Published in final edited form as:

Nat Chem Biol. ; 8(1): 65–71. doi:10.1038/nchembio.710.

Deciphering the regulatory logic of a stimulon

Byung-Kwan Cho¹, Stephen Federowicz, Young-Seoub Park², Donghyuk Kim, Karsten Zengler, and Bernhard Ø. Palsson*

Department of Bioengineering, University of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0412

Abstract

Although metabolic networks have been reconstructed on a genome-scale, the corresponding reconstruction and integration of governing transcriptional regulatory networks has not been fully achieved. Here we reconstruct such an integrated network for amino acid metabolism in *Escherichia coli*. Analysis of ChIP-chip and gene expression data for the transcription factors ArgR, Lrp, and TrpR showed that ~82% of the genes they regulate are directly involved in amino acid metabolism. Further analysis shows that 19/20 amino acid biosynthetic pathways are either directly or indirectly controlled by these three transcription factors. Classifying the regulated genes into three functional categories of transport, biosynthesis, and metabolism leads to elucidation of regulatory motifs constituting the integrated network's basic building blocks. The regulatory logic of these motifs was determined based on the relationships between transcription factor binding and changes in transcript levels in response to exogenous amino acids. Remarkably, the resulting logic shows how amino acids are differentiated as signaling and nutrient molecules, and thus revealing the overarching regulatory principles of this stimulon.

Transcriptional regulatory networks (TRN) in bacteria govern metabolic flexibility and robustness in response to environmental signals¹. Thus, causal relationships between transcript levels for metabolic genes and the direct association of transcription factors (TFs) at the genome-scale is fundamental to fully understand bacterial responses to their environment^{2,3}. In particular, the molecular interaction between small molecules ranging from nutrients to trace elements and TFs governs the TRN and ultimately regulates the related metabolic pathways. From the causal relationships, a small set of recurring regulation patterns, or network motifs^{3,4} were identified and reconstructed to describe the design principles of complex biological systems. One primary discovery from this effort was the connected feedback circuit which coordinates influx (biosynthesis and transport) and efflux (metabolism) pathways that are jointly regulated by a TF sensing the relevant small molecule³. For example, a part of the global TRN is comprised of certain TFs (ArgR, Lrp, and TrpR) that sense the presence of exogenous amino acids (arginine, leucine, and tryptophan, respectively) and, in response, regulate the expression of a number of target genes⁵. Upon addition of these amino acids to the environment, the TFs exhibit enhanced, reversed, or unaffected regulatory modes^{3,6-8}. These TF responses make these amino acids not just nutrients but also signaling molecules⁹.

*Corresponding author: Bernhard Ø. Palsson, Department of Bioengineering, University of California, San Diego, CA 92093.

¹Current addresses: Byung-Kwan Cho, Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea;

²Current addresses: Young-Seoub Park, Mogam Biotechnology Research Institute, Yongin 449-799, Korea.

Author contributions: B.-K.C. and B.O.P. conceived the study. B.-K.C. and Y.-S.P. designed and performed experiments. B.-K.C. and S.F. analyzed data with contributions from K.Z. The manuscript was written by B.-K.C., S.F., K.Z., and B.O.P.

Previously discovered network motifs^{3,4} represent a significant step forward in our understanding of complex biological behavior. However, they fail to appropriately elucidate the system wide response since they were either based upon incomplete information⁴, or were only specific to a single transcription factor and regulon³. This has resulted in an inability to appropriately understand complex regulatory phenomena existing across multiple transcription factors and regulatory signals. Hence, it is necessary to achieve a full elucidation of these interactions with systematic and integrated experimental analysis. Comprehensive elucidation of the causal relationships is achievable by integrated analysis of expression data obtained from microarray or sequencing (e.g., RNA-seq)¹⁰ with direct TF-binding information from chromatin immunoprecipitation coupled with microarrays or sequencing (ChIP-chip or ChIP-seq)^{3,11} under appropriate environmental conditions. Thus, we obtain and integrate genome-scale data from ChIP-chip for each TF and gene expression profiling to reconstruct regulons involved in amino acid metabolism at the genome-scale.

The elucidated regulatory logic falls into two categories that differentiate the role of amino acids as signaling and as nutrient molecules. Therefore, the reconstruction of the regulatory logic of the network motif allows us to establish the physiological role of each TF regulon and to determine how they govern the amino acid regulation in *E. coli*. Then, the integration of these multiple regulons into a unified network led to the first full bottom-up genome-scale reconstruction of a stimulon.

Results

Genome-wide identification of TF-binding regions: Regulatory code analysis

ArgR, Lrp, and TrpR are TFs involved in amino acid metabolism in *E. coli*^{6,7,12}, responding to arginine, leucine, and tryptophan, respectively. The binding of the small effector molecule (here being the amino acids) to these TFs carries out the genome's regulatory code by enhancing or decreasing the TFs affinity for a specific genomic region and concurrently modulating the transcription of downstream genes. In the case of Lrp, the direct analysis of *in vivo* binding was fully described³ using chromatin immunoprecipitation coupled with microarrays (ChIP-chip) experiments. A total of 141 binding regions were analyzed, representing coverage of 74% of the previously identified regions³. However, similar genome-scale data for the other two major TFs in amino acid metabolism, ArgR and TrpR were unavailable. To determine their binding regions on a genome-wide level in an unbiased manner, we employed the ChIP-chip approach to *E. coli* cells harboring 8×myc-tagged ArgR or TrpR protein¹³. The resulting log₂ ratios obtained from the ChIP-chip experiments identify the genomic regions enriched in the IP-DNA sample compared with the mock IP-DNA sample and thereby represent a genome-wide map of *in vivo* ArgR- and TrpR-binding regions (Fig. 1a).

Using a previously described binding region detection algorithm¹⁴, 61 and 8 unique and reproducible ArgR- and TrpR-binding regions were identified, respectively (Supplementary Table 1 and Supplementary Table 2). The 61 ArgR-binding sites detected included 13 sites previously characterized by DNA-binding experiments *in vitro* and mutational analyses *in vivo*^{15,16}. For example, the ArgR-arginine complex transcriptionally represses *gltBD*, *artPIQM* operon, and *artJ* gene encoding arginine transport systems^{17,18}. Our results confirmed that the ArgR-arginine complex binds to each of these promoter regions (Fig. 1b). In addition, the ArgR occupancy level at the promoter of the *artJ* gene is greater than that of *artPIQM* operon in the presence and absence of exogenous arginine (Supplementary Table 1). This result is in good agreement with the de-repression/repression ratio of 28 for P_{artJ} and 3.2 for P_{artP} previously reported for repressibility of the *artJ* and *artP* promoters¹⁸. Also, this result is consistent with recent microarray and qPCR experiments showing a significant arginine and ArgR-dependent down-regulation of both the *artJ* (about 50-fold) and *artPIQM*

mRNA levels (about three to six-fold)¹⁷. In the case of TrpR, a total of five associations have been determined by DNA-binding experiments *in vitro* and mutational analyses *in vivo*^{7,19}, all of which were also identified in our study (Fig. 1a and Supplementary Table 2). For instance, TrpR directly binds to the promoter regions of *aroH* and *mtr* involved in biosynthesis and transport of aromatic amino acids (Fig. 1b). Against the current genome annotation¹⁴, all of the ArgR- and TrpR-binding regions were observed within intergenic regions, i.e., promoter and promoter-like regions. The same preference was observed for Lrp-binding sites (Supplementary Table 1 and 2)³. DNA sequence motifs for each of the transcription factors were also re-derived based solely upon the ChIP binding regions and were in full agreement with previously described motifs (Supplementary Fig 2). Based on the fact that the increase in the intracellular arginine and tryptophan levels enhances ArgR and TrpR binding to its DNA targets^{20,21}, the confirmation of previously discovered sequence motifs, and the full coverage of the known binding regions in our data we concluded that ArgR- and TrpR-binding regions identified here are *bona fide* binding sites.

Interestingly, as with *gltBD*, *artPIQM*, *potFGHI*, and *mtr* (Fig. 1b), we observed that Lrp directly binds to nine ArgR- and one TrpR-binding regions (Fig. 1c and Supplementary Fig. 1). For example, the direct binding of Lrp to the promoter region of the *gltBD* operon encoding glutamate synthase resulted in the activation of its transcription. In contrast, the role of ArgR-binding represents the negative regulation of the operon. Integrating binding regions and changes in transcript levels, the reciprocal mode³ in the transcriptional regulation of ArgR and Lrp was observed for cellular functions including putrescine transport (*potFGHI*), arginine transport (*artPIQM*), leucine response protein (*lrp*), arginine biosynthesis and utilization (*argA* and *astCADBE*), the formation of nucleoid (*stpA*), as well as glutamate biosynthesis and transport (*gltBD* and *gltP*). While Lrp activates the tryptophan transport (*mtr*), TrpR represses its transcription. In addition to confirming previously identified ArgR- and TrpR-binding regions, we found 47 and 3 novel ArgR- and TrpR-binding regions, which include the promoter region of *potFGHI*, encoding putrescine ABC transporter (Fig. 1b).

Identification of regulons: topological analysis

A regulon is defined as a group of genes whose transcription is controlled by a transcriptional regulator. The arginine regulon describing the genetic and regulatory organization of the genes involved in arginine biosynthesis in *E. coli* was used as an example in proposing the definition of the regulon in 1964^{17,22}. However, it has not been included in the definition of regulon whether each regulation is direct or indirect. So far, a total of 37, 56, and 10 genes have been characterized as members of regulons directly regulated by ArgR, Lrp, and TrpR, respectively^{15,16}. Based upon regulatory codes described above, we significantly extended the size of these regulons and obtained 140, 283, and 15 target genes for each regulon. Since ArgR directly controls the transcription of *lrp*, the regulon size of each transcription factor can be described as ArgR (423) > Lrp (283) > TrpR (15). These regulons represent a hierarchical structure that can be used to identify the indirect effect of the TFs. For example, *thrLABC* operon involved in the threonine biosynthesis is directly activated by Lrp, either in the absence or presence of exogenous leucine. We observed that ArgR indirectly represses this operon in response to exogenous arginine; i.e., transcriptional repression without the direct binding of ArgR. It is therefore possible to partially elucidate the indirect regulation by ArgR based on the hierarchical regulatory network. ArgR represses Lrp leading to the indirect repression of the *thrLABC* operon. As shown in this example, integrated analysis of ChIP-chip and expression profiles allows us to fully understand the hierarchical TRN including the indirect regulatory effects.

Next, we classified the 438 target genes based on their functional annotation and found that most of these functions (~82%) were assigned to amino acid metabolism and transport, as

well as carbohydrate, nucleotide, and energy metabolism (Fig. 2). We are then able to show (Fig. 3) that 19/20 amino acid biosynthetic pathways are directly or indirectly controlled by these three TF's. To do this we first mapped the directly regulated genes to known amino acid biosynthetic pathways and transport systems to determine their direct metabolic roles (Fig. 3a, b). ArgR directly regulates the transcription of all genes involved in the biosynthesis of arginine and histidine. It also regulates *gltBD*, *aroB*, *aroK*, and *dapE* involved in glutamate, aromatic amino acids, and lysine biosynthesis, respectively. The genes encoding the enzymes for the biosynthesis of branched chain amino acids are comprehensively regulated by Lrp, which also controls the transcription of *gltBD* and *gdhA* encoding glutamate synthase and glutamate dehydrogenase (glutamate biosynthesis), *serC* and *serB* encoding phosphoserine transaminase and phosphatase (serine biosynthesis), *thrABC* operon for aspartate kinase, homoserine kinase, and threonine synthase (threonine biosynthesis), *argA* for N-acetylglutamate synthase (arginine biosynthesis), and *aroA* for 3-phosphoshikimate-1-carboxyvinyltransferase (the chorismate formation for aromatic amino acid biosynthesis). TrpR regulates the transcription of genes involved in tryptophan biosynthetic pathway (*trpLEDCBA* operon), as well as *aroH* and *aroL*. In addition, it has been determined that TyrR directly regulates several genes in the aromatic amino acid biosynthesis (*aroF*, *aroG*, *aroK*, *aroA*, *tyrA*, and *tyrB*) in response to exogenous tyrosine^{15,16}. Taken together, these four TFs control the biosynthesis of 12 amino acids. Furthermore, the biosynthesis of proline, glutamine, glycine, cysteine, and methionine is through branched biosynthetic pathways of glutamate, serine and aspartate (Fig. 3a). The remaining three amino acids (i.e., alanine, aspartate, and asparagine) are synthesized from glutamate as an amino donor (green dots in Fig. 3a). Therefore, biosynthetic pathways for all amino acids are directly or indirectly controlled by these four TFs.

Next, we classified the amino acids into ten groups based on the substrate specificity of each transport system, which are A (tyrosine, phenylalanine, tryptophan), B (arginine, histidine, lysine), C (glutamate, aspartate), D (leucine, isoleucine, valine), E (alanine, serine, glycine, threonine), F (proline), G (methionine), H (cysteine), I (asparagine), and J (glutamine) (Fig. 3b). As expected, the amino acids in the same group have a similar chemical structure, e.g. aromatic amino acids and branched chain amino acids in group A and group D, respectively. Transport systems for groups G-J are highly specific and were therefore classified into individual groups.

Determination of causal relationships: functional analysis

In general, genes for amino acid biosynthesis are repressed by each corresponding TF, whereas catabolic operons such as *astCADBE*, *tdh-kbl*, and *gcvTHP* are induced in response to the exogenous amino acids^{12,23}. To determine the causal relationships between binding of a TF and the changes in RNA transcript levels of genes in the regulons, we integrated the binding regions of ArgR, TrpR, Lrp, and TyrR with the publicly available transcriptomic data (Fig. 4)^{3,17}. We then determined activation or repression based upon the regulatory modes described previously³. Among genes in the ArgR regulon, about 18% genes were directly activated in response to the exogenous arginine, which include *aroP* and *gltP* genes encoding aromatic amino acids and glutamate/aspartate transporters. On the other hand, ArgR represses about 70% of its regulon members, including *potFGHI*, *artJ*, *artPIQM*, and *hisJQMP* encoding putrescine, arginine, lysine, ornithine, and histidine ABC transporters (Fig. 4). ArgR represses genes involved in the arginine and glutamate biosynthesis pathways, and unexpectedly, it directly down-regulates genes involved in histidine, aromatic amino acids, and lysine biosynthesis pathways. In case of amino acid utilization, ArgR induces *astCADBE* and *puuEB* operons encoding the metabolic pathways for arginine and putrescine, respectively. The remaining 12% of its regulon members had a direct association

with ArgR without differential gene expression. Most of the remaining genes are currently annotated as genes of unknown function (Supplementary Table 1).

Gene expression profiles validated that Lrp directly regulates 283 genes. 45% and 55% of the Lrp-regulated genes were repressed and activated in response to the addition of the exogenous leucine³. As expected, Lrp controls the transport, biosynthetic and utilization pathways more globally than other transcription factors do. Lrp represses the transport systems for branched chain amino acids (*brnQ*, *livKHMGE*, and *livJ*), dipeptides (*dppABCD*), and lipoproteins (*lolCDE*) but it activates a whole set of other transporters. Transporters that are activated by Lrp are aromatic amino acids (*tyrP* and *mtr*), arginine (*artMQIP*), glutamate (*gltP*), alanine, serine, glycine and threonine (*cycA*, *tdcC*, *sdaC*, and *sstT*), proline (*proY*), putrescine (*potFGHI*), dipeptide (*dtpB*), and oligopeptides (*oppABCD*) (Fig. 4). In terms of amino acid biosynthetic pathways, Lrp represses all genes but the *thrLABC* operon for threonine biosynthesis. For amino acid utilization, Lrp activates all pathways for aromatic amino acids, arginine, aspartate, branched chain aromatic amino acids, alanine, glycine, serine, threonine, methionine, and putrescine. In case of the TrpR regulon, a total of 15 genes are directly regulated, of which 13 genes are repressed (Supplementary Table 2)^{16,24}. TrpR also represses *mtr* encoding the tryptophan transporter as well as *aroH*, *aroL*, and *trpABCDE* involved in the tryptophan biosynthesis pathway. While TyrR activates the transport systems for aromatic amino acids (*aroP*, *tyrP*, and *mtr*), it represses tyrosine biosynthetic pathway comprising of *aroG*, *aroL*, *aroF*, *tyrA*, and *tyrB* (Fig. 4).

The function of a stimulon: elucidation of overarching regulatory logic

Based on the integrated analysis of TF-binding locations and gene expression profiles, we were able to connect transport, biosynthesis, and utilization of amino acids, and generate the connected bidirectional circuits (Fig. 5a). In the left feed-back circuit, TF-amino acid (TF-AA) complexes regulate the transcription of the transporters (**T**) and biosynthesis pathways (**B**), facilitating the influx of the amino acid molecules (**AA_{in}**) from amino acids in the media (**AA_{out}**) and precursors (**AA_{pre}**). In the right feed-forward circuit, TF-AA complexes control transcription of utilization genes (**U**) responsible for converting **AA_{in}** into metabolites (**M**). Thus, the logical structures of the connected bidirectional circuit motifs can be described by a notation that uses three signs indicating repression (**R**) or activation (**A**) for each of **T**, **B**, and **U** (Fig. 5b). For example, the **A-R-A** circuit motif indicates that the transcription of transport, biosynthesis, and metabolic genes are activated, repressed, and activated, respectively, whereas the **R-R-A** circuit motif demonstrates that the transcription of both transport and biosynthesis are repressed and the metabolic genes are activated. The possible logical structures of the connected circuit motifs can be characterized depending on how the TF-AA complex activates or represses both influx (**T** and **B**) and efflux (**U**) in response to the exogenous amino acids. Based on the connected circuit motifs, we analyzed the behavior of logical structures of the transcription of transport, biosynthesis, and metabolic genes in responses to the exogenous arginine and leucine (Fig. 5b).

Surprisingly, there are only three influx-efflux combinations found between amino acid groups and TFs (Fig. 5c). For example, the connected circuit motif controlled by ArgR-arginine complex shows the **R-R-A** logical structure for group B amino acids (lysine, histidine, and arginine), whereas the logical structure of the motif is switched to **A-R-R** for glutamate and aspartate and **A-R-A** for other amino acids. On the other hand, the connected motif controlled by Lrp-leucine complex indicates the **R-R-A** logical structure for group D (valine, leucine, and isoleucine) and is again switched to **A-R-R** for glutamate and aspartate and **A-R-A** for other amino acids. For glutamate our primary observation was that the utilization was repressed given its role as a substrate for nine biosynthetic pathways (Fig.

3,4). However we acknowledge that the regulation is highly complex and not universally repressed. This logically follows from the critical and centralized role it plays throughout the metabolome²⁵. Overall, we conclude that for two global transcription factors (ArgR and Lrp) in amino acid regulation, the connected circuit motif has an **R-R-A** logical structure for signaling molecules (i.e., arginine for ArgR and leucine for Lrp) and the **A-R-A** and **A-R-R** logical structures for other amino acids (Fig. 5c).

Discussion

We reconstructed the regulons of ArgR, Lrp, and TrpR in *E. coli* individually and then integrated them to form the first genome-scale reconstruction of a stimulon. First, we set out to comprehensively establish the TF-binding regions on the *E. coli* genome experimentally and furthermore to elucidate any DNA sequence motif(s) correlated with the TF regulatory action. Second, we significantly extended the size of each regulon and obtained 140, 283, and 15 target genes for each regulon. Third, using changes in transcript levels on a genome-scale, we identified the regulatory modes for individual gene governed by each TF in responses to exogenous arginine, leucine, and tryptophan. The integrated analyses indicate that the functional assignment of the regulated genes is strongly enriched in the amino acid metabolism-related functions. As suggested previously, many of these genes are likely to be involved in the “feast or famine” adaptation for survival in nutrient-rich or depleted environments^{3,9}. Fourth, we assigned the regulated target genes to three functional categories; transport, biosynthesis, and metabolism of amino acids. The classification allowed us to identify the connected circuit motif as a basic building block of the integrated network. Finally, we determined the regulatory logic of the connected circuit motif based on the causal relationships between the association of TFs and changes in transcript levels. These fall into two categories and thus allow for the differentiation between amino acids as signaling and nutrient molecules.

In general, transport systems along with biosynthetic and metabolic pathways convert external resources to basic building blocks to sustain life. The coordinated regulation of this primary process underlies expression of optimized metabolic states under different external conditions. Thus, we examined the logical structures of the metabolite-regulation connected circuit in response to the changes in the external amino acid availability in the reconstructed stimulon. We uncovered three unique logical structures that govern the amino acid biosynthesis and metabolism. The **R-R-A** logical structure was observed for signaling molecules whereas the **A-R-A** and **A-R-R** logical structures were determined for other amino acids severing as nutrient source (Fig. 5a, b). In principle, every metabolic pathway that includes transport, biosynthesis, and utilization functions could follow these logical structures. For example, the purine metabolism in *E. coli* contains a wide range of genes whose functions are transport (*yieG*), biosynthesis (*cvpA-purF-ubiX*, *purHD*, *purMN*, *purT*, *purL*, *purEK*, *purC*, *hflD-purB*, *purA*, and *guaAB*), utilization (*apt*), and a transcriptional regulator (*purR*). The metabolic functions of regulon members of PurR enriched into the purine metabolism and the connected circuit motif indicated the logical structures for signaling molecule in response to the exogenous purine²⁶. It can be therefore envisioned that other potential metabolic pathways follow similar logical structures as determined for the amino acid metabolism in bacteria.

Bacterial cells import essential nutrients and inorganic ions such as galactose and iron due to the absence of the biosynthesis pathway. It is therefore of interest that the simple feedback circuit (SFL) motif, a connected circuit motif of transporter and utilization pathway by TF, is often observed in the regulatory circuits for these molecules²⁷. If we assume the feedback circuit composed of influx and efflux combination, the logical structures of **R-R-A**, **A-R-A**, and **A-R-R** in the CFL motif can be reduced to **R-A**, **A-A**, and **A-R**, respectively. In *E. coli*,

the galactose metabolic pathway is controlled by the galactose repressor (GalR) and galactose isorepressor (GalS), whereas iron homeostasis is controlled by the ferric uptake regulator (Fur)^{28,29}. In the case of galactose metabolism, both GalR and GalS directly repress the transcription of *galP* encoding galactose permease. In a similar way, GalR partially represses the *mgIBAC* operon encoding high-affinity, ABC-type transport system. When galactose is available in the medium, the DNA-binding by both GalR and GalS is inhibited, followed by the activation of those genes along with the genes for galactose utilization²⁹. Therefore, the SFL motif exhibits the **A-A** logical structure, confirming the exogenous galactose as nutrient. In the iron homeostasis system in *E. coli*, intracellular iron binds to Fur, forming the active TF complex, which in turn activates the production of iron-using metabolic enzymes and also shuts down expression of iron transporters. Interestingly, the SFL motif for Fur regulon exhibits the **R-A** logical structure, similar to amino acids serving as signaling molecules described above. Therefore, we can conclude that iron acts as signaling molecule rather than nutrient.

In summary, we have described an integrative analysis of genome-scale data sets to comprehensively understand the basic principles governing a stimulon in the TRN of *E. coli*. The overarching regulatory principle elucidated enabled us to differentiate between metabolites as signaling and nutrient molecules. This important distinction between seemingly similar metabolites is non-intuitive and represents a triumph of genome-scale systems analysis. Similar analysis of other stimulons and large-scale regulatory networks may reveal that this regulatory principle is general. Thus, this approach to the analysis of regulation at the network level may reveal other fundamental non-obvious regulatory principles at work in genome-scale regulatory networks.

Methods

Bacterial strains and growth conditions

All strains used are *E. coli* K-12 MG1655 and its derivatives. The *E. coli* strains harboring ArgR-8myc, Lrp-8myc, and TrpR-8myc were generated as described previously¹³. Glycerol stock of ArgR-8myc strains were inoculated into W2 minimal medium containing 2 g/L glucose and 2g/L glutamine, and cultured overnight at 37 °C with constant agitation³⁰. The cultures were inoculated into 50 mL of the fresh W2 minimal media in either the presence or absence of 1 g/L arginine and continued to culture at 37 °C with constant agitation to an appropriate cell density. *E. coli* strains harboring Lrp-8myc and TrpR-8myc were grown in glucose (2 g/L) minimal M9 medium supplemented with or without 20 mg/L tryptophan or 10 mM leucine, respectively^{3,31}.

Chromatin immunoprecipitation and microarray analysis (ChIP-chip)

To identify ArgR-, Lrp-, and TrpR-binding regions *in vivo*, we isolated the DNA bound to ArgR protein from formaldehyde cross-linked *E. coli* cells harboring ArgR-8myc by chromatin immunoprecipitation with the specific antibodies that specifically recognizes myc tag (9E10, Santa Cruz Biotech)³². Cells were harvested from the exponential growth conditions in the presence or absence of exogenous arginine or tryptophan. The immunoprecipitated DNA (IP-DNA) and mock immunoprecipitated DNA (mock IP-DNA) were hybridized onto the high-resolution whole-genome tiling microarrays, which contained a total of 371,034 oligonucleotides with 50-bp tiles overlapping every 25-bp on both forward and reverse strands^{3,14}. A ChIP-chip protocol previously described was used^{32,33} and microarray hybridization, wash, and scan were performed in accordance with manufacturer's instruction (Roche NimbleGen).

qPCR

To monitor the enrichment of promoter regions, 1 μ L immunoprecipitated DNA was used to carry out gene-specific qPCR³. The quantitative real-time PCR of each sample was performed in triplicate using iCycler™ (Bio-Rad Laboratories) and SYBR green mix (Qiagen). The real-time qPCR conditions were as follows: 25 μ L SYBR mix (Qiagen), 1 μ L of each primer (10 pM), 1 μ L of immunoprecipitated or mock-immunoprecipitated DNA and 22 μ L of ddH₂O. All real-time qPCR reactions were done in triplicates. The samples were cycled to 94 °C for 15 s, 52 °C for 30 s and 72 °C for 30 s (total 40 cycles) on a LightCycler (Bio-Rad). The threshold cycle values were calculated automatically by the iCycler™ iQ optical system software (Bio-Rad Laboratories). Primer sequences used in this study are available on request.

ChIP-chip and expression data analysis

To identify TF-binding regions, we used the peak finding algorithm built into the NimbleScan™ software. Processing of ChIP-chip data was performed in three steps: normalization, IP/mock-IP ratio computation (log base 2), and enriched region identification. The log₂ ratios of each spot in the microarray were calculated from the raw signals obtained from both Cy5 and Cy3 channels, and then the values were scaled by Tukey bi-weight mean³⁴. The log₂ ratio of Cy5 (IP DNA) to Cy3 (mock-IP DNA) for each point was calculated from the scanned signals. Then, the bi-weight mean of this log₂ ratio was subtracted from each point. Each log ratio dataset from duplicate samples was used to identify TF-binding region using the software (width of sliding window = 300 bp). Our approach to identify the TF-binding regions was to first determine binding locations from each data set and then combine the binding locations from at least five of six datasets to define a binding region using the recently developed MetaScope software^{14,35}.

Motif searching

The ArgR-, Lrp-, and TrpR-binding motif analysis was completed using the MEME and FIMO tools from the MEME software suite³⁶. We first determined the proper binding motif and then scanned the full genome for its presence. The elicitation of the motif was done using the MEME program on the set of sequences defined by the ArgR-, Lrp-, and TrpR-binding regions respectively³⁷. Using default settings the previously determined ArgR³⁸, Lrp³, and TrpR⁷ motif were recovered and then tailored to the correct size by setting the width parameter to 18-bp, 15-bp, and 8-bp respectively. We then used these motifs and the PSPM (position specific probability matrix) generated for each by MEME to rescan the entire genome with the FIMO program. The sequence logo generated from these sites.

Raw experimental data

All raw data files can be downloaded from <http://systemsbiology.ucsd.edu/publications> or Gene Expression Omnibus through accession numbers GSE26054.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Marc Abrams and Joshua Lerman for critical reading of the manuscript. The National Institutes of Health, through Grant GM062791, and The Office of Science-Biological and Environmental Research, U.S. Department of Energy, DE-FOA-0000143 supported this work.

References

- Joyce AR, Palsson BO. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol.* 2006; 7:198–210. [PubMed: 16496022]
- Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature.* 2004; 429:92–6. [PubMed: 15129285]
- Cho BK, Barrett CL, Knight EM, Park YS, Palsson BO. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2008; 105:19462–7. [PubMed: 19052235]
- Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet.* 2007; 8:450–61. [PubMed: 17510665]
- Hershberg R, Margalit H. Co-evolution of transcription factors and their targets depends on mode of regulation. *Genome Biol.* 2006; 7:R62. [PubMed: 16859509]
- Charlier D, et al. Arginine regulon of *Escherichia coli* K-12. A study of repressor-operator interactions and of in vitro binding affinities versus in vivo repression. *J Mol Biol.* 1992; 226:367–86. [PubMed: 1640456]
- Yang J, et al. In vivo and in vitro studies of TrpR-DNA interactions. *J Mol Biol.* 1996; 258:37–52. [PubMed: 8613990]
- Pittard J, Camakaris H, Yang J. The TyrR regulon. *Mol Microbiol.* 2005; 55:16–26. [PubMed: 15612913]
- Tani TH, Khodursky A, Blumenthal RM, Brown PO, Matthews RG. Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A.* 2002; 99:13471–6. [PubMed: 12374860]
- Levin JZ, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods.* 2010
- Faith JJ, et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 2007; 5:e8. [PubMed: 17214507]
- Calvo JM, Matthews RG. The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol Rev.* 1994; 58:466–90. [PubMed: 7968922]
- Cho BK, Knight EM, Palsson BO. PCR-based tandem epitope tagging system for *Escherichia coli* genome engineering. *Biotechniques.* 2006; 40:67–72. [PubMed: 16454042]
- Cho BK, et al. The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol.* 2009; 27:1043–9. [PubMed: 19881496]
- Gama-Castro S, et al. RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Sensor Units). *Nucleic Acids Res.* 2011; 39:D98–105. [PubMed: 21051347]
- Keseler IM, et al. EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.* 2009; 37:D464–70. [PubMed: 18974181]
- Caldara M, Charlier D, Cunin R. The arginine regulon of *Escherichia coli*: whole-system transcriptome analysis discovers new genes and provides an integrated view of arginine regulation. *Microbiology.* 2006; 152:3343–54. [PubMed: 17074904]
- Caldara M, Minh PN, Bostoen S, Massant J, Charlier D. ArgR-dependent repression of arginine and histidine transport genes in *Escherichia coli* K-12. *J Mol Biol.* 2007; 373:251–67. [PubMed: 17850814]
- Jeeves M, Evans PD, Parslow RA, Jaseja M, Hyde EI. Studies of the *Escherichia coli* Trp repressor binding to its five operators and to variant operator sequences. *Eur J Biochem.* 1999; 265:919–28. [PubMed: 10518785]
- Van Duyne GD, Ghosh G, Maas WK, Sigler PB. Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J Mol Biol.* 1996; 256:377–91. [PubMed: 8594204]
- Zhang RG, et al. The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature.* 1987; 327:591–7. [PubMed: 3600756]
- Maas WK. Studies on the Mechanism of Repression of Arginine Biosynthesis in *Escherichia coli*. II. Dominance of Repressibility in Diploids. *J Mol Biol.* 1964; 8:365–70. [PubMed: 14168690]

23. Kiupakis AK, Reitzer L. ArgR-independent induction and ArgR-dependent superinduction of the *astCADBE* operon in *Escherichia coli*. *J Bacteriol.* 2002; 184:2940–50. [PubMed: 12003934]
24. Khodursky AB, et al. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2000; 97:12170–5. [PubMed: 11027315]
25. Bennett BD, et al. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol.* 2009; 5:593–599. [PubMed: 19561621]
26. Cho BK, et al. The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res.* 2011
27. Krishna S, Semsey S, Sneppen K. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proc Natl Acad Sci U S A.* 2007; 104:20815–9. [PubMed: 18093927]
28. Semsey S, et al. Genetic regulation of fluxes: iron homeostasis of *Escherichia coli*. *Nucleic Acids Res.* 2006; 34:4960–7. [PubMed: 16982641]
29. Semsey S, Krishna S, Sneppen K, Adhya S. Signal integration in the galactose network of *Escherichia coli*. *Mol Microbiol.* 2007; 65:465–76. [PubMed: 17630975]
30. Powell BS, et al. Novel proteins of the phosphotransferase system encoded within the *rpoN* operon of *Escherichia coli*. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an *erats* mutant. *J Biol Chem.* 1995; 270:4822–39. [PubMed: 7876255]
31. Klig LS, Oxender DL, Yanofsky C. Second-site revertants of *Escherichia coli* *trp* repressor mutants. *Genetics.* 1988; 120:651–5. [PubMed: 3066681]
32. Cho BK, Knight EM, Barrett CL, Palsson BO. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.* 2008; 18:900–10. [PubMed: 18340041]
33. Cho BK, Knight EM, Palsson BO. Genomewide identification of protein binding locations using chromatin immunoprecipitation coupled with microarray. *Methods Mol Biol.* 2008; 439:131–45. [PubMed: 18370100]
34. Mosteller F, Tukey JW. *Data Analysis and Regression: A second Course in Statistics.* 1977:203–209.
35. Kim DH, Palsson B, Cho BK. MetaScope: A visualization and data integration tool for analyzing genome-scale data of transcription unit architecture. Submitted. 2010
36. Bailey TL, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37:W202–8. [PubMed: 19458158]
37. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006; 34:W369–73. [PubMed: 16845028]
38. Makarova KS, Mironov AA, Gelfand MS. Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol.* 2001; 2:RESEARCH0013. [PubMed: 11305941]

Figure 1. Genome-wide distribution of ArgR- and TrpR-binding regions (regulatory code analysis)

(a) An overview of ArgR- and TrpR-binding profiles across the *E. coli* genome in the presence of exogenous arginine (blue track) and tryptophan (red track). Enrichment fold on the y-axis was calculated from Cy5 (IP-DNA) and Cy3 (mock control) signal intensity of each probe and plotted against each location on the 4.64 Mb *E. coli* genome. Dots indicate the binding regions previously identified (black) and newly determined (white). (b) Examples of genuine ArgR- (blue track), TrpR- (red track), and Lrp-binding (green track) regions on the selected regions. The *gltBD* operon is regulated by both ArgR and Lrp, whereas the *aroH* gene is solely regulated by TrpR. In the case of *mtr*, both TrpR and Lrp regulate its transcription. Both ArgR and Lrp regulate the *potFGHI* and *artPIQM* operons, however ArgR only regulates *artJ*. (c) Overlaps between ArgR-, Lrp-, and TrpR-binding regions.

Figure 2. Functional classification of genes directly regulated by ArgR, Lrp, and TrpR
The functions of regulon members are strongly enriched into amino acid, carbohydrate, membrane transport (mostly amino acid related), and energy metabolism (yellow bars).

Figure 3. Delineation of amino acid biosynthetic pathways and transport systems in *E. coli* (topological analysis)

(a) The amino acid biosynthetic pathways. The genes are directly regulated by ArgR (blue), Lrp (green), and TrpR (red), respectively. The orange colored genes (*gltB* and *gltD* in glutamate biosynthesis, and *argA* in arginine biosynthesis) are regulated by both ArgR and Lrp. Green dots indicate the biosynthetic reactions, which use glutamate as an amino donor.

(b) The amino acid transport systems. The genes encoding each transport system can be classified into ten groups (A ~ J) based on the amino acid specificity. The transport systems in green color are directly regulated by ArgR, Lrp, or TrpR. For others (gray), the transcriptional regulation has not been determined.

Figure 4. Causal relationships between direct associations of transcription factors and the changes in RNA transcript levels of genes (functional analysis)

*Here we include *tnaA* as an indirectly regulated gene with significant differential expression but no ChIP enrichment in order to fully capture the utilization response to arginine.

**For class C we note that the direct targets of utilization for glutamate are in fact the biosynthetic genes that are shown in that section and pointed out in figure 3.

Figure 5. Reconstruction of regulatory motif and the logical structures of connected circuit motifs in response to the exogenous amino acids (network analysis)
(a) Schematic diagram for the regulatory motif reconstruction in feedback circuit. (b) Logical structures of the regulatory motif in response to the exogenous amino acids. (c) The classification of function of amino acids derived from the logical structures of the connected feedback circuit motif. *The utilization patterns for glutamate and aspartate are highly complex. Here we conclude that the overall trend is repression given the role of glutamate as a substrate for nine biosynthetic pathways (Fig 3,4).