y. P.; Lin, H. H.; Yang, C. D.; Huang, S. H.; Tseng, C. P., Regulatory role of cAMP receptor protein over Escherichia coli fumarase genes. *J Microbiol* **2012,** 50, (3), 426-33.
23. Valgepea, K.; Adamberg, K.; Nahku, R.; Lahtvee, P. J.; Arike, L.; Vilu, R., Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst Biol* **2010,** 4, 166.
24. Nam, T. W.; Park, Y. H.; Jeong, H. J.; Ryu, S.; Seok, Y. J., Glucose repression of the Escherichia coli sdhCDAB operon, revisited: regulation by the CRP*cAMP complex. *Nucleic Acids Res* **2005,** 33, (21), 6712-22.
25. Gorke, B.; Stulke, J., Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **2008,** 6, (8), 613-24.
26. Stulke, J.; Hillen, W., Carbon catabolite repression in bacteria. *Curr Opin Microbiol* **1999,** 2, (2), 195-201.
27. Zhang, Z.; Gosset, G.; Barabote, R.; Gonzalez, C. S.; Cuevas, W. A.; Saier, M. H., Jr., Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in Escherichia coli. *J Bacteriol* **2005,** 187, (3), 980-90.
28. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2006,** 2, 2006 0008.
29. Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U.; Speed, T. P., Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **2003,** 4, (2), 249-64.
30. Boyle, E. I.; Weng, S.; Gollub, J.; Jin, H.; Botstein, D.; Cherry, J. M.; Sherlock, G., GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **2004,** 20, (18), 3710-5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

Figures

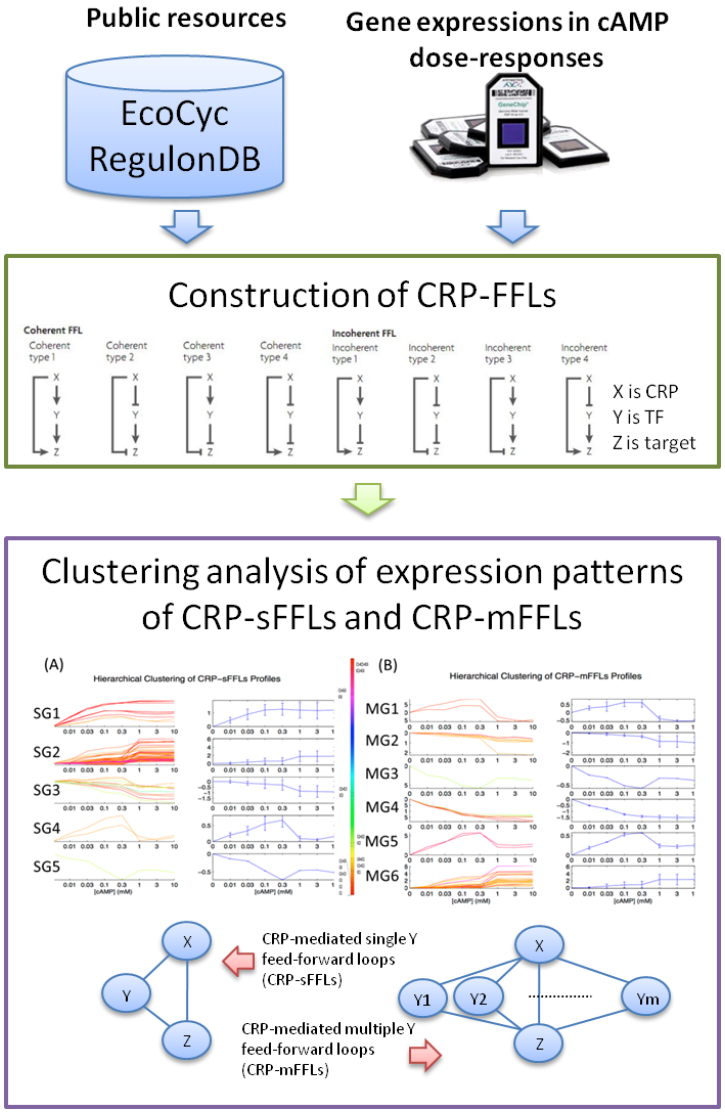

Figure 1. The data analysis flow of CRP-FFLs. Data analysis proceeds in two steps: (i) utilization of public resources and gene expressions in cAMP dose-responses for construction of CRP-FFLs [adapted from reference (11)] and (ii) hierarchical clustering analysis of gene expressions to reveal hidden patterns in both CRP-sFFLs and CRP-mFFLs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

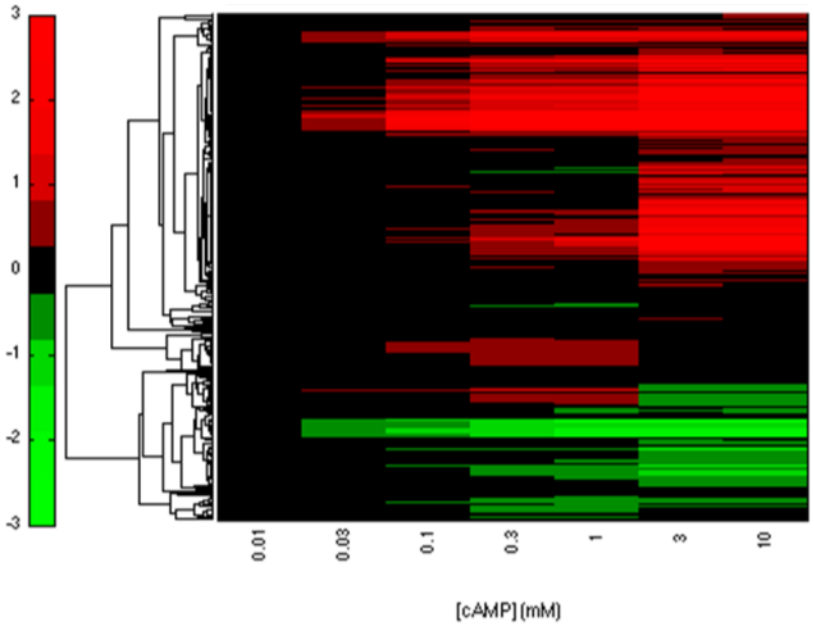

Figure 2. Heatmap for CRP-regulated genes. The clustering microarray data of the reported CRP-regulated genes is summarized in the Heatmap. The expression profile of genes was used to construct CRP-FFLs for further analysis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

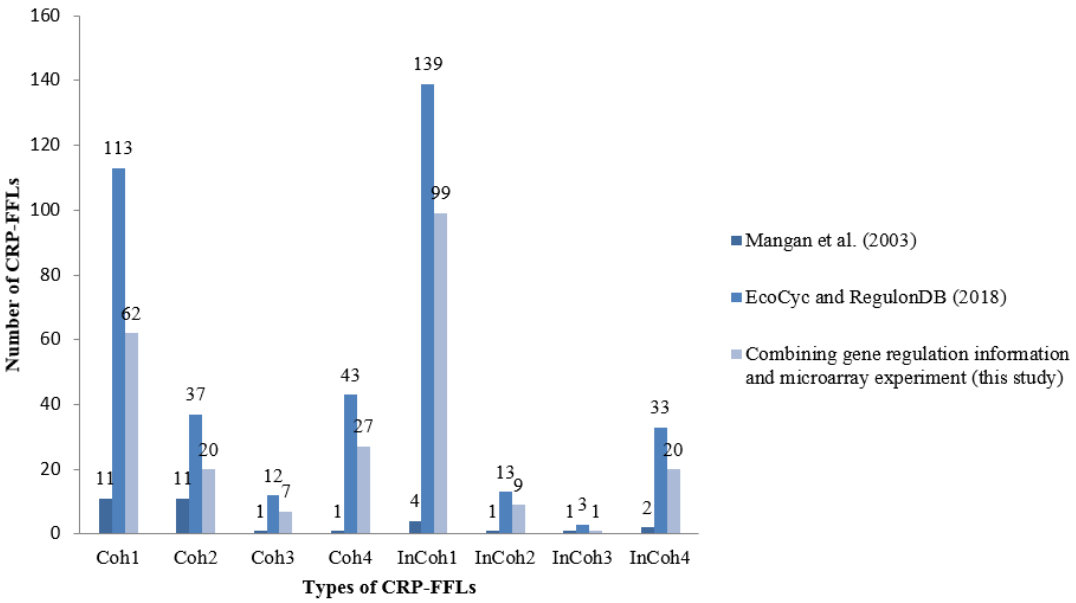

Figure 3. Distribution of CRP-FFLs in *E. coli*. Regulation information was based on statistics sourced from Mangan *et al.*, EcoCyc and RegulonDB, and CRP-FFLs generated in this study. Some CRP-regulated genes reported in the literature and public databases did not show significant differential expression in our experimental conditions, probably because these genes require other transcription factors that are only expressed in specific conditions for a cooperative response to cAMP.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

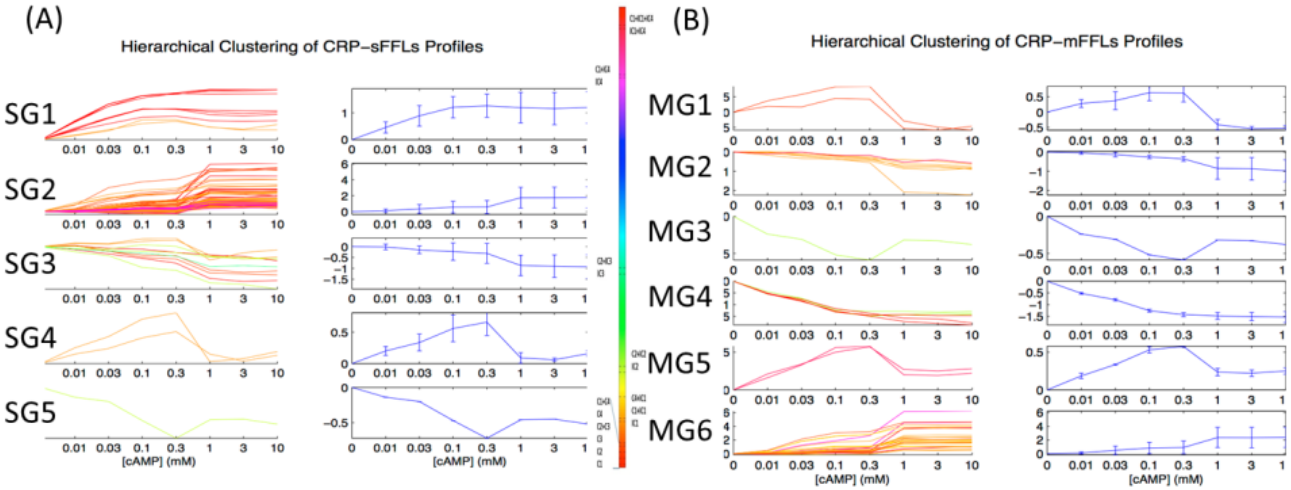

Figure 4. Expression profiles of CRP-sFFLs and CRP-mFFLs. Colored CRP-FFLs are used to distinguish between the various CRP-FFL types. The CRP-FFLs were divided into five groups (SG1 to SG5) in CRP-sFFLs and six groups (MG1 to MG6) in CRP-mFFLs according to distinct cAMP dose-response profiles. Hierarchical clustering results for (A) CRP-sFFLs and (B) CRP-mFFLs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

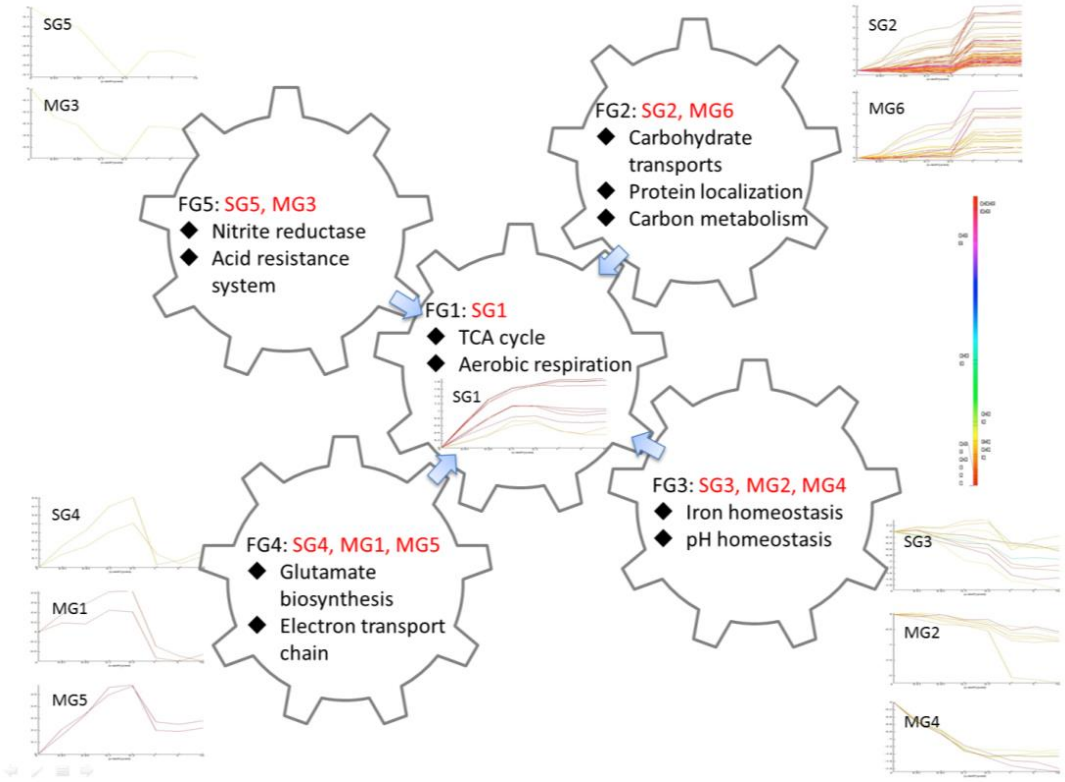

Figure 5. Biological interactions of CRP-FFLs. Expression patterns of target genes can be divided into five functional groups (FG1 to FG5). Each group combines single- or multi-TFs CRP-FFLs involved in different biological functions such as carbon metabolism, transport, and iron and pH homeostasis. Most functional groups in this study surround and point toward the TCA cycle.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

Tables

Table 1. Data statistics for each type of CRP-FFLs. The detailed numbers of CRP-FFLs, TFs, and genes are listed for each type of CRP-FFL. A number of target genes regulated by multiple TFs and TF-regulated multiple genes are also listed. Target genes for each type of CRP-FFL are listed for further analysis.

Type	Number of CRP-FFLs	Number of TF	Number of Gene	Genes
Coh1	49	13	49	*araF, araG, dctA, fecA, fecB, fecC, fecD, fecE, fucA, fucI, fucK, fucO, fucP, fucR, fumC, hyfH, idnD, idnK, idnO, lamB, malE, malF, malG, malK, malM, malS, marB, ompF, prpB, prpC, prpD, prpE, rhaR, sdhA, sdhB, sdhC, sdhD, srlA, srlB, srlD, srlE, sucA, sucB, sucC, sucD, xylA, xylF, xylG, xylH*
Coh2	15	3	10	*gadA, gadB, gadB, gadC, gadC, gadE, gadE, gadX, gltD, gltF, mdtE, mdtE, mdtF, mdtF, proP*
Coh3	2	1	2	*gltD, gltF*
Coh4	26	3	26	*acs, actP, aldB, fadL, flhC, flhD, glcC, guaA, guaB, hlyE, hupB, mglA, mglC, mtlA, mtlD, mtlR, nanA, nanE, nanK, nanT, nmpC, xylF, xylG, xylH, yiaK, yjcH*
InCoh1	87	22	73	*cdd, chbB, cirA, cyoD, cyoD, cyoE, cyoE, cytR, entD, entH, fecA, fecB, fecC, fecD, fecE, fepA, fiu, flhC, flhD, fumC, galK, galK, galP, galP, galS, glcC, glpD, glpF, glpK, glpT, gntK, gntP, gntP, gntU, grcA, grcA, lsrA, lsrB, lsrC, lsrD, lsrF, lsrG, lsrK, lsrR, malI, malX, manX, manY, manZ, marB, mglA, mglA, mglA, mglB, mglB, mglB, mglC, mglC, mglC, mtlA, mtlD, mtlR, nagB, nagE, nupC, nupG, prpR, rbsA, rbsB, rbsC, rbsD, rbsK, rbsR, srlA, srlB, srlD, srlE, tsx, udp, uidA, uidB, uidC, uxuA, uxuA, uxuB, uxuB, xylA*
InCoh2	8	3	8	*bhsA, gadA, gadB, gadC, gadX, nirB, osmY, yiaJ*
InCoh3	1	1	1	*araJ*
InCoh4	14	4	14	*csgD, csgE, csgF, csgG, cyoD, cyoE, exuT, glpB, grcA, malE, malF, malG, marB, nrdB*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

Table 2. Biological process ontology of CRP-FFLs. Associated Gene Ontology terms and biological functions for each functional group are listed. Most functional groups with a small P-value are statistically significant.

Functional Group ID	Group ID	GO Term	P-value	Number of Gene	Summary of function
FG1	SG1	GO:0006099 tricarboxylic acid cycle	5.02E-12	8	● TCA cycle ● Aerobic respiration
		GO:0009060 aerobic respiration	8.00E-11	8	
		GO:0045333 cellular respiration	3.13E-09	8	
		GO:0015980 energy derivation by oxidation of organic compounds	2.18E-08	8	
		GO:0006091 generation of precursor metabolites and energy	2.25E-07	8	
		GO:0055114 oxidation-reduction process	7.12E-03	8	
FG2	SG2 MG6	GO:0008643 carbohydrate transport	3.84E-26	37	● Carbohydrate transports ● Protein localization ● Carbon metabolism
		GO:0034219 carbohydrate transmembrane transport	6.99E-18	27	
		GO:0071702 organic substance transport	4.69E-12	44	
		GO:0006810 transport	2.93E-08	49	
		GO:0051234 establishment of localization	3.75E-08	49	
		GO:0015749 monosaccharide transport	5.37E-08	14	
		GO:0044765 single-organism transport	2.23E-07	46	
		GO:0051179 localization	4.36E-07	49	
		GO:0015768 maltose transport	4.45E-05	6	
		GO:0009401phosphoenolpyruvate-dependent sugar phosphotransferase system	1.03E-04	10	
		GO:0055085 transmembrane transport	1.20E-04	36	
		GO:0015750 pentose transport	2.71E-04	7	
		GO:0044724 single-organism carbohydrate catabolic process	6.79E-04	21	
		GO:0015766 disaccharide transport	7.07E-04	7	
		GO:0042956 maltodextrin transport	7.40E-04	5	
		GO:0015772 oligosaccharide transport	1.61E-03	7	
		GO:0016052 carbohydrate catabolic process	2.08E-03	21	
		GO:0006004 fucose metabolic process	3.19E-03	6	
		GO:0019521 D-gluconate metabolic process	7.58E-03	6	
		GO:0044275 cellular carbohydrate catabolic process	8.56E-03	11	
FG3	SG3 MG2 MG4	GO:0055080 cation homeostasis	7.34E-06	8	● Iron homeostasis ● pH homeostasis
		GO:0050801 ion homeostasis	9.10E-06	8	
		GO:0048878 chemical homeostasis	1.37E-05	8	
		GO:0042592 homeostatic process	3.33E-04	8	
		GO:0055072 iron ion homeostasis	7.44E-03	5	
		GO:0045852 pH elevation	9.50E-03	3	
		GO:0051454 intracellular pH elevation	9.50E-03	3	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 June 2018 doi:10.20944/preprints201806.0100.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2335; doi:10.3390/ijms19082335

FG4	SG4 MG1 MG5	No significant GO term	● Glutamate biosynthesis (gltDF) ● Electron transport chain (cyoDE) ● Ferric enterobactin transport (fepA) ● Iron transport (fiu)
FG5	SG5 MG3	No significant GO term	● Nitrite reductase (nirB) ● Acid resistance system (gadX)

Table 3. Detailed information on CRP-sFFLs and CRP-mFFLs. CRP-sFFLs and CRP-mFFLs are respectively composed of five and six groups involved with various types of FFLs (checked types). Five functional groups are listed using functional group identifier. Each group is listed according to the expression pattern.

Type	Group ID	Number of CRP-FFLs (Gene)	Coh1	Coh2	Coh3	Coh4	InCoh1	InCoh2	InCoh3	InCoh4	Functional Group ID
CRP-sFFLs	SG1	10 (10)	V				V				FG1
	SG2	78 (78)	V			V	V			V	FG2
	SG3	10 (10)	V	V		V	V	V	V		FG3
	SG4	2 (2)					V				FG4
	SG5	1 (1)						V			FG5
CRP-mFFLs	MG1	4 (2)		V	V						FG4
	MG2	15 (7)	V				V			V	FG3
	MG3	2 (1)		V				V			FG5
	MG4	14 (6)		V				V			FG3
	MG5	6 (2)					V			V	FG4
	MG6	60 (27)	V			V	V			V	FG2