BMC
Systems Biology

PROCEEDINGS   **Open Access**

Comparison on extreme pathways reveals nature of different biological processes

Yanping Xi[1], Yue Zhao[1], Li Wang[2], Fei Wang[1*]

Abstract

Background: Constraint-based reconstruction and analysis (COBRA) is used for modeling genome-scale metabolic networks (MNs). In a COBRA model, extreme pathways (ExPas) are the edges of its conical solution space, which is formed by all viable steady-state flux distributions. ExPa analysis has been successfully applied to MNs to reveal their phenotypic capabilities and properties. Recently, the COBRA framework has been extended to transcriptional regulatory networks (TRNs) and transcriptional and translational networks (TTNs), so efforts are needed to determine whether ExPa analysis is also effective on these two types of networks.

Results: In this paper, the ExPas resulting from the COBRA models of E.coli's MN, TRN and TTN were horizontally compared from 5 aspects: (1) Total number and the ratio of their amount to reaction amount; (2) Length distribution; (3) Reaction participation; (4) Correlated reaction sets (CoSets); (5) interconnectivity degree. Significant discrepancies in above properties were observed during the comparison, which reveals the biological natures of different biological processes. Besides, by demonstrating the application of ExPa analysis on E.coli, we provide a practical guidance of an improved approach to compute ExPas on COBRA models of TRNs.

Conclusions: ExPas of E.coli's MN, TRN and TTN have different properties, which are strongly connected with various biological natures of biochemical networks, such as topological structure, specificity and redundancy. Our study shows that ExPas are biologically meaningful on the newborn models and suggests the effectiveness of ExPa analysis on them.

Background

Many large-scale biological networks, including metabolic networks (MNs) [1], signaling networks [2], transcriptional regulatory networks [3] and transcriptional and translational networks [4] have been reconstructed along with the development of high-throughput technology in the past decades. These networks are then transformed into mathematical models for further analysis. Constraint-Based Reconstruction and Analysis (COBRA) is one of the most commonly used frameworks introduced to model and analyze steady-state biochemical networks [5]. In the past two decades, it has been successfully applied on MNs to study various phenotypes

[6-9]. Recently, the same principles were also extended to other types of biochemical networks mentioned above [2-4,10].

All the possible phenotypes, i.e. the flux distributions of feasible steady states, of a constraint-based biochemical model form a high-dimensional cone. Network-based pathways such as Extreme Pathways (ExPas) [11] are defined to study this cone. ExPas are vectors of fluxes that lie on the edges of the cone [12]. They constitute the minimal and unique vector set which generates the space of all feasible steady states through non-negative linear combination. Since ExPas characterize the limits on the capabilities of a cell's metabolic system [13], ExPa analysis will reveal systemic properties of metabolism [14]. ExPa analysis as an approach to characterize the fundamental and time-invariant topological properties of a given network [15] has been successfully applied to MNs,

* Correspondence: wangfei@fudan.edu.cn
[1]Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, 200433, China
Full list of author information is available at the end of the article

such as those of human red blood cells [16], *Escherichia coli* [17-19], *Sacchoromryces cerevisiae* [20,21], *Helicobacter pylori* [22,23], *Haemophilus influenzae* [11,24] and *Methylobacterium extroguens* [25]. Besides, network models respectively describing a prototypic signaling system [10] and the JAK-STAT signaling system in the human B-Cell [2] have also been studied through ExPa analysis.

Recently, there emerged two COBRA models of biochemical systems with different types: *E.coli* transcriptional regulatory network (TRN) [3] and *E.coli* transcriptional and translational network (TTN) [4]. What should be clarified is whether ExPa analysis is still useful for new types of networks and whether ExPas of TRN or TTN show some properties different from those of MN. These questions are biologically significant because the answers determines whether we can rely on the existing analysis approaches to obtain novel and biologically meaningful findings in a brand new field. In this paper, we try to provide an anwser by comparing properties of ExPas among the *E.coli* TRN, MN and TTN. In the comparison, differences between biological processes were observed from multiple perspectives, including network structure, reaction participation, specificity and redundancy. The results indicate that ExPa analysis can be extended to biochemical systems of TRN and TTN, which helps researchers to further understand the corresponding biological systems. Besides an improved method was introduced to simplify the calculation and interpretation of ExPas on TRN models [3], which could also be useful.

Results

Firstly, we calculated extreme pathways of the three biological networks mentioned above. Since the number of ExPas grows exponentially with a networks' complexity [15], the enumeration of ExPas on the highly complex ones such as *E.coli* MN and TTN is computationally intractable. Fortunately ExPa calculation will be much more manageable if a MN or a TTN is divided into smaller sub-networks. Therefore, we chose the sub-networks with relatively complete and independent functions as the representatives of their belonging biologic systems. For the *E.coli* MN, two sub-networks were chosen: (1) Amino acid, Carbohydrate and Lipid metabolism (sACL) and (2) Membrane and Murein metabolism (sMM). For *E.coli* TTN, the two sub-networks were: (1) transcription (sTC) and (2) translation (sTL).

Then ExPa analysis was performed on each network/subnetwork and the properties from different aspects were obtained, including the total number of ExPas, the number-based ratio of ExPa to reaction, ExPa length distribution, reaction participation distribution, correlated reaction sets (CoSets), and the inter-connectivity

of ExPas. Finally, a horizontal comparison on the properties was made among the five networks/subnetworks.

Moreover, some incompleteness and incorrectness in the *E.coli* TRN model which were stumbled through ExPa analysis are also reported in this section. This findings illustrate that ExPa analysis is capable of directing model refinement.

E.coli TRN model

The *E.coli* TRN model was published by Gianchandani et al. in 2009 [3]. It contains 147 environmental stimuli, 125 transcriptional factors and 503 downstream target genes which are represented in a matrix $\mathbf{R}^*$ [3]. The TRN model was improved to enhance the efficiency of ExPa calculation (Details are provided in *Materials and Methods*). The final TRN model contains 1009 components, 1106 internal regulatory reactions, and 1009 exchange reactions each corresponding to a component. All the extracellular metabolites were considered as inputs and all protein products were considered as outputs. There were 1599 ExPas, of which 9 were biologically infeasible because they employed conflicting input fluxes, and thus they were excluded from the ExPa set used in analysis.

In *E.coli* TRN, 16 reactions do not participate in any ExPa; namely they are never used to form a transcriptional state of the network. These unused reactions were categorized into two types as listed in Table 1 and Table 2 respectively.

Reactions in Table 1 all relate to *NOT_BirA* (absence of protein *BirA*). However, no regulatory rule corresponds to the presence or absence of *BirA*, and therefore, the initial steps are unknown. As a result, the internal reactions using *NOT_BirA* (*b0774_1*, *b0775_1*, *b0776_1* and *b0778_1*) and the corresponding exchange reactions (*Ex_b0774*, *Ex_b0775*, *Ex_b0776* and *Ex_b0778*) will never be initiated. Furthermore, proteins *BirA* and the

Table 1 Unused reactions in the *E.coli* TRN (Type I - Regulatory rules missing).

Reactions	Reuglatory rules	Reaction type
NOT_BirA	–	internal
b0774_1	NOT(BirA)	internal
b0775_1	NOT(BirA)	internal
b0776_1	NOT(BirA)	internal
b0778_1	NOT(BirA)	internal
Ex_b0774	–	exchange
Ex_b0775	–	exchange
Ex_b0776	–	exchange
Ex_b0778	–	exchange

NOT_BirA represents the regulatory reaction that leads to the inhibition of gene transcription generating BirA. '-' in the second column represents the deficiency of corresponding regulatory rules in the model.

Table 2 Unused reactions in the *E.coli* TRN (Type II - Contradictory regulatory rules)

Reactions	Regulatory rules	Reaction type
b1814	(gly(e)>0) **AND** (leu-L(e)>0) leu-L(e)>0 (**NOT** o2(e)>0) **AND** (leu-L(e)>0) (gly(e)>0) **AND** (**NOT** leu-L(e)>0) **AND** (leu-L(e)>0) (leu-L(e)>0) **AND** (**NOT** leu-L(e)>0) (NOT o2(e)>0) **AND** (**NOT** leu-L(e)>0) **AND** (leu-L(e)>0)	Internal
b3942_1	(Growth>0) **AND** (h2o2(e)>0) **AND** (**NOT** Growth>0)	Internal
b4111_1	(**NOT**(Crp)) **AND** (Growth>0) **AND** (**NOT** Growth>0)	internal
Ex_b3942	–	Exchange
Ex_b4111	–	Exchange

Crp appears in the second column is a transcription factor and others are extracellular metabolites. The contradictory regulatory rules are labeled with shade in the table and these reactions never occur.

gene products of *b0774, b0775, b0776* and *b0778* do not participate in any other reaction except those in Table 1, so their invalidation will not affect other reactions in the network. In a word, these 9 reactions do not participate in any ExPa because their relevant reactions (either producing their substrates or consuming their products) are unavailable in the network. The unused reactions in Table 1 show the incompleteness of the *E.coli* TRN model and necessitate further refinement.

For the reactions in Table 2, the regulatory rule of *b1814* can be divided by simple logical transformation into 6 rules, of which 3 contradict with each other (the shaded parts in Table 2). Since there are still 3 operational regulatory rules relating to the transcription of *b1814*, its corresponding exchange reaction can be initiated. Similarly, the regulatory rules of *b3942* and *b4111* are both contradictory and cannot be used in any ExPa. These reactions may imply some incorrect information in the model. Therefore, new biological knowledge is needed to improve *E.coli* TRN.

E.coli MN and TTN model

The MN model of *E.coli* K-12 MG1655, iAF1260, was published by Feist et al, in 2007 [26]. It includes the activities of 1260 open reading frames (ORFs). It consists of 1688 metabolites and 2382 reactions. The *E.coli* TTN model was published by Thiele et al. in 2009 [4]. It consists of 11991 components and 13694 reactions which give rise to 423 functional gene products [4]. Given the critical inherent problem of combinatorial explosion during ExPa calculation, *E.coli* MN and TTN were divided into small sub-networks depending on the reactions' functions [11]. Sub-networks as representatives of important biological processes were chosen.

The *E.coli* MN was divided into 6 discrete sub-networks with different functions: one for exchange reactions which transfer metabolites in and out of the metabolic system and the others for internal reactions. Each reaction was assigned to one of the six sub-networks, whose details are listed in Table 3. Two sub-networks, Amino acid, Carbohydrate and Lipid metabolism (sACL) and Membrane and Murein metabolism (sMM), lie in the central part of *E.coli* MN and form the basis of other biological processes, and therefore they were chosen as the representatives of *E.coli* MN for ExPa analysis.

The *E.coli* TTN model comprises of 27 biological processes and the details are provided in [4]. Each process was treated as a discrete sub-network. The largest two sub-networks, *Transcription* and *Translation*, were chosen for further ExPa analysis.

ExPa counting

The total numbers of ExPas and the number-based ratios of ExPa to reaction (P/R) are listed in Table 4. P/R depicts the proportionality of the numbers of ExPas and reactions in a network. Table 4 shows that the P/Rs of sACL (33.44) and sMM (32.40) are much higher than those of TRN (0.75), sTC (0.12) and sTL (0.25), which are a consequence of the linear structures of TRNs and TTNs [3,4]. In contrast, MNs are in more complex interconnection with a large number of alternative pathways, and thus their P/Rs are much higher. The redundancy of ExPas increases a metabolic system's flexibility and fitness to sudden environmental changes [23,27]. These results illustrate the fundamental differences in topological structure and redundancy among the three types of networks.

ExPa length

The length of an ExPa equals to the number of reactions that participate in it [13]. Figure 1 shows the histograms of ExPa length distribution for each network/sub-network above. The details are listed in Table 5.

The length distributions of ExPas corresponding to those biological processes are very diverse. The longest ExPas consists 51, 82 32 and 109 reactions in sACL, sMM, sTC and STL, respectively, which is much longer than that in TRN (21). Reactions in *E.coli* TRN represent transcriptional regulatory rules rather than real biochemical reactions as in MN and TTN, and thus the ExPa length in TRN depicts the number of regulatory rules used for expressing certain genes. A regulatory rule describes how environmental stimuli affect transcriptional factors, which in turn affect downstream target genes. Therefore, the ExPa in TRN is reasonably shorter as the biological network has a relatively flat hierarchical structure [3]. Given the number of reactions, the ratio of average ExPa length to reaction

Table 3 Sub networks of the *E.coli* MN

# Sub network	Name	Containing subsystems	No. of reactions
1	Amino acid, Carbohydrat and Lipid metabolism (sACL)	Alanine and Aspartate Metabolism, Cysteine Metabolism, Folate Metabolism, Glutamate Metabolism, Glycine and Serine Metabolism, Histidine Metabolism, Methionine Metabolism, Threonine and Lysine Metabolism, Tyrosine Tryptophan and Phenylalanine Metabolism, Valine Leucine and Isoleucine Metabolism, Arginine and Proline Metabolism, Citric Acid Cycle, Pentose Phosphate Pathway, Pyruvate Metabolism, Glyoxylate Metabolism, Methylglyoxal Metabolism, Glycolysis and Gluconeogenesis, Oxidative Phosphorylation, Glycerophospholipid Metabolism, Anaplerotic Reactions	741
2	Nitrogen and Nucleotide metabolism	Cofactor and Prosthetic Group Biosynthesis, Nitrogen Metabolism, Nucleotide Salvage Pathway, Purine and Pyrimidine Biosynthesis, tRNA Charging	445
3	Membrane and Murein metabolism (sMM)	Cell Envelope Biosynthesis, Lipopolysaccharide Biosynthesis Recycling, Membrane Lipid Metabolism, Murein Biosynthesis, Murein Recycling	358
4	Transport	Transport Inner Membrane,Transport Outer Membrane,Transport Outer Membrane Porin	849
5	Others	Alternate Carbon Metabolism; Inorganic Ion Transport and Metabolism; Unassigned	408
6	Exchange	Exchange reactions	304

number (L/R) was calculated for each biological network or subnetwork (Table 5). The L/Rs of the two representatives in MN are higher than those in TRN and their counterparts in TTN. Since ExPas convert substrates into products, ExPa length relates to how many reaction steps are needed to carry out the corresponding function. ExPa length can be characterized as the size and complexity of the corresponding flux distribution map [13]. The results indicate that the flux distribution map in MN is much more complex than those in TRN and TTN.

Reaction participation

The reaction participation rate (RPR) is defined as the percentage of ExPas in which a given reaction participates [13]. Figure 2 shows the distribution of RPRs for each biological network/sub-network. Most reactions participate in less than 10% of ExPas, especially in TRN, sTC and sTL, but a few active reactions participate in many ExPas. Although the high-RPR reactions are most exchange reactions, some of them are internal reactions which usually play a more important role in determining the phenotypic potentials of the five biological processes. Given this, RPR can be reasonably considered as a metric for evaluating the importance of a reaction to implement the corresponding biological function [13].

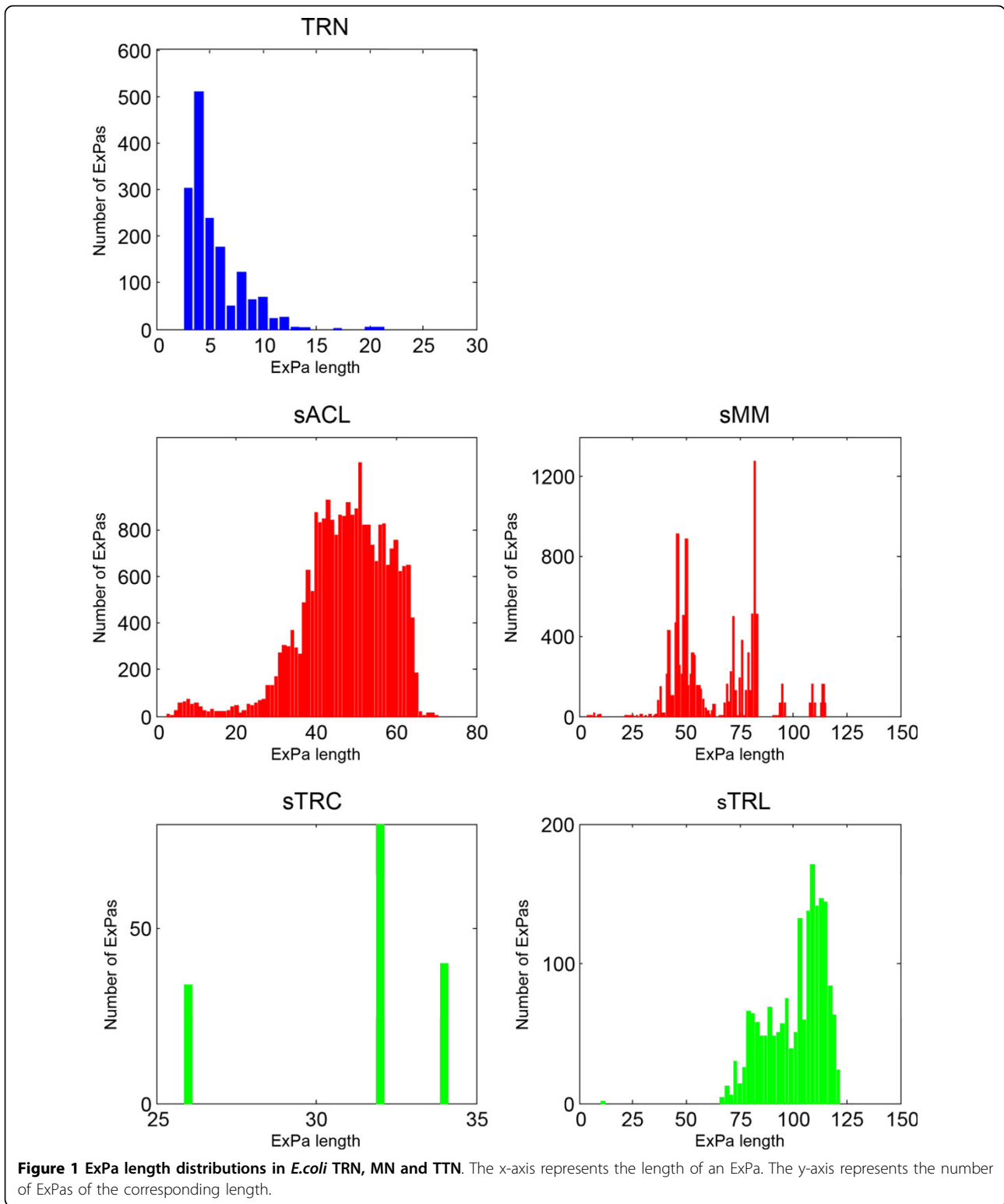**Table 4 Network characteristics and ExPa calculation results**

Network name	No. of reaction	No. of ExPa	P/R ratio
TRN	2115	1590	0.75
ACL	741	24778	33.44
sMM	358	11600	32.40
sTC	1276	154	0.12
sTL	7386	1871	0.25

Abbreviations for subsystems: TRN, transcriptional regulation; sACL, amino acid, carbohydrate and lipid metabolism; sMM, membrane and murein metabolism; sTC, transcription; sTL, translation.

Here the top 10 internal reactions with the highest RPRs of each process are sorted in a descending order (Table 6). Several reactions of vital importance were found, and representatives were chosen for detailed study.

In TRN, the two most active reactions *CRP_noGLC_1* and *Crp_1* relate to the regulation rules of the transcription factor (TCF) C-reactive protein (CRP). Other high rank reactions *Fis_1, Lrp_1, Fnr_1*, and *NOT_ArcA_1* relate to the regulation rules of the TCFs Fis, Lrp, Fnr and ArcA, respectively. In *E.coli*, the above TCFs belong to the seven global regulators that control most of the regulated genes [28]. The reaction *NOT_Cra_1* is relevant to the regulation rules of the TCF Cra, a pleiotropic regulatory protein that controls carbon and energy fluxes in enteric bacteria [29,30]. The reaction *NOT_PdhR_1* concerns the regulation rules of PdhR, a TCF that controls the respiratory electron transport system in *E.coli*. Its regulation target, the pyruvate dehydrogenase (PDH) multienzyme complex, plays a key role in the metabolic interconnection between glycolysis and the citric acid cycle [31].

In sACL, the most active reaction is *ASPTA*. It transfers oxoglutarate and aspartate to corresponding ketoacid, which are indispensable in glyoxylate cycle, an anabolic metabolic pathway occurring in *E. coli* [32]. The second one is *ASAD* which is the second step in the biosynthesis of amino acids in prokaryotes, fungi, and some higher plants. *ASAD* forms an early branch point in the metabolic pathway producing lysine, methionine, leucine and isoleucine from aspartate as well as diaminopimelate which plays an essential role in bacterial cell wall formation [33]. Deletion of gene *asd* (encoding *ASAD*) is lethal to the organism as demonstrated by experiments with *Legionella pneumophila, Salmonella typhimurium*, and *Streptococcus mutans*, which indicates that *ASAD* may also be an essential reaction in the metabolism of *E.coli* [34]. Another active reaction is *ASPK*, which is the

Figure 1 ExPa length distributions in *E.coli* TRN, MN and TTN. The x-axis represents the length of an ExPa. The y-axis represents the number of ExPas of the corresponding length.

commitment step in the pathway to the synthesis of lysine, methionine, threonine and isoleucine.

In sMM, the reaction *ACCOAC* is most active. It is a rate-determining step in the fatty acid synthetic pathway

and may play a pivotal role in regulating fatty acid oxidation [35]. The second most active reaction *MCOATA* transfers Malonyl CoA to acyl-carrier proteins (ACPs). The product Malonyl ACP provides malonyl groups for

Table 5 Summary of the statistical analysis of ExPa lengths

Network	No. of reactions	ExPa length				L/R Ratio(%)
		Avg	Max	Min	Most	
TRN	2115	5	21	3	4	0.24
sACL	741	47	70	3	51	6.40
sMM	358	64	115	4	82	17.97
sTC	1276	31	34	26	32	2.43
sTL	7386	101	121	11	109	1.37

Abbreviations for subsystems: TRN, transcriptional regulation; sACL, amino acid, carbohydrate and lipid metabolism; sMM, membrane and murein metabolism; sTC, transcription; sTL, translation. 'Most' represents the ExPa length with the highest frequency. L/R ratio is the ratio of average ExPa length to reaction amount.

biosynthesis of fatty acid and polyketide. On the other hand, Malonyl CoA, the substrate of *MCOATA*, is a highly-regulated molecule in fatty acid synthesis as it inhibits the rate-limiting step in beta-oxidation of fatty acids [36]. Flux change in *MCOATA* affects the consistency of Malonyl CoA and guarantees the biosynthesis of fatty acid.

In sTC, all the top reactions relate to the formation of the transcription elongation complex, an extremely complicated and highly regulated molecular machine that can sense signals coming from numerous regulatory protein factors, as well as those encoded in the DNA sequence. They are the basis of transcription elongation, because transcription can run smoothly and continuously only depending on their precise work.

In sTL, the reactions *IF2_RECHARG*, *Rib_30_ini_-FORM* and *Rib_70_DISS* are used by all ExPas. *IF2_RECHARG* recharges the initiation factor 2 (IF2) with GTP and *Rib_30_ini_FORM* produces 30S translation initiation complex which consists of 30S subunit, IF1, IF2-GTP and IF3. In bacteria, the correct mRNA starting site and the reading frame are selected when, with the help of IF1, IF2 and IF3, the initiation codon is decoded in the peptidyl site of the 30S ribosomal subunit by the anticodon fMet-tRNAfMet. Furthermore, *Rib_30_ini_-FORM* is also proved to be the intermediate step in the formation of 70S initiation complex (70SIC) which regulates translation initiation, the rate-limiting step in protein synthesis [37]. The other reaction *Rib_70_DISS* dissociates 70S ribosomes to 30S ribosomal subunit/IF1/IF3 complex (rib_30_IF1_IF3) and 50S ribosomal subunit (rib_50_inact). This is an essential step before a ribosome can participate in a new round of translation since the initiation complex for protein synthesis involves a 30S subunit. The dissociation of 70S ribosomes contributes to the efficiency and sustainability of protein synthesis [38].

Reportedly, RPRs help to find important reactions in MN [13]. Our results further indicate that RPR can also be extended to TRN and TTN to evaluate the relative importance of a given reaction.
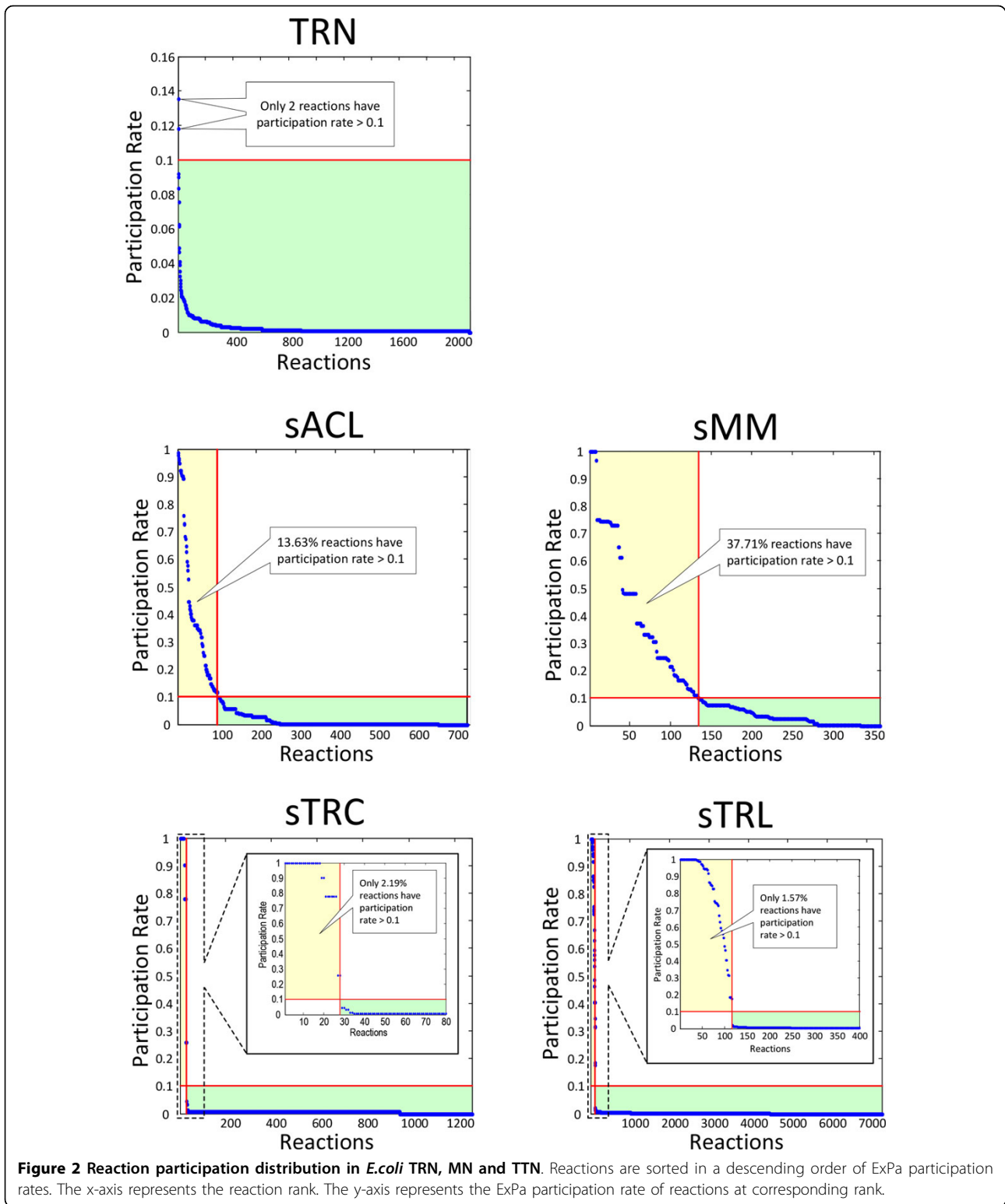
Correlated reaction set

A correlated reaction set (CoSet) comprises reactions that always participate in the same ExPa set in a given network [13]; namely if one reaction functions, the others in the same CoSet function simultaneously.

A CoSet can be transformed to a graph by treating each reaction as a node and adding an edge between two reactions that involve a common substance. In a certain CoSet, some member reactions are topologically connected while others are not. The correlationship of the second type of reactions often indicates a transcriptional coregulation by the corresponding genes [11] while that of the first type has relatively trivial biological meaning. Therefore, a CoSet is defined as a trivial set if all its member reactions are connected in topology. A trivial CoSet provides less novel information, and thus it is unworthy of deep study. In this paper, the adjacent ratio is used to represent the percentage of trivial CoSets.

CoSets were calculated for each biological network/sub-network about which several features, including the adjacent ratio, were stretched and shown in Table 7. The adjacent ratios of TRN, sTC and sTL are much higher than those of sACL and sMM, which indicates that almost all the CoSets obtained in the former three networks are due to the linear structure. For the metabolic netowrk, more CoSets consist of reactions which are not adjacent in topology. The results suggest that CoSet analysis may be more useful in study of MNs.

Crosstalk analysis

Crosstalk analysis was first raised to illustrate the relationship between multiple inputs or outputs of a signaling pathway [39]. The whole ExPa set was compared pairwise to build the simplest form of crosstalk [2,10]. A pair of ExPas may have identical, overlapped or disjoint inputs (or outputs). There are 9 categories of crosstalk with their biological meanings described in [10]. Here, crosstalk analysis is applied to other biological processes to detect the relationships between fundamental functional states. Various forms of crosstalk in the five networks/sub-networks above were characterized. As several exchange reactions participate in most ExPas of sACL, MN, sTC and sTL, almost all of the ExPa pairs have overlapped inputs or outputs. A close look at the highly participating exchange reactions reveals that most of them relate to small molecules such as H_2O, ATP and NADP commonly seen in various biochemical reactions. In order to further elucidate the difference in crosstalk between ExPa pairs, all the exchange reactions in the four sub-networks were sorted in a descending order depending on RPR and the top 20%

Figure 2 Reaction participation distribution in *E.coli* TRN, MN and TTN. Reactions are sorted in a descending order of ExPa participation rates. The x-axis represents the reaction rank. The y-axis represents the ExPa participation rate of reactions at corresponding rank.

ExPa pairs were neglected in the subsequent crosstalk analysis.

As shown in Figure 3, more than 90% of the ExPa pairs have disjoint inputs and disjoint outputs in TRN, sTC and sTL in contrast to sACL and sMM. A higher disjoint input/disjoint output rate implies that each ExPa has more specific functions and cannot be replaced easily by others. This indicates that the

Table 6 The top 10 most frequently participated internal reactions

Part 1

The top 10 reactions with highest participation rate in the TRN

Order	Reaction	Participation Rate (%)
1	CRP_noGLC_1	9.18
2	Crp_1	8.99
3	NOT_PdhR_1	7.55
4	Fis_1	6.16
5	Lrp_1	4.09
6	Fnr_1	3.90
7	NOT_PurR_1	3.27
8	NOT_ArcA_1	3.02
9	NOT_Cra_1	2.83
10	NOT_Lrp_1	2.45

The top 10 reactions with highest participation rate in the sACL

Order	Abbr.	Reaction Name	Participation Rate (%)
1	ASPTA	aspartate transaminase	91.42
2	ASAD	aspartate semialdehyde dehydrogenase	90.40
3	ASPK	aspartate kinase	90.40
4	HSDy	homoserine dehydrogenase NADPH	90.31
5	GHMT2r	glycine hydroxymethyltransferase reversible	75.82
6	GLYAT	glycine C acetyltransferase	68.22
7	HSK	homoserine kinase	67.27
8	THRS	threonine synthase	67.27
9	THRD	L threonine dehydrogenase	59.23
10	FUM	Fumarase	57.71

The top 10 reactions with highest participation rate in the sMM

Order	Abbr.	Reaction Name	Participation Rate (%)
1	ACCOAC	acetyl CoA carboxylase	75.08
2	MCOATA	Malonyl CoA ACP transacylase	75.08
3	3HAD100	3 hydroxyacyl acyl carrier protein dehydratase n C100	74.48
4	3HAD40	3 hydroxyacyl acyl carrier protein dehydratase n C40	74.48
5	3HAD60	3 hydroxyacyl acyl carrier protein dehydratase n C60	74.48
6	3HAD80	3 hydroxyacyl acyl carrier protein dehydratase n C80	74.48
7	3OAR100	3 oxoacyl acyl carrier protein reductase n C100	74.48
8	3OAR40	3 oxoacyl acyl carrier protein reductase n C40	74.48
9	3OAR60	3 oxoacyl acyl carrier protein reductase n C60	74.48
10	3OAR80	3 oxoacyl acyl carrier protein reductase n C80	74.48

Part 2

The top 10 reactions with highest participation rate in the sTC

Order	Abbr.	Reaction Name	Participation Rate (%)
1	tscr_elo_TU-8389_ini	Formation complex for elongation of TU-8389'	0.65
2	tscr_elo_TU-8390_ini	Formation complex for elongation of TU-8390'	0.65
3	tscr_elo_TU-8397_ini_rho_dep	Formation complex for elongation of TU-8397 (RHO DEPENDENT TERMINATION)'	0.65
4	tscr_elo_TU-8407_ini_rho_dep	Formation complex for elongation of TU-8407 (RHO DEPENDENT TERMINATION)'	0.65
5	tscr_elo_TU0-1181_ini_stab	Formation complex for elongation of TU0-1181 (stable RNA)'	0.65
6	tscr_elo_TU0-1182_ini_stab	Formation complex for elongation of TU0-1182 (stable RNA)'	0.65
7	tscr_elo_TU0-1183_ini_stab	Formation complex for elongation of TU0-1183 (stable RNA)'	0.65
8	tscr_elo_TU0-1186_ini_stab	Formation complex for elongation of TU0-1186 (stable RNA)'	0.65
9	tscr_elo_TU0-1187_ini_stab	Formation complex for elongation of TU0-1187 (stable RNA)'	0.65
10	tscr_elo_TU0-1189_ini_stab	Formation complex for elongation of TU0-1189 (stable RNA)'	0.65

Table 6 The top 10 most frequently participated internal reactions (Continued)

	The top 10 reactions with highest participation rate in the sTL		
Order	Abbr.	Reaction Name	Participation Rate (%)
1	IF2_RECHARG	recharge of IF2 with GTP'	100
2	Rib_30_ini_FORM	formation of 30S translation initiation complex (30S subunit, IF1, IF2-GTP, IF3)'	100
3	Rib_70_DISS	70S ribosome dissociation'	100
4	EF-G_RECHARG	recharge of EF-G with GTP'	99.95
5	tl_elo_b3461_16_rib1	Translation elongation 1 b3461 16 ribosome(s)'	0.75
6	tl_elo_b3461_16_rib2	Translation elongation 2 b3461 16 ribosome(s)'	0.75
7	tl_elo_b3461_1_rib1	Translation elongation 1 b3461 1 ribosome(s)'	0.75
8	tl_elo_b3461_1_rib2	Translation elongation 2 b3461 1 ribosome(s)'	0.75
9	tl_elo_b3461_8_rib1	Translation elongation 1 b3461 8 ribosome(s)'	0.75
10	tl_elo_b3461_8_rib2	Translation elongation 2 b3461 8 ribosome(s)'	0.75

biological processes in *E. coli* TTN and TRN are more deterministic than those in MN. Reportedly, a large number of genes are regulated by only a few independent regulatory rules in *E.coli* TRN [3], and the majority of the associated functions in *E.coli* TTN have only one coding gene in the genome [4]. These facts indicate that the specificity of TRN and TTN is much higher than MN. In order to function normally, cells have to respond accurately to the environmental signals with the help of precise transcriptional regulations and subsequently produce necessary gene products through accurate transcription and translation systems.

Except sTC, the other networks/sub-networks all have ExPa pairs with identical inputs and identical outputs. These ExPas are redundant pathways which fulfill completely identical function through systemically independent routes. ExPa redundancy was demonstrated in genome-scale MNs [23,24], as well as a prototypic signaling network [10] and the JAK-STAT signaling network [2]. The redundant ExPas in *E.coli* TRN can be attributed to the fact that the transcription of some genes can be stimulated by different transcriptional factors. For example, two redundant ExPas shown in Figure 4 stimulate the expression of gene *b2243* in the same environment, but they employ the regulatory rules of '*CRP_noRIB AND Fnr AND NOT(GlpR)*' and '*CRP_noRIB AND ArcA AND NOT (GlpR)*', respectively. From Figure 3, the percentage of

ExPa pairs with overlapped inputs and overlapped outputs in the biological processes of MN is much higher than those in TTN and TRN. These results indicate that *E.coli* MN is more flexible than TTN and TRN.

Discussion

ExPa analysis were applied to two new models, the *E. coli* TRN and TTN. A horizontal comparison was performed for the five networks/sub-networks: TRN, sACL, sMM, sTC, sTL from five aspects: (1) Total number of ExPas and the P/R ratios; (2) ExPa length distribution and L/R ratios; (3) Reaction participation rates; (4) Correlated reaction sets and adjacent ratios; (5) Inter-connectivity of ExPas.

Reactions in TTN represent actual biochemical reactions like those in MN, and thus, ExPas in TTN characterize the steady-states of the corresponding biological systems. In contrast, columns in TRN represent the transcriptional regulatory rules and coefficents only reflect the qualitative information describing the presence or absence of the corresponding components rather than the quantitative information describing reaction stoichiometries as in TTN and MN. Therefore, an ExPa in TRN characterizes a specific transcriptional regulatory state, namely which transcriptional regulatory rules are activated and which genes are expressed in a specific environmental state.

Table 7 Summary of CoSets

Network	No. of reactions	No. of CoSet	Size of CoSet				Adjacent Ratio (%)
			Max	Min	Avg	Most	
TRN	2115	481	8	2	2	2	98
sACL	741	157	11	2	2	2	78
sMM	358	49	25	2	4	3	47
sTC	1276	161	18	2	6	6	93
sTL	7386	1718	23	2	2	2	85

Adjacent ratio is defined as the percentage of such CoSets in which all the reactions are topologically connected, that is, if all the reactions in a CoSet form a graph, the graph only contains one connected component.

Figure 3 Crosstalk analysis of *E.coli* TRN, MN and TTN. Given a pair of ExPas, the relationship of their input sets falls into one of the following three cases: disjoint, partially overlapped and identical. And so does that of their output sets. Thus, all pairs of ExPas can be classified into 9 categories according to the relationship of input/output sets. Classification results of different networks are shown in this figure by 3x3 matrices in which the number in each cell represents percentages of ExPa pairs falling into this category.

Figure 4 Example ExPas with identical inputs and identical outputs. The ExPa is represented by a DAG (Directed Ascyclic Graph), in which each node represents a component and each edge represents a reaction. The shaded part indicates the difference between two ExPas.
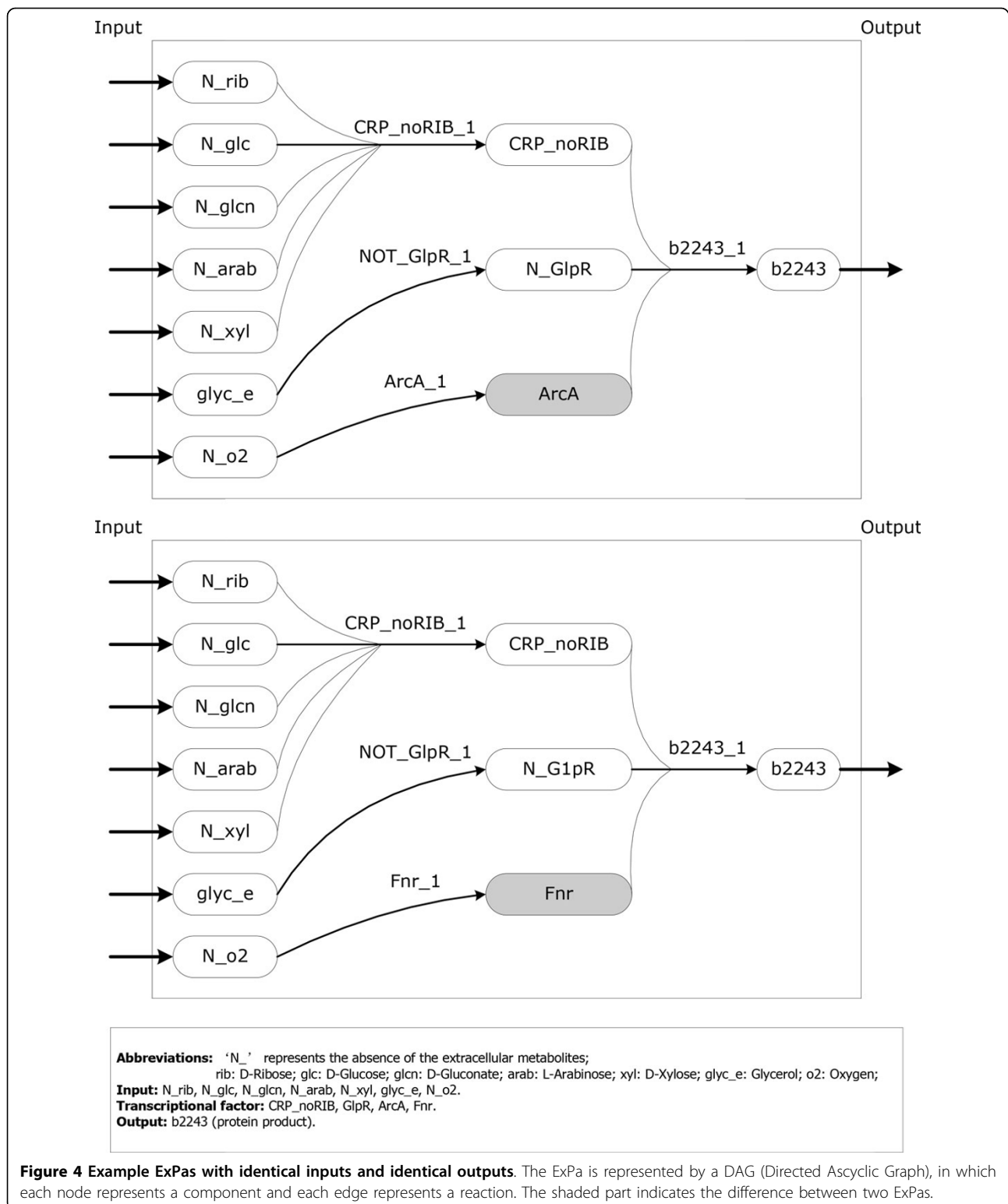
ExPa analysis emphasizes the functional and systemic properties of biologcial process as ExPas are systemically independent functional units. The total number of ExPas and the P/R ratios characterize the flexibility of the networks/sub-networks. ExPa length corresponds to the reaction steps needed to form a steady state, therefore showing a close relation to network complexity. Crosstalk enables the analysis of pathway redundancy and network

determinacy. Comparisons from these aspects indicate that MN is more flexible but less deterministic than TRN and TTN. Environmental cues affect transcriptional regulation, which controls the following transcription and translation processes. Then the resulting gene products (enzymes) enter the metabolic system to catalyze the corresponding reactions. It is necessary for a cell to respond accurately to the environment and produce the required enzymes. MN is more robust to environmental changes, which reflects the struggle of a cell to achieve an alternative steady-state to provide substance support for TRN and TTN and maintain life.

The distributions of reaction participation in the five networks/sub-networks are similar except that there are more reactions participating in more than 10% ExPas in sACL and sMM. Only a small percent of the reactions participate in a large number of ExPas, which indicates the phenotypic potentials of TRN, TTN and MN are affected greatly by a small number of important reactions. Evaluations on the representatives show that reactions with high participation rates often play an important role in certain biological processes. These reactions are the relatively weak part of the networks because a large number of ExPas will be destroyed when these reactions become invalid, which may cause the loss of various functions. These reactions may be used as drug targets and further direct the design of new drugs.

CoSets were identified via the calculation of reaction participation. Besides the expected topological connections, the topologically unconnected reactions in a CoSet may indicate the information of transcriptional coregulation in MN. However, most Cosets of TRN and TTN are trivial, and thus have few chances to be a clue giving novel information like in MN.

Last but not least, an improved approach was introduced to calculate the ExPas on TRN models. Compared to the existing method, the biggest advantage of ours is the high efficiency in calculating all the extreme pathways of a TRN, especially for the one which may work under huge amount of environmental conditions. For example, the *E.coli* TRN model which we studied in the paper has 776 components whose availability (i.e., presence or absence) constitute the environmental condition, including environmental stimuli, transcription factors or proteins. It is impossible to enumerate all the possible conditions due to "combination explosion" without mentioning the calculation of the ExPas under each condition. However, using the approach we proposed, it took only about 45 seconds to computing the whole ExPa set on a PC with four 3.2-GHz Intel(R) XEON processors and 16GB RAM (in fact, only one processor and 15MB RAM are used for the calculation).

We believe that this approach could be helpful for readers who are also interested in the ExPas of TRNs.

Conclusions

This study presents the first horizontal comparison among the E.coli TRN, MN and TTN through ExPa analysis. The results show that ExPa also has biological meanings in TRN and TTN. Different properties of ExPas reflect the biological nature of each biological process. Along with the the increase of reconstructed models on TRNs and TTNs as well as the development of new methods, ExPa analysis may reveal more biological properties and get larger space of application in the medical and biochemical fields.

Methods

COBRA framework and ExPa analysis

The COBRA framework stoichiometrically represents a biochemical network as a matrix $\mathbf{S}$, whose rows and columns correspond to components and reactions respectively. COBRA is capable of predicting and understanding the achievable cellular function, namely the phenotypic behavior of a biochemical network. With the hypothesis of steady state and certain constraints, all possible flux distributions lie in the null space of $\mathbf{S}$:

$$\mathbf{S}\mathbf{v} = \mathbf{0}, \ v_i^{\min} \leq v_i \leq v_i^{\max}, \ i = 1, \cdots, n$$

where $\mathbf{S}_{m \times n}$ is the stoichiometric matrix of a biochemical network with m components and n reactions and $\mathbf{v}_{n \times 1}$ is a vector of the fluxes through each reaction in the system [40].

Given the reversibility of reactions, an internal reversible reaction can be divided into a forward and a backward sub-reactions, each taking a non-negative flux. The model's solution space is now a convex polyhedral cone in high-dimensional space [19,40], which can be demarcated by an ExPa set $\mathbf{p}_i(i = 1, \cdots, k)$ [11,41]. All steady-states lie in the cone and each can be represented by a nonnegative linear combination of ExPas:

$$\mathbf{v} = \sum \alpha_i \mathbf{p}_i \quad \text{where} \quad \alpha_i \geq 0$$

For a given network, the ExPa set has the following properties: (1) It is unique; (2) Each ExPa uses fewest reactions to form a function unit; (3) It is systemically independent which means an ExPa cannot be represented by a nonnegative linear combination of other ExPas [42,43].

ExPa calculation on the MN and TTN

ExPas were calculated using an open source tool '*expa*' [44]. The *E.coli* MN and TTN models were divided into small sub-networks using the method proposed in [11].

## An improved approach to compute the ExPas of TRN models

A TRN is composed of a set of transcriptional regulatory rules which describe cells' transcriptional responses to environmental signals. A regulatory network matrix $\mathbf{R}$ was used by Gianchandani et al. to represent the components (environmental cues, metabolites, genes and proteins) and reactions (regulatory rules and exchange reactions of products) of a TRN [3]. It was further combined with an environmental matrix $\mathbf{E}$, which characterizes a particular environmental state, yielding a complete regulatory state matrix $\mathbf{R}^* = [\mathbf{R}|\mathbf{E}]$. Each column of $\mathbf{E}$ delineates the availability of a unique environmental cue, transcription factor, target gene or protein [3,45]. Different environmental states correspond to different $\mathbf{E}$s, thus forming different $\mathbf{R}^*$s.

For example, given a toy TRN with three regulatory rules:

A + B → Protein 1; C → Protein 2; D → Protein 2;
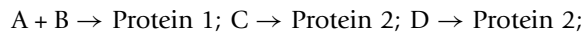
where A, B, C and D are four metabolites enacting as signalling stimuli.

The corresponding converses are:

A → Protein 1; B → Protein 1; C + D → Protein 2;
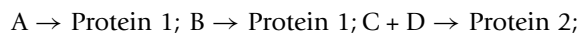
The matrix $\mathbf{R}^*$ is illustrated in Figure 5A under the environmental condition that A and D are present while B and C are absent. The shaded columns represent the inputs of environmental cues. Any steady state of TRN under the given environmental cues lies in the space which satisfies $\mathbf{R}^*\mathbf{v} = \mathbf{0}$ and $\forall i, v_i \geq 0$. The convex basis of the right null space of $\mathbf{R}^*$ forms the ExPa set under the given environmental state.

In order to calculate all the ExPas of the TRN, all the environmental states, namely all possible $\mathbf{E}$s, need be enumerated. Then ExPas participating in each possible environmental state are generated and the unique ones are grouped to form the complete ExPa set. Since the number of possible environmental states grows exponentially with the number of extracellular metabolites, it is inefficient to enumerate all possible environmental states for a TRN with numerous envionmental cues [45]. Therefore, an improved method is introduced here to simplify the ExPa calculation on the COBRA model of TRN.

The gist of the method is to improve Gianchandani's method by employing two columns instead of one to delineate the presence and absence of a unique envionment cue respectively, by which a new environment matrix $\mathbf{E}_{new}$ is constructed. The matrix $\mathbf{E}_{new}$ covers all possible environmental states. Without loss of generality, we assume that the top n rows in $\mathbf{R}$ and $\mathbf{E}_{new}$ represents the present state of n environmental inputs m_e ($e = 1, \cdots, n$) one-to-one and the following n rows

represents the absent state of them. The original regulatory state matrix is $\mathbf{R}^* = [\mathbf{R}|\mathbf{E}] = [\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_k|\mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}]$ and the new matrix is $\mathbf{R}^*_{new} = [\mathbf{R}|\mathbf{E}_{new}] = [\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_k|\mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}|\mathbf{r}_{k+n+1}, \mathbf{r}_{k+n+2}, \cdots, \mathbf{r}_{k+2n}]$ (k is the number of columns in $\mathbf{R}$, and $\mathbf{E}_{new} = [\mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}|\mathbf{r}_{k+n+1}, \mathbf{r}_{k+n+2}, \cdots, \mathbf{r}_{k+2n}]$). For an input m_e, column $\mathbf{r}_{k+e}$ represents its presence and column $\mathbf{r}_{k+n+e}$ represents its absence under the environmental condition, where $\mathbf{r}_{k+e}(e)$ and $\mathbf{r}_{k+n+e}(n+e)$ equal to 1 and the other elements are all zeros. For example, the $\mathbf{R}^*_{new}$ matrix of the above toy model is illustrated in Figure 5B. The shaded columns constitute $\mathbf{E}_{new}$. Obviously, the space and time complexity for constructing $\mathbf{E}_{new}$ is O(n), where n is the number of components of a TTN model. The convex basis of the right null space of $\mathbf{R}^*_{new}$ comprises the ExPa set of the TRN which could then be enumerated by the tool '*expa*' [44].

Notably, some infeasible steady states employing contradictory inputs may be involved in the right null space of $\mathbf{R}^*_{new}$. For example, Figure 6A shows an infeasible steady sate of the TRN described in Figure 5B. The two shaded elements of $\mathbf{v}$ both equal to 1. This means metabolite A is both present and abscent in the environment, which is obviously impossible. If an ExPa proves to be an infeasible steady state, it should be removed from the ExPa set.

Figures 6B and 6C show two ExPas resulting from the matrixs in Figures 5A and 5B respectively. The two vectors represent the same steady state of the TRN in which gene G1 is inhibited because of lack of metabolite B. In Figure 6B, the exact meaning of "1" in element B_{AV} cannot be decided directly from ExPa without referring to the shaded part of matrix in Figure 5A. However, in Figure 6C, "1" in column B_{AB} clearly means the absence of metabolite B. Namely, the interpretation of an ExPa resulting from the improved method is independent from the environmental matrix, which makes an ExPa easier to understand.

Validation of the approach of ExPa calculation on TRNs

Given n environmental cues, there are 2^n possible environmental states, each corresponding to a matrix $\mathbf{E}_i$ and the corresponding $\mathbf{R}^*_i$ ($\mathbf{R}^*_i = [\mathbf{R}|\mathbf{E}_i]$, $i = 1, \cdots, 2^n$). The ExPa set obtained from $\mathbf{R}^*_i$ is denoted as P_i and the feasible ExPa set calculated from $\mathbf{R}^*_{new}$ is denoted as P_{new}. Since the meaning of the environmental part of P_i is dependent on the environmental states, ExPas of different environmental states should be normalized to eliminate the dependence before being grouped up. We normalized a ExPa $\mathbf{p}^j_i$ in the set P_i by expanding its dimension of the input part from n ($\mathbf{p}^j_i$) to $2n$ ($\hat{\mathbf{p}}^j_i$). Details of the normalization are described in *Algorithm 1*.

Data: $\mathbf{p}^j_i$, $\mathbf{A}^i$, $\bar{\mathbf{A}}^i$.
// $\mathbf{p}^j_i$ represents the jth ExPa in the ith environment, where $i = 1, \cdots, 2^n$;

A

Inputs of environmental cues

	G1$_{AC}$	G2a$_{AC}$	G2b$_{AC}$	G1a$_{IN}$	G1b$_{IN}$	G2$_{IN}$	P1$_{PR}$	P2$_{PR}$	P1$_{AB}$	P2$_{AB}$	A$_{AV}$	B$_{AV}$	C$_{AV}$	D$_{AV}$
A$_P$	-1	0	0	0	0	0	0	0	0	0	1	0	0	0
B$_P$	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
C$_P$	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
D$_P$	0	0	-1	0	0	0	0	0	0	0	0	0	0	1
A$_a$	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
B$_a$	0	0	0	0	-1	0	0	0	0	0	0	1	0	0
C$_a$	0	0	0	0	0	-1	0	0	0	0	0	0	1	0
D$_a$	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
P1$_P$	1	0	0	0	0	0	-1	0	0	0	0	0	0	0
P2$_P$	0	1	1	0	0	0	0	-1	0	0	0	0	0	0
P1$_a$	0	0	0	1	1	0	0	0	-1	0	0	0	0	0
P2$_a$	0	0	0	0	0	1	0	0	0	-1	0	0	0	0

B

Inputs of environmental cues

	G1$_{AC}$	G2a$_{AC}$	G2b$_{AC}$	G1a$_{IN}$	G1b$_{IN}$	G2$_{IN}$	P1$_{PR}$	P2$_{PR}$	P1$_{AB}$	P2$_{AB}$	A$_{PR}$	B$_{PR}$	C$_{PR}$	D$_{PR}$	A$_{AB}$	B$_{AB}$	C$_{AB}$	D$_{AB}$
A$_P$	-1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B$_P$	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
C$_P$	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
D$_P$	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
A$_a$	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
B$_a$	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0
C$_a$	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	1	0
D$_a$	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1
P1$_P$	1	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
P2$_P$	0	1	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
P1$_a$	0	0	0	1	1	0	0	0	-1	0	0	0	0	0	0	0	0	0
P2$_a$	0	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0	0

Rows(components):

> **p:** presence of network component (in a regulatory reaction)
> **a:** absence of network component (in a regulatory reaction)

Columns (gene transcription rules, state of environmental conditions and protein products)

> **AC:** rule describing activation of transcription of gene
> **IN:** rule describing inhibition of transcription of gene
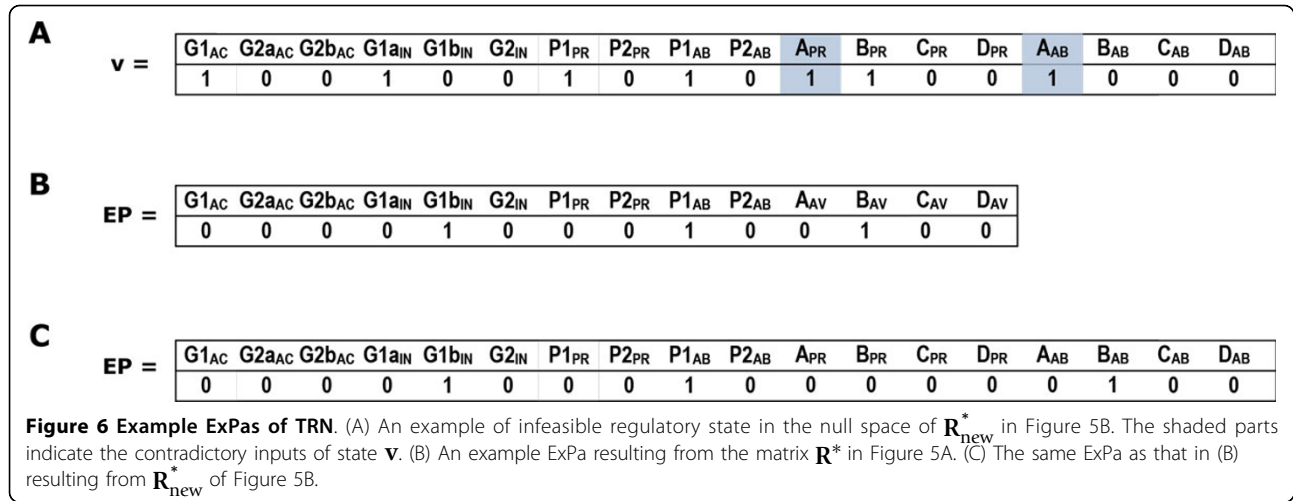> **PR:** presence of components within an environment
> **AB:** absence of components within an environment
> **AV:** availability (i.e., presence or absence) of components within an environment

Figure 5 Matrix formalism of the TRN model. (A) The regulatory state matrix $\mathbf{R}^*$ of the toy model in which the regulatory rules are: $C \rightarrow \text{Protein 2}; C \rightarrow \text{Protein 2}; D \rightarrow \text{Protein 2}$. The environmental state of $\mathbf{R}^*$ is that metabolites A and D are present while B and C are absent. (B) The corresponding $\mathbf{R}^*_{\text{new}}$ of $\mathbf{R}^*$.

// $\mathbf{p}_i^j = [v_1, v_2, \cdots, v_k | v_{k+1}, v_{k+2}, \cdots, v_{k+n}]$;
// $\mathbf{A}^i$ is a set which consists of all the absent inputs;
// $\bar{\mathbf{A}}^i$ is a set which consists of all the present inputs;

Result: $\hat{\mathbf{p}}_i^j$. // $\hat{\mathbf{p}}_i^j = [v_1, v_2, \cdots, v_k | v_{k+1}, v_{k+2}, \cdots, v_{k+n}, v_{k+n+1}, v_{k+n+2}, \cdots, v_{k+2n}]$;
For $q = 1$ **to** k **do** $\hat{\mathbf{p}}_i^j(q) = \mathbf{p}_i^j(q)$; **End for**
For $q = 1$ **to** n **do**

Figure 6 Example ExPas of TRN. (A) An example of infeasible regulatory state in the null space of $\mathbf{R}_{new}^*$ in Figure 5B. The shaded parts indicate the contradictory inputs of state **v**. (B) An example ExPa resulting from the matrix $\mathbf{R}^*$ in Figure 5A. (C) The same ExPa as that in (B) resulting from $\mathbf{R}_{new}^*$ of Figure 5B.

If $\mathbf{p}_i^j(k+q) == 0$ **do**
  $\hat{\mathbf{p}}_i^j(k+q) = 0; \hat{\mathbf{p}}_i^j(k+n+q) = 0;$
Else if $\mathbf{p}_i^j(k+q) == 1$ **do**
  If $q \in \mathrm{A}^i$ **do**
    $\hat{\mathbf{p}}_i^j(k+q) = 0; \hat{\mathbf{p}}_i^j(k+n+q) = 1;$
  Else if $q \in \bar{\mathrm{A}}^i$
    $\hat{\mathbf{p}}_i^j(k+q) = 1; \hat{\mathbf{p}}_i^j(k+n+q) = 0;$
  End if
End if
End for
Algorithm 1: Procedure of normalizing $\mathbf{p}_i^j$ to $\hat{\mathbf{p}}_i^j$ by dimension expanding.

In a normalized ExPa $\hat{\mathbf{P}} = [v_1, v_2, \cdots, v_{k+2n}]$, "1" on v_{k+e} $(e = 1, \cdots, n)$ indicates that m_e is present on the ExPa while "1" on v_{k+n+e} $(e = 1, \cdots, n)$ indicates that m_e is absent, and 0 indicates that m_e does not affect the transcriptional states characterized by this Expa. The normalized ExPa set of P_i is denoted as $\hat{P}_i$ and the union of $\hat{P}_i$ $(i = 1 \ldots 2^n)$ is denoted as $\hat{p}$. As explained above, the ExPas in set P_{new} are already in the normalized form, hence no normalization are needed.

Here we prove that $\hat{P}_{new}$ equals to $\hat{p}$:

Statement 1: each ExPa in $\hat{p}$ can be obtained from $\mathbf{R}_{new}^*$.

Proof: given n extracellular metabolites $m_p (p = 1, \cdots, n)$, each $\mathbf{R}_i^*$ can be transformed to $\hat{\mathbf{R}}_i^*$ as follows (**Algorithm 2**):

Data: $\mathbf{R}_i^*$, $\mathbf{A}^i$, $\bar{\mathbf{A}}^i$.
// $\mathbf{R}_i^*$ represents TRN in the ith environment, where $i = 1, \cdots, 2^n$;
// $\mathbf{R}_i^* = [\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_k | \mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}]$;
// $\mathbf{A}^i$ is a set which consists of all the absent inputs;
// $\bar{\mathbf{A}}^i$ is a set which consists of all the present inputs;
Result: $\hat{\mathbf{R}}_i^*$.
//
$\hat{\mathbf{R}}_i^* = [\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_k | \mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}, \mathbf{r}_{k+n+1}, \mathbf{r}_{k+n+2}, \cdots, \mathbf{r}_{k+2n}]$;
For $q = 1$ **to** $2n$ **do** $\mathbf{r}_{k+q} = \mathbf{0}$; **End for**

For $q = 1$ **to** n **do**
  If $q \in \mathbf{A}^i$ **do**
    $\mathbf{r}_{k+n+q}(n+q) = 1;$
  Else if $q \in \bar{\mathbf{A}}^i$
    $\mathbf{r}_{k+q}(q) = 1;$
  End if
End for
Algorithm 2: Procedure of transforming $\mathbf{R}_i^*$ to $\hat{\mathbf{R}}_i^*$.

For $\hat{\mathbf{R}}_i^* = [\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_k | \mathbf{r}_{k+1}, \mathbf{r}_{k+2}, \cdots, \mathbf{r}_{k+n}, \mathbf{r}_{k+n+1}, \mathbf{r}_{k+n+2}, \cdots, \mathbf{r}_{k+2n}]$ $(i = 1, \cdots, 2^n)$ resulted from Algorithm 2, if $\exists j \in \{k+1, \cdots, k+2n\}$ such that $\mathbf{r}_j = \mathbf{0}$, then a constraint $v_j = 0$ is added. Then the resulting network is a sub-network of that represented by $\mathbf{R}_{new}^*$. As proven in [46], **G** and **G'** are two MNs whose reactions are all irreversible and whose ExPa sets are EP and EP', respectively. If EP' is a sub-network of EP, then EP' $\subseteq$ EP. Therefore $\hat{P}_i \subseteq \hat{P}_{new}$, because $P = \bigcup_{i=1}^{n} \hat{P}_i$, $\hat{P} \subseteq \hat{P}_{new}$.

Statement 2: each feasible ExPa in $\hat{P}_{new}$ can be obtained by some $\hat{\mathbf{R}}_i^*$.

Proof: Since any environmental cue is impossible to be both present and absent in a specific environment, $v_{k+e} \times v_{k+n+e} = 0$ $(e = 1, \cdots, n)$ is true for each ExPa in P_{new}. For any ExPa $\mathbf{p} \in \hat{P}_{new}$, let $\mathbf{T} = \mathbf{R}_{new}^*$. For any e, **T** is modified as follows: (1) If $v_{k+e} = 0$ and $v_{k+n+e} \neq 0$, $\mathbf{t}_{k+e} = \mathbf{0}$; (2) If $v_{k+n+e} = 0$ and $v_{k+e} \neq 0$, $\mathbf{t}_{k+n+e} = \mathbf{0}$; (3) If $v_{k+e} = 0$ and $\mathbf{t}_{k+e} = \mathbf{0}$, $\mathbf{t}_{k+e} = \mathbf{0}$, where $\mathbf{t}_i$ is the ith column of **T**. As can be shown easily, **p** is an ExPa of the right null space of **T**. According to *Algorithm 2*, a legal $\hat{\mathbf{R}}_i^*$ contains one zero column and one non-zero column corresponding to the two input reactions of a certain input component respectively. Therefore, **T** is a legal $\hat{\mathbf{R}}_i^*$, and each ExPa in $\hat{P}_{new}$ can be obtained by some $\hat{\mathbf{R}}_i^*$, or in other words, $\hat{P}_{new} \subseteq \hat{P}$.

From statements (1) and (2), we conclude that $\hat{P}_{new} = \hat{P}$, and thus all possible ExPas of a TRN can be obtained using our new representation.

Classification of ExPas

ExPas fall into three classes, in which class III stands for internal reaction cycles with no exchange flux [12]. Class III ExPas were proven to be thermodynamically infeasible [47] and thus were not considered in our analysis.

List of abbreviations used

COBRA: Constraint-based Reconstruction and Analysis; MN: Metabolic Network; ExPa: Extreme Pathway; TRN: Transcriptinal Regulatory Network; TTN: Transcriptional and Translational Network; sACL: The subnetwork of Amino acid, arbohydrate and Lipid metabolism; sMM: The subnetowrk of Membrane and Murein metabolism; sTC: The subnetwork of Transcription in the TTN; sTL: The subnetwork of Translation in the TTN; ORF: Open Reading Frame; P/R: the Number-based Ratios of ExPa to Reaction; L/R: the Ratio of Average ExPa Length to Reaction Number; RPR: the Reaction Participation Rate; TCF: Transcription Factor; CRP: C-reactive Protein; PDH: Pyruvate Dehydrogenase; ACP: Acyl-carrier Protein; IF: Initiation Factor; 70SIC: 70S Initiation Complex; rib_30_IF1_IF3: 30S Ribosomal Subunit/IF1/IF3 Complex; rib_50_inact: 50S Ribosomal Subunit; CoSet: Correlated Reaction Set.

Authors' contributions

YX conceived and designed the study, participated in drafting and revising the manuscript. YZ carried out the analysis and drafted the manuscript. LW interpreted the results biologically. FW supervised the study, participated in its design and to revise the manuscript. All authors read and approved the final manuscript.

Authors' details

[1]Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, 200433, China. [2]Department of Life Science, Fudan University, Shanghai, 200433, China.

Published: 24 January 2014

References

1. Reed JL, Famili I, Thiele I, Palsson BO: **Towards multidimensional genome annotation.** *Nature Reviews Genetics* 2006, **7**:130-141.
2. Papin JA, Palsson BO: **The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis.** *Biophys J* 2004, **87**:37-46.
3. Gianchandani EP, Joyce AR, Palsson BO, Papin JA: **Functional states of the genome-scale Escherichia coli transcriptional regulatory system.** *PLoS Comput Biol* 2009, **5**:e1000403.
4. Thiele I, Jamshidi N, Fleming RMT, Palsson BO: **Genome-Scale Reconstruction of Escherichia coli's Transcriptional and Translational Machinery: A Knowledge Base, Its Mathematical Formulation, and Its Functional Characterization.** *PLoS Comput Biol* 2009, **5**:e1000312.
5. Palsson B: *Systems biology: properties of reconstructed networks* Cambridge Univ Pr; 2006.
6. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886-897.
7. Covert MW, Schilling CH, Famili I, Edwards JS, Selkov E, Palsson BO: **Metabolic modeling of microbial strains in silico.** *Trends Biochem Sci* 2001, **26**:179-186.
8. Edwards JS, Covert M, Palsson B: **Metabolic modelling of microbes: the flux-balance approach.** *Environ Microbiol* 2002, **4**:133-140.
9. Reed JL, Palsson BO: **Thirteen years of building constraint-based in silico models of Escherichia coli.** *J Bacteriol* 2003, **185**:2692-2699.
10. Papin JA, Palsson BO: **Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk.** *J Theor Biol* 2004, **227**:283-297.
11. Schilling CH, Letscher D, Palsson BO: **Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *Journal of Theoretical Biology* 2000, **203**:229-248.
12. Schilling CH, Schuster S, Palsson BO, Heinrich R: **Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era.** *Biotechnol Prog* 1999, **15**:296-303.
13. Papin JA, Price ND, Palsson BO: **Extreme pathway lengths and reaction participation in genome-scale metabolic networks.** *Genome Research* 2002, **12**:1889-1900.
14. Papin JA, Reed JL, Palsson BO: **Hierarchical thinking in network biology: the unbiased modularization of biochemical networks.** *Trends in Biochemical Sciences* 2004, **29**:641-647.
15. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Metabolic pathways in the post-genome era.** *Trends in Biochemical Sciences* 2003, **28**:250-258.
16. Wiback SJ, Palsson BO: **Extreme pathway analysis of human red blood cell metabolism.** *Biophysical Journal* 2002, **83**:808-818.
17. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED: **Metabolic network structure determines key aspects of functionality and regulation.** *Nature* 2002, **420**:190-193.
18. Liao JC, Hou SY, Chao YP: **Pathway analysis, engineering, and physiological considerations for redirecting central metabolism.** *Biotechnol Bioeng* 1996, **52**:129-140.
19. Schilling CH, Edwards JS, Letscher D, Palsson BØ: **Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems.** *Biotechnology and Bioengineering* 2000, **71**:286-306.
20. Forster J, Gombert AK, Nielsen J: **A functional genomics approach using metabolomics and in silico pathway analysis.** *Biotechnol Bioeng* 2002, **79**:703-712.
21. Carlson R, Fell D, Srienc F: **Metabolic pathway analysis of a recombinant yeast for rational strain development.** *Biotechnol Bioeng* 2002, **79**:121-134.
22. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO: **Genome-scale metabolic model of Helicobacter pylori 26695.** *Journal of Bacteriology* 2002, **184**:4582-4593.
23. Price ND, Papin JA, Palsson BO: **Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis.** *Genome Res* 2002, **12**:760-769.
24. Papin JA, Price ND, Edwards JS, Palsson BB: **The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy.** *J Theor Biol* 2002, **215**:67-82.
25. Van Dien SJ, Lidstrom ME: **Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph Methylobacterium extorquens AM1, with application to reconstruction of C(3) and C(4) metabolism.** *Biotechnol Bioeng* 2002, **78**:296-312.
26. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol Syst Biol* 2007, **3**:121.
27. Thiele I, Price ND, Vo TD, Palsson BO: **Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet.** *J Biol Chem* 2005, **280**:11683-11695.
28. McLeod SM, Johnson RC: **Control of transcription by nucleoid proteins.** *Curr Opin Microbiol* 2001, **4**:152-159.
29. Reshamwala S, Noronha S: **Biofilm formation in <i>Escherichia coli</i><i>cra</i> mutants is impaired due to down-regulation of curli biosynthesis.** *Archives of Microbiology* 2011, **193**:711-722.
30. Perrenoud A, Sauer U: **Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli.** *J Bacteriol* 2005, **187**:3171-3179.

31. Ogasawara H, Ishida Y, Yamada K, Yamamoto K, Ishihama A: **PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in Escherichia coli.** *J Bacteriol* 2007, **189**:5534-5541.
32. Kornberg HL: **The role and control of the glyoxylate cycle in Escherichia coli.** *Biochem J* 1966, **99**:1-11.
33. Hadfield A, Kryger G, Ouyang J, Petsko GA, Ringe D, Viola R: **Structure of aspartate-beta-semialdehyde dehydrogenase from Escherichia coli, a key enzyme in the aspartate family of amino acid biosynthesis.** *J Mol Biol* 1999, **289**:991-1002.
34. Harb OS, Abu Kwaik Y: **Identification of the aspartate-beta-semialdehyde dehydrogenase gene of Legionella pneumophila and characterization of a null mutant.** *Infect Immun* 1998, **66**:1898-1903.
35. Cohen G: **The common pathway to lysine, methionine, and threonine.** *Biotechnology Series[BIOTECHNOL SER] 1983* 1983.
36. Szafranska AE, Hitchman TS, Cox RJ, Crosby J, Simpson TJ: **Kinetic and mechanistic analysis of the malonyl CoA:ACP transacylase from Streptomyces coelicolor indicates a single catalytically competent serine nucleophile at the active site.** *Biochemistry* 2002, **41**:1421-1427.
37. Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP: **Structure of the 30S translation initiation complex.** *Nature* 2008, **455**:416-420.
38. Bade EG, Gonzalez NS, Algranati IS: **Dissociation of 70S ribosomes: some properties of the dissociating factor from Bacillus stearothermophilus and Escherichia coli.** *Proc Natl Acad Sci USA* 1969, **64**:654-660.
39. Schwartz MA, Baron V: **Interactions between mitogenic stimuli, or, a thousand and one connections.** *Current opinion in cell biology* 1999, **11**:197-202.
40. Covert MW, Palsson BO: **Constraints-based models: regulation of gene expression reduces the steady-state solution space.** *Journal of Theoretical Biology* 2003, **221**:309-325.
41. Schilling CH, Palsson BO: **Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis.** *Journal of Theoretical Biology* 2000, **203**:249-283.
42. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO: **Comparison of network-based pathway analysis methods.** *Trends Biotechnol* 2004, **22**:400-405.
43. Price ND, Reed JL, Papin JA, Famili I, Palsson BO: **Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices.** *Biophysical Journal* 2003, **84**:794-804.
44. Bell SL, Palsson BO: **Expa: a program for calculating extreme pathways in biochemical reaction networks.** *Bioinformatics* 2005, **21**:1739-1740.
45. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO: **Matrix formalism to describe functional states of transcriptional regulatory systems.** *PLoS Comput Biol* 2006, **2**:e101.
46. Xi YP, Chen YPP, Cao M, Wang WR, Wang F: **Analysis on relationship between extreme pathways and correlated reaction sets.** *BMC Bioinformatics* 2009, **10**:S58.
47. Price ND, Famili I, Beard DA, Palsson BO: **Extreme pathways and Kirchhoff's second law.** *Biophys J* 2002, **83**:2879-2882.