microbial biotechnology

Open Access

Escherichia coli segments its controls on carbon-dependent gene expression into global and specific regulations

Qing Pan,[1,2] Zongjin Li,[1,†] Xian Ju,[1,†] Chaofan Hou,[1] Yunzhu Xiao,[1] Ruoping Shi,[1] Chunxiang Fu,[2] Antoine Danchin[3] and Conghui You[1,*]

[1]*Shenzhen Key Laboratory of Microbial Genetic Engineering, College of Life Sciences and Oceanology, Shenzhen University, Shenzhen, Guangdong, China.*

[2]*Shandong Provincial Key Laboratory of Energy Genetics, Key Laboratory of Biofuels, Qingdao Engineering Research Center of Biomass Resources and Environment, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, China.*

[3]*Kodikos Labs/Stellate Therapeutics, Institut Cochin, 24 rue du Faubourg Saint-Jacques, Paris, 75014, France.*

Summary

How bacteria adjust gene expression to cope with variable environments remains open to question. Here, we investigated the way global gene expression changes in *E. coli* correlated with the metabolism of seven carbon substrates chosen to trigger a large panel of metabolic pathways. Coarse-grained analysis of gene co-expression identified a novel regulation pattern: we established that the gene expression trend following immediately the reduction of growth rate (GR) was correlated to its initial expression level. Subsequent fine-grained analysis of co-expression demonstrated that the Crp regulator, coupled with a change in GR, governed the response of most GR-dependent genes. By contrast, the Cra, Mlc and Fur regulators governed the expression of genes responding to non-glycolytic substrates, glycolytic substrates or phosphotransferase system transported sugars following an idiosyncratic way. This work allowed us to expand additional genes in the panel of gene complement regulated by each regulator and to elucidate the regulatory functions of each regulator comprehensively. Interestingly, the bulk of genes controlled by Cra and Mlc were, respectively, co-regulated by Crp- or GR-related effect and our quantitative analysis showed that each factor took turns to work as the primary one or contributed equally depending on the conditions.

Introduction

Working as an important *chassis* in industry, the Enterobacterium *Escherichia coli* constantly meets variable environments, including frequent nutrient depletion. Extensive expression profiling studies carried out for various types of nutrient-limiting conditions allowed identification of primary regulators involved in the corresponding adaptation processes (Oh, *et al*., 2002; Gosset, *et al*., 2004; Hua, *et al*., 2004; Gyaneshwar, *et al*., 2005; Liu, *et al*., 2005; Kang, *et al*., 2013; Folsom, *et al*., 2014). The target genes of many of these regulators have been further characterized by additional methods, such as the *in vitro* SELEX system (Shimada, *et al*., 2011a) or the *in vivo* assay of ChIP-seq (Seo, *et al*., 2014). Despite these efforts, the study of adaptation of the bacteria to their environments remained incomplete. Available work could only provide limited knowledge because the gene expression alterations observed in such conditions could rarely be explicitly ascribed to the role of specific regulators (Liu, *et al*., 2005; Kochanowski, *et al*., 2017). It is now critical to obtain more in-depth experimental details on the regulatory network that *E. coli* puts into play to cope with environmental changes.

It has long been known that, when the carbon source is scarce, the cyclic AMP (cAMP) receptor protein Crp acts as the leading regulator. It activates the expression of many carbon catabolic genes to balance the limitations imposed on carbon metabolism (Kolb, *et al*., 1993; You, *et al*., 2013). However, this is far from the whole story. Several global regulators are also differentially expressed during carbon limitation, blurring the picture (Oh, *et al*., 2002; Liu, *et al*., 2005). Moreover, when the carbon influx flow rate slows down, the cell's growth rate (GR) conforms, making elucidation of the details of the regulatory network problematic as ribosome-dependent

Received 23 October, 2020; revised 5 February, 2021; accepted 5 February, 2021.
*For correspondence. E-mail cyou@szu.edu.cn; Tel. +86-755-8657 3250; Fax +86-755-26534274.
† Z.L. and X.J. contributed equally to this work.
Microbial Biotechnology (2021) **14**(3), 1084–1106
doi:10.1111/1751-7915.13776

protein synthesis is highly sensitive to GR (Hernandez and Bremer, 1990; Scott, *et al*., 2010; Lemke, *et al*., 2011; Danchin, *et al*., 2020). Obviously, this must have important consequences on gene expression (Shahrezaei and Marguerat, 2015; Thomas, *et al*., 2018). The relative contribution of these processes to regulation remained obscure until our recent work (Li, *et al*., 2019) that gave a systematic overview of how genes were controlled globally when glucose was limiting. We discovered that, to cope with glucose scarcity, *E. coli* superimposed adaptative responses, including Crp-mediated control of gene expression and GR reduction. This finding could hardly be used as the norm, however, as it was based on the relatively simple model condition of glucose limitation. This is quite different from the situation which prevails in most ecological niches (Bertin, *et al*., 2013; Jimenez, *et al*., 2020) where various types of carbon substrates are required. For example, the metabolism of individual carbon substrates showed profound effects on the colonization or fitness of *E. coli* (Chang, *et al*., 2004; Alteri, *et al*., 2015). These observations make it critical to study the regulatory network when *E. coli* faces a broader range of carbon substrates.

Here, we studied the global regulation strategy evoked in *E. coli* when the bacteria grow in batch culture using a series of seven carbon substrates chosen among a panel of individual carbon sources expected to trigger expression of widely different pathways appearing in the metabolic network from top to bottom. The metabolism of these seven carbon substrates (Fig. 1A) shows distinct features in: (i) the final outcome of growth as indicated by the changes of the GR, varying from 1.0 h^{-1} to 0.5 h^{-1} (Fig. S1); (ii) the points where they enter the metabolic network with the glycolytic substrates [glucose (Glc), lactose (Lac), glucose-6-phosphate (G6P), sorbitol (Sor) and glycerol (Gly)] entering the glycolysis and with the non-glycolytic substrates [acetate (Ace) and succinate (Suc)] entering the TCA cycle; (iii) the ways they are transported *via* PTS (phosphotransferase system) (Deutscher, 2008) (Glc and Sor) or non-PTS (G6P, Lac, Gly, Suc and Ace). We developed a systematic analysis of the seven gene expression profiles according to their pattern of co-expression at both coarse- and fine-grained levels (Fig. 1B). This allowed us to discover that genes could be clustered into consistent classes responding to the features just outlined. First, the coarse-grained analysis identified a novel rule of regulation based on the striking correlation between the expression of relevant genes and the GR. Subsequently, fine-grained analysis identified the regulation strategy that an individual gene exercised, based on its distinct expression pattern. Crp-mediated and GR-related effects governed the expression of the bulk of genes responding to GR, making them the actual agents that drive the expression trend of the relevant gene, whereas Cra, Mlc and Fur individually governed the expression of genes responding to non-glycolytic substrates (Suc and Ace), glycolytic substrates (Glc and Lac) and PTS sugars (Sor and Glc). In this process, we identified many of the genes likely regulated by Crp, Cra, Mlc or Fur, based on the observation that they co-expressed with those previously reported as belonging to relevant regulons. Here, their expression was further studied *in vitro* and *in silico* in order to obtain an explicit identification of all genes submitted to the direct regulation of each regulator. This allowed us to expand the list of the members of each regulon controlled by each one of the four regulators. We are now in a position to propose a comprehensive clarification of their regulatory functions. Finally, the bulk of the genes under specific regulation mediated by Cra and Mlc were, respectively, identified to be co-regulated by Crp or GR. Through this way, we quantified the individual contribution of each factor to the coordinated regulations and showed how they competed with each other in various conditions.

Results

The expression level of a gene correlates with its GR-dependent expression trend

The transcriptomes of a wild-type *E. coli* K-12 strain grown in each of the seven carbon substrates as a sole carbon source were studied. As a first step, we compared every pair of the seven transcriptomes to identify the expressed genes (DEGs) that displayed differing expression. A total of 1522 CDSs (protein-coding sequences) genes and 35 ncRNA (non-coding RNA) accounting for one third of the genome were retained as significant among the seven conditions (Data S1). Given that ~ 30% of the DEGs were with unknown function (*y*-genes), we manually updated the annotations of all the corresponding genes based on published literature in order to gain better insight in the cell's response to these carbon substrates. This allowed us to identify several novel carbohydrate catabolism operons showing increased expression in carbon substrates with low growth rate, such as genes required for sulfoquinovose [*squUTS* (*yihUTS*)] or kojibiose [*koj* (*ycj*) operon] degradation. To understand the processes that control gene expression under these conditions, we performed a coarse-grained co-expression analysis of all the DEGs of the retained CDSs. As shown in the heatmap (Fig. 2A), these DEGs of CDS could be classified into two groups at a coarse-grained level, with genes showing negative correlation with GR in GR-negative group which showed increased expression when GR reduced and with genes showing positive correlation with GR in GR-positive group which showed decreased expression when GR

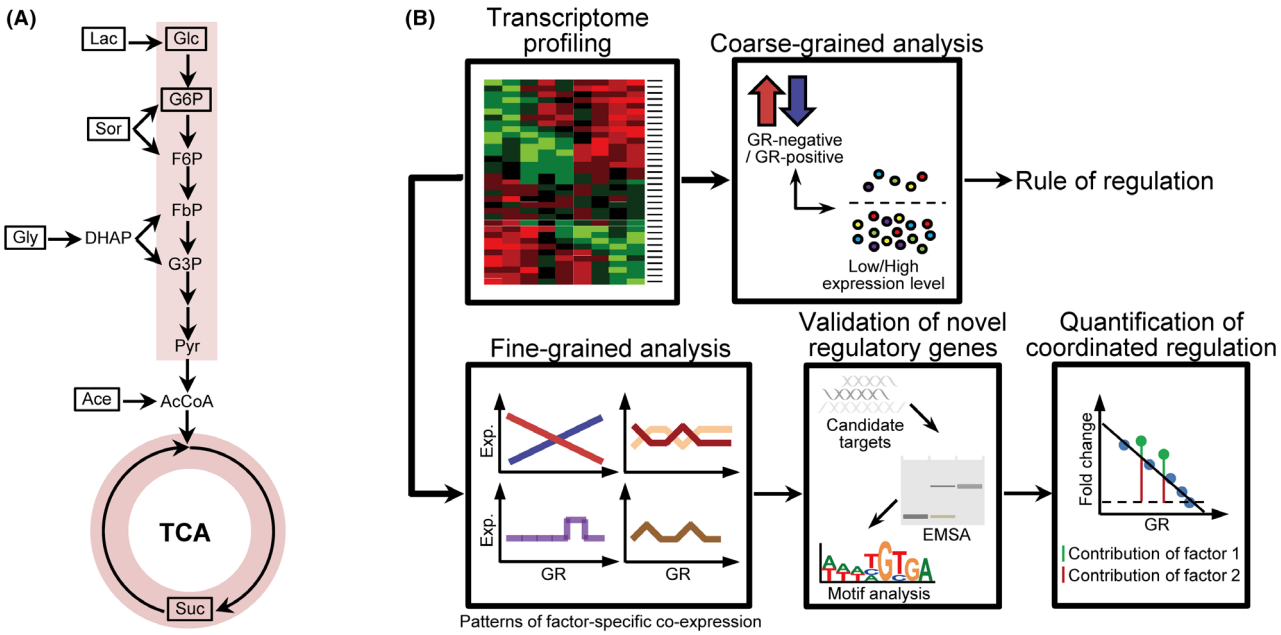

Fig. 1. Scheme of the workflow for uncovering regulatory strategies of *E. coli* grown in carbon substrates with distinct features.
A. Entries of the seven carbon substrates into the central carbon metabolism pathways.
B. Scheme of the systematic workflow. The workflow consists of four main procedures: (i) high resolution RNA-seq for transcriptome profiling; (ii) systematic analysis of the co-expression of the differentially expressed genes identified in the seven expression profiles in the coarse-grained or fine-grained scale, respectively; (iii) validation of novel genes regulated by key transcription factors; (iv) quantification of the individual contribution of each factor to genes under their coordinated regulations.

reduced. There are more genes in the GR-negative group ($n = 1113$) than that in the GR-positive group ($j = 409$). In each group, the average fold change of all the DEGs (EA representing GR-negative genes and EB representing GR-positive genes) in each carbon substrate was plotted against the corresponding GR (λ). The symbol of λ was used to represent GR with a unit in h^{-1}. A linear relation was observed for each group (Fig. 2B, 2C), indicating that these genes responded to the GR in general. An average tenfold increased expression and an average 2.3-fold decreased expression were, respectively, observed for the genes in the GR-negative group and GR-positive group when GR decreased from 1.0 to 0.5 h^{-1}. Thus, the bulk of the genes showed increased expression and a small subset showed decreased expression when the GR was reduced. Moreover, genes showing increased expression displayed a larger fold change of expression as compared to genes showing decreased expression. What would *E. coli* try to tell us by these intriguing facts?

To explore this subject matter, we turned to a proteome partition model (You, *et al.*, 2013; Basan, *et al.*, 2015; Hermsen, *et al.*, 2015). Given the tight coupling of transcription and translation in *E. coli* (Taniguchi, *et al.*, 2010; Elgamal *et al.*, 2016), the average protein mass of GR-negative (A) and GR-positive (B) genes at various growth conditions could be described by linear relations

$$A = (K\lambda + a)M \quad (1)$$
$$B = (L\lambda + b)N \quad (2)$$

with parameters K and L shown in Figure 2B, C, and with M and N, respectively, denoting the average protein mass of GR-negative and GR-positive genes when cells were grown in G6P. In line with the fact that the mass of the total proteome was constant under various GR (You, *et al.*, 2013; Hermsen, *et al.*, 2015), the total increased amount of GR-negative proteins when GR decreased from λ_1 to λ_2 should be equal to the total decreased amount of GR-positive proteins (Fig. 2D). Under this premise, we got the value of M versus N (M/N), so that the ratio of A to B (A/B) can be described by an equation with only GR (λ) as a variable and now be plotted against GR (λ) (Fig. 2E, Supplementary Note). As shown in Fig. 2E, the ratio of A to B is always smaller than 1 in carbon substrates with high GR, indicating that when cells were grown in carbon substrate leading to high GR, the GR-negative genes had a lower expression level when compared to the GR-positive genes. Indeed, the expression level of GR-negative genes was significantly lower than that of GR-positive genes in substrates giving high GR (Fig. 2F). As a result, the expression level of a gene determined its GR-dependent expression trend: a gene with a high expression level tends to show

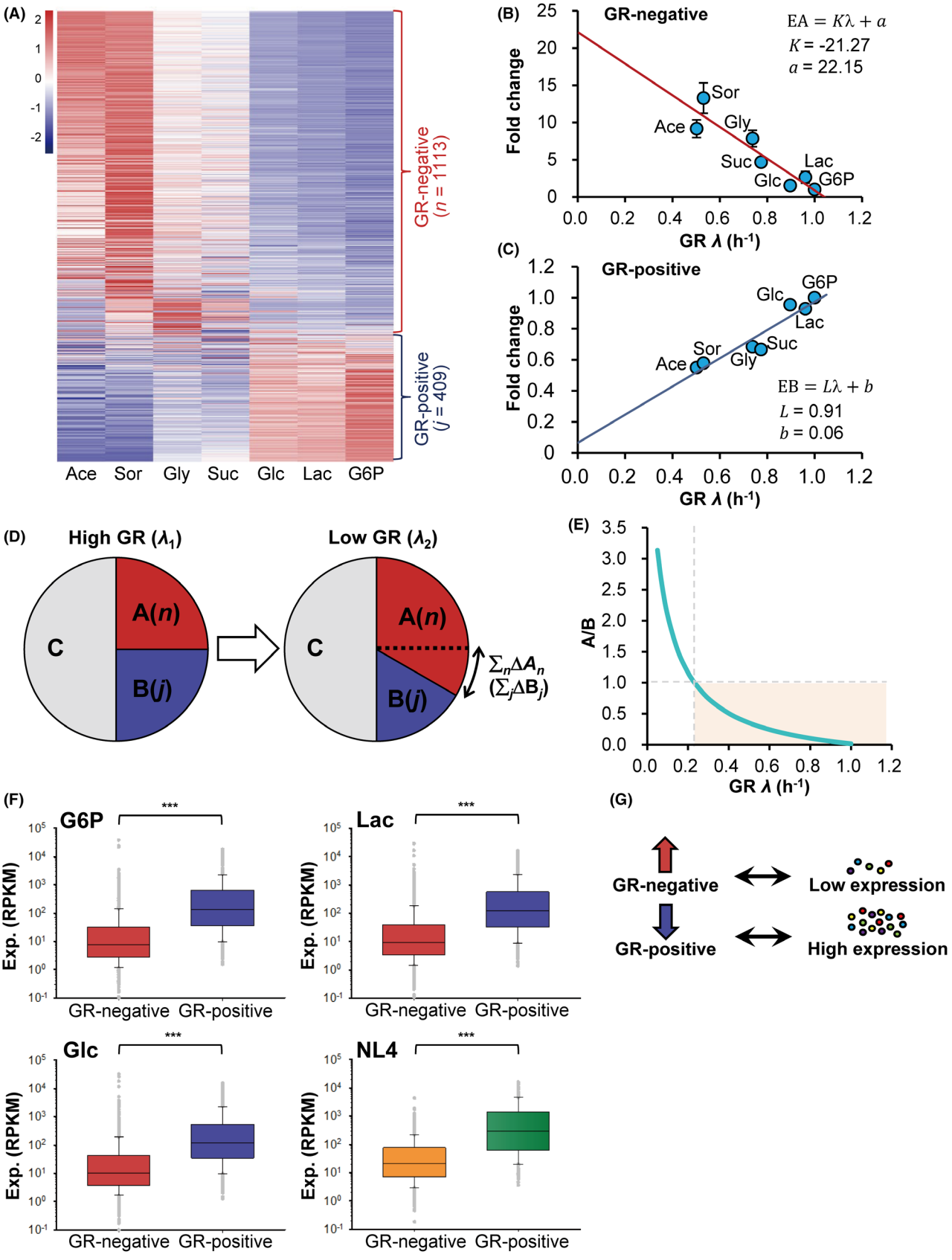

Fig. 2. Correlation between gene expression level and its GR-dependent expression trend.
A. Heatmap of expression of 1522 up- and down-regulated DEGs of CDS in seven carbon substrates. Ordered transcriptional levels (RPKM) of selected genes were visualized by the 'Heatmap' program, normalized using z-score at row. See Materials and Methods.
B, C. Correlation of the change fold of GR-negative (panel B) and GR-positive (panel C) genes with GR. Gene expression in G6P was normalized to 1 and that in other carbon substrates was each determined relative to this value to obtain fold change. In each carbon substrate, the average fold change of the 1113 GR-negative genes of CDS (panel B) or 409 GR-positive genes of CDS (panel C) was calculated and expressed as average ± SEM. The linear lines and equations with slopes (*K* and *L*) and y-axis intercepts (*a* and *b*) described the best fits of the data.
D. Illustration of the proteome partition model. When GR reduced, the protein mass of a GR-negative protein A increased, whereas the protein mass of a GR-positive protein B decreased, with *n* and *j* representing the numbers of proteins. Proteins that remained constant were indicated by C. Basing on the model, the total increased amount of GR-negative proteins ($\Sigma_n \Delta A_n$) is equal to the total decreased amount of GR-positive proteins ($\Sigma_j \Delta B_j$); see Note S1.
E. Ratio of the average protein mass of GR-negative genes A to that of GR-positive genes B at various GRs. The shaded region indicates where A/B < 1; see equation (II) in Note S1.
F. Expression levels of the GR-negative and GR-positive genes in G6P, Lac, Glc and NL4 [a nitrogen rich condition with GR at 0.9 h^{-1} (Li, *et al*., 2019)]. '***' indicates *P*-value < 0.001 by Student's *t*-test.
G. Diagram of the correlation between gene expression level and its GR-dependent expression trend: a highly expressed gene tends to show GR-positive expression or lower its expression, and a lowly expressed gene tends to show GR-negative expression or enhance its expression.

GR-positive expression or lower its expression, and a gene with a low expression level tends to show GR-negative expression or enhance its expression when GR reduced (Fig. 2G). Supporting this conclusion, a similar phenomenon was observed when *E. coli* was grown in nitrogen-limited conditions (Fig. 2F, Fig. S2). Thus, the expression level correlated to the changing trend of gene expression could be a novel rule of regulation.

Crp- and GR-related effect control the GR-dependent expression trend of DEGs

We now know that the trend in gene expression that followed the reduction of GR in various carbon substrates was correlated to the initial gene expression level. This allowed us to group together 1101 genes, accounting for 70% of all the DEGs, as responding to the GR of substrates by showing a significant correlation with GR (*P*-value <0.05) (Data S2). How were they regulated actually to ensure this expression trend? To answer this question, we next identified the actual actor of the regulation in an individual gene level.

On one hand, as the GR dependence of Crp-regulated genes was well established by the response of genes to carbon limitation as exampled by several catabolic genes and by a hundred of proteins identified by the proteomics study which studied ~ 1000 proteins (You, *et al*., 2013; Hui, *et al*., 2015), Crp could be one of the actors. On the other hand, the GR-related effect was recently identified to be another key factor controlling the regulation of many others genes in addition to the well-known ribosomal genes which responded to GR during carbon limitation (Li, *et al*., 2019), and the GR-related effect could govern the expression of some of the genes identified here that responding to GR. However, the relative contribution of Crp- and the GR-related effect to the regulation of genes responding to GR remained obscure.

This made it necessary to explore the contribution of Crp- and the GR-related effect to the whole genome at the individual gene level. We therefore identified genes regulated by Crp or sensitive to GR by comparing their expression in the wild type to that in a *crp* deletion strain, using a method reported previously (Li, *et al*., 2019) (Data S2). The key feature of this approach was to integrate the knowledge of the change trends of gene expression that paralleled GR together with the principles now described and summarized in Fig. 3A. In brief, the logic of the analysis is that the expression of a gene regulated by Crp in the *crp* deletion background, visualized in a plot of expression versus GR, will be far off its trend of expression changes following GR obtained when the wild-type strain is grown in various carbon substrates. By contrast, for the gene that is controlled by a GR-related effect, these data would fit well with its trend of expression changes following GR in the same conditions. This allowed us to applied specific selection criteria to identify these genes.

First, for a gene showing significant negative correlation with GR, if its expression in G6P (Exp_{G6P}) was higher than its expression in the *crp* deletion strain grown in glucose (Exp_{crp-}) – Exp_{crp-} < Exp_{G6P} – then this gene was deemed to be activated by Crp, retaining the evidence that G6P gave the highest GR among the seven substrates and cells grown in G6P contained the lowest level of cAMP resulting into the lowest amount of functional Crp (Epstein, *et al*., 1975; Deutscher, 2008; You, *et al*., 2013). As a consequence, we privileged the comparison between Exp_{G6P} and Exp_{crp-} over other comparisons. This allowed us to identify 83 genes in the cognate group. Yet, this criterion may be too strict, as genes with Exp_{crp-}, when equal to or similar to Exp_{G6P}, would be filtered out. Hence, we compared gene expression in the *crp* deletion strain to that of the wild type grown in the carbon substrate with similar GR. Given that the GR of wild-type cells grown in sorbitol (0.53 h^{-1}) was close to that of the *crp* deletion strain grown in glucose (0.41 h^{-1}) (Li, *et al*., 2019), the gene expression data in sorbitol (Exp_{Sor}) were chosen.

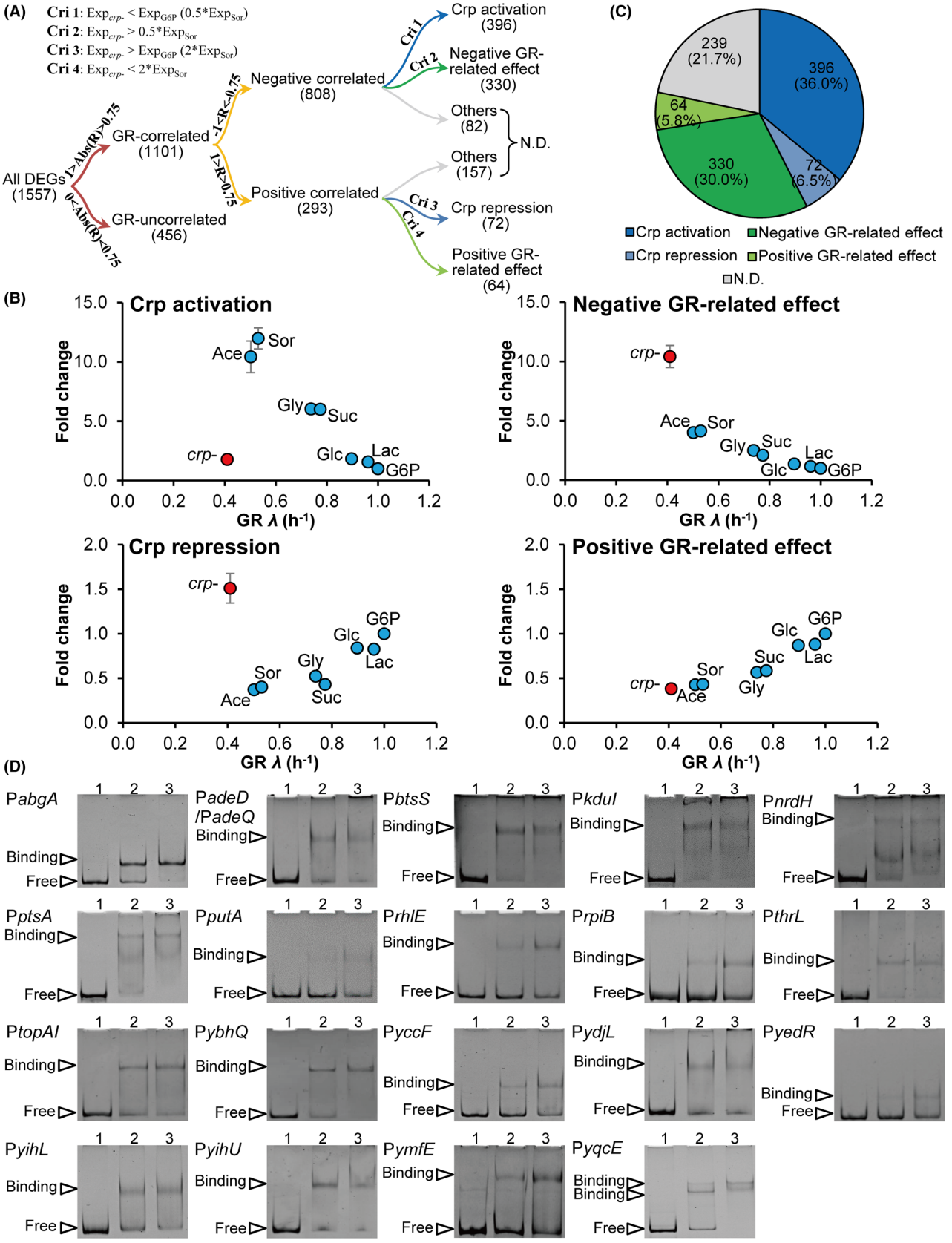

Fig. 3. Crp- and GR-related effect controlled genes and their characteristics.
A. A diagram outlining the 'decision tree' used to classify which genes are Crp- or GR-related effect-dependent. 'Abs' means absolute value. 'N.D.' means not determined. 'Cri' means criterion.
B. The fold change of genes controlled by Crp- or GR-related effect against GR. Gene expression was normalized as in Fig. 2B.
C. Distribution of 1101 genes responding to the GR categorized in each regulatory group. See Data S2.
D. Validation of Crp-cAMP binding promoters by EMSA. 0.25 pmol of each promoter fragment was incubated with 0, 3, 6 pmol Crp plus 1 mM cAMP from lane 1 to 3 as described in Materials and Methods. Free and binding probes were, respectively, indicated by triangles next to the image. At least two independent repeats were performed for each assay. See Data S4 and S5.

We expected that if any gene expression Exp_{Sor} was higher than its Exp_{crp-}, then this gene was activated by Crp. Using twofold change as a cut-off, the selection criterion was $Exp_{crp-} < 0.5*Exp_{Sor}$, and 393 genes were identified accordingly. In contrast, if Exp_{crp-} was higher than 0.5-fold of Exp_{Sor} – implying that Exp_{crp-} was not smaller than Exp_{Sor} – Crp was not expected to play a role. In this situation, a gene was identified as controlled by a negative GR-related effect among those that showed at least twofold of expression changes within the range of GR studied. Finally, along the same line of reasoning but in reverse, for the gene showing significant positive correlation with GR, if Exp_{crp-} was higher than Exp_{G6P} or twofold of Exp_{Sor}, this gene was deemed repressed by Crp, whereas if Exp_{crp-} was smaller than twofold of Exp_{Sor}, it was assumed that this gene was controlled by a positive GR-related effect among those that showed at least twofold of expression changes within the range of GR studied.

The average fold change of gene expression in the seven carbon substrates and in the *crp* deletion background was plotted against the corresponding GR (Fig. 3 B). As expected, in the case of the genes identified as regulated by Crp, their average fold change in the *crp* deletion background was far off their fold change trend following GR substantiating a Crp-dependent response. In contrast, for the genes identified as controlled by a GR-related effect, these data fitted very well with that of other conditions indicative of a Crp-independent but GR-related response. In summary, the Crp- and GR-related effects controlled the expression of ~80% of the genes that responded to GR (Fig. 3C), with approximately half controlled by Crp and the other half controlled by GR-related effect indicative of their equal regulatory role. As a large number of genes were controlled by Crp- or GR-related effect, both factors were regarded as exhibiting global regulations.

In total, 468 genes were identified to be activated (396 genes) or repressed (72 genes) by Crp directly or indirectly (Fig. 3C, Data S2). These Crp-controlled genes included the entire gene list in each of the transcriptional unites of the 33 polycistrons, as exemplified by all the genes in the *lsrACDBFG* operon and in the *ddpXABCDF* operon, suggesting the high quality of the global data that we obtained and the high sensitivity of the selection criteria we applied. Approximately half of the 468 genes belonged to the previously reported Crp regulon, either being documented in public databases (Gama-Castro, *et al*., 2016) or being identified by the *in vitro* ROMA assay (Zheng, *et al*., 2004) and the *in vitro* SELEX system (Shimada, *et al*., 2011a). However, many of the Crp regulon genes identified by the *in vitro* ROMA assay could not be identified by the *in vivo* transcriptional profiles in the same report (Zheng, *et al*., 2004) and most of the genes identified by the *in vitro* SELEX system had not been tested *in vivo* leaving the regulatory mode unrevealed (Shimada, *et al*., 2011a). Hence, our results provided the *in vivo* proof of the Crp-controlled regulation of these genes. We further revealed the regulatory mode (activation/repression) of Crp on the 22 Crp regulon genes as Crp binding sites were identified on these genes only by the *in vitro* SELEX system (Shimada, *et al*., 2011a) (Data S3). Moreover, we found that the reported regulatory mode of Crp on several of the genes (*gpmM*, *typA*, *grcA*, *serA*) was questionable. Contrary to be activated by Crp as reported (Yang, *et al*., 2002; Zheng, *et al*., 2004; Shimada, *et al*., 2011a; Choi and Hwang, 2018), all of them showed decreased expression when the GR was reduced and their expression level in the *crp* null background was far off the trend of their expression changes following GR in the *crp* wild-type background (Fig. S3) indicative of Crp repression in the conditions we studied.

Intriguingly, the present study established that the expression of 226 genes previously not known to be Crp-dependent was likely to be regulated by Crp (Data S4). To substantiate this hypothesis, we performed electrophoretic mobility shift assays (EMSA) of Crp for 52 genes (~1/4), including all the 34 genes belonging to 20 polycistrons, a random selection of 17 genes in monocistrons and 1 ncRNA. The promoters of 13 polycistrons and 7 monocistrons were indeed binding Crp (32 genes in total) (Fig. 3D, Table 1, Data S5), indicating that these transcription units were regulated by Crp directly. This makes ~ 60% of the promoters we assayed. The putative Crp binding site of each promoter was predicted *in silico* (Data S5). It matched the consensus recognition sequence of Crp ($TGTGAN_6TCACA$) (Busby and Ebright, 1999; Ishihama, 2010) (Fig. S4).

Based on the newly expanded Crp regulon, we could differentiate in an explicit way the physiological roles of the positive and negative regulation of Crp. All the Crp-

Table 1. Novel and validated genes in Crp, Cra, Mlc and Fur regulons.

Operon	Type	Regulation	Operon	Type	Regulation
Crp regulon			Cra regulon		
abgABT	CDS	Crp activation	*asnB*	CDS	Cra repression
adeD	CDS	Crp activation	*aspA-dcuA*	CDS	Cra activation
adeQ	CDS	Crp activation	*csrC*	ncRNA	Cra activation
btsSR	CDS	Crp activation	*dapD*	CDS	Cra repression
kduI	CDS	Crp activation	*dcuB-fumB*	CDS	Cra activation
nrdHIEF	CDS	Crp activation	*dhaKLM*	CDS	Cra repression
ptsA-fsaB-gldA	CDS	Crp activation	*dsrB*	CDS	Cra repression
putA	CDS	Crp activation	*dtpD*	CDS	Cra repression
rhlE	CDS	Crp repression	*gltBDF*	CDS	Cra repression
rpiB	CDS	Crp activation	*lsrACDBFG-tam*	CDS	Cra activation
thrLABC	CDS	Crp repression	*mdh*	CDS	Cra activation
topAI-yjhQ	CDS	Crp activation	*panD*	CDS	Cra repression
ybhQ	CDS	Crp activation	*rihA*	CDS	Cra activation
yccFS	CDS	Crp repression	*AA953_ RS06005*	CDS	Cra repression
ydjLKJIHG	CDS	Crp activation	*AA953_ RS09660*	CDS	Cra activation
yedRJ	CDS	Crp repression	*sstT-ygjV*	CDS	Cra activation
yihLM	CDS	Crp activation	*sthA*	CDS	Cra activation
yihUTS	CDS	Crp activation	*ucpA*	CDS	Cra activation
ymfED	CDS	Crp activation	*ychH*	CDS	Cra activation
yqcE-ygcE	CDS	Crp activation	*yejG*	CDS	Cra repression
			hadM (*yidA*)	CDS	Cra repression
Mlc regulon			Fur regulon		
manXYZ-yobD	CDS	Mlc repression	*shiA*	CDS	Fur repression

CDS, Coding sequence; ncRNA, non-coding RNA.

regulated genes identified here were further projected on the KEGG pathway map (Fig. S5). Genes activated by Crp were enriched in catabolism of carbon substrates to maintain sufficient carbon source catabolism when the GR is reduced during carbon limitation, as reported previously (Perlman, *et al*., 1969; Kolb, *et al*., 1993; Zheng, *et al*., 2004; Deutscher, 2008; You, *et al*., 2013; Franchini, *et al*., 2015; Li, *et al*., 2019), and genes repressed by Crp were enriched in biosynthesis of molecules of intermediary metabolism such as amino acids, a less known role for Crp, including the newly identified *thrLABC* operon which encode enzymes functioning in threonine biosynthesis and the *serA* which was identified to be repressed by Crp in this work and encodes an enzyme functioning in serine biosynthesis.

The genes regulated by Cra exhibit distinct responses to glycolytic and non-glycolytic carbon substrates

Besides the observation that these seven carbon substrates gave various GR, they could be classified into two groups, with five glycolytic substrates (Glc, Lac, G6P, Sor and Gly) in the glycolytic group and two non-glycolytic substrates (Ace and Suc) in the non-glycolytic group (Fig. 1A). As we have identified genes responding to the changes of GR, we next tried to identify genes that would respond differently to glycolytic and non-glycolytic substrates. Given GR varied among these carbon substrates (Fig. S1) and GR affected gene expression (Fig. 3B, C), we compared gene expression in glycolytic substrates versus non-glycolytic substrates that gave similar GR. As a result, gene expression in Ace (Exp_{Ace}) was compared to gene expression in Sor (Exp_{Sor}) as the GR in Sor was closest to that in Ace (Fig. S1). In parallel, gene expression in Suc (Exp_{Suc}) was compared to gene expression in Gly (Exp_{Gly}) and Glc (Exp_{Glc}) as these two glycolytic substrates gave most close GR to Suc with Glc giving a bit higher GR and Gly giving a bit smaller GR than Suc (Fig. S1). As a result, DEGs matched the three criteria simultaneously [$Exp_{Ace} > Exp_{Sor}$ (resp. $Exp_{Ace} < Exp_{Sor}$) (P-value < 0.05), $Exp_{Suc} > Exp_{Gly}$ (resp. $Exp_{Suc} < Exp_{Gly}$) (P-value < 0.05) and $Exp_{Suc} > Exp_{Glc}$ (resp. $Exp_{Suc} < Exp_{Glc}$) (P-value < 0.05)] were regarded as responding positively (resp. negatively) to the two non-glycolytic substrates. In total, 102 DEGs were identified as responding to acetate and succinate by showing expression peaks or valleys in the two non-glycolytic substrates (Data S6). Among those, 49 genes showed expression peaks in acetate and succinate (Fig. 4A), and the remaining 53 genes showed expression valleys in these conditions (Fig. 4B).

Given the important regulatory role of Cra in distinguishing glycolytic carbon sources from non-glycolytic (gluconeogenic) carbon sources (Kim, *et al*., 2018; Okano, *et al*., 2020), these 102 DEGs responding to acetate and succinate might be controlled by Cra. Indeed, 45

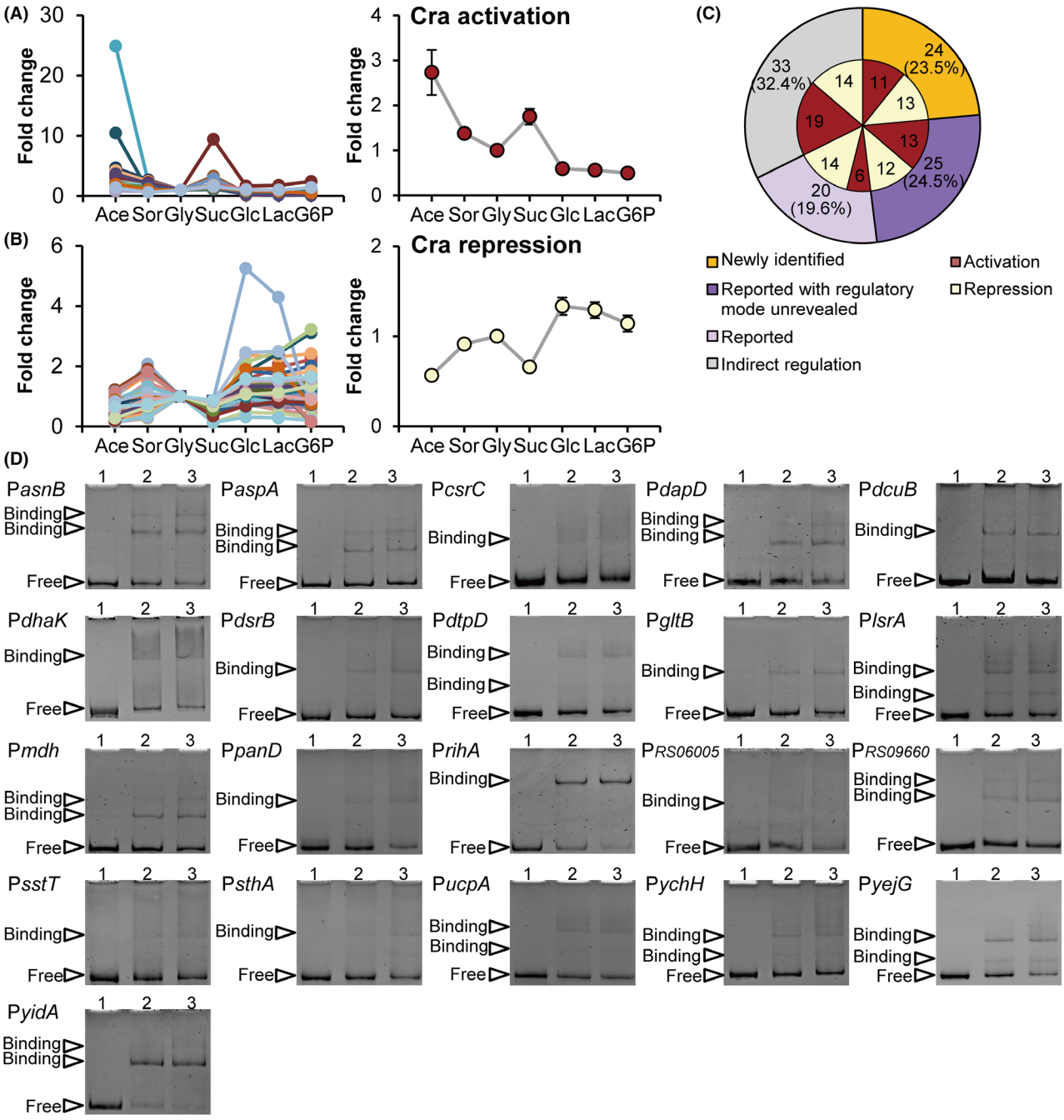

Fig. 4. Genes regulated by Cra and their characteristics.
A,B. The fold change of genes showing expression peaks (panel A) or valleys (panel B) in acetate (Ace) and succinate (Suc) (plot on the left) in various carbon substrates. In each panel, gene expression in glycerol was normalized to 1 and gene expression in other carbon substrates was determined relative to this value to obtain fold change. The average fold change of genes in each panel was calculated and expressed as average ± SEM (plot on the right).
C. Distribution of all 102 genes regulated by Cra into groups of newly identified, reported with regulatory mode unrevealed, reported and indirect regulation, as well as the classification of Cra regulatory strategy identified in this study. See Data S6.
D. Validation of Cra binding promoters by EMSA. 0.25 pmol of each promoter fragment was incubated with 0, 1, 2 pmol Cra from lane 1 to 3 as described in Materials and Methods. Free and binding probes were, respectively, indicated by triangles next to the image. At least two independent repeats were performed for each assay. See Data S6 and S7.

of these 102 genes belonged to the Cra regulon, being identified either by global methods as *in vitro* SELEX system and *in vivo* ChIP-exo technique, or by individual evidences documented in the public database (Shimada, *et al.*, 2011b; Gama-Castro, *et al.*, 2016; Keseler, *et al.*, 2017; Kim, *et al.*, 2018) (Fig. 4C, Data S6).

However, the regulatory mode (activation/repression) of Cra was known for a mere 20 genes, among the 45 genes identified here, with effects on the remaining genes undefined. This situation blurred the interpretation of the physiologic role of Cra regulation. Thus, we next explored the regulatory mode of Cra on the remaining 25 genes.

It has been discovered that the metabolite of fructose-1-phosphate (F1P) in the glycolysis pathway functioned as a Cra inhibitor (Chavarria, *et al*., 2014; Bley Folly, *et al*., 2018; Chavarria and de Lorenzo, 2018). The intracellular level of F1P was found to be the lowest in succinate and acetate when it was compared to that in the glycolytic substrates (Kochanowski, *et al*., 2017). This observation was further confirmed here by our analysis of the intracellular level of fructose-1,6-bisphosphate (FbP) (Fig. S6), given the level of FbP has been found to correlate well with that of F1P (Kochanowski, *et al*., 2017). Moreover, the expression of Cra showed no significant change in these carbon substrates (Fig. S7A, B). As a consequence, Cra-dependent genes appeared to be more activated when cells were grown in acetate or succinate. This can be interpreted as showing that the regulatory mode of Cra on genes exhibiting expression peaks in acetate and succinate (Fig. 4A) would be activation, whereas the regulatory mode of Cra on genes exhibiting expression valleys (Fig. 4B) would be repression.

In the case of the 20 genes with reported regulatory mode of Cra, our findings based on their response to acetate and succinate matched their reported mode well, indicating our method was robust (Data S6). Interestingly, most of the 20 genes were characterized previously by both the *in vitro* SELEX system (Shimada, *et al*., 2011b) and the *in vivo* ChIP-exo technique (Kim, *et al*., 2018) with one exception of *pck* encoding phosphoenolpyruvate carboxykinase of the gluconeogenesis pathway that was identified only by the *in vitro* method. Here, our study showed the evidence that *pck* was activated by Cra *in vivo*. Supporting the well-known regulatory interactions of Cra in the carbon metabolism, most of these genes activated by Cra functioned in the TCA cycle (*acnB*), the glyoxylate shunt pathway (*aceA*) and the gluconeogenesis pathway (*pck*, *fbp* and *ppsA*), but most of these genes repressed by Cra functioned in the glycolysis pathway (*gpmM*, *pykF*, *ppc* and *zwf*), the Entner–Doudoroff pathway (*edd*) and the transporting systems of sugars *via* the PTS (*fruAB* and *ptsHI*).

This allowed us to establish the regulatory mode of the 25 remaining genes with unknown regulation (Data S6). The majority of the binding sites of Cra to these genes were identified by the *in vitro* SELEX system without the *in vivo* proof (Shimada, *et al*., 2011b). With this *in vivo* data, we identified interesting genes repressed by Cra such as *thrLABC* encoding enzymes functioning in threonine biosynthesis, *pntAB* encoding pyridine nucleotide transhydrogenase enzymes, favouring reactions consuming NADH to generate NADPH (Sauer, *et al*., 2004), *ptsG* encoding a glucose-specific PTS enzyme, *yfbT* (*hxpA*) encoding a putative enzyme functioning in the carbon metabolism of glycolysis, and the two transcriptional factor genes of *appY* and *abrB*. Interesting genes activated by Cra included *gltA* and the operons of *sdh* and *suc* encoding enzymes functioning in the TCA cycle, *ppnN* (*ygdH*) encoding an enzyme functioning in carbon metabolism of AMP degradation, *lld* operon encoding lactate/glycolate:H+ symporter and the two transcriptional factor genes of *melR* and *mhpR*.

Moreover, as 45 of the 102 genes responding to acetate and succinate were Cra regulon genes (Fig. 4C), Cra might be a primary regulator, again possibly controlling the remaining gene set (directly or indirectly). This prompted us to perform EMSA experiments with Cra on the remaining 57 genes, not previously known to be regulated by Cra, in order to identify additional genes under direct regulation of Cra. We found that Cra binds to the promoter region of 24 of these genes (42% of the genes we assayed) (Fig. 4D, Data S6), including 21 transcription units (Table 1). The putative Cra binding site of each promoter was predicted *in silico* (Data S7). The Cra recognition sequence of these 24 genes (Fig. S8) matches the reported logo of GCTGAANCGNTTCA (Shimada, *et al*., 2011b; Kim, *et al*., 2018).

The newly identified genes repressed by Cra included amino acids biosynthetic genes [*asnB* (asparagine Asn biosynthesis), *gltBD* (glutamate biosynthesis), *dapD* (lysine biosynthesis) and *panD* (β-alanine biosynthesis)], the genes functioning in the glycolysis pathway [*hadM* (*yidA*) encoding a putative hexitol phosphatase] and in the glycerol degradation pathway (*dhaKLM* encoding enzymes functioning in glycerol degradation) with the generated metabolite of glycerone phosphate entering the glycolysis pathway, and a gene encoding an oligopeptide transporter (*dtpD*). The newly identified genes activated by Cra included two genes (*fumB* and *mdh*) encoding fumarate hydratase and malate dehydrogenase of the TCA cycle. We found Cra activated the expression of genes encoding enzymes functioning in catabolism of carbon substrates including aspartate (*aspA*), autoinducer-2 (*lsr* operon) and pyrimidine (*rihA*) with most of the generated metabolite entering the TCA cycle as exemplified by the aspartate degradation *via* AspA generating fumarate and autoinducer-2 degradation *via* Lsr enzymes generating acetyl-CoA. Moreover, Cra activated the expression of *sthA*. The gene of *sthA* encoding a pyridine nucleotide transhydrogenase enzyme favours the reaction consuming NADPH to generate NADH (Sauer, *et al*., 2004), which was a reversed

reaction catalyzed by PntAB with its regulatory mode being identified to be repressed by Cra in this work. The gene of *sstT* encoding sodium: serine/threonine symporter, the gene of *ucpA* encoding an oxidoreductase and one ncRNA gene (*csrC*) which modulates regulation of carbon storage (Weilbacher, *et al.*, 2003) were all identified to be activated by Cra.

The featured functions of the newly identified genes and genes with the regulatory mode determined in this work were now summarized (Fig. 5). Intriguingly, our study identified novel interactions of Cra regulation: Cra-repressed genes functioned in the amino acids biosynthesis, in the catabolic pathways of carbon substrates generating intermediates further metabolized by the

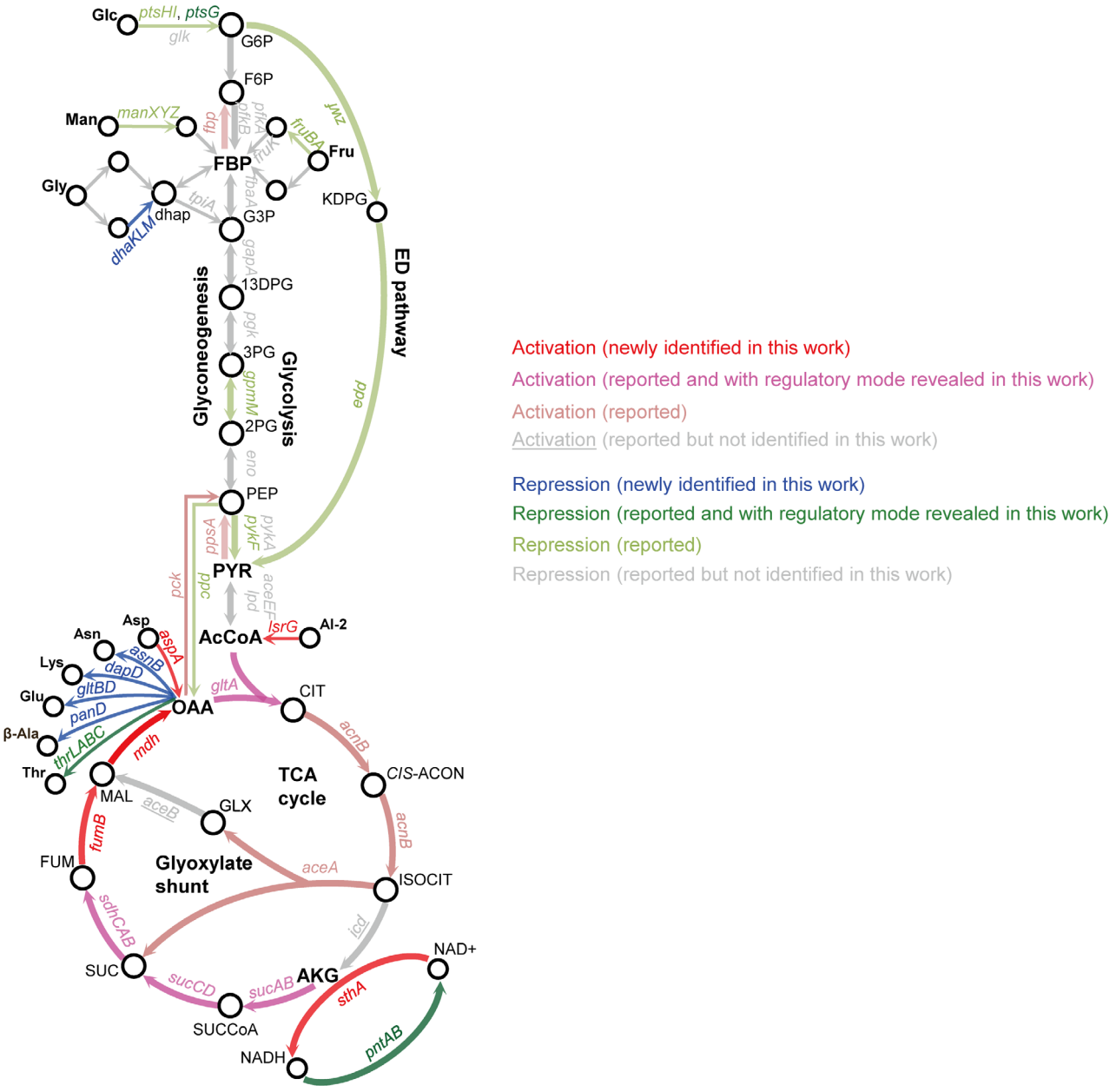

Fig. 5. The expanded regulatory interactions of Cra in *E. coli*, based on the function of 69 Cra directly regulatory genes. Cra-activated genes were enriched in the TCA cycle, glyoxylate shunt, gluconeogenesis, catabolism of carbon substrates which generate intermediate metabolites going into the TCA cycle and the reaction consuming NAD + to generate NADH; Cra-repressed genes were enriched in glycolysis, PTS, other degradation pathways generating intermediates that goes into the glycolysis pathway, Entner–Doudoroff (ED) pathway, amino acid biosynthesis and the reaction consuming NADH to generate NAD+. See Data S8.

glycolysis pathway and in the pathway of NADH consumption; Cra-activated genes functioned in the catabolic pathways of carbon substrates generating intermediate metabolites that go into the TCA cycle, and in the pathway of NADH generation. In parallel, our work complemented the reported interactions of Cra regulation in the carbon metabolism by our additional input of genes functioning in the glycolysis pathway, in the TCA cycle and in the PTS systems.

In summary, 69 out of these 102 genes responding to the two non-glycolytic substrates belonged to the Cra regulon (Data S8). With the newly identified interactions of Cra additional to its reported ones which were confirmed or complemented by our additional input of genes, we were able to delineate a fairly complete picture of the regulatory interactions of Cra in *E. coli* physiology (Fig. 5, Data S8). When cells are grown in non-glycolytic carbon substrates, Cra is more active and it activates the expression of enzymes in the glyoxylate shunt and gluconeogenesis to generate building blocks carrying 5-6 carbon atoms (Kornberg, 1966). At the same time, Cra represses the expression of enzymes in the glycolysis pathway, in the PTS systems and in other degradation pathways generating intermediates that goes into the glycolysis pathway, while activating the expression of enzymes of the TCA cycle, enzymes functioning in catabolism of carbon substrates which generate intermediate metabolites going into the TCA cycle, and enzymes catalyzing NADH generation to promote electron transfer *via* NADH to generate ATP. This is metabolically relevant because the TCA cycle is more efficient to generate ATP as compared to the glycolysis pathway (Basan, *et al*., 2015). Moreover, to ensure both generation of ATP and of building blocks, Cra represses the expression of the enzymes consuming intermediate metabolites of the TCA cycle to perform amino acid biosynthesis, and the enzymes catalyzing reactions consuming NADH. In summary, Cra mainly ensures the most efficient pathways to generate ATP and building blocks but weakens any other antagonistic pathways of metabolism.

Mlc regulon genes and Fur regulon genes display peaks of expression in specific glycolytic/PTS carbon substrates

As we have identified genes responding to GR and non-glycolytic substrates, we tried to identify genes that would respond to other features of these carbon substrates. The glycolytic substrates drew our attention. Among them, glucose and lactose give similar GR and the degradation of lactose generates glucose. Given these similarities, genes might respond specifically to glucose and lactose. Another pair of substrates, glucose and sorbitol, have also distinctive properties. Although they give a large GR difference, they are the only two sugars that are both transported *via* the PTS. As a consequence, genes might display a common feature responding to glucose and sorbitol as well. To explore this hypothesis, we applied distinct selection criteria (see Material and Method) to identify genes which would exhibit expression peaks in these conditions.

We identified that seven genes showed expression peaks in glucose and lactose, including *manXYZ*, *ptsG*, *ptsHI*, *yobD* (Fig. 6A, Data S9). Except for *yobD*, these genes were all reported to be repressed by Mlc (Schiefner, *et al*., 2005). Given the genomic location of *yobD* is adjacent to the *manXYZ* operon, we suspected that *yobD* could form an operon with *manXYZ* and was regulated by Mlc accordingly. Indeed, the 5′-RACE experiment demonstrated that *yobD* formed an operon with *manXYZ* (Fig. 6B, Fig. S9, Table 1). As a result, the predicted inner membrane protein YobD could take part in mannose catabolism.

Meanwhile, we identified 22 genes showing peaks of expression in the two PTS sugars (sorbitol and glucose) (Fig. 6C, Data S10). Interestingly, 18 of the 22 genes were repressed by Fur directly and functioned in iron metabolism (Seo, *et al*., 2014). We did EMSA of Fur on the remaining four genes, finding that Fur bound to the promoter region of one more gene of *shiA* (Table 1, Fig. 6D), and the remaining genes could be controlled by Fur indirectly due to the similar reasons as reported (Seo, *et al*., 2014). The putative Fur binding site of *shiA* was predicted as CCATACTAATTATAA by holo-Fur-repressed mode (Seo, *et al*., 2014). The newly identified *shiA* was also involved in iron metabolism as *shiA* encodes a high-affinity transporter of shikimate, a metabolic precursor of siderophore ferric enterobactin (Porcheron, *et al*., 2014).

The regulations of Cra and Mlc coordinate with regulations of Crp- or GR-related effect

To go beyond these interpretations, we looked into the possibility of coordinated regulations exhibited by the factors identified here. As shown previously, genes controlled by Cra exhibited a general increased (Fig. 4A) or reduced (Fig. 4B) changing trend of expression when the GR was reduced as a function of the carbon source, from G6P to Ace. This highlighted a GR-dependent response. Interestingly, genes controlled by Mlc also exhibited a similar response when GR was reduced (Fig. 6A). These observations are consistent with the simultaneous involvement of a coordinated regulation of a global behavior resulting from Crp- or GR-related effects in combination with their specific regulations *via* Cra or Mlc. To characterize these global regulations, we

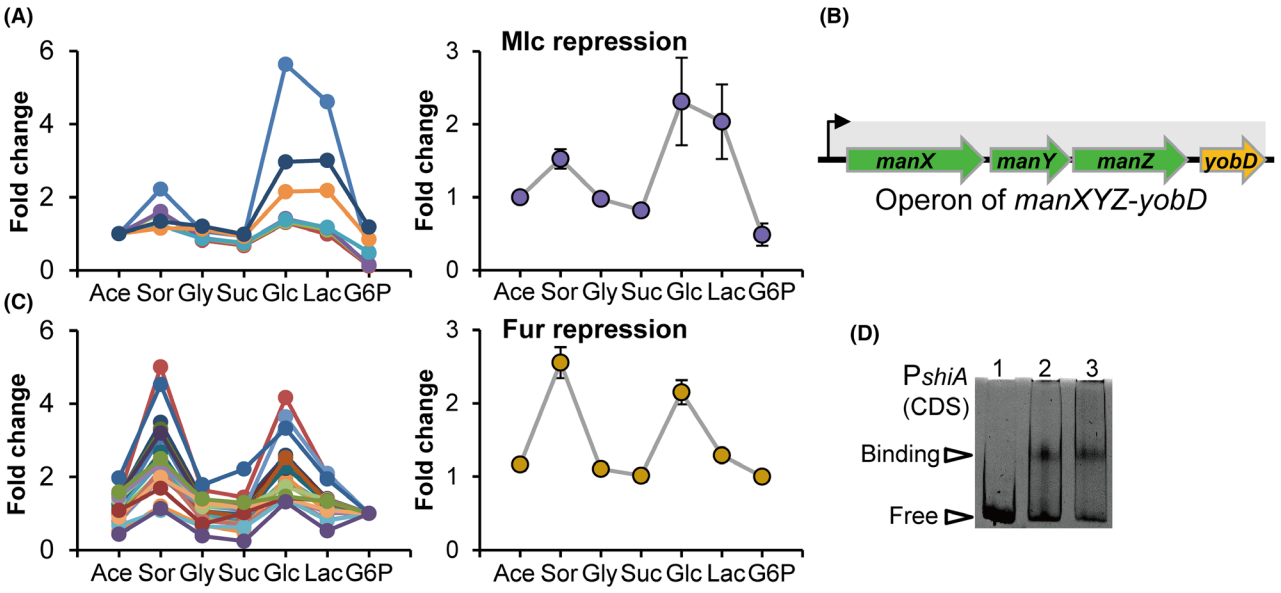

Fig. 6. Genes regulated by Mlc or Fur, and their characteristics.
A. The fold change of genes responding specifically to glucose and lactose (plot on the left) in various carbon substrates. Gene expression in acetate (Ace) was normalized to 1 and gene expression in other carbon substrates was determined relative to this value to obtain fold change. The average fold change of genes was calculated and expressed as average ± SEM (plot on the right). See Data S9.
B. Diagram of *manXYZ-yobD* operon. The black arrow indicates the TSS (Transcription start site) of the operon (in gray).
C. The fold change of genes responding specifically to sorbitol (Sor) and glucose (Glc) in various carbon substrates (plot on the left). Gene expression in G6P was normalized to 1 and gene expression in other carbon substrates was determined relative to this value to obtain fold change. The average fold change of genes was calculated and expressed as average ± SEM (plot on the right). See Data S10.
D. Validation of Fur binding promoters by EMSA. 0.25 pmol of each promoter fragment was incubated with 0, 20, 40 pmol Fur in the presence of 200 μM $CoCl_2$ from lane 1 to 3 as described in Materials and Methods. Free and binding probes were, respectively, indicated by triangles next to the image. At least two independent repeats were performed for each assay. See Data S10.

re-fitted each of their correlations of gene expression to GR by omitting the expression data involving succinate and acetate in the case of Cra regulated genes or by omitting the expression data involving glucose and lactose in the case of Mlc-regulated genes, as these data points showing expression peaks or valleys would affect the value R of the correlation. Thus, for those showing a significant correlation with GR (P-value < 0.05), we identified genes controlled by Crp- or GR-related effect by applying the method described in the previous section.

In total, 37 out of the 69 Cra regulon genes (Data S11) were identified to be co-regulated by Crp- or GR-related effect. For the genes activated by Cra, 21 genes (Fig. 7A, Data S11) were co-activated by Crp, with 19 genes under direct regulation of Crp and two genes under its indirect regulation, and four genes were co-regulated by the negative GR-related effect (Data S11). In the case of the genes repressed by Cra, nine genes were co-repressed by Crp directly (seven genes) or indirectly (two genes) (Fig. 7B, Data S11), and three genes were co-regulated by negative or positive GR-related effect (Data S11). In summary, the bulk of the genes under coordinated regulation were co-regulated by Crp and a small subset were co-regulated by GR-related effect.

Interestingly, most of the genes co-activated by Crp and Cra functioned in carbon metabolism, including the genes encoding enzymes of the TCA cycle (*acnB*, *gltA*, *suc-sdh* operon), the *pck* gene encoding an enzyme of the gluconeogenesis, and the two genes (*aspA* and *lsrG*) encoding enzymes functioning in catabolism of other carbon substrates. The genes co-repressed by Crp and Cra included amino acid biosynthetic genes, *thrABC* and *gltBD*, and the two genes, *gpmM* and *pykF*, encoding glycolysis enzymes. However, there was no documented evidence that *pykF* was repressed by Crp directly and we also failed to identify any Crp binding site in its promoter region by EMSA, indicative of indirect regulation of Crp. We noted that the co-activation of Crp and Cra on *acnB* was also identified in a previous report (Kim, *et al.*, 2018).

In the seven genes repressed by Mlc (Data S9), the four genes of the entire *manXYZ-yobD* operon were identified to be co-activated by Crp (Fig. 7C, Data S11), which fitted to the reported Crp regulation of this operon (Plumbridge, 1998). For the other three genes (*ptsHI* and *ptsG*) repressed by Mlc, although Crp binding sites were reported in their promoters (Shin, *et al.*, 2003; Zheng, *et al.*, 2004), *ptsHI* was filtered out because the expression of *ptsHI* showed less than twofold change in

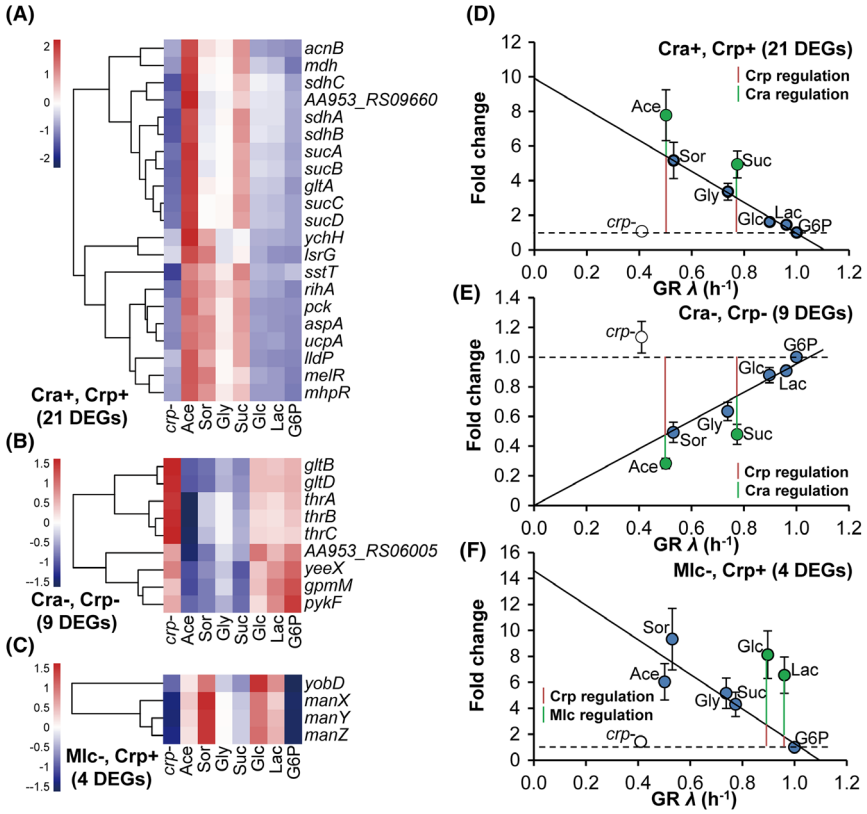

Fig. 7. Genes under coordinated regulations.
A–C. Heatmap of the expression of genes co-activated by Cra and Crp (panel A), co-repressed by Cra and Crp (panel B) and co-regulated by Mlc and Crp (panel C). See Data S11. Genes expression of the wild type grown in the seven carbon substrates and of the *crp*-null strain grown in glucose was shown. Data were normalized as in Fig. 2A and clustered.
D–F. Quantification of the contributions of Cra, Mlc and Crp to genes under their coordinated regulations. Gene expression in G6P was normalized to 1, and that in other conditions was each determined relative to this value to obtain fold change. The average fold change of the 22 genes co-activated by Cra and Crp in panel A, the 9 genes co-repressed by Cra and Crp in panel B and the four genes co-regulated by Mlc and Crp was individually calculated, expressed as average ± SEM and was plotted against GR in each growth condition. The solid black and linear lines described the best fits of the data of Sor, Gly, Glc, Lac and G6P (panel D and panel E) and the data of Ace, Sor, Gly, Suc and G6P (panel F). The dashed lines indicated the average fold change in G6P. The individual contributions of Cra and Crp in Suc or in Ace to the increased expression (panel D) or to the decreased expression (panel E) when compared to that in G6P were indicated by the lengths of green and red lines. The individual contributions of the Mlc and Crp in Glc or in Lac to the increased expression when compared to that in G6P was indicated by the lengths of green and red lines (panel F).

the five carbon substrate when omitting the data points in glucose and lactose indicative of GR-independent response in the conditions we studied, and *ptsG* was filtered out due to the observation that its expression showed non-significant correlation with GR (P-value > 0.05) among the five carbon substrates. Thus, our selection criteria of Crp-controlled regulation were stringent.

We quantified the contribution of each factor to the coordinated regulations. The average fold change of these genes under coordinated regulations of Cra and Crp or coordinated regulations of Mlc and Crp in each condition was plotted against the corresponding GR (Fig. 7D-F). In the case of genes under coordinated regulation of Cra and Crp, the expression trend of Crp-controlled gene expression would follow the linear line fitted by the expression data of the five carbon substrates (Sor, Gly, Glc, Lac, G6P) against GR. So that in Suc (resp. Ace), the contribution of Crp-controlled expression could be calculated. Accordingly, the contribution of Cra-controlled expression could be obtained by simply subtracting the contribution of Crp-controlled expression from the total measured expression in Suc (resp. Ace). Similarly, in the case of genes under coordinated regulation of Mlc and Crp, the contribution of Crp-controlled expression in Glc (resp. Lac) could be calculated basing on the linear line fitted by the expression data of the five carbon substrates (Ace, Sor, Gly, Suc, G6P) against GR. The contribution of Mlc-controlled expression could be obtained by subtracting the contribution of Crp-controlled expression from the total measured expression in Glc (resp. Lac).

As shown in Fig. 7D-E, in both cases of co-activation and co-repression controlled by Cra and Crp, when comparing gene expression among the three conditions (Suc, Ace and G6P), Cra took similar contributions in succinate and acetate, whereas the contribution of Crp enhanced in acetate when GR was further reduced. Thus, in succinate, Crp and Cra took almost equal contributions to the regulation of these genes, but in acetate, the contribution of Crp took over that of Cra. In contrast, for the genes under coordinated regulation of Mlc and Crp (Fig. 7F), due to the mild changes of GR, the contribution of Crp increased in glucose when compared to that in lactose, but the contribution of Mlc was still much bigger than that of Crp in both conditions indicative of the overriding regulatory activity of Mlc over Crp in these conditions.

Similarly, we also quantified the contribution Cra- and GR-related effect to the seven genes under their coordinated regulations (Fig. S10). Consistent with what we observed in the situation of genes under coordinated regulation of Crp and Cra, the contribution of global regulation of GR-related effect increased when GR reduced and it took over that of Cra in acetate. The only exception was the gene of *aceA* in the *ace* operon, which encoded enzymes in the glyoxylate shunt, showing an extremely high expression in acetate. It could be that the *ace* operon is highly sensitive to Cra in acetate. However, this seems less reasonable given most of Cra-regulated genes showed similar response to Cra in acetate and succinate based on the quantified contributions of Cra in the two conditions (Fig. 7D, E). Thus, the most reasonable explanation would be that a third factor overrode the regulation of *ace* operon in acetate. Interestingly, the highly accumulated stimulator of *ace* operon, the glyoxylate generated *via* acetate metabolism (Lorca, *et al*., 2007), could take the overriding role. We noted that the *ace* operon was identified recently to be under antagonist regulation of Crp and Cra (Kim, *et al*., 2018). This finding was based on the documented literature that *ace* was repressed by Crp. However, in fact, Crp was reported to repress the *ace* operon only in the Fur deleted background (Zhang, *et al*., 2005), different to the Fur wild-type background studied here. We found that, except for Ace, it was the negative GR-related effect but not the Crp that took the primary role based on its trend of expression changes in the conditions we studied (Fig. S10A). This evidence can be understood as implying that *E. coli* applied various regulation strategies to tightly control the expression of *ace* operon in different background.

Discussion

Understanding how bacteria control gene expression to cope with environmental changes remains an open challenge. This challenge is essentially limited by the lack of large-scale collection of data linked to the physiology of bacteria. Here, we addressed this limitation by collecting data on a significant panel of carbon sources that are metabolized by *E. coli via* various pathways. This allowed us to discover novel global regulation strategies that involved four global regulators, Crp, Cra, Mlc and Fur, in combination with GR-related effect, and to reveal a novel rule of regulation.

At a coarse-grained level, all the DEGs identified in this work were regarded as responding either positively or negatively to the reduction of GR (Fig. 2A–C). By analyzing the expression changes of genes in this two groups, we found that the regulation of gene expression was constrained by the gene's initial expression level basing on a simple model of re-allocation of proteomic resource (You, *et al*., 2013; Hermsen, *et al*., 2015) (Fig. 2D). This is also true when *E. coli* was grown in nitrogen-limited conditions (Fig. 2F). What makes the trend of gene expression correlating with its initial expression level? A comprehensive explanation would be that any expression change is constrained by the fact that the mass of the total proteome remains constant, no matter how much the GR reduces (You, *et al*., 2013; Hermsen, *et al*., 2015) (Fig. 2D). We observed more GR-negative genes showing larger expression changes as compared to GR-positive genes when GR was reduced (Fig. 2A–C). This implies that the GR-negative genes are further expressed at a lower level while those GR-positive genes are expressed at a higher level, maintaining the total proteome constant, which we described mathematically (Figure 2E). Thus, we could predict, just knowing its expression level, whether a gene will be up-regulated or down-regulated when there is a change in the environment (Fig. 2G). The actual expression level could contain some code of regulation. Unfortunately, many studies of the global behavior of gene expression (Liu, *et al*., 2005; Brauer, *et al*., 2008), whether in prokaryotes or in eukaryotes, reported various trends of changes in gene expression of multiple genes, but they did not focus on their actual expression level. This makes it interesting to explore whether the rule of regulation we have discovered exists in other organisms.

Performing fine analysis of co-expression, we further identified that Crp- and GR-related effects were the primary factors that contributed equally and drove the expression trend correlated to their initial expression levels by controlling the expression of the bulk of these genes, mediating global regulation. By contrast, Cra-, Mlc- and Fur-mediated specific regulations by individually controlling genes responding specifically to a particular carbon substrate. These findings were based on the observations that genes controlled by a certain factor showed a factor-specific pattern of co-expression.

Accordingly, new gene partners of many regulons were identified (Table 1) and the presumably complete regulon of each of these four regulators was reconstructed. As a result, our approach of the fine analysis of co-expression shows the advantage in discovering regulon of regulators comparable to the widely-used ChIP-seq technique (Rhee and Pugh, 2012), which could be used to discover similar regulations in other organisms.

Here we found that Crp- and GR-related effects contribute almost equally to the regulation of the expression of genes responding to the GR (Fig. 3C). On one hand, in spite of 50 years of extensive study on Crp, with the expanded Crp regulon, our work identified new features of Crp regulation. We found that the primary role of Crp regulation is to coordinate carbon catabolism with biosynthesis. This new feature of Crp regulation might be used for further understanding of unknown biological functions. As an example, the *squUTS* (*yihUTS*) operon involved in sulfoquinovose degradation was activated by Crp when the GR was reduced, suggesting that the 'unusual' sulfoquinovose sugar (Li *et al*., 2020) could actually be one of the most widespread carbon sources in the niche of *E. coli* (Haange, *et al*., 2020). This is indeed consistent with the widespread presence of sulfoquinovose in the environment, which is a major plant-derived metabolite (Benning, 1998) and comprises a major portion of the organo-sulphur in nature (Denger, *et al*., 2014). To our knowledge, the induction of the *squ* operons during carbon limitation has never been reported by other global assays (Zheng, *et al*., 2004; Hui, *et al*., 2015). In the absence of the inducer sulfoquinovose (Denger, *et al*., 2014), these genes are expressed at a very low level, e.g. from ~ 1 to several RPKM in the conditions of our study (Data S2). In parallel, other well-known Crp regulon genes which are expressed at a low level in the absence of inducers in the cultures and failed to be detected previously (Zheng, *et al*., 2004), e.g. the *als* operon encoding enzymes transporting the sugar of allose, were also identified here (Data S2). It seems therefore reasonable to assume that the very low expression level of these genes was the major reason that made them undetectable in previous works. These observations indicate that the method that we applied has a high resolution. On the other hand, the GR-controlled and Crp-independent response to GR (Fig. 3B), is in line with the evidence of a GR effect on gene expression (Shahrezaei and Marguerat, 2015; Thomas, *et al*., 2018). By comparing the present data to the data of a recent genome-wide analysis of ppGpp regulation (Sanchez-Vazquez, *et al*., 2019), we found that the expression of half of the genes controlled by a positive GR-related effect and one third of the genes controlled by a negative GR-related effect had been reported to depend on ppGpp binding to RNA polymerase (Fig. S11). This supports the hypothesis that ppGpp is the primary regulatory factor mediating the Crp-independent effect of GR, in addition to its well-known role of controlling synthesis of ribosome and amino acid biosynthesis (Hernandez and Bremer, 1990; Scott, *et al*., 2010; Lemke, *et al*., 2011; Danchin, *et al*., 2020). It is worth noticing that these ppGpp-controlled genes included several global regulators, such as RpoS and RpoE (Lemke, *et al*., 2011; Gopalkrishnan, *et al*., 2014) (Data S2, Fig. S12), which could therefore control indirectly the expression of the remaining genes that were not regulated by ppGpp directly. As a result, ppGpp could be another important signal in addition to cAMP (You, *et al*., 2013; Li, *et al*., 2019) that mediates global regulation when the GR declines. Further studies are required to clarify this issue.

In addition to Crp, Cra is another well-known factor controlling carbon metabolism, but their relative contributions to the physiology of *E. coli* remain, as yet, obscure (Kim, *et al*., 2018). Given the uncontrolled variation between different biological experiments, it will be more accurate to compare their relative contributions within the same study. Although some studies have monitored the activities of the two factors, we were able to do this comparison simultaneously as, besides Crp, our experiments allowed us to monitor the activity of Cra as well. This was possible because the level of F1P—functioning as the Cra inhibitor—was much lower in the non-glycolytic substrates (Kochanowski, *et al*., 2017) (Fig. S6), so that genes controlled by Cra would exhibit different responses to glycolytic and non-glycolytic substrates. Our study analyzed the effect of both types of substrates. We identified genes regulated by Cra by characterizing genes responding distinctly to non-glycolytic substrates. Approximately 70% of these genes (69/102) were demonstrated to be controlled by Cra directly (Fig. 4C), highlighting the sensitivity of our selection method. These 69 genes included genes in the previously reported Cra regulon (45 genes), together with genes newly identified by our *in vitro* EMSA (24 genes) (Fig. 4D). We noted that the regulatory modes of Cra on 25 of the 45 reported Cra regulon genes had not been determined previously. This allowed us to amend their regulatory modes accordingly, based on the observation of their increased or decreased expression (Fig. 4C). Knowing the physiological functions of the genes now identified to be controlled by Cra, we are in a better position to propose a presumably comprehensive view of the regulatory interactions of Cra (Fig. 5). The regulatory role of Crp is more global than that of Cra, and the role of Cra appears to be limited to the central carbon metabolism and other pathways flowing into or out of it. The specific importance of Cra in central carbon metabolism can be further judged by the much lower growth rate of

the *cra*-null strain as compared to that of the wild type when cells were grown in non-glycolytic carbon substrates. No significant growth difference was observed when cells were grown in the glycolytic carbon substrate (Kim, *et al*., 2018) (Fig. S7C). This can be easily understood as, when cells are grown in non-glycolytic carbon substrates, they require much activated Cra regulators to ensure the efficient generation of ATP and building blocks *via* the TCA cycle and the gluconeogenesis in order to maintain the optimal growth.

As an unexpected aside, we identified that two more factors, Mlc and Fur, were involved in the metabolism of the carbon substrates of interest, with their targeted genes, respectively, showing expression peaks in glucose/lactose and in the two PTS sugars (glucose/sorbitol) (Fig. 6A, C). This finding highlighted the primary control of Mlc on the transport of glucose and mannose (Schiefner, *et al*., 2005) and of Fur on iron homeostasis (Seo, *et al*., 2014) based on the functions of the genes in the expanded two regulons. Why did Mlc regulon genes respond specifically to glucose and lactose? Mlc was sequestered and repressed by un-phosphorylated PtsG (Nam, *et al*., 2001). Glucose, transported via phosphorylated PtsG (PtsG-P), generated a large amount of un-phosphorylated PtsG (Deutscher, 2008) which inactivated Mlc (Fig. S13). As a consequence, genes repressed by Mlc showed high expression. The increased expression of Mlc-controlled genes in lactose supported that the intracellular glucose generated *via* lactose degradation could also be phosphorylated by phosphate-PtsG, generating a large amount of un-phosphorylated PtsG (Postma, *et al*., 1993) which inactivated Mlc (Fig. S13). Thus phospho-PtsG could act as a novel hexokinase (Meyer, *et al*., 1997). This interpretation needs further investigation. Yet another very interesting question emerges. What causes iron depletion when cells grow in PTS sugars, as indicated by the induction of Fur-repressed genes? A tentative response is as follows. Quite intriguingly, in the two PTS sugar transporters (PtsG and SrlA), cysteine is the key residue phosphorylated during sugar transport (Nuoffer, *et al*., 1988; Nguyen, *et al*., 2006). Cysteine is sensitive to reactive oxygen species (Hillion and Antelmann, 2015; Lee, *et al*., 2016), and this is doomed to affect cysteine phosphorylation (Scotto, *et al*., 1998). As a consequence, cells grown with these two PTS sugars may require more iron-sulfur clusters for example (Py, *et al*., 2018) to maintain a proper level of cysteine phosphorylation, which results into the depletion of free iron. Hence, the very role of cysteine phosphorylation (Sun, *et al*., 2012) in *E. coli* is worth further exploration.

Finally and intriguingly, we found that more than half of the genes under specific regulation of Cra and Mlc were also subject to coordinated regulation by Crp or the GR-related effect (Fig. 7). Supporting our findings, Kim *et al*. (2018) reported recently that Crp and Cra co-regulated the expression of several transcriptional units. Here, with the expanded genes list discovered by our systematic assay (Data S11), we could summarize the physiological roles of the coordinated regulation of Crp and Cra at a higher resolution, based on the functions of their co-regulated genes. In consistence with their individual role to the regulation of carbon metabolism, the focus of the coordinated regulation of Crp and Cra is in the central carbon metabolism. Crp and Cra work in concert to co-activate the most efficient pathways to generate ATP (*via* the TCA cycle) and building blocks (*via* gluconeogenesis) but they co-repress other antagonistic pathways (the glycolysis pathway and the pathway of amino acid biosynthesis) of metabolism (Fig. S14). Importantly, we quantified the individual contributions of Crp-, Cra-, Mlc- and GR-related effect to the regulation of genes under their coordinated control by dividing the total expression into expression, respectively, controlled by each of the factors (Fig. 7, Fig. S10). Crp and Cra took nearly equal contributions when cells were grown in succinate, but the contribution of Crp took over that of Cra when GR reduced further in acetate (Fig. 7D, E). Similarly, we also found that GR-related effect increased when GR reduced and it took over that of Cra in acetate in the case of genes under their coordinated regulation (Fig. S10). Taken together, it seems that the global regulations win when GR reduced. But the situation is different in the case of genes co-regulated by Mlc and Crp (Fig. 7F). Mlc took the overriding role in both glucose and lactose although glucose gave a lower GR than lactose. By contrast, in other carbon substrates which gave a much lower GR such as acetate, Crp took the primary role as an activator when Mlc worked as a repressor as there were no un-phosphorylated PtsG in acetate (Fig. S13) (Deutscher, 2008). Thus, this quantitative analysis draws a dynamic picture of the process of coordinated regulations, in which regulatory factors take turns to work as the primary one or contribute equally depending on the conditions, in order to generate the desired yield of proteins to cope with the cellular metabolism when environment changes.

In summary, we deciphered the global regulatory strategy that *E. coli* applies to cope with the metabolism of an array of carbon substrates with distinctive features. In this process, we disentangled the individual and coordinated contributions of various control responses (Crp, Cra, Mlc, Fur and GR-related effect) and identified a novel rule of regulation. All these notable findings were achieved using well-chosen physiological conditions combined with systematic analyses (You, *et al*., 2013; Basan, *et al*., 2015; Hermsen, *et al*., 2015) of the high-resolution transcriptome data (Li, *et al*., 2019). This could

hardly have been performed previously. The comprehensive knowledge of gene regulation revealed here makes *E. coli* an ideal organism for biotech fermentation.

Experimental procedures

Construction of bacterial strains and plasmids

Wild-type *E. coli* K-12 strain NCM3722 (You, *et al*., 2013) was used in the transcriptome study, and its derivative of CY189 (Δ*cra*) and CY940 (*cra-flag*) constructed by λ-Red system (Datsenko and Wanner, 2000) was applied for growth and detection of Cra protein by Western blotting, respectively. Strain CY134 for over-expression of Crp was constructed previously (Li, *et al*., 2019). For over-expression of Cra and Fur, full-length *cra* and *fur* genes from NCM3722 were amplified by PCR and both cloned into pET24a via *Nde* I/*Xho* I, generating pET24a-cra and pET24a-fur, respectively. The verified plasmids were then transformed into strain BL21 (DE3), achieving strains CY923 and CY668.

Growth of cell cultures

Batch cultures were grown in the N^-C^- minimal medium (You, *et al*., 2013) supplemented with 20 mM NH_4Cl as the nitrogen source, and one of the seven carbon substrates [20 mM glucose-6-phosphate (G6P), 0.2% (w/v) lactose (Lac), 0.4% (w/v) glucose (Glc), 20 mM succinate (Suc), 0.4% (w/v) glycerol (Gly), 20 mM sorbitol (Sor), 60 mM acetate (Ace)] as a sole carbon source. All growths were carried out in three steps in a 37°C water bath shaker as described previously (You, *et al*., 2013). Samples were collected at OD_{600} ~ 0.4 for the following transcriptome study, western blotting or FbP assay.

RNA-seq assay and analysis of transcriptome data

Total RNA extraction and RNA-seq assay were performed as described previously (Li, *et al*., 2019). Three independent total RNA extractions and transcriptome analyses by RNA-seq were performed for NCM3722 grown in each carbon substrate. DEGs between a comparison of randomly picked two transcriptomes among the seven were characterized by DEseq2 [twofold cut-off and adjusted *P*-value (*q*-value) <0.01] and followed by ANOVA analysis to identify genes that were statistically differentially expressed among the seven growth conditions (degrees_freedom1 = 6 and degrees_freedom2 = 14, *P*-value < 0.05). RPKM method (Reads Per Kilo bases per Million Reads) (Mortazavi, *et al*., 2008) was used to calculate expression of each gene. Pearson correlation coefficient R between the expression of all the DEGs identified at the seven carbon substrates and the corresponding GR were then calculated. In case of the coarse-grained analysis, DEGs with $R < 0$ were grouped in GR-negative group and DEGs with $R > 0$ were grouped in GR-positive group. In case of the fine-grained analysis, genes regulated by Crp or effect of GR were screened among all DEGs that showed significant correlation with GR ($R<-0.75$ or $R > 0.75$ with P-value < 0.05) using the method as described in the main text. Genes showed increased (resp. decreased) expression in acetate versus sorbitol paralleled by increased (resp. decreased) expression in succinate versus glycerol and succinate versus glucose, were also characterized by using the method as described in the main text. Genes showed expression peaks in glucose and lactose were characterized by $Exp_{Glc}/Exp_{Suc} > 1.5$, $Exp_{Glc}/Exp_{G6P} > 1.5$, $Exp_{Glc}/Exp_{Gly} > 1.5$, $Exp_{Lac}/Exp_{G6P} > 1.5$, $Exp_{Glc}/Exp_{Lac} < 1.5$, $Exp_{Sor} > Exp_{Ace}$ and $Exp_{Sor} > Exp_{Gly}$. Genes showed expression peaks in sorbitol and glucose were characterized by $Exp_{Sor}/Exp_{Ace} > 1.5$, $Exp_{Sor}/Exp_{Gly} > 1.5$, $Exp_{Sor}/Exp_{Suc} > 1.5$, $Exp_{Glc} > Exp_{Gly}$, $Exp_{Glc} > Exp_{Suc}$, $Exp_{Glc} > Exp_{G6P}$, $Exp_{Glc} > Exp_{Lac}$, $Exp_{Sor}/Exp_{Glc} < 2$, $Exp_{Ace}/Exp_{G6P} < 2$.

In order to identify genes under coordinated regulation of Cra and Crp/GR-related effect, the Pearson correlation coefficient R between the expression of all the 69 Cra regulon genes identified at the five carbon substrates (Sor, Gly, Glc, Lac, G6P) and the corresponding GR was re-calculated. Then the genes showing significant correlation with GR (P-value < 0.05) were further judged using the criteria of characterizing the regulations of Crp- or GR-related effect. To identify genes under coordinated regulation of Mlc- and Crp/GR-related effect, we did similar analysis for the seven Mlc regulon genes but with the Pearson correlation coefficient R re-calculated between their expression at the five carbon substrates (Ace, Sor, Gly, Suc, G6P) and the corresponding GR.

The transcriptome data of NL1 (in a nitrogen limited condition with a GR of 0.2 h^{-1}) and NL4 (in a nitrogen rich condition with a GR of 0.9 h^{-1}) from a nitrogen-limited chemostat was obtained as described previously (Li, *et al*., 2019).

Heatmap visualization and KEGG pathway-enriched analysis

Both tools for heatmap visualization and KEGG pathway-enriched analysis of the transcriptome data were conducted with OmicShare online platform (www.omicshare.com/tools) (GENE DENOVO). For heatmap, ordered transcriptional levels (RPKM) of selected genes were visualized by the 'Heatmap' tool, normalized using z-score at row, and clustered if necessary. For pathway analysis, gene sequences of NCM3722 (NZ_CP011495.1 and NZ_CP011496.1) were locally blast with sequences from KEGG GENES Database (www.kegg.jp/kegg/genes.html)

to obtain matched genes in the database. The matched genes were next mapped with KEGG PATHWAY Database (www.kegg.jp/kegg/pathway.html) to obtain the relationship of gene-pathway. Our selected genes in each subgroup were categorized and enriched by the 'Pathway Enrichment Analysis' tool, using a gene-pathway corresponding document as the background set.

Overexpression and purification of Crp, Cra and Fur

Strains for over-expression of His-tagged proteins were all cultured in LB broth with kanamycin at 37°C, 250 rpm to the log phase (OD_{600} ~ 0.6). Subsequently, with addition of 1 mM IPTG, cells were kept growing under different temperatures and time slots prior to harvest. For Crp, CY134 was used, and kept growing at 30°C overnight; for Cra, CY923 was used, and kept growing at 37°C for over 3 h; for Fur, CY668 was used, and kept growing at 28 °C overnight.

Procedures of His-tagged Crp, Cra and Fur purification were all following the instruction of Capturem His-Tagged Purification Miniprep Kit (Clontech, Mountain View, CA, USA). Dialyzed His-tagged proteins were concentrated by 3K MWCO protein concentrators, and the concentration and purity were, respectively, determined by Bradford assay and Coomassie stains of SDS-PAGE gels. Proteins with purity larger than 95% were achieved for following assays.

EMSA of Crp, Cra and Fur

The promoter region of each target gene was PCR amplified using the primers listed in Table S1 and purified to serve as a probe in EMSA. DNA-binding reaction by Crp (10 μl) was set up as follows: 0.25 pmol DNA, 0, 3, 6 pmol Crp proteins in 1 × binding buffer [10 mM Tris-HCl (pH 7.2), 150 mM NaCl, 3 mM MgAc, 5% glycerol] plus 1 mM cAMP. DNA-binding reaction by Cra (10 μl) was set up as follows: 0.25 pmol DNA, 0, 1, 2 pmol Cra proteins in 1 × TGED binding buffer [10 mM Tris-HCl (pH 8.0), 5% (v/v) glycerol, 0.1 mM EDTA, 1 mM dithiothreitol]. DNA-binding reaction by Fur (20 μl) was set up as follows: 0.25 pmol DNA, 0, 20, 40 pmol Fur proteins in 1 × binding buffer [20 mM BisTris/borate (pH 7.0), 40 mM KCl, 1 mM $MgCl_2$, 200 μM $CoCl_2$ and 5% glycerol]. The reaction mix was incubated at 37 °C for 15 min, mixed with 5 × Native gel-loading buffer, separated on a 6% (For Crp and Cra) or 5% (For Fur) native polyacrylamide gel in 0.5 × TBE (For Crp), 1 × TG buffer [48 mM Tris, 39 mM glycine (pH 8.9)] (For Cra), or 1 × running buffer [20 mM BisTris/borate (pH 7.0)] (For Fur) at 120 V for 60 min (For Crp and Cra) or 220 V for 40 min (For Fur). After electrophoresis, the gel was stained in corresponding buffer with GelRed dye and visualized by Gel Doc XR+ (Bio-Rad). Two to six independent repeats were performed for each target gene. Prediction of Crp, Cra and Fur binding sites *in silico* were conducted with MEME and FIMO program in MEME suite 5.1.0 (Bailey *et al.*, 2009). Consensus of Crp and Cra binding sequences were generated by WebLogo 3 (Schneider and Stephens, 1990).

Quantification of Cra by Western blotting

Strain CY940 (*cra-flag*) was grown in Glc, Suc, Gly, Sor and Ace as a sole carbon source to OD_{600} ~ 0.4. Then 5×10^7 cells were aliquot and subjected to Western blot analysis. Cell pellets were resuspended in SDS sample buffer, boiled and separated by SDS-PAGE. Separated proteins on a gel were next electrically transferred to a PVDF membrane (GE, Boston, MA, USA) at 100 V for 40 min. For simultaneous detection of Flag-tagged Cra and a loading control RpoB, the membrane was excised into two parts, and treated individually. After blocking with 5% non-fat milk in TBST for 1 h, the membranes for Cra and RpoB were, respectively, incubated with polyclonal anti-DDDDK (FLAG) tag (Abcam, Cambridge, MA, USA) and monoclonal anti-RpoB antibody (BioLegend, San Diego, CA, USA) for 1.5 h, washed 3 times with TBST, subsequently blotting with the goat anti-rabbit IgG and goat anti-mouse IgG HRP-conjugated secondary antibody (Abmart, Berkeley Heights, NJ, USA) for 1.5 h, and washed three times with TBST. Following a treatment with Super ECL Detection Reagent (Yeasen, Shanghai, China), the membranes were exposed in Gel Doc XR+ (Bio-Rad, Hercules, CA, USA) to obtain images of target proteins. Signals corresponding to each protein were quantified with ImageJ (NIH) and presented as a relative amount to the level of that in Glc. Results are averages from two independent assays.

5′-RACE assay

The 5′ end of *yobD* mRNA transcript was identified as described previously following the instruction of SMARTer RACE 5′/3′ Kit (Clontech) (Li, *et al.*, 2019) with RACE-ready cDNA from 1.0 to 10 μg of total RNA as the template. PCR products obtained by using primers listed in Table S1 were sub-cloned into T-vector pMD19 simple (Takara), and single colonies were picked and subjected to sequencing. The transcriptional start site was determined by mapping the sequencing data with genomic sequences of NCM3722 (NZ_CP011495.1).

Quantitative analysis of FbP

Levels of FbP for NCM3722 grown in succinate, sorbitol or acetate as a sole carbon source were analyzed by a

targeted mass spectrometry-based quantitative metabolomic approach in biological duplicate with four samples per group of assay. Briefly, for each sample, cell pellets ($3–4\times10^9$) were resuspended in 3:1 MeOH:H_2O (v/v), subjected to sonication with 500 W for 10 min, and next centrifuged for 10 min at 4 °C. The supernatant was collected and dried in a lyophilizer. The powder was resuspended in 2:8 MeOH:H_2O (v/v) and centrifuged for 10 min at 4°C, and then, the supernatant was ready for detection. FbP analysis was performed using 10 μl of sample loaded into a Nexera UPLC (SHIMADZU) with an ACQUITY UPLC C18 column (100 × 2.1 mm, 1.7 μm, Waters) and underwent gradient elution procedures; then the sample was ionized by ESI and detected by QE high-resolution mass spectrometer (Thermo) with Xcalibur 4.0 workstation in the negative ion mode, scanning precursor ion m/z 338.99 for FbP. The amount of FbP in unknown samples was calculated by peak areas when compared with that of an FbP standard.

Acknowledgements

We are grateful to Chaoliang Wei for help on the RNA-seq data analysis.

Funding Information

This work was supported by National Natural Science Fund of the People's Republic of China (NSFC) (31970072), by Major Projects in Basic and Applied Research of Guangdong Province Office of Education (2017KZDXM074), by Shenzhen Peacock Project (KQTD2016112915000294), and by the 1000 Youth Plan Fund.

Conflict of interest

None declared.

Data Availability Statement

The whole dataset of RNA-seq has been deposited to GEO with the accession number of GSE156143.

References

Alteri, C.J., Himpsl, S.D., and Mobley, H.L. (2015) Preferential use of central metabolism *in vivo* reveals a nutritional basis for polymicrobial infection. *PLoS Pathog* **11:** e1004601.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., *et al*. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208.

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J.R., and Hwa, T. (2015) Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* **528:** 99–104.

Benning, C. (1998) Biosynthesis and function of the sulfolipid sulfoquinovosyl diacylglycerol. *Annu Rev Plant Physiol Plant Mol Biol* **49:** 53–75.

Bertin, Y., Chaucheyras-Durand, F., Robbe-Masselot, C., Durand, A., de la Foye, A., Harel, J., *et al*. (2013) Carbohydrate utilization by enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. *Environ Microbiol* **15:** 610–622.

Bley Folly, B., Ortega, A.D., Hubmann, G., Bonsing-Vedelaar, S., Wijma, H.J., van der Meulen, P., *et al*. (2018) Assessment of the interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcription factors CggR and Cra. *Mol Microbiol* **109:** 278–290.

Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., *et al*. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* **19:** 352–367.

Busby, S., and Ebright, R.H. (1999) Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293:** 199–213.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., *et al*. (2004) Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci USA* **101:** 7427–7432.

Chavarria, M., Durante-Rodriguez, G., Krell, T., Santiago, C., Brezovsky, J., Damborsky, J., and de Lorenzo, V. (2014) Fructose 1-phosphate is the one and only physiological effector of the Cra (FruR) regulator of *Pseudomonas putida*. *FEBS Open Bio* **4:** 377–386.

Chavarria, M., and de Lorenzo, V. (2018) The imbroglio of the physiological Cra effector clarified at last. *Mol Microbiol* **109:** 273–277.

Choi, E., and Hwang, J. (2018) The GTPase BipA expressed at low temperature in *Escherichia coli* assists ribosome assembly and has chaperone-like activity. *J Biol Chem* **293:** 18404–18419.

Danchin, A., Sekowska, A., and You, C. (2020) One-carbon metabolism, folate, zinc and translation. *Microb Biotechnol* **13:** 899-925.

Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97:** 6640–6645.

Denger, K., Weiss, M., Felux, A.K., Schneider, A., Mayer, C., Spiteller, D., *et al*. (2014) Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature* **507:** 114–117.

Deutscher, J. (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* **11:** 87–93.

Elgamal, S., Artsimovitch, I., and Ibba, M. (2016) Maintenance of transcription-translation coupling by elongation factor P. *mBio* **7:** e01373-16.

Epstein, W., Rothman-Denes, L.B., and Hesse, J. (1975) Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc Natl Acad Sci USA* **72:** 2300–2304.

Folsom, J.P., Parker, A.E., and Carlson, R.P. (2014) Physiological and proteomic analysis of *Escherichia coli* iron-limited chemostat growth. *J Bacteriol* **196:** 2748–2761.

Franchini, A.G., Ihssen, J., and Egli, T. (2015) Effect of global regulators RpoS and cyclic-AMP/CRP on the catabolome and transcriptome of *Escherichia coli* K12 during carbon- and energy-limited growth. *PLoS One* **10:** e0133793.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muniz-Rascado, L., Garcia-Sotelo, J.S., *et al*. (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res* **44:** D133–143.

Gopalkrishnan, S., Nicoloff, H., and Ades, S.E. (2014) Coordinated regulation of the extracytoplasmic stress factor, sigmaE, with other *Escherichia coli* sigma factors by (p) ppGpp and DksA may be achieved by specific regulation of individual holoenzymes. *Mol Microbiol* **93:** 479–493.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A., and Saier, M.H. Jr (2004) Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J Bacteriol* **186:** 3516–3524.

Gyaneshwar, P., Paliy, O., McAuliffe, J., Popham, D.L., Jordan, M.I., and Kustu, S. (2005) Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses. *J Bacteriol* **187:** 1074–1090.

Haange, S.B., Groeger, N., Froment, J., Rausch, T., Burkhardt, W., Gonnermann, S., *et al*. (2020) Multiplexed quantitative assessment of the fate of taurine and sulfoquinovose in the intestinal microbiome. *Metabolites* **10:** 430.

Hermsen, R., Okano, H., You, C., Werner, N., and Hwa, T. (2015) A growth-rate composition formula for the growth of *E.coli* on co-utilized carbon substrates. *Mol Syst Biol* **11:** 801.

Hernandez, V.J., and Bremer, H. (1990) Guanosine tetraphosphate (ppGpp) dependence of the growth rate control of *rrnB* P1 promoter activity in *Escherichia coli*. *J Biol Chem* **265:** 11605–11614.

Hillion, M., and Antelmann, H. (2015) Thiol-based redox switches in prokaryotes. *Biol Chem* **396:** 415–444.

Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004) Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Appl Environ Microbiol* **70:** 2354–2366.

Hui, S., Silverman, J.M., Chen, S.S., Erickson, D.W., Basan, M., Wang, J., *et al*. (2015) Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Syst Biol* **11:** 784.

Ishihama, A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev* **34:** 628–645.

Jimenez, A.G., Ellermann, M., Abbott, W., and Sperandio, V. (2020) Diet-derived galacturonic acid regulates virulence and intestinal colonization in enterohaemorrhagic *Escherichia coli* and *Citrobacter rodentium*. *Nature microbiology* **5:** 368–378.

Kang, A., Tan, M.H., Ling, H., and Chang, M.W. (2013) Systems-level characterization and engineering of oxidative stress tolerance in *Escherichia coli* under anaerobic conditions. *Mol BioSyst* **9:** 285–295.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martinez, C., Caspi, R., *et al*. (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45:** D543–D550.

Kim, D., Seo, S.W., Gao, Y., Nam, H., Guzman, G.I., Cho, B.K., and Palsson, B.O. (2018) Systems assessment of transcriptional regulation on central carbon metabolism by Cra and CRP. *Nucleic Acids Res* **46:** 2901–2917.

Kochanowski, K., Gerosa, L., Brunner, S.F., Christodoulou, D., Nikolaev, Y.V., and Sauer, U. (2017) Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol Syst Biol* **13:** 903.

Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62:** 749–795.

Kornberg, H.L. (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem J* **99:** 1–11.

Lee, S.J., Lee, I.G., Lee, K.Y., Kim, D.G., Eun, H.J., Yoon, H.J., *et al*. (2016) Two distinct mechanisms of transcriptional regulation by the redox sensor YodB. *Proc Natl Acad Sci USA* **113:** E5202–E5211.

Lemke, J.J., Sanchez-Vazquez, P., Burgos, H.L., Hedberg, G., Ross, W., and Gourse, R.L. (2011) Direct regulation of *Escherichia coli* ribosomal protein promoters by the transcription factors ppGpp and DksA. *Proc Natl Acad Sci USA* **108:** 5712–5717.

Li, J., Epa, R., Scott, N. E., Skoneczny, D., Sharma, M., Snow, A. J. D., *et al*. (2020) A sulfoglycolytic Entner-Doudoroff pathway in *Rhizobium leguminosarum* bv. trifolii SRDI565. *Appl Environ Microbiol* **86:** e00750–20.

Li, Z., Pan, Q., Xiao, Y., Fang, X., Shi, R., Fu, C., *et al*. (2019) Deciphering global gene expression and regulation strategy in *Escherichia coli* during carbon limitation. *Microb Biotechnol* **12:** 360–376.

Liu, M., Durfee, T., Cabrera, J.E., Zhao, K., Jin, D.J., and Blattner, F.R. (2005) Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. *J Biol Chem* **280:** 15921–15927.

Lorca, G.L., Ezersky, A., Lunin, V.V., Walker, J.R., Altamentova, S., Evdokimova, E., *et al*. (2007) Glyoxylate and pyruvate are antagonistic effectors of the *Escherichia coli* IclR transcriptional regulator. *J Biol Chem* **282:** 16476–16491.

Meyer, D., Schneider-Fresenius, C., Horlacher, R., Peist, R., and Boos, W. (1997) Molecular characterization of glucokinase from *Escherichia coli* K-12. *J Bacteriol* **179:** 1298–1306.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.

Nam, T.W., Cho, S.H., Shin, D., Kim, J.H., Jeong, J.Y., Lee, J.H., *et al*. (2001) The *Escherichia coli* glucose transporter enzyme IICB(Glc) recruits the global repressor Mlc. *EMBO J* **20:** 491–498.

Nguyen, T.X., Yen, M.R., Barabote, R.D., and Saier, M.H. Jr (2006) Topological predictions for integral membrane permeases of the phosphoenolpyruvate:sugar phosphotransferase system. *J Mol Microbiol Biotechnol* **11:** 345–360.

Nuoffer, C., Zanolari, B., and Erni, B. (1988) Glucose permease of *Escherichia coli*. The effect of cysteine to serine

mutations on the function, stability, and regulation of transport and phosphorylation. *J Biol Chem* **263:** 6647–6655.

Oh, M.K., Rohlin, L., Kao, K.C., and Liao, J.C. (2002) Global expression profiling of acetate-grown *Escherichia coli*. *J Biol Chem* **277:** 13175–13183.

Okano, H., Hermsen, R., Kochanowski, K., and Hwa, T. (2020) Regulation underlying hierarchical and simultaneous utilization of carbon substrates by flux sensors in *Escherichia coli*. *Nat Microbiol* **5:** 206–215.

Perlman, R.L., De Crombrugghe, B., and Pastan, I. (1969) Cyclic AMP regulates catabolite and transient repression in *E. coli*. *Nature* **223:** 810–812.

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.

Porcheron, G., Habib, R., Houle, S., Caza, M., Lepine, F., Daigle, F., *et al*. (2014) The small RNA RyhB contributes to siderophore production and virulence of uropathogenic *Escherichia coli*. *Infect Immun* **82:** 5056–5068.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57:** 543–594.

Py, B., Gerez, C., Huguenot, A., Vidaud, C., Fontecave, M., Ollagnier de Choudens, S., and Barras, F. (2018) The ErpA/NfuA complex builds an oxidation-resistant Fe-S cluster delivery pathway. *J Biol Chem* **293:** 7689–7702.

Rhee, H. S., and Pugh, B. F. (2012) ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol* Chapter **21:** Unit 21 24.

Sanchez-Vazquez, P., Dewey, C.N., Kitten, N., Ross, W., and Gourse, R.L. (2019) Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc Natl Acad Sci USA* **116:** 8310–8319.

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., and Fischer, E. (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* **279:** 6613–6619.

Schiefner, A., Gerber, K., Seitz, S., Welte, W., Diederichs, K., and Boos, W. (2005) The crystal structure of Mlc, a global regulator of sugar metabolism in *Escherichia coli*. *J Biol Chem* **280:** 29073–29079.

Schneider, T.D., and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18:** 6097–6100.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* **330:** 1099–1102.

Scotto, C., Mely, Y., Ohshima, H., Garin, J., Cochet, C., Chambaz, E., and Baudier, J. (1998) Cysteine oxidation in the mitogenic S100B protein leads to changes in phosphorylation by catalytic CKII-alpha subunit. *J Biol Chem* **273:** 3901–3908.

Seo, S.W., Kim, D., Latif, H., O'Brien, E.J., Szubin, R., and Palsson, B.O. (2014) Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. *Nat Commun* **5:** 4910.

Shahrezaei, V., and Marguerat, S. (2015) Connecting growth with gene expression: of noise and numbers. *Curr Opin Microbiol* **25:** 127–135.

Shimada, T., Fujita, N., Yamamoto, K., and Ishihama, A. (2011a) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS One* **6:** e20081.

Shimada, T., Yamamoto, K., and Ishihama, A. (2011b) Novel members of the Cra regulon involved in carbon metabolism in *Escherichia coli*. *J Bacteriol* **193:** 649–659.

Shin, D., Cho, N., Heu, S., and Ryu, S. (2003) Selective regulation of *ptsG* expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. *J Biol Chem* **278:** 14776–14781.

Sun, F., Ding, Y., Ji, Q., Liang, Z., Deng, X., Wong, C.C., *et al*. (2012) Protein cysteine phosphorylation of SarA/MgrA family transcriptional regulators mediates bacterial virulence and antibiotic resistance. *Proc Natl Acad Sci USA* **109:** 15461–15466.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., *et al*. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329:** 533–538.

Thomas, P., Terradot, G., Danos, V., and Weisse, A.Y. (2018) Sources, propagation and consequences of stochasticity in cellular growth. *Nat Commun* **9:** 4528.

Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., *et al*. (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* **48:** 657–670.

Yang, L., Lin, R.T., and Newman, E.B. (2002) Structure of the Lrp-regulated *serA* promoter of *Escherichia coli* K-12. *Mol Microbiol* **43:** 323–333.

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C.W., *et al*. (2013) Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500:** 301–306.

Zhang, Z., Gosset, G., Barabote, R., Gonzalez, C.S., Cuevas, W.A., and Saier, M.H. Jr (2005) Functional interactions between the carbon and iron utilization regulators, Crp and Fur, in *Escherichia coli*. *J Bacteriol* **187:** 980–990.

Zheng, D., Constantinidou, C., Hobman, J.L., and Minchin, S.D. (2004) Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res* **32:** 5874–5893.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. The growth rate (GR) of *E. coli* NCM3722 cultured in one of the seven carbon substrates.

Fig. S2. Changing trend of genes when *E. coli* grown in nitrogen-limited (NL) conditions.

Fig. S3. Correlation of the expression of *gpmM*, *grcA* (*yfiD*), *serA* and *typA* with GR.

Fig. S4. A consensus sequence logo of the Crp-cAMP binding site for genes identified in this work

Fig. S5. The KEGG pathway-enriched analysis with 468 Crp regulatory genes identified in this work.
Fig. S6. Relative abundance of intracellular FbP (fructose-1,6-bisphosphate).
Fig. S7. The expression level of *cra*, protein level of Cra and the growth of Δ*cra* variant in different carbon sources.
Fig. S8. A consensus sequence logo of the Cra binding site for genes identified in this work.
Fig. S9. The 5′-RACE of *yobD*.
Fig. S10. Genes co-regulated by Cra and GR-related effects.
Fig. S11. The number of known ppGpp controlled genes among genes regulated by negative GR- or positive GR-related effect.
Fig. S12. Correlation of the expression of *rpoS* and *rpoE* with GR.
Fig. S13. Mode of Mlc repression by un-phosphorylated PtsG in glucose and lactose.
Fig. S14. The prime roles of coordinated regulations of Crp and Cra in *E. coli* physiology.
Table S1. Primers used in this study.
Data S1. All identified DEGs.
Data S2. DEGs responding to the GR and their regulatory strategies.
Data S3. Regulatory mode of Crp regulon genes unrevealed previously.
Data S4. Candidate Crp regulatory genes.
Data S5. Newly validated Crp regulatory genes.
Data S6. Candidate Cra regulatory genes.
Data S7. Newly validated Cra regulatory genes.
Data S8. Functional categories of Cra regulatory genes.
Data S9. Candidate Mlc regulatory genes.
Data S10. Candidate Fur regulatory genes.
Data S11. Genes co-regulated by Cra-, Mlc-, Crp- and GR-related effects.

