Supplementary Online Material for the article:

" Aggregation of topological motifs in the *E. coli* transcriptional regulatory network"

R. Dobrin, Q.K. Beg, A.-L. Barabási & Z.N. Oltvai

Table of content

I. The biology of motif clusters

II. Additional analyses

THE BIOLOGY OF MOTIF CLUSTERS

A. Feed-forward motif clusters

1. The first *feed-forward* motif cluster, comprising the **"flagellar-motor module"**, contains a set of more than 50 genes divided into at least 17 operons further classified into 3 different classes, which are involved in the synthesis and function of the flagellar and chemotaxis system of *E. coli* (Figure S1a). These genes are organized in an ordered cascade in which the expression of a gene located at a given level requires the transcription of another one at a higher level. The system is subject to complex regulation by environmental factors and regulatory proteins (Soutourina *et al.,* 2002). For example, flagellum biosynthesis is sensitive to catabolite repression, and it is also repressed under stress conditions, such as high temperature and osmolarity. The expression of many genes of the flagellum-chemotaxis regulon is positively regulated by the cyclic AMP-catabolite activator protein (cAMP-CAP) complex. However, recent studies have also demonstrated the involvement of H-NS (histone-like nucleoid-associated protein) in bacterial motility (Bertin *et al.*, 1994; Soutourina *et al.*, 2002). These studies have shown that the regulation of flagellar master operon, flhDC, is dependent on the H-NS protein. After expression of the first master operon (flhDC) by crp and hns, the genes of the flagellar-motor module function in a temporal order where the class 1 master operon (flhDC) encodes the transcriptional activator of class 2 operons (fliFGHIJK, fliLMNOPQR, fliE, flhBAE, flgBCDEFGHIJ, fliAZY, flgAMN) followed by transcription of class 3 genes (flgKL, fliDST, flgMN, fliC, meche, mocha and tsr) (Chilcott and Hughes 2000; Kalir *et al.*, 2002). The functions of the different genes of the flageller-motor motif cluster is listed in Table SM1.

Table SM1: Functions and properties of different genes of flagellar-motor module

Gene	Function(s)
fliA	Flagellar biosynthesis; alternative sigma factor 28; regulation of flagellar operons
fliC	Flagellar biosynthesis; flagellin; filament structural protein
fliD	Flagellar biosynthesis; filament capping protein; enables filament assembly
fliE	Flagellar biosynthesis; basal-body component; possibly at (MS-ring)-rod junction
fliF	Flagellar biosynthesis; basal-body MS (membrane and supramembrane)-ring and collar protein
fliL	Flagellar biosynthesis
flhB	Putative part of export apparatus for flagellar proteins
flhD	Regulator of flagellar biosynthesis; acting on class 2 operons; transcriptional

	initiation factor
flgA	Flagellar biosynthesis; assembly of basal-body periplasmic P ring
flgB	Flagellar biosynthesis; cell-proximal portion of basal-body rod
flgK	Flagellar biosynthesis; hook-filament junction protein 1
flgM	Anti-FliA (anti-sigma) factor; also known as RflB protein
hns	DNA-binding protein HLP-II (HU; BH2; HD; NS); pleiotropic regulator
tar (meche)	Methyl-accepting chemotaxis protein II; aspartate sensor receptor
motA (mocha)	Proton conductor component of motor; no effect on switching
tsr	Methyl-accepting chemotaxis protein I; serine sensor receptor

2. The *feed-forward* motif cluster, comprising the "aerobic-anaerobic switch module", contains the coding gene for the central anaerobic regulatory transcriptional factor, FNR (Fumarate-nitrate reductase regulator), which globally regulates the gene expression in response to oxygen deprivation in *E. coli*. In *E. coli,* the transcriptional regulation in response to oxygen is effected by two transcriptional regulators arcAB (aerobic respiratory control) and fnr. These two global regulators are the major controlling factors of catabolic gene expression and in most cases operate coordinately to fine-tune the catabolism in response to oxygen. The FNR, along with other sets of genes and/or operons (Figure S1b) comprise a distinct module in *E. coli*. FNR controls transcription of many genes whose functions facilitate adaptation of a cell to grow under O_2-limiting conditions. Thus, responding to the changes in the state of the growth medium, the FNR protein functions as the main regulator at the time of aerobic to anaerobic switch during growth of *E. coli* (Kiley and Beinert 1999; Sawers 1999). Low oxygen tensions activate the FNR protein, which is expressed constitutively and is inactive under aerobic conditions. The results of chemostat measurements and theoretical predictions in a report have shown that the genes for aerobic respiration (the sdh gene encoding succinate dehydrogenase) are expressed efficiently only at O_2 tensions above 5 mbar ($\cong$5 μM O_2), genes for microaerobic or anaerobic respiration are expressed between 1-5 mbar, while genes for fermentation are expressed or functional at O_2 tensions below 1 mbar (Unden 1998). In a recent study by Gunsalus' group (Salmon *et al.,* 2003), DNA microarray technology was used to identify genes involved in the regulatory networks that facilitate the transition of *E. coli* K-12 cells from an aerobic to an anaerobic growth state. They have identified several genes regulated by FNR and demonstrated that the expression of over one-third of the genes expressed during growth under aerobic conditions are altered when *E. coli* cells transition to an anaerobic growth state. They also showed that the expression of 49% of these genes is either directly or indirectly modulated by FNR.

In another recent study by the Gunsalus group (Wang and Gunsalus 2003), the role of NarL and NarP genes in regulation of genes of formate dehydrogenase operon (fdnGHI and fdhF) in response to nitrate, nitrite, and formate in anaerobic chemostat cultures has also been shown. Another major part of this motif cluster (the two-component arcA/arcB system) is the major control system for the regulation of expression of genes encoding enzymes involved in both aerobic and catabolic pathways. ArcA is a cytoplasmic regulatory protein and ArcB is a transmembrane protein, which senses unfavorable respiratory conditions. Upon simulation, ArcB undergoes two autophosphorylations and transfer a phosphoryl group to ArcA, which than becomes functional (Iuchi and Lin 1992). However, a recent study in wild type *E. coli* and in a mutant lacking arcA regulator using glucose-limited chemostat cultures at controlled levels of oxygen availability (ranging from full aerobiosis to complete anaerobiosis) have also shown that requirement of arcAB system as redox regulator in microaerobic conditions, but not in aerobic or anaerobic conditions (Alexeeva *et al.,* 2003). Table SM2 lists the functions and properties of different genes of aerobic-anaerobic transcriptional network.

Table SM2: Functions and properties of different genes of aerobic-anaerobic switch

Gene	Function(s)
fnr	Transcriptional regulation of aerobic; anaerobic respiration; osmotic balance
arcA	Aerobic respiration sensor-response protein; histidine protein kinase/phosphatase; sensor for arcA
appC	Probable third cytochrome oxidase; subunit I
appY	Regulatory protein affecting appA and other genes
cydA	ATP-binding component of cytochrome-related transport; Zn sensitive
cyoA	Cytochrome o ubiquinol oxidase subunit II
icdA	Isocitrate dehydrogenase; specific for $NADP^+$
focA	Probable formate transporter (formate channel 1)
nuoA	NADH dehydrogenase I chain A
glpA	sn-glycerol-3-phosphate dehydrogenase (anaerobic); large subunit
caiF	Transcriptional regulator of cai operon
dcuB	Anaerobic dicarboxylate transport
dmsA	Anaerobic dimethyl sulfoxide reductase subunit A
fdnG	Formate dehydrogenase-N; nitrate-inducible; alpha subunit
frdA	Fumarate reductase; anaerobic; flavoprotein subunit
hypA	Pleiotrophic effects on 3 hydrogenase isozymes
narG	Nitrate reductase 1; alpha subunit
nirB	Nitrite reductase [NAD(P)H] subunit
sdhC	Succinate dehydrogenase; cytochrome b556
acs	Acetyl-CoA synthetase

ansB	Periplasmic L-asparaginase II
aspA	Aspartate ammonia-lyase (aspartase)
ndh	Respiratory NADH dehydrogenase

3. The third *feed-forward* cluster (Figure S1c) comprises of operons and genes of **"nitrogen regulation and formate regulon"** with rpoN as central positive regulator of all other operons. The rpoN gene encodes for the sigma factor 'sigma 54'. In *E. coli*, the glnA gene encodes for glutamine synthetase and is transcribed from two promoters (glnAp1 and glnAp2). The glnAp1 is a sigma 70-dependent promoter that is activated by cAMP-repressor protein, crp (Figure S1c). The downstream of glnAp2 is dependent on sigma 54 factor (encoded by rpoN) and is activated by glnG. The study on CRP-cAMP complex and downregulation of glnAp2 promoter has established a regulatory link between carbon and nitrogen assimilation in *E. coli*. Tian *et al.,* (2001) have shown that the glnAp2 expression is clearly affected by different carbon and nitrogen sources and the CRP-cAMP complex complex inhibits the glnAp2 activity. Another important gene in this *E. coli* transcriptional network of nitrogen regulation is the nac gene (nitrogen assimilation control gene). The nac gene of *E. coli* is transcriptionally active, and its expression is dependent on sigma 54 factor and product of glnG (or ntrC) (Camarena *et al.,* 1998). Under nitrogen-limiting conditions the nac gene is involved in transcriptional repression of gdhA gene (encoding glutamate dehydrogenase) except when L-glutamine is used as substrate (Camarena *et al.,* 1998). In some recent studies, the nac gene has also been shown to regulate expression of cytosine deaminase (codBA) operon (Muse *et al.,* 2003) and asnC and asnA transcription (Poggio *et al.,* 2002) in *E. coli*. Another complex regulatory mechanism in this *feed-forward* cluster is regulation of formate dehydrogenylase (hyc operon) and selenium-dependent formate dehydrogense (fdhF) components of *E. coli* also called as **"formate regulon"**. The products of many genes of formate dehydrogenylase (FHL) system in *E. coli* are coordinately regulated in response to variations in the oxygen and nitrate concentrations and the pH of the culture medium (Mnatsakanyan *et al.,* 2002; Rossmann *et al.,* 1991). Transcription of both the fdhF and hyc operons requires the activator, fhlA, as well as formate and molybdate (Self and Shanmugam 2000; Self *et al.,* 2001). *E. coli* growing under anaerobic conditions produces H_2 and CO_2 by enzymatic cleavage of formate catalyzed by formate dehydrogenylase consisting of molybdoenzyme formate dehydrogenase H (fdhF), hydrogenase 3 (hyc operon) and intermediate electron carriers of hyc operon. All these components (fdhF, hyp and hyc) of formate regulon are also regulated by

transcription from sigma 54-dependent promoters. Table SM3 lists the functions and properties of different genes of the nitrogen regulation and formate regulon transcriptional network.

Table SM3: Functions and properties of different genes of feed-forward cluster 4 consisting of genes of nitrogen metabolism and formate regulon

Gene	Function(s)
FdhF	Selenopolypeptide subunit of formate dehydrogenase H
FhlA	Formate hydrogen-lyase transcriptional activator for fdhF; hyc and hyp operons
GlnA	Glutamine synthetase
GlnH	Periplasmic glutamine-binding protein; permease
HycA	Probable small subunit of hydrogenase-3; iron-sulfur protein [part of formate hydrogenlyase (FHL) complex]
Nac	Nitrogen assimilation control protein
RpoN	RNA polymerase; sigma (54 or 60) factor; nitrogen and fermentation regulation

4. *Feed-forward* cluster four in the *E. coli* transcriptional networks comprises of genes of ompR/envZ two component regulatory system called **"osmoregulated porin gene expression"**, which senses environmental osmolarity, and also regulates bacterial adhesion and biofilm formation (Figure S1d; Table SM4). The ompR234 protein has been shown to promote biofilm formation by binding to the promoter region of csgD promoter region and activating its transcription (Prigent-Combaret *et al.,* 2001). The csgG gene encodes the transcription regulator CsgD, which in turn activates the transcription of csgBA operon encoding curli, the extracellular structures involved in bacterial adhesion. The ompR is a response regulator of two-component regulatory system along with envZ (Mattison *et al.,* 2002; Yoshida *et al.,* 2002). Here, the osmosensor, EnvZ, has a dual enzymatic functions with OmpR kinase and OmpR-P phosphatase. The envZ is a sensory histidine kinase, which regulates phosphorylation of ompR, its conjugate response regulator, and both are required for the expression of genes of outer membrane porin proteins (ompF and ompC) in response to external osmotic changes. The ompR is a positive regulator specific for ompF and ompC genes. In one study, it has also been shown that the phosphorylated form of ompR, generated by either osmolarity change or the internal level of acetyl phosphate in the medium, negatively regulates the expression of flagellar master operon, flhDC (Shin and Park 1995). OmpR is also shown to confer increased single- or multidrug resistance in *E. coli* (Hirakawa *et al.*, 2003). Another element of this *feed-forward* cluster is himA, which codes for integration host factor (IHF) of *E. coli.* The IHF in *E. coli* is an asymmetric histone-like protein that binds and bends the DNA at specific sequences

resulting in formation of 'nucleo-protein' complexes (Goosen and van de Putte 1995). It functions as an accessory factor in variety of processes, such as replication, site-specific recombination and simulate transcription via a direct interaction with RNA polymerase.

Table SM4: Functions and properties of different genes of feed-forward cluster 5 comprising of genes of osmoregulated porin gene expression and biofilm formation

Gene	Function(s)
csgB	Minor component of curlin subunit precursor; located on the bactraisl surface; similar to CsgA
csgA	Curlin major subunit protein of the fibre; coiled surface structures; cryptic
csgD	Putative 2-component transcriptional regulator for second curli operon
himA	Integration host factor (IHF), alpha subunit
ompC	Outer membrane protein 1b (Ib;c)
ompF	Outer membrane protein 1a (Ia;b;F)
ompR	Osmolarity-sensing response regulator (sensor; envZ) affecting transcription of ompC and ompF: outer membrane protein synthesis

5. Another *feed-forward* cluster in the *E. coli* transcriptional networks comprises of genes of connectivity between different regulatory pathways in the organism's **"oxidative stress response"** (Figure S1e; Table SM5). Two main components of this network soxS and marA, are the direct regulators of superoxide (soxRS) and multiple antibiotic resistance (mar) regulons in *E. coli*. The soxRS regulon, in part, mediates transcriptional activation of an oxidative stress response, which protects the cell against both superoxide-generating (redox-cycling) agents and nitric oxide. The synthesis of both these regulons is induced by specific environmental stimuli and elicits a clear defense mechanism. The multiple antibiotic resistance in *E. coli* is also mediated by induction of soxS or marA protein, triggered by oxygen radicals (in soxRS regulon) or certain antibiotics (in marRAB regulon), respectively. Another member of this network, rob (right oriC binding), is regarded as constitutively expressed protein and it activates many other genes of the network (Figure S1g). Michan *et al.*, (2002) by doing RT-PCR showed that the transcriptional level of rob are strongly down-regulated in response to the superoxide-generating agent paraquat (PQ) and the magnitude of rob repression was comparable to that of the induction quantified for the most sensitive soxS targets. Rob affects the basal expression of many genes with a broad range of functions including antibiotic resistance, acid adaptation, carbon metabilism, cell wall synthesis, central intermediary mechanism and transport (Bennik *et al.*, 2000). The promoters of mar/sox/rob regulon of *E.*

coli contain a binding site (marbox) for the homologous transcriptional activators marA, soxS and rob. MarA is the negative regulator of *E. coli* multiple antibiotic resistance (marRAB) operon. The transcription of the marRAB operon is autorepressed by marR and is autoactivated by marA at a site in marbox that is also activated by soxS (Martin *et al.*, 1996). The marR family is a group of regulatory factors whose activity is modulated in response to signals from phenolic compounds, many of which are plant derived (Sulavik *et al.*, 1995). Another member of this motif cluster, fumarase C (fumC), is also controlled by soxRS regulon. The function of fum C in *E. coli* is to catalyze the stereo-specific interconversion of fumarate to L-malate as part of the citric acid cycle. Fumarase C was shown to be strongly induced by paraquat in a parental strain of *E. coli*, but was not induced in a strain lacking the soxRS response (Liochev and Fridovich 1992). The transcription of another gene in the network, sodA (encoding manganese-containing superoxide dismutase) is governed by other regulators: the products of soxRS, rob and marRAB positively regulate sodA; whereas products of arcA (aerobic respiratory control), fur2 (ferric uptake regulation), and himA (IHF) negatively regulate sodA expression. The expression of another gene, zwf, that encodes a glucose-6-phosphate dehydrogenase, is coordinated with cellular growth and is induced by products of superoxide-generating agents (soxRS, rob and marRAB regulons) in response to superoxide stress. Similarly these three agents also activate expression of nfo gene, encoding endonuclease IV, a DNA repair enzyme for free radical damages, which protects *E. coli* from lethal and mutagenic effects of nitric oxide and near-UV irradiaion.

Table SM5: Functions and properties of different genes of feed-forward cluster 6 consisting of genes of oxidative stress response

Gene	Function(s)
fumC	Fumarase C; fumarate hydratase Class II; isozyme
marR	Multiple antibiotic resistance protein; repressor of mar operon
nfo	Endonuclease IV
rob	Right origin-binding protein
sodA	Superoxide dismutase; manganese
soxS	Regulation of superoxide response regulon
zwf	Glucose-6-phosphate dehydrogenase

6. The sixth *feed-forward* cluster of *E. coli* comprises of set of several genes and regulons with crp gene, which codes for cyclic AMP receptor protein, as their central regulator of all regulons and op-

erons (Figure S1g; Table SM6). The transcription of crp gene is also under autogenous regulation by CRP-cAMP complex that binds to a specific site located downstream from transcriptional site (Hanamura and Aiba 1991). The **arabinose regulon** (Figure S1g) is positively regulated by activity of crp gene. In *E. coli*, the products of araC gene positively and negatively regulate the expression of proteins required for uptake and catabolism of sugar L-arabinose. The cAMP receptor protein and AraC protein activate transcription of araFGH and araJ promoters. The AraC protein also represses its own transcription from its own promoter and when associated with arabinose, activates transcription of two more promoters of araBAD and araE. Similarly, the activation and positive co-regulation of two divergent *E. coli* structural operons, caiTABCDE and fixABCX, involved in an-aerobic carnitine metabolism is co-dependent on crp protein and caiF (the transcriptional regulator of cai operon). Both operons are coexpressed during anaerobic growth in the presence of carnitine, respond to environmental stimuli, like glucose and nitrate, and are also modulated by the same regulators (crp, fnr, narL and hns) of caiF (Buchet *et al.*, 1998; 1999). The fur (ferric uptake regulator) protein, encoded by fur gene, is a global negative regulator of the iron absorption, and of several other siderophore-mediated high affinity iron transport systems in *E. coli*. The fur protein uses iron as a co-factor to bind to specific DNA sequences. A variety of genes involved in various mechanism, such as oxidative and acid stresses are also controlled by fur (D'Autreaux *et al.*, 2002). The synthesis of colicin I receptor protein, encoded by cir gene, is also under the dual control of crp regulatory system and fur. Another *E. coli* regulon responsible for utilization of **L-fucose** as carbon and energy source consists of fucAO operon, positively regulated by crp and fucP genes. The fucPIK operon in *E. coli*, encoding for fucose permease, fucose isomerase and fucose kinase, activates the other gene fucA, encoding fucolase-1-phosphate aldolase. Another *feed-forward* motif in this crp-dependent cluster is regulation of **gal (galactose) regulon** involving gal genes and genes of mgl (β-methyl galactosidase transport) operon. The galR binds strongly to the operator of galS, and galS binds strongly to the mgl operon, and represses the activity of enzymes of galactose metabolism. The **maltose regulon** (for maltose biosynthesis) of *E. coli* consists of several operons that direct the synthesis of proteins and enzymes required for transport and metabolism of maltose and maltodextrins. The transcription of **mal regulon** is positively regulated by malT gene and this malT gene is under catabolite repression of crp-cAMP complex (Figure S1f). Its activity is also regulated by ATP and maltotriose (the inducer), which positively controls its multimerization. Another operon of this *feed-forward* cluster comprises of genes for N-acetylglucosamine (NAG)

utilization. The turnover and recycling of cell wall murein is a major metabolic pathway in *E. coli*. *E. coli* recycles NAG and anhydro-N-acetylmuramic acid (present in cell wall murein) degradation products for de novo murein and lipopolysaccharide synthesis. This involves the regulatory step of NAG metabolism, which involves isomerization and deamination of glucose-6-phosphate to form fructose-6-phosphate and ammonia (Park 2001; Horjales *et al.*, 1999). The enzymes required for NAG utilization in *E. coli* are enzyme IInag (gene nagE), N-acetylglucosamine-6-phosphate deacetylase (gene nagA) and glucosamine-6-phosphate acetylase (gene nagB). The gene glmS, encoding enzyme glucosamine-6-phosphate synthase controls the activity of genes of nag regulon. Another element of this small cluster, the manXYZ operon, encodes a sugar transporter of the phosphoenol pyruvate (PEP)-dependent phosphotransferase system (PTS), which helps in transportation of many sugars, *viz.*, glucose, mannose, amino sugars, glucosamine and N-acetylglucosamine. The transcription of manXYZ is also positively regulated by crp-cAMP complex (Plumbridge 1998). The **melibiose synthesis** in *E. coli* is regulated by activity of melR (encoding MelR protein) and melAB genes. The MelR protein belongs to the family of AraC family of transcriptional factors, it activates the expression of melAB operon in response to presence of melibiose in the environment, and is also regulated by crp-cAMP complex.

Table SM6: Functions and properties of different genes of feed-forward cluster 7 regulated by a global regulator crp-cAMP complex

Gene	Function(s)
araB	L-ribulokinase
araC	Transcriptional regulator for ara operon
araE	Low-affinity L-arabinose transport system proton symport protein
araF	L-arabinose-binding periplasmic protein
araJ	Involved in either transport or processing of arabinose polymers
caiF	Transcriptional regulator of cai operon
caiT	Probable carnitine transporter
cirA	Outer membrane receptor for iron-regulated colicin I receptor; porin; requires tonB gene product
crp	Cyclic AMP receptor protein
fixA	Probable flavoprotein subunit; carnitine metabolism
fucA	L-fuculose-1-phosphate aldolase
fucP	Fucose permease
fur	Negative regulator of iron absorption
galS	mgl repressor; galactose operon inducer
malE	Periplasmic maltose-binding protein; substrate recognition for transport and chemotaxis

malK	ATP-binding component of transport system for maltose
malS	Gene encoding the periplasmic α-amylase
malT	Positive regulator of mal regulon
malX	PTS system; maltose and glucose-specific II ABC
manX	PTS enzyme IIAB; mannose-specific
melA	α-galactosidase
melR	Regulator of melibiose operon
mglB	Galactose-binding transport protein; receptor for galactose taxis
nagB	Glucosamine-6-phosphate deaminase
nagE	PTS system; N-acetylglucosamine-specific enzyme IIABC

7. One *feed-forward* motif is not part of a cluster (Figure S1f), and its components play role in methionine biosynthesis, which is under complex regulation in *E. coli*. The repression of the biosynthetic pathway by methionine is mediated by a repressor protein metJ, which is the repressor of met genes in *E. coli*, and S-adenosyl methionine, which functions as corepressor for the metJ protein (Weissbach and Brot 1991). The metR, a *lys R* family protein, is a positive regulator for several genes in the met regulon, and homocysteine, a methionine pathway intermediate, serves as a coregulator. MetR positively regulate the expression of glyA and metH genes (Lorenz and Stauffer 1996). Table SM7 lists the functions of the three genes within this motif.

Table SM7: Functions and properties of different genes of methionine biosynthesis module

Gene	Function(s)
metA	Homoserine transsuccinylase; methionine biosynthesis from homoserine; first enzyme in the methionine biosynthetic pathway
metJ	Repressor of all met genes; negative DNA-binding transcriptional regulator for methionine biosynthesis; the protein family is MetJ
metR	Regulator for metE and metH; dual DNA-binding transcriptional regulator for methionine biosynthesis; the protein family is LysR

B. Bi-fan motif cluster

1. One of the *bi-fan* motif clusters (Figure S2a; Table SM8) consists of genes of the **sigma E regulon and a two-component signal transduction system** (cpxAR). In *E. coli*, the sigma factors are the key regulators of bacterial transcription. The bacterial heat shock response is under the control of two alternative sigma factors sigma 32 and sigma E. The sigma E factor is encoded by the rpoE gene. This heat shock-inducible sigma factor and the cpx two component signal transduction

system (cpxAR) controls the synthesis of periplasmic protein-folding enzymes in *E. coli*. The rpoE gene (sigma E regulon), is induced upon heat stress due to misfolding of proteins in the periplasm or the outer membrane and this transcriptional factor is positively autoregulated at the transcriptional level. The second gene of this operon, rseA, encodes an anti-sigma E activity (Missiakas *et al.*, 1997). When the stress is removed by the temperature downshift, the sigma E activity is strongly repressed and is slowly returned to levels maintained in the unstressed cells. The cpxAR signal transduction system, which consists of an inner-membrane sensor (CpxA) and a cognate response regulator (CpxR), is stimulated by outer-membrane protein, NlpD (encoded by nlpD). In response to these cytoplasmic stimuli, the sigma E and cpxAR together are involved in regulating activities of other genes (Figure S2a) responsible for both protein-turnover and protein folding activities in bacterial envelope. The cpxAR gene is also directly involved in the negative regulation of genes, tsr (encoding the serine protease receptor) and motABcheAW (specifying motility and chemotaxis) of *feed-forward* cluster 1. Therefore, this regulon also has a role in controlling bacterial motility in the context of a "stress response network" in which the unifying element is coordination of protein turnover and energy conservation. (De Wulf *et al.,* 1999).

Table SM8: Bi-fan motif cluster 1

Gene	Function(s)
cpxA	The sensor protein CpxA is part of the two-component CpxA/CpxR signal transduction system. The system senses and responds to aggregated and misfolded proteins in the bacterial envelope.
ecfI	Conserved hypothetical protein
fliA	Flagellar biosynthesis; alternative sigma factor 28; regulation of flagellar operons
htrA	Periplasmic serine protease; heat shock protein HtrA
motA	Proton conductor component of motor; no effect on switching
rpoE	RNA polymerase; sigma-E factor; heat shock and oxidative stress
skp	Histone-like protein, located in outer membrane or nucleoid
tsr	Methyl-accepting chemotaxis protein I; serine sensor receptor
xprB (xerD)	Site-specific recombinase

2. The second *bi-fan* motif cluster comprises a very large number of nodes, with involvement in multiple cellular functions. Thus, we will not attempt the detailed analysis of this cluster.

3. An additional *bi-fan* motif does not participate in a motif cluster (Figure S2b; Table SM9), but its gene products play a role in **"aromatic amino acid biosynthesis"**. The tyrR gene (tyrosine repressor) is the major transcriptional regulator of a group of genes that are essential for aromatic amino acid biosynthesis and transport in *E. coli*. The TyrR protein is involved in both repression and activation of other transcriptional units of the tyrR regulon, which itself can function either as activator or as a repressor, and also has an intrinsic ATPase and phosphatase activity (Zhao *et al.,* 2000). The products of trpR (tryptophan repressor) controls transcription initiation in the tryptophan biosynthetic operon and four other operons in *E. coli*, and the expression of trpR is also autoregulatory. Both TyrR protein and TrpR repressor are together responsible for regulation of other genes of aromatic amino acid transport. Site-directed mutagenesis studies confirmed that TyrR protein in association with phenylalanine or tyrosine is responsible for mtr gene activation, whereas the Trp repressor in conjunction with tryptophan serves to repress mtr gene (tryptophan specific permease gene) (Sarsero *et al.,* 1991). However, both these proteins positively regulate the expression of aroL gene encoding enzyme shikimate kinase II in *E. coli*.

Table SM8: Bi-fan motif cluster 1 comprising genes of "aromatic amino acid biosynthesis"

Gene	Function(s)
aroL	Shikimate kinase II
mtr	Tryptophan-specific transport protein
trpR	Regulator for trp operon and aroH; trp aporepressor
tyrR	Transcriptional regulation of aroF, aroG, tyrA and aromatic amino acid transport

Additional analyses

Network Statistics

For the *E. coli* transcriptional regulatory network, we have also determined the in-degree and out-degree connectivity distributions. As seen in Figure S3 the out-degree distribution is consistent with a scale-free distribution, the slope of the solid line being g = -1.5, and is identical to the solid line on Figure 2E. However, the in-degree distribution is exponential, consistent to that demonstrated previously for the transcriptional regulatory network in the yeast, *S. cerevisiae,* by Guelzim et al. (44). We expect that the exponent value might change if one is using a more complete database. When considering the exponents it is important to note that the network has only 423 nodes and 578 links, which is insufficient for a reliable determination of the connectivity distribution. Also, because of its nature (regulatory network) the in-degree and out-degree are quite dissimilar, most operons being regulated by only one transcription factor, while there are no highly regulated operons as we can see also see from figure S3. Similar measurements on the network have been made by Alon et. al. (Refs. 3,4).

Effects of data incompleteness and database errors

The database used is likely incomplete and has errors. In order to verify these effects on our results we have randomly removed 10% or 20% of the interactions, which accounts for the effect of false negative interactions. In the case of removing ~10% (50 links) of the links we have obtained on average 31±2 *feed-forward* motifs and 145±11 *bi-fan* motifs, respectively. Compared to the original results, the number of total *feed-forward* motifs is reduced by 26%, and a 30% reduction of *bi-fan* motifs is seen. If we increase the number of removed links to ~20% (100 links) we detect on average 22±3 *feed-forward* motifs and 92±13 *bi-fan* motifs, a loss of ~47% in the number of *feed-forward* motifs and 56% in the number of the *bi-fan* motifs, respectively. These measurements indicate that despite the relative decrease in their numbers, the observed motifs are still statistically highly significant when compared with a randomized version of the networks.

To mimic false positives we can randomly add 10% or 20% new links to the original network. This task is much harder because the nature of the transcriptional network would be dramatically altered. It is unclear at this moment if by adding links randomly the results would have any relevance. Adding hypothetical regulation randomly could change the network function, which, at this moment, has not been deciphered.

References:

1. Alexeeva S, Hellingwerf KJ, Teixeira de Mattos MJ. (2003). Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *J. Bacteriol.* 185: 204-209.

2. Bennik MH, Pomposiello PJ, Thorne DF, Demple B. (2000). Defining a rob regulon in *Escherichia coli* by using transposon mutagenesis. *J. Bacteriol.* 182: 3794-3801.

3. Bertin P, Terao E, Lee EH, Lejeune P, Colson C, Danchin A, Collatz E. (1994). The H-NS protein is involved in the biogenesis of flagella in *Escherichia coli*. *J. Bacteriol.* 176: 5537-40.

4. Buchet A, Eichler K, Mandrand-Berthelot MA. (1998). Regulation of the carnitine pathway in *Escherichia coli*: investigation of the cai-fix divergent promoter region. *J. Bacteriol.* 180: 2599-2608.

5. Buchet A, Nasser W, Eichler K, Mandrand-Berthelot MA. (1999). Positive co-regulation of the *Escherichia coli* carnitine pathway cai and fix operons by CRP and the CaiF activator. *Mol. Microbiol.* 34: 562-575.

6. Camarena L, Poggio S, Garcia N, Osorio A. (1998). Transcriptional repression of gdhA in *Escherichia coli* is mediated by the Nac protein. *FEMS Microbiol. Lett.* 167: 51-56.

7. Chilcott GS, Hughes KT. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 64: 694-708.

8. D'Autreaux B, Touati D, Bersch B, Latour JM, Michaud-Soret I. (2002). Direct inhibition by nitric oxide of the transcriptional ferric uptake regulation protein via nitrosylation of the iron. *Proc. Natl. Acad. Sci. USA.* 99: 16619-16624.

9. De Wulf P, Kwon O, Lin EC. (1999). The CpxRA signal transduction system of Escherichia coli: growth-related autoactivation and control of unanticipated target operons. *J. Bacteriol.* 181: 6772-6778.

10. Goosen N, van de Putte P. (1995). The regulation of transcription initiation by integration host factor. *Mol. Microbiol.* 16: 1-7.

11. Hanamura A, Aiba H. (1991). Molecular mechanism of negative autoregulation of *Escherichia coli* crp gene. *Nucleic Acids Res.* 19: 4413-4419.

12. Hirakawa H, Nishino K, Hirata T, Yamaguchi A. (2003). Comprehensive studies of drug resistance mediated by overexpression of response regulators of two-component signal transduction systems in *Escherichia coli*. *J. Bacteriol.* 185: 1851-1856.

13. Horjales E, Altamirano MM, Calcagno ML, Garratt RC, Oliva G. (1999). The allosteric transition of glucosamine-6-phosphate deaminase: the structure of the T state at 2.3 A resolution. *Structure Fold Des.* 7: 527-537.

14. Iuchi S, Lin EC. (1992). Mutational analysis of signal transduction by ArcB, a membrane sensor protein responsible for anaerobic repression of operons involved in the central aerobic pathways in *Escherichia coli*. *J. Bacteriol.* 174: 3972-3980.

15. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292: 2080-2083.

16. Kiley PJ, Beinert H. (1999). Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster. *FEMS Microbiol. Rev.* 22: 341-352.

17. Liochev SI, Fridovich I. (1992). Fumarase C, the stable fumarase of *Escherichia coli*, is controlled by the soxRS regulon. *Proc. Natl. Acad. Sci. USA.* 89: 5892-5896.

18. Lorenz E, Stauffer GV. (1996). MetR-mediated repression of the glyA gene in *Escherichia coli*. *FEMS Microbiol. Lett.* 144: 229-233.

19. Martin RG, Jair KW, Wolf RE, Rosner JL. (1996). Autoactivation of the marRAB multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli*. *J. Bacteriol.* 178: 2216-2223.

20. Mattison K, Oropeza R, Kenney LJ. (2002). The linker region plays an important role in the interdomain communication of the response regulator OmpR. *J. Biol. Chem.* 277: 32714-37221.

21. Michan C, Manchado M, Pueyo C. (2002). SoxRS down-regulation of rob transcription. *J. Bacteriol.* 184: 4733-4738.

22. Missiakas D, Mayer MP, Lemaire M, Georgopoulos C, Raina S. (1997). Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol. Microbiol.* 24: 355-71

23. Mnatsakanyan N, Vassilian A, Navasardyan L, Bagramyan K, Trchounian A. (2002). Regulation of *Escherichia coli* formate hydrogenlyase activity by formate at alkaline pH. *Curr. Microbiol.* 2002 45: 281-286.

24. Muse WB, Rosario CJ, Bender RA. (2003). Nitrogen regulation of the codBA (cytosine deaminase) operon from *Escherichia coli* by the nitrogen assimilation control protein, NAC. *J. Bacteriol.* 185: 2920-2926.

25. Park JT. (2001). Identification of a dedicated recycling pathway for anhydro-N-acetylmuramic acid and N-acetylglucosamine derived from *Escherichia coli* cell wall murein. *J. Bacteriol.* 183: 3842-3847.

26. Plumbridge J. (1998). Control of the expression of the manXYZ operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol. Microbiol.* 27: 369-380.

27. Poggio S, Domeinzain C, Osorio A, Camarena L. (2002). The nitrogen assimilation control (Nac) protein represses asnC and asnA transcription in *Escherichia coli. FEMS Microbiol. Lett.* 206:151-156.

28. Prigent-Combaret C, Brombacher E, Vidal O, Ambert A, Lejeune P, Landini P, Dorel C. (2001). Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the csgD gene. *J. Bacteriol.* 183: 7213-7223.

29. Rossmann R, Sawers G, Bock A. (1991). Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Mol. Microbiol.* 5: 2807-2814.

30. Salmon K, Hung SP, Mekjian K, Baldi P, Hatfield GW, Gunsalus RP. (2003). Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *J. Biol. Chem.* 278: 29837-29855.

31. Sarsero JP, Wookey PJ, Pittard AJ. (1991). Regulation of expression of the *Escherichia coli* K-12 mtr gene by TyrR protein and Trp repressor. *J. Bacteriol.* 173: 4133-4143.

32. Sawers G. (1999). The aerobic/anaerobic interface. *Curr. Opin. Microbiol.* 2:181-187.

33. Self WT, Hasona A, Shanmugam KT. (2001). N-terminal truncations in the FhlA protein result in formate- and MoeA-independent expression of the hyc (formate hydrogenlyase) operon of *Escherichia coli*. Microbiol. 147: 3093-3104.

34. Self WT, Shanmugam KT. (2000). Isolation and characterization of mutated Fh1A proteins which activate transcription of the hyc operon (formate hydrogenlyase) of *Escherichia coli* in the absence of molybdate(1). *FEMS Microbiol. Lett.* 184: 47-52.

35. Soutourina OA, Krin E, Laurent-Winter C, Hommais F, Danchin A, Bertin PN. (2002). Regulation of bacterial motility in response to low pH in *Escherichia coli*: the role of H-NS protein. *Mol. Microbiol.* 148: 1543-1551.

36. Shin S, Park C. (1995). Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J. Bacteriol.* 177: 4696-4702.

37. Sulavik MC, Gambino LF, Miller PF. (1995). The MarR repressor of the multiple antibiotic resistance (mar) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. *Mol. Med.* 1: 436-446.

38. Tian ZX, Li QS, Buck M, Kolb A, Wang YP. (2001). The CRP-cAMP complex and downregulation of the glnAp2 promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in *Escherichia coli. Mol. Microbiol.* 41: 911-924.

39. Unden G. (1998). Transcriptional regulation and energetics of alternative respiratory pathways in facultatively anaerobic bacteria. *Biochim. Biophys. Acta.* 1365: 220-224.

40. Wang H, Gunsalus RP. (2003) Coordinate regulation of the *Escherichia coli* formate dehydrogenase fdnGHI and fdhF genes in response to nitrate, nitrite, and formate: roles for NarL and NarP. *J Bacteriol.* 185:5076-85.

41. Weissbach H, Brot N. (1991). Regulation of methionine synthesis in *Escherichia coli*. *Mol. Microbiol.* 5: 1593-1597.

42. Yoshida T, Cai S, Inouye M. (2002). Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator. *Mol. Microbiol.* 46: 1283-1294.

43. Zhao S, Zhu Q, Somerville RL. (2000). The sigma (70) transcription factor TyrR has zinc-stimulated phosphatase activity that is inhibited by ATP and tyrosine. *J. Bacteriol.* 182: 1053-1061.

44. Guelzim N, Bottani S, Bourgine P, Kepes F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31: 60-3.

Figure Legend

Figure S1 The *feed-forward* motif clusters

The operons found in *feed-forward* motif clusters are the nodes colored in either green, if the link between them is shared by two or more motifs and/or the operon is shared by more than one motif, or blue otherwise. The red nodes represent TFs outside of the motif cluster, but which regulate operons that are part of the motif. The black nodes denote operons or group of operons outside the motif cluster but that are regulated by TFs that are part of the motif cluster. Since we have three types of regulation we have colored the links as follows: black for activation, red for repression and double-arrow link for either activation or repression. The groups of outside genes regulated by TFs inside the motifs for each figure from (a) to (g) are: **(a)** Group 1: flgKL, flgMN, fliC, fliDSC, motABcheAW, tarTapcheRBYZ, tsr; **(b)** Group 1: acs, ansB, aspA, caiF, dcuB_fumB, dmsABC, fdnGHI, frdABCD, hypABCDE, narGHIJ, nirBDC_cysG, nark, tdcABCDEFG; Group 2: aceBAG, betIBA, dctA, fadBA, fumA, fumC, gltA, glcDEFGB, lcdPRD, mdh, sodA; **(c)** Group 1: atoC, actA, hypA, nycA, pspABCDE, rtcR, zraP; **(d)** Group 1: aceBAK, dps, ecp_htrE, focA_pflB, glcDEFGB, hycABCDEFGH, hypABCDE, narGHJI, narK, pspABCDE, tdcABCDEFG; Group 2: caiTABCDE, himD, nuoABCDEFGHIJKLMN, osmE, sodA; **(e)** Group 1: aslB, inaA, mdlA, ybaO, ybiS, yfhD; **(g)** Group 1: entCEBA, fecIR, fepA,_entD, fepB, fepDGC, fhuACDB, sodA, tonB; Group 2: cyaA, dctA, dcuB_fumB, fadL, ompA, spec; Group 3: acs, aldB, ansB, cpdB, dadAX, ebgAC,epd_pgk, flhDC, focA_pflB, glgCAP, glgS, glnALG, glpACB, glpD, glpFK, glpTQ, gltA, ivbL_ilvBN, lacZYA, rhaT, sdhCDAB_b0725_sucABCD, srlAEBD_gutM_srlR_gutQ, tdcABCDEFG, tnaLAB, ubiG, udp, uhpT, yiaKLMNOPQRS; Group 4: deoCABD, galETKM, nupG, ppiA, proP, ptsHI_crr, rhaBAD, rpoH, tsx, yhfA.

Figure S2 The bi-fan motif clusters

We have detected 3 bi-fan motif clusters, however only 2 of them are detailed in this figure, the third one, as seen in Fig. 2C, is too large for attempting to index all of its operons. The color codes are the same as in Figure S1. The gene groups are as follows: **(a)** Group 1: cutC, dapA_nlpB_purA,

ecfABC, ecfD, ecfF, ecfG, ecfH, ecfJ, ecfK, ecfLM, fkpA, ksgA_epaG_epaH, lpxDA_fabZ, mdoGH, nlpB_purA, ostA_surA_pdxA, rfaDFLC, rpoD, rpoH, uppS_cdsA_ecfE; Group 2: flgBCDEFGHIJK,flgKL, flgMN, flhBAE, fliC, fliDST, fliE, fliFGHIJK, fliLMNOPQR

Figure S3 The degree distribution for incoming and outgoing links in the *E. coli* transcriptional regulatory network

(A) The incoming (red squares) and (B) outgoing (black circles) degree distributions are shown. The solid black line in panel (A) is an exponential, while the dashed-line in panel (B) has a power-law with exponent $\gamma = -1.5$. The out-degree distribution can be regarded as a power law, similar to the degree distribution from Figure 2E, where we have considered the original network as undirected. However, the in-degree distribution closely resembles an exponential distribution (see the text for details).

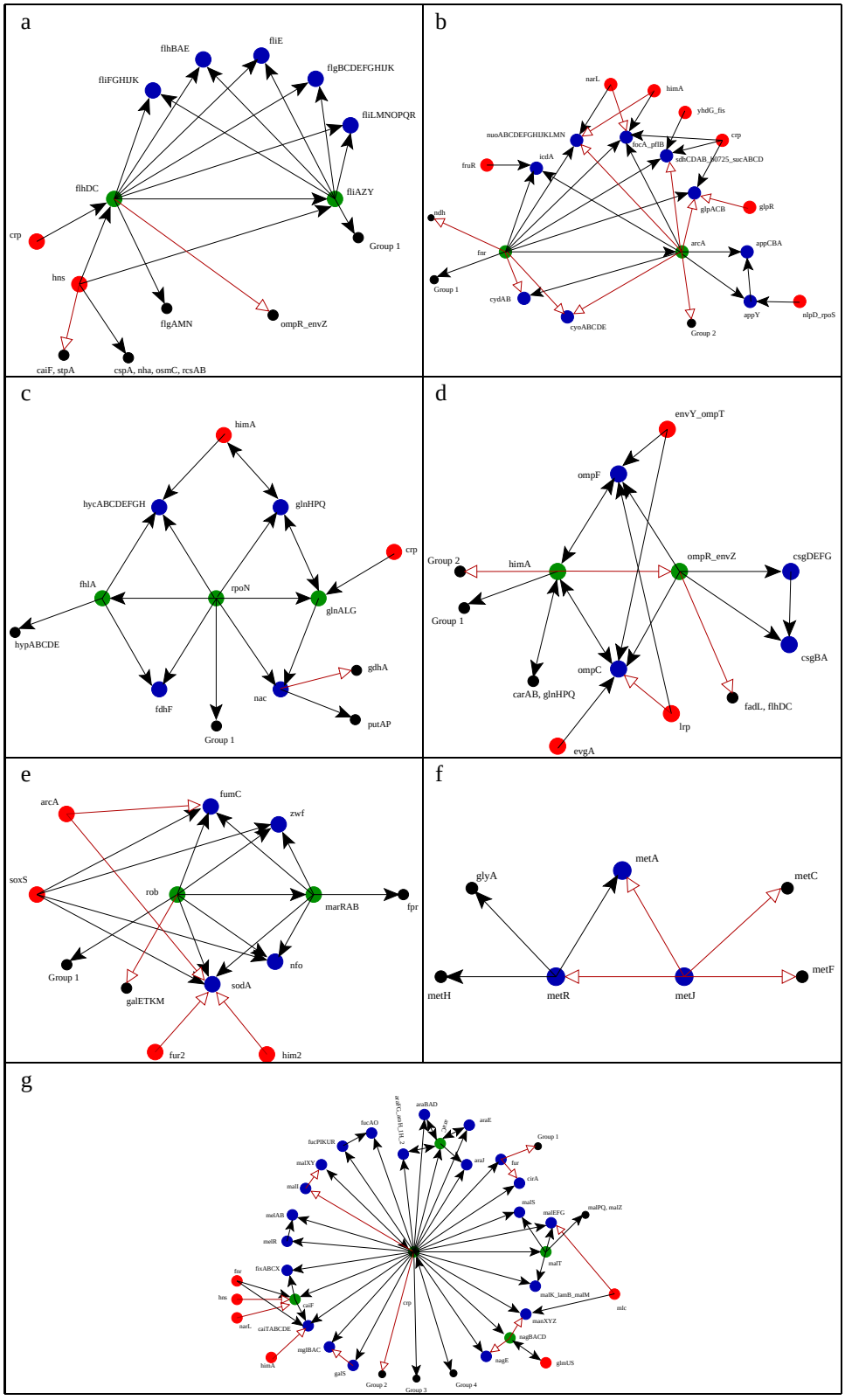
a
flhBAE
fliE
fliFGHIJK
flgBCDEFGHIJK
fliLMNOPQR
flhDC
fliAZY
crp
Group 1
hns
flgAMN
ompR_envZ
caiF, stpA
cspA, nha, osmC, rcsAB
b
narL
himA
yhdG_fis
nuoABCDEFGHIJKLMN
tocA_pflB
crp
fruR
icdA
sdhCDAB_b0725_sucABCD
glpACB
glpR
ndh
fnr
arcA
appCBA
Group 1
cydAB
appY
nlpD_rpoS
cyoABCDE
Group 2
c
himA
hycABCDEFGH
glnHPQ
crp
fhlA
rpoN
glnALG
hypABCDE
gdhA
nac
fdhF
putAP
Group 1
d
envY_ompT
ompF
Group 2
himA
ompR_envZ
csgDEFG
Group 1
csgBA
ompC
carAB, glnHPQ
fadL, flhDC
lrp
evgA
e
arcA
fumC
zwf
soxS
rob
marRAB
fpr
Group 1
nfo
sodA
galETKM
fur2
him2
f
glyA
metA
metC
metF
metH
metR
metJ
g
araBAD
araC
fucAO
araE
fucPIKUR
Group 1
malXY
araJ
fur
malI
cirA
malS
malEFG
malPQ, malZ
melAB
melR
malT
fixABCX
fnr
hns
caiF
malK_lamB_malM
nlc
narL
caiTABCDE
manXYZ
mglBAC
nagBACD
himA
nagE
glmUS
galS
Group 2
Group 3
Group 4

Figure S1

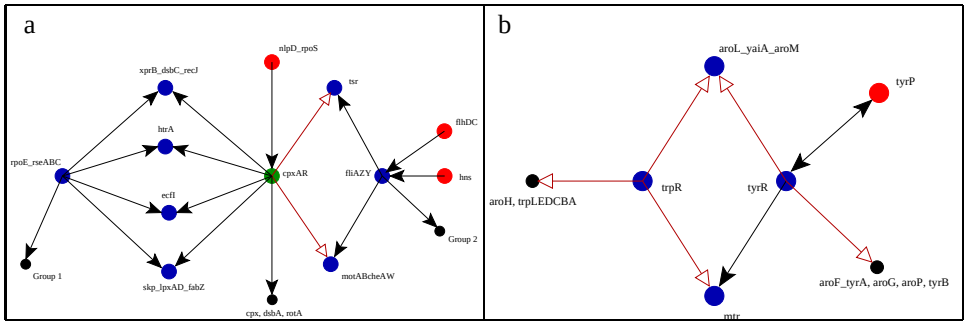
a
nlpD_rpoS
xprB_dsbC_recJ
tsr
flhDC
htrA
rpoE_rseABC
cpxAR
fliAZY
hns
ecfI
Group 2
Group 1
skp_lpxAD_fabZ
motABcheAW
cpx, dsbA, rotA
b
aroL_yaiA_aroM
tyrP
aroH, trpLEDCBA
trpR
tyrR
aroF_tyrA, aroG, aroP, tyrB
mtr

Figure S2

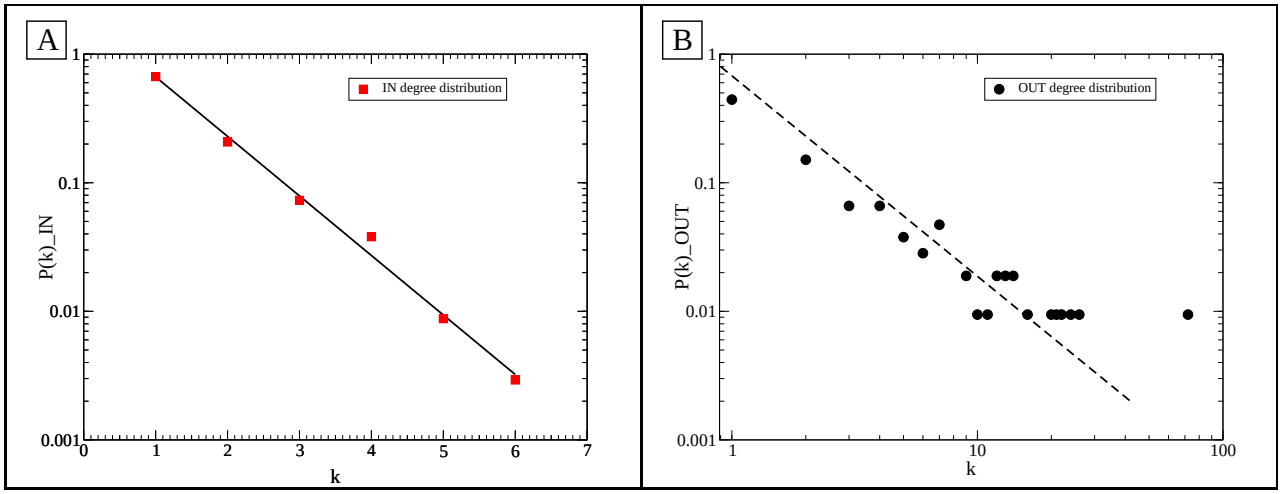
A
IN degree distribution
P(k)_IN
k
1
0.1
0.01
0.001
0
1
2
3
4
5
6
7
B
OUT degree distribution
P(k)_OUT
k
1
0.1
0.01
0.001
1
10
100

Figure S3