

RESEARCH ARTICLE

A system-wide network reconstruction of gene regulation and metabolism in *Escherichia coli*

Anne Grimbs¹, David F. Klosik², Stefan Bornholdt², Marc-Thorsten Hütt^{1*}

1 Computational Systems Biology, Department of Life Sciences & Chemistry, Jacobs University, Bremen, Germany, **2** Institute for Theoretical Physics, University of Bremen, Bremen, Germany

* m.huett@jacobs-university.de



OPEN ACCESS

Citation: Grimbs A, Klosik DF, Bornholdt S, Hütt M-T (2019) A system-wide network reconstruction of gene regulation and metabolism in *Escherichia coli*. PLoS Comput Biol 15(5): e1006962. <https://doi.org/10.1371/journal.pcbi.1006962>

Editor: Laurence Yang, University of California San Diego, UNITED STATES

Received: June 25, 2018

Accepted: March 18, 2019

Published: May 3, 2019

Copyright: © 2019 Grimbs et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: S.B. and M.-T.H. acknowledge the support of Deutsche Forschungsgemeinschaft (DFG), grants BO 1242/6 and HU 937/9. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Genome-scale metabolic models have become a fundamental tool for examining metabolic principles. However, metabolism is not solely characterized by the underlying biochemical reactions and catalyzing enzymes, but also affected by regulatory events. Since the pioneering work of Covert and co-workers as well as Shlomi and co-workers it is debated, how regulation and metabolism synergistically characterize a coherent cellular state. The first approaches started from metabolic models, which were extended by the regulation of the encoding genes of the catalyzing enzymes. By now, bioinformatics databases in principle allow addressing the challenge of integrating regulation and metabolism on a system-wide level. Collecting information from several databases we provide a network representation of the integrated gene regulatory and metabolic system for *Escherichia coli*, including major cellular processes, from metabolic processes via protein modification to a variety of regulatory events. Besides transcriptional regulation, we also take into account regulation of translation, enzyme activities and reactions. Our network model provides novel topological characterizations of system components based on their positions in the network. We show that network characteristics suggest a representation of the integrated system as three network domains (regulatory, metabolic and interface networks) instead of two. This new three-domain representation reveals the structural centrality of components with known high functional relevance. This integrated network can serve as a platform for understanding coherent cellular states as active subnetworks and to elucidate crossover effects between metabolism and gene regulation.

Author summary

Networks—the compact representation of systems in terms of nodes and links—are an efficient data structure for biological information. They also allow us to establish relationships between network structure and dynamical function and thus hold the potential of implementing a systems-level view on biological processes. Using the formal language of networks and careful manual curation, we unite information from a range of publicly

available databases, in order to provide a metabolic-regulatory network model for the gut bacterium *Escherichia coli*, which allows us to provide novel topological characterizations of system components based on their positions in the entire network. From the network representation we derive a new partition of the system into three network domains, one predominantly associated with gene regulation, a second, which covers all metabolic processes and a third domain, containing protein interactions and serving as an interface between the two other domains. This has consequences for the topological prediction of the biological relevance of the system components. We discuss specific examples, where this new three-domain representation reveals the structural centrality of components with known high functional relevance. This integrated network can serve as a platform for understanding biological phenomena jointly mediated by gene regulation and metabolism and thus provide insight relevant for genetic and metabolic engineering.

Introduction

So far, metabolic processes and gene regulatory events are typically considered individually in system-level investigations. However, ample evidence exists that the majority of cellular processes involves both, metabolism and gene regulation, and thus requires their joint examination [1]. One of the best-investigated individual examples in *Escherichia coli* (*E. coli*) is the phosphoenolpyruvate–carbohydrate phosphotransferase system (PTS) which is responsible for import and phosphorylation of sugars [2]. Additionally, the PTS is involved in the regulation of the import process depending on the available carbohydrate mixtures in the growth medium. By carbon catabolite repression and inducer exclusion, primarily the uptake of a preferred carbon source to be metabolized, such as glucose, is selected from other carbohydrates present in the growth medium. In order to understand the underlying principles, not only the effects of both ‘layers’, metabolism and regulation, need to be taken into account, but also their interface [3].

On a more qualitative level, the importance of the interface of metabolism and gene regulation can be illustrated by having a closer look at their most prominent representatives, namely, enzymes and metabolic transcriptional regulators. Both examples are proteins and can be thought of as a component type organizing the interplay of genes and metabolic reactions (Fig 1). For enzymes the connection is straightforward: The majority of metabolic reactions can only take place if the corresponding genes of the catalyzing enzymes are expressed. These genes, in turn, are often involved in regulatory processes, especially if they are associated with central biochemical reactions. In contrast, metabolic transcriptional regulators can be illustrated by looking at transcription factors, the probably best-investigated transcriptional regulators. Some of them require the binding of a metabolite to be active and are therefore called metabolic transcriptional regulators. In the context of the integrative view discussed here, it is noteworthy that only the interaction with a metabolic component enables their functionality as gene expression regulators.

Conventional reconstructions of *E. coli*'s metabolism as well as of its gene regulation thoroughly describe the process itself but usually lack information on interacting elements of the other biological system. While there are numerous genome-scale metabolic reconstructions available [4–9], only a few large-scale transcriptional regulatory networks exist that are mainly based on the information from RegulonDB [10]. First attempts to integrate both cellular processes started from metabolic reconstructions which were expanded by regulatory genes and stimuli of the associated encoding metabolic genes [11, 12]. Both studies started from the

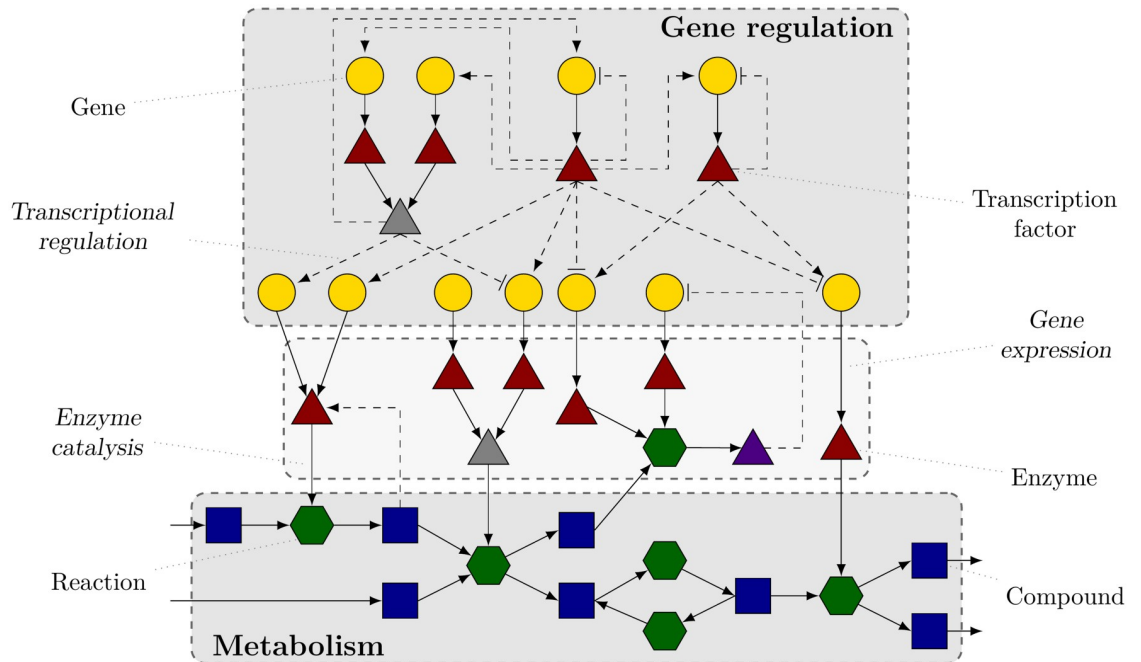


Fig 1. Schematic representation of the involved processes and biological elements in the integrative metabolic-regulatory *E. coli* network. Gene regulatory processes primarily comprise genes (yellow circles) and several proteins (monomers (brown triangles) as well as complexes (gray triangles)), mainly transcription factors. In contrast, metabolic processes are predominantly defined by small molecules (blue squares) and the catalyzing biochemical reactions (green hexagons). The interactions between regulatory and metabolic processes can be mainly characterized by proteins (also modified proteins (purple triangles)) serving as enzymes and regulators, respectively. The symbols are also explained in Fig 2. While regulatory links are represented as dashed lines, the encoding and reaction-associated links are shown as solid lines. Note that this scheme does not cover all aspects of the biological categories and their classification (see Methods for details).

<https://doi.org/10.1371/journal.pcbi.1006962.g001>

metabolic model of [5] and include 104 regulatory genes and 583 regulatory rules regulating approximately 50% of the metabolic genes. In this manner, the close proximity of regulatory events was captured but more far-reaching and global effects, *e.g.*, self-contained regulatory dynamics among genes, could not be considered. Further approaches examine the regulatory processes of the metabolic network based on the aforementioned pioneering attempts [13, 14]. For this purpose, the information about regulatory events was assembled in terms of Boolean rules as a variant of Boolean network models.

More recently, [15] introduced a method called probabilistic regulation of metabolism, a new variant of regulatory flux balance analysis, *i.e.*, the class of approaches behind some of the pioneering integrative models discussed above [11, 12]. The state of the many variants of integrating regulatory information into flux-balance analysis models has been reviewed in [16] and [17]. The necessity of achieving such data integration, even on the network level, has recently been discussed in [18]. To a certain degree all these studies consider the regulation of metabolism but only cover the proximity rather than a genome scale.

A very recent example of a successful topological characterization of biological networks, in order to understand the interplay of gene regulation and metabolism, is the analysis of gene regulatory—metabolic feedback loops [19]. In [19] joint representations of gene regulatory and metabolic networks have been compiled for two organisms, *E. coli* and *B. subtilis*. These network representations are then analyzed with a focus on the hierarchical structure of the resulting network and on feedback loops between gene regulation and metabolism. These feedback loops are then further characterized in terms of their potential impact (defined as the

number of genes downstream of the transcription factor receiving the feedback from metabolism), their possible function (*e.g.*, in the processing of environmental information) and other properties [19]. In this way, the authors obtain insight in the algorithmic features of the interface between gene regulation and metabolism.

Understanding the interplay of metabolism and gene regulation will help to gain insight in cellular, system-wide responses such as to changing environmental conditions. Here, we present the database-assisted reconstruction of an integrative *E. coli* network capturing metabolic as well as regulatory processes. The attribution of network components (in terms of individual vertices) to the metabolic and regulatory domains, as well as the protein interface enables the further characterization of the network in terms of its modular organization, its path statistics and the vertex centrality.

In particular, we formulate a new measure by evaluating *domain-traversing paths*, in order to quantitatively assess the role of components in the interface domain and thus identify cross-systemic key elements contributing to both regulatory and metabolic processes. In all cases, these topological assessments highlight system components and functional subsystems, which are well known for their biological relevance, thus emphasizing the predictive power of network topology. Employing observations on the topological (structural, network-architectural) level, in order to identify components in the system of particular functional relevance has a long tradition in network biology (and in network science in general).

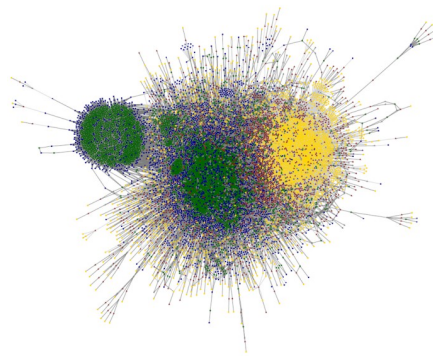
The main results of our investigation are: We present an integrated network representation of gene regulation and metabolism of *E. coli* and illustrate how it is a promising starting point for the structural investigation of system-wide phenomena. In particular, the network perspective suggests the explicit consideration of a protein interface between the genetic and metabolic realms of the cell. Employing network metrics we (1) argue that a three-domain partitioning is architecturally and functionally plausible, and (2) show that prominent components of the network according to the structural investigation tend to be of evident biological importance. Especially, the evaluation of possible paths through the interface domain of the network reconstruction yields well-known functional subsystems. The overlap of structural and biological relevance here suggests that a careful analysis of such a structural model can guide biological investigations by focusing on a limited number of structurally outstanding components. This network model can also serve as a starting point for a range of topological analyses with methods developed in statistical physics (see, *e.g.*, [20] for a recent review).

Summarizing, in contrast to the separate analyses of (*e.g.*, the metabolic or gene regulatory) subsystems, we expect that the integrative network model shown here will draw the attention to system-wide feedback loops not contained in the individual subsystems and to different roles of individual components, which become only visible from the perspective of interdependent networks.

Results

Database-assisted network reconstruction

By now, the dramatic growth of bioinformatics databases [21], both in content and in diversity, allows addressing the challenge of integrating regulation and metabolism on a system-wide level. We devised a semi-automated framework to integrate information from EcoCyc database [22] and RegulonDB [10] into a network for *E. coli* including major cellular processes, from metabolic processes via protein modifications to a variety of regulatory events (see [Methods](#)). Networks are an efficient data structure for integrating this wealth of information [23–25]. In this way, the vast amount of data contained in the bioinformatics databases provide an ‘architectural embedding’ for metabolic-regulatory networks and guides subsequent steps of



Vertex composition		iMC1010*	
●	Reaction	4693	569/ 767
■	Compound	2681	557/ 615
●	Gene	2545	971/1010
▲	Protein monomer	1917	
▲	Protein-protein complex	929	771/ 817
▲	Protein-compound complex	100	
▲	Protein-RNA complex	3	
		12868	2868/3209

* accounted only for enzymatic reactions and unique metabolites (1076 and 762 in total)

Fig 2. Spring-block graph representation and vertex composition of the integrative *E. coli* network. A scalable force directed placement algorithm has been used. The coverage of the pioneer model from [11] is provided in column iMC1010.

<https://doi.org/10.1371/journal.pcbi.1006962.g002>

model refinement and validation. We augmented and validated the resulting network based on existing reconstructions of metabolic [6, 8, 26–28] as well as of gene regulatory processes [10].

The integrative *E. coli* network constructed here comprises the three major biological components, genes, proteins, and metabolites, as well as the metabolizing reactions summing up to more than 12,000 components. Represented as a graph, the network has seven types of vertices depicting the major biological components (Fig 2, Table A in S1 Text) and seven different types of edges including two types of encoding associations, *i.e.*, transcription and translation processes, four types describing the associations within biochemical reactions, and one type summarizing regulatory relations (Table B in S1 Text). Two small annotated biological examples are shown in Fig B in S1 Text. The graph representation facilitates the mapping of reactions and their catalyzing enzymes, as both are depicted as vertices. In contrast, metabolic systems are often represented as hypergraphs to illustrate the Boolean ‘AND’ association of reaction educts and the fixed stoichiometric ratio of the involved metabolites. Those aspects are assigned explicitly as edge properties in the graph representation. For the purpose of measuring the propagation of perturbations through the network, for example, the following logical assignments are helpful (see [29] for details on these definitions): Besides the associations of reaction educts, the encoding relations of protein complexes are of Boolean ‘AND’ type, termed *conjunct* links. On the contrary, associations representing isoforms of protein subunits, isoenzymes as well as reaction products are implemented by Boolean ‘OR’ links, called *disjunct*. The third linkage type, *regulation*, covers approximately 7,300 regulatory associations, *i.e.*, transcriptional, translational as well as metabolic ones (Table C in S1 Text).

The metabolic and regulatory processes

The comparison with existing models reveals that the presented integrative network is a comprehensive representation of the metabolic and regulatory processes in *E. coli*. The very first approach of embedding metabolic processes in the regulatory context of [11], the iMC1010 model, started from a metabolic model which was extended by the regulation of the encoding genes of the catalyzing enzymes. For the purpose of determining the overlap of the integrative metabolic-regulatory network and the iMC1010 model, transport reactions as well as the artificial biomass reaction have been disregarded and, moreover, only unique metabolites (neglecting compartmentation) have been taken into account. Else, the different levels of details of the transport systems such as PTS as well as of the compound compartmentation would render a

correct mapping impossible. Overall, the *iMC1010* model is covered by our model to more than 89% (Fig 2, see Table D in S1 Text, column 3).

To assess the coverage of *E. coli*'s metabolic processes, the embedded metabolic processes of the integrative *E. coli* network have been associated to the ones of an established *E. coli* metabolic reconstruction, namely the *iAF1260* model from [6]. About 67% of the involved biochemical reactions, compounds and genes could be mapped directly (see Table D in S1 Text, column 4). Particularly, these two thirds capture almost all biologically relevant components in terms of *in silico* viability. Using flux balance analysis for simulating the biomass production capacity of the *iAF1260* model and taking the overlap with mapped components of the integrative *E. coli* network revealed that for the default medium setup approximately 75% of the essential reactions (to yield 1% biomass) are covered by the integrative *E. coli* network.

Analogous to the metabolic processes, the coverage of *E. coli*'s gene regulation has been determined using the transcriptional regulatory network from RegulonDB [10]. This model has been assembled in a similar fashion but is accounting only for transcription factors and their regulated genes. With a coverage of more than 98%, the transcription-related regulatory processes are considered as completely recorded in the integrative *E. coli* network (see Table D in S1 Text, column 5). Apart from that, for this assessment of overlap a comparison of regulatory processes associated with RNA translation as well as metabolic regulatory events is not possible since the RegulonDB transcriptional regulatory network does not consider protein and metabolic interaction processes.

The interface of metabolic and regulatory processes

The most conspicuous links between metabolic and gene regulatory processes are metabolic transcription factors, *i.e.*, gene expression regulators binding metabolites, and metabolic genes, *i.e.*, genes with significant and coordinated response on the metabolic level such as encoding enzymes. Intuitively, the interface is considered so far as the direct interactions of metabolic elements and gene regulatory elements, and the integrative *E. coli* network can be partitioned into metabolic and regulatory domain (MD—RD).

However, by examining those interactions in more detail the topological role of proteins becomes apparent. Regarding the metabolic transcription factors, the respective metabolite binds to a protein and this metabolite-protein complex then subsequently regulates the gene expression. In the case of metabolic genes, ultimately the respective gene encodes a protein which either by itself or as a complex serves as an enzyme. In line with this, the interface of metabolic and gene regulatory processes should be considered as the series of interactions of metabolites and genes, respectively, with proteins and subsequent protein modifications. Thus, the interface does not only comprise interactions (edges) but also components (vertices), and the integrative *E. coli* network will in the following be divided into a metabolic domain, a protein interface and a regulatory domain (MD—PI—RD).

In the next section, the plausibility of the three-domain partition (and the set of biologically motivated rules devised to create it) will be assessed in comparison to the likewise proposed two-domain (MD—RD) representation.

The interface structure—A matter of network partition. In order to assess the large-scale structure of the reconstructed network we apply a set of rules that assign each vertex of the network to one of two and three domains, respectively, by considering the biological types of the vertices themselves as well as those of their neighbors (as outlined in the Methods section). Since these rules have been designed to group together vertices connected to the same biological processes we expect them to result in biologically plausible network partitions.

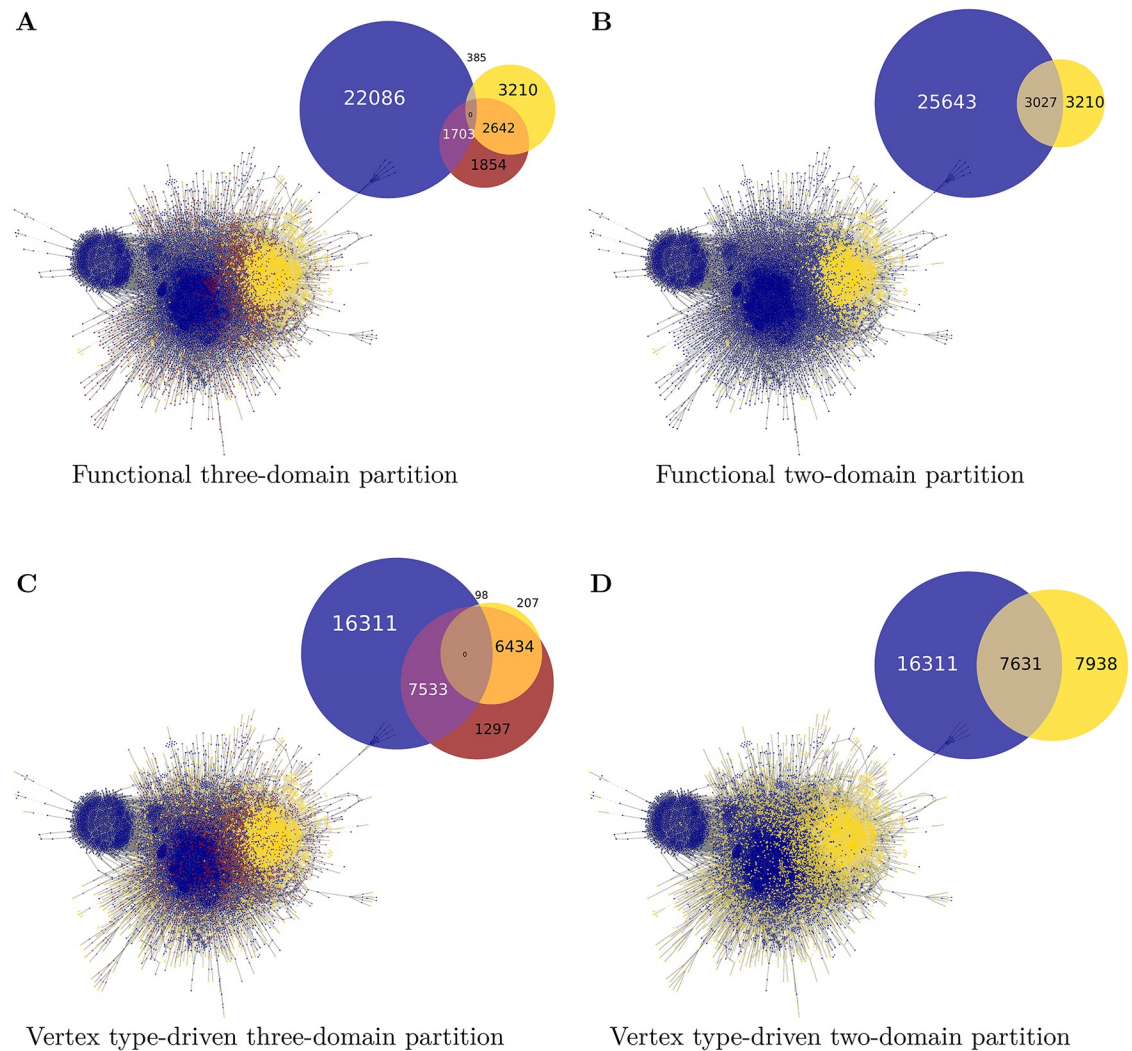


Fig 3. Graph snapshots of the four partitions. The functional three-domain partition into metabolic and regulatory domains and protein interface (MD—PI—RD) (A), the functional two-domain partition into metabolic and regulatory domains (MD—RD) (B), vertex-driven three-domain partition into compounds/reactions, proteins and genes (C), vertex-driven two-domain partition into compounds/reactions, and proteins/genes (D). Vertices are colored according to their domain-affiliation: yellow—(pseudo) regulatory and gene-focused domain, respectively, and blue—(pseudo) metabolic and compound-focused domain, respectively. The interface domain in the three-domain partitions are drawn in red. The diagrams in the top right corners of each panel show the edge composition of the system in terms of intra-domain and inter-domain edges.

<https://doi.org/10.1371/journal.pcbi.1006962.g003>

To complement the two *functional* partitions, MD—RD and MD—PI—RD, two partitions that solely take into account the vertex types have been analyzed, also representing a metabolic-regulatory division into two and three domains, respectively. For the *vertex-driven* two-domain partition, the sets of gene and protein vertices denote the regulatory processes while in the three-domain partition regulation is given by the set of genes and the interface domain only consists of the protein vertices. In both cases, metabolism is represented by the sets of reactions and compounds. In the three-domain case, the vertex-driven three-domain partition, the vertex set of proteins form an interface similar to the MD—PI—RD partition (Fig 3). The functional and vertex-driven three-domain partitions are of roughly similar size in terms of vertex count, while the respective two-domain partitions have a metabolic-regulatory vertex ratio of 5:1 and 4:3, respectively (see Fig 4).













		Functional partitions			Vertex-driven partitions		
		MD – PI – RD	MD – RD	   	   	   	
		I – II – III	I – III	I – II – III	I – II – III	I – III	
Vertices	I	8369	10655		7374	7374	
	II	2286			2949		
	III	2213	2213		2545	5494	
Inter-module edge fraction	I	0.086	0.106		0.319	0.319	
	II	0.701			0.915		
	III	0.485	0.485		0.969	0.49	
Modularity, M	Total	<u>0.287</u>	0.157		0.081	0.226	
	I	0.166	0.079		0.113	0.113	
	II	0.042			-0.027		
	III	0.078	0.079		-0.005	0.113	

Fig 4. Topological properties of the functional and vertex type-driven network partitions. The functional partitions are denoted by the respective modules, metabolic domain (MD), regulatory domain (RD) and protein interface (PI). The vertex type-driven partitions are represented by the comprising vertex types, reaction (green hexagon), compound (blue square), gene (yellow circle), and protein (brown triangle). For each property, the module-specific coefficients and contributions (I, II, III) are presented, respectively. For the modularity, M , the overall network coefficient (Total) is shown as well as the best coefficient is underlined, the module-specific values correspond to the terms in the sum of Eq (1).

<https://doi.org/10.1371/journal.pcbi.1006962.g004>

First, the two three-domain partitions will be compared, *i.e.*, the functional partition, MD—PI—RD, and the vertex-driven partition. In the following, we will argue that the additional third domain acts as an interface between the regulatory and metabolic domains in the functional partition, while we will see that the vertex-driven partition fails to give a coherent picture of the domain-level organization of the biological system.

Especially, it will become clear, also in later sections, that the interface domain in the functional partition contains processes that are known to play prominent roles in system-scale communication within the cell, and may therefore be considered an important component of the large-scale organizational structure of the combined regulation and metabolism of *E. coli*.

A simple quantity to illustrate the domain-level picture is the fraction of inter-module edges (linking to a vertex of a different domain) over all edges connected to vertices of a specific domain (*i.e.*, external and internal edges). Of course, there is no objectively ‘correct’ partition the result of our procedure could be measured against, but there are a number of fundamental properties that a biologically plausible partition in the given context should possess. On the one hand, a proper interface provides the main means of communication between the regulatory and the metabolic processes, *i.e.*, the majority of paths between the outer two domains should run through the interface. Indeed, the interface of the functional partition shows a considerably larger inter-module edge fraction than the remaining domains (0.7 compared to 0.5 and 0.1, Fig 4), stressing its special character as a bridging module. A high inter-module edge fraction of the interface is also found in the vertex-driven partition, however, its regulatory domain shows an even higher inter-module edge fraction which indicates an entanglement between the two groups rather than one domain acting as a bridging module to another domain. This exactly gives rise for the second criteria, that the domains should capture actual processes (here, structures on the level of several vertices). Unambiguously, regulatory or metabolic processes should be contained within the respective domain so that system-wide interaction takes place between processes. In the following chapter, Interface characterization, it will be shown that this actually is also the case for the interface in the functional partition. In contrast, in the vertex-driven partition already the regulatory domain shows deficiencies with

respect to that criterion. Since this regulatory domain solely contains gene-gene interactions the intermediate transcription factor steps are not within the domain which become visible in the almost exclusively inter-module edges, linking it to the interface domain.

Next, we compare the three-domain partitions with the two-domain partitions. While the introduction of a third domain allows to study the system in terms of an explicit interface, the partitions into two domains is much closer to common biological intuition. The question which needs to be answered is whether metabolism and gene regulation are solely interfaced by the linking processes such as gene expression, and activation or inhibition of transcription factors and genes, so that the system can appropriately be described with two domains. Or whether there is an actual interface that preferably comprises entire processes additionally including protein modifications and suchlike. Here, this question will be assessed from a topological perspective.

A relevant topological quantity is the network modularity [30] of a given network partition. For a biologically meaningful classification, one would expect on the network level that the regulatory and the metabolic domains show high intra-module connectivity (a large number of links are within a domain) and sparse inter-module linkages (a small number of links are between domains). Accordingly, the network modularity should be high for a successful partition. The results for the modularity are listed in Fig 4. The functional partitions clearly outperform the vertex type-driven partitions. Also, when going from MD—RD to MD—PI—RD there is a notable increase in the modularity of the network ($M_2 = 0.157$, $M_3 = 0.287$). Note that here we consider specific candidates for biologically plausible partitions, while a purely topological analysis of the module structure of this large network yields a much larger set of significant modules. Here, a detailed biological interpretation is still missing and will be discussed elsewhere.

Altogether, the functional partition into metabolic domain, protein interface and regulatory domain reflects a biologically reliable classification in two delimited domains linked by a bridging module. Reinforced by the topological properties, the interface structure including full protein modification processes will be used subsequently.

Interface characterization. The interface of metabolic and gene regulatory processes of the integrative *E. coli* network comprises, as expected, predominantly proteins, *i.e.*, monomers and complexes (Table A in S1 Text), and mainly protein modification processes such as protein translation, protein complex formation and biochemical protein conversion (Table B in S1 Text). On closer examination, the covered processes can be divided in internal and peripheral ones. According to the bridging role of the interface, the majority of these are peripheral processes (Fig 3, Table B in S1 Text). The peripheral processes, in turn, can be subdivided according to their directionality meaning from regulatory to metabolic domain (subsequently termed ‘downwards’) and from metabolic to regulatory domain (‘upwards’), respectively. To enumerate the portion of peripheral processes forming complete paths across the interface, direct downwards and upwards links and the new topological concept of *domain-traversing paths* (or short: *traversing paths*) have to be considered. A traversing path connects regulatory and metabolic domain via the protein interface, whereby only starting and end vertex are not affiliated to the bridging domain and the path direction is considered carefully (see Methods).

Examination of the downwards-upwards subdivision, especially the traversing paths, reveals a considerable (though biologically expected) asymmetry of the interface (Fig 5): The downwards interface is much more pronounced comprising predominantly the transcription of enzymes, *i.e.*, metabolic genes, and the formation of enzymatic protein complexes. On the contrary, the upwards interface is comparably sparse with roughly half the direct (102/283) and quarter the traversing paths (4,070/18,904) connections of the downwards interface. These

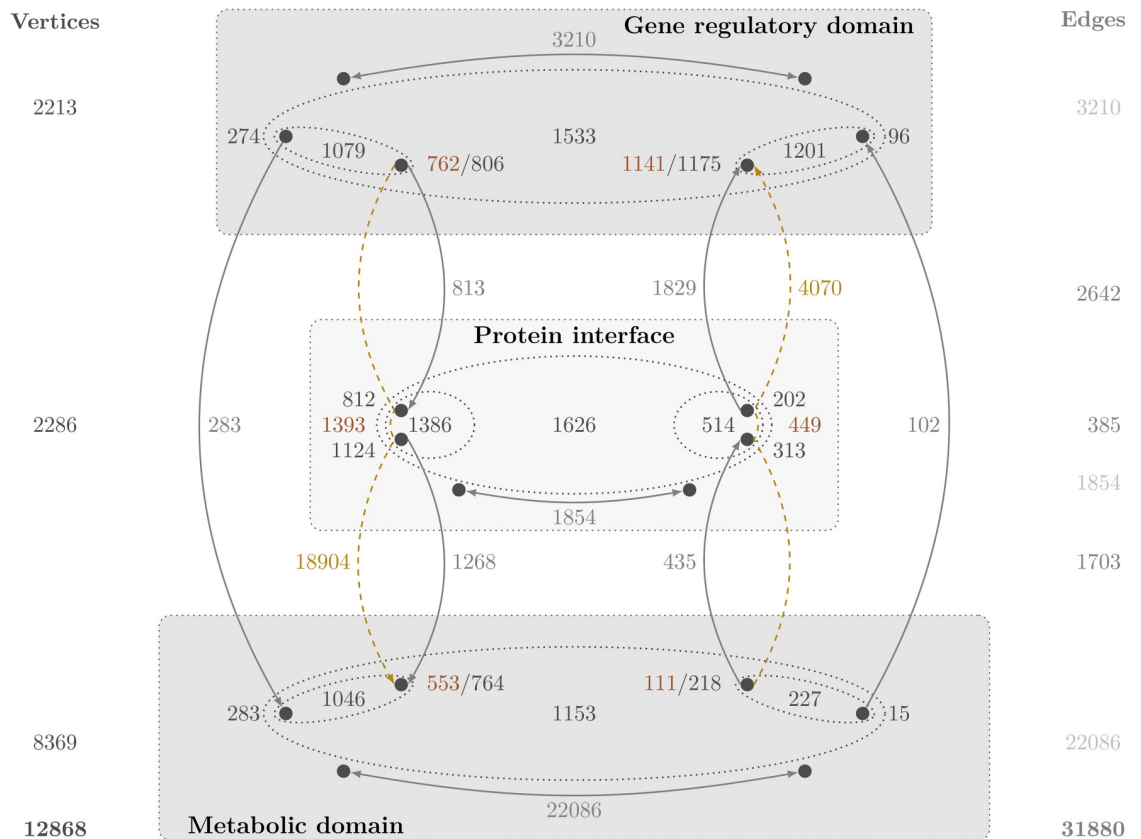


Fig 5. Schematic overview of the components and connections of the integrative *E. coli* network, especially those involved in the protein interface. The information about edges are presented in gray and about traversing paths are shown in gold while the number of vertices are shown in dark blue and the traversing paths-related ones are given in dark brown, in addition. The solid lines denote direct link connections while the dashed lines the traversing paths connections. In the regulatory domain, for example, 274 vertices are directly linked to 283 vertices in the metabolic domain via 283 edges. Also, 806 vertices have a directed edge to 812 vertices in the protein interface (summing up to 813 edges). Similarly, the regulatory domain has 3210 edges residing in this domain. The total number of vertices in the regulatory domain that exchange an incoming or outgoing link with one of the other domains are 1533. The number of edges displayed on the right-hand side of the figure is the total count of edges between two domains or within a single domain. The number 385, for example, is the total count of edges between the regulatory domain and the metabolic domain. The number 1854 is the total count of edges within the interface domain. In gold (as numbers and dashed lines) the information about traversing paths is given. The total number of downwards traversing paths (18904) and of upwards traversing paths (4070) is indicated in gold along these dashed lines.

<https://doi.org/10.1371/journal.pcbi.1006962.g005>

few upwards processes mainly include the formation of metabolic regulators, especially transcription factors, and the corresponding regulatory events.

In addition to confirming the interface asymmetry, the traversing paths reveal the bottleneck characteristic of the interface. First indications for this special property are (1) the low number of involved vertices and (2) the distribution of traversing path lengths. For both, downwards and upwards traversing paths, the number of distinct interface vertices in the traversing paths is low compared to the total number, *i.e.*, 1,393 and 449 interface vertices of 2,286 in total, respectively (Fig 5). On the other hand, for both, downwards and upwards traversing paths, emerges a remarkable clustering of paths of length 8–10 and four, six, and 9–11, respectively (Fig 6). This is in contrast to a smooth distribution one would expect in random graphs. By enumerating the involved vertices it is striking that approximately 46% of traversing paths contain one of five three-vertex-combinations, respectively. The respective combinations of downwards and upwards traversing paths pertain to three functional systems,

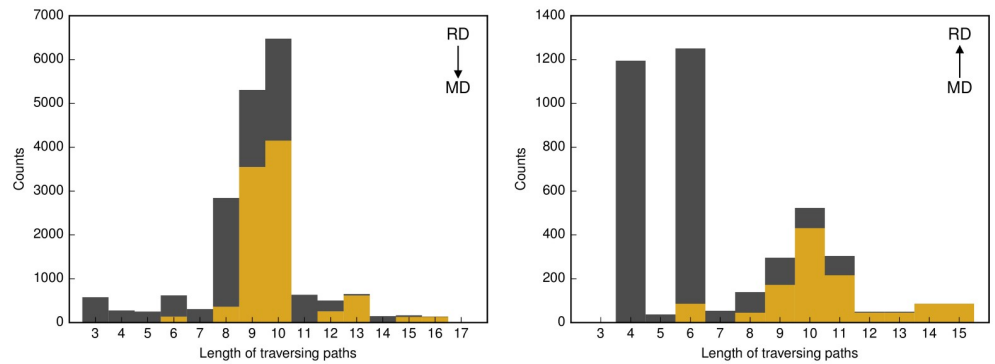


Fig 6. Distribution of the path lengths. Lengths are provided for the downwards (RD → MD) and upwards traversing paths (MD → RD), respectively (dark blue). The golden bars represent the fraction of downwards and upwards traversing paths comprising the PTS and RNR, and the NtrBC system associated vertices.

<https://doi.org/10.1371/journal.pcbi.1006962.g006>

the phosphoenolpyruvate-dependent sugar phosphotransferase system, PTS, the ribonucleotide reducing system, RNR system, as well as the nitrogen regulation two-component signal transduction system, NtrBC system (Table E in [S1 Text](#)).

All three biological subsystems, the PTS [2, 31], the RNR [32–35] as well as the NtrBC system [36–38] are well-studied with respect to their functionality and their cellular context. A schematic representation of the three subsystems is provided in [Fig 7](#). The PTS is an enzymatically active protein complex involved in the transport and phosphorylation of several sugars, so-called PTS-sugars [2]. In the integrative *E. coli* network more than 18 different sugars serve as potential substrates which are imported from peroxisome to cytosol at the same time (Table F in [S1 Text](#)). The substrate variety together with the manifold usage of the associatively produced pyruvate point out the key role of the PTS in *E. coli*'s metabolism and, moreover, suggest that the PTS acts as a bottleneck in the interface.

The RNR system, the second system dominating the downwards traversing paths, provides the major DNA building blocks [34]. Each of the different core enzyme classes, ribonucleotide reductase class I–III, are capable of catalyzing the reduction of all four nucleotides. Its transcriptional and metabolic regulation ensures the balanced supply and, thus, avoid the increase of mutation rates and the loss of DNA replication fidelity [39]. The central cellular role which is reflected in its regulatory embedding, together with its alternate substrates point to its special position in the interface.

The NtrBC system is a two-component signal transduction system initiating the nitrogen starvation response regulation. More precisely, depending on the nitrogen availability NtrB can autophosphorylate and the transfer of the NtrB phosphate group activates the global transduction regulator, NtrC. In *E. coli*, more than 40 genes known to be activated are involved in the nitrogen-response reaction such as active transport and mobilization of nitrogen in terms of N-containing compounds (for integrative *E. coli* network see Table G in [S1 Text](#)). The extensive regulatory function and the linkage to metabolism due to the allocation of ATP for NtrB autophosphorylation indicate that also the NtrBC system acts as a bottleneck in the interface, in the opposite direction to the PTS and RNR system.

The three central traversing paths systems and their biological relevance suggest that a topologically prominent position can be indicative of a biologically important functional entity. To corroborate the general validity of this indication, in the following section different topological properties have been analyzed and the prominent elements have been further characterized from a functional perspective.

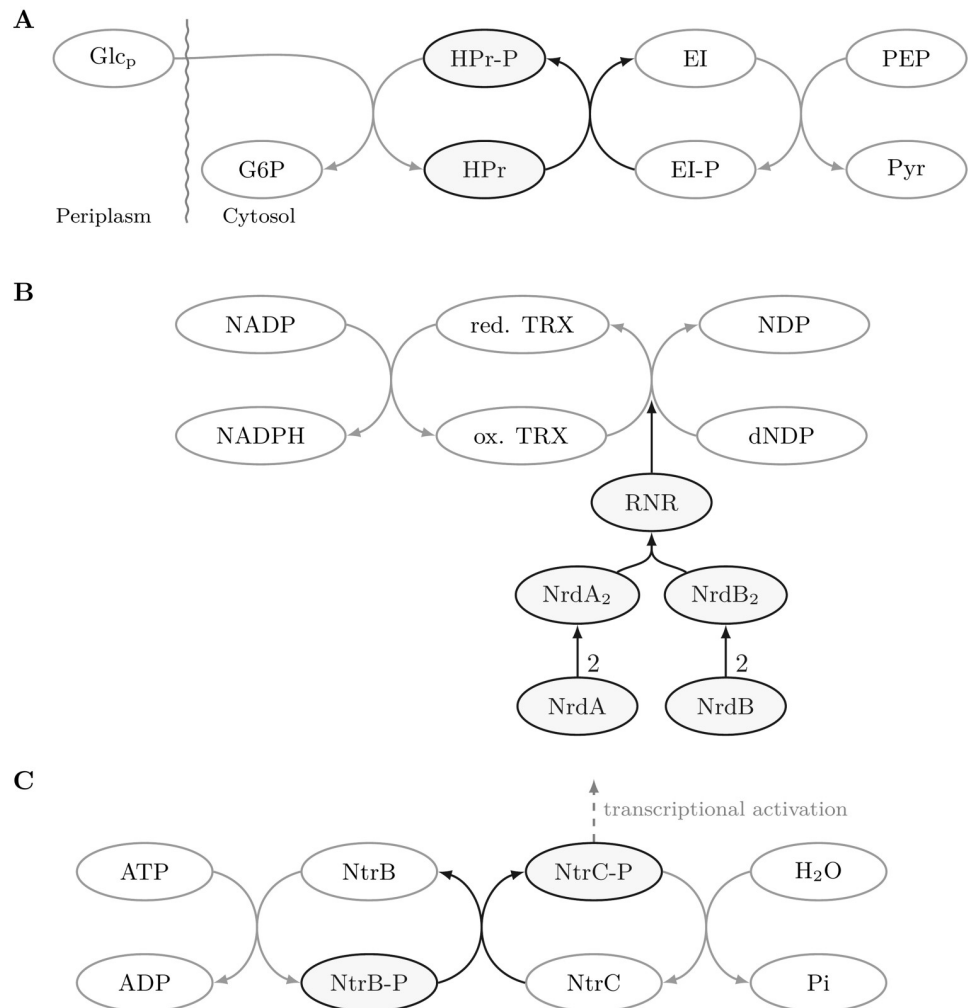


Fig 7. Classical representation of the three major interface systems of the integrative *E. coli* network. The systems shown here are the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS, **A**), the ribonucleotide reducing system (RNR system, **B**) and the nitrogen regulation two-component signal transduction system (NtrBC system, **C**). The edges represent biochemical reactions and the vertices denote the involved compounds and proteins. The reactions and proteins highlighted in dark blue are the most abundant vertices determining nearly half of the traversing paths (Table E in [S1 Text](#)).

<https://doi.org/10.1371/journal.pcbi.1006962.g007>

In order to also assess these traversing paths on a statistical level, we studied the percentage of traversing paths passing through specific vertices, pairs of vertices and triples of vertices (*i.e.*, three-vertex combinations). Fig E in [S1 Text](#) shows the percentage of downwards (RD → MD) and upwards (MD → RD) traversing paths for each of these three cases. The vertices and vertex combinations listed in Table E in [S1 Text](#) are highlighted in red. The histograms in Fig E in [S1 Text](#) show that, while the vast majority of vertices or vertex combinations is only involved in a small fraction of paths, some vertices or vertex combinations are involved in a much larger fraction of paths. These ‘outlier’ vertex groups (large number of paths are explained by these groups of vertices) also appear on the level of pairs and individual vertices, but on the level of triples they become biologically meaningful. Note that large parts of each of the domain lie in the largest strongly connected component (SCC), thus tightly coupling the three domains (Fig F in [S1 Text](#)). Regarding the domain-traversing paths we observe that

about 96% of the downwards and 67% of the upwards traversing paths are fully contained in the largest SCC.

Is the ‘interface’ nature of the protein domain also visible on a purely structural level? Fig G in S1 Text shows that the average betweenness centrality of the vertices in the interface domain is typically higher than the average betweenness centralities of randomly chosen subsets of vertices from the whole network of the same size as the interface domain, even though this distribution has a long tail to high values going beyond the average value of the interface domain.

Prompted by the recent study [19], we analyzed the feedback loops formed by our upwards and downwards paths. We passed from our sets of upwards and downwards traversing paths to traversing feedback loops by searching for combinations of upwards and downwards paths linked both, on the gene regulatory domain and in the metabolic domain. A closure of the loop in the regulatory domain is, for example, a transcription factor at the end of the upwards path regulating the gene, which serves as the starting node of the downwards path. A closure of the loop in the metabolic domain can be a direct path between the compound ending the downwards path and the compound starting the upwards paths. Surprisingly, the downwards paths (from RD to MD) contributing to loops are also dominated by the same triples, pairs of vertices and individual vertices as listed in Table E in S1 Text (and highlighted in the further statistical analysis in Fig E in S1 Text), while the upwards paths (from MD to RD) contributing to feedback loops deviate visibly from the set obtained by analyzing the upwards paths alone (*i.e.*, as listed in Table E in S1 Text). Fig H in S1 Text summarizes this observation in the same format as Fig E in S1 Text.

Cross-systemic key elements of *E. coli*

The integration of metabolic and regulatory events allows us to determine the key elements of *E. coli*, especially those beyond the individual processes. In particular, the functional three-domain partition facilitates to recover network components (in terms of individual vertices) of evident biological relevance, *e.g.*, by means of simple centrality measures. In the following, two different aspects of centrality have been examined [40]: degree centrality depicting the direct linkage of a vertex, and betweenness centrality which can be thought of as the participation of a vertex in the network flow [41].

Starting with the prominent local vertex structure, the so-called hubs (here, vertices with a total degree larger than 50), it is noticeable that they are primarily compounds and proteins, in particular protein complexes and appear in all three domains (see Table H in S1 Text, columns 3–5). In the metabolic domain, hubs include trivial compounds such as H⁺ and H₂O and, so-called, currency metabolites, *e.g.*, ATP, NAD(P)H and coenzyme A, while hubs of regulatory processes are obviously global regulators which characteristically exhibit a remarkably strong asymmetry of in-degree and out-degree. Particularly, well-known transcriptions factors top this list such as FNR (fumarate and nitrate reduction) [42], Fis (factor for inversion stimulation) and H-NS (histone-like nucleoid structuring protein) [43]. As stated above, hubs predominantly occur in metabolic and gene regulatory domain while only a few are affiliated to the protein interface. However, it was not to be expected to identify cross-systemic elements solely based on their degree.

To assess/detect cross-systemic key elements an extended approach of degree centrality has been used that additionally accounts for the domain boundaries. The *intra-domain degree fraction* ξ , also termed embeddedness [44], denotes the ratio of the internal degree of a vertex, within a domain, and the total degree in the network. This measure very clearly distinguishes between, on the one hand, metabolic and regulatory hubs which show intra-domain degree fractions $\xi > 0.87$ (except one single compound with $\xi = 0.185$) and hubs in the interface

which in contrast have $\xi \leq 0.06$ (see Table H in [S1 Text](#), last column). Thus, while metabolic and regulatory hubs are embedded in their respective domains, hubs in the protein interface are mainly connected to vertices in the neighboring domains. In total, seven hubs show a significant low intra-domain degree fraction pointing to their prevalent interactions with the other two domains (Fig A in [S1 Text](#) and Table K in [S1 Text](#), column 5). Six of them are affiliated to the protein interface exhibiting numerous interactions with the regulatory domain. Their linkages to the metabolic domain become visible when considering their composition, in case of the protein complexes, and their modes of action, respectively. The former involve the four protein-compound complexes Crp-cAMP (cyclic-AMP receptor protein binding cyclic-AMP) [31, 45, 46], DksA-ppGpp (dnaK suppressor binding guanosine 3'-diphosphate 5'-diphosphate) [47–49], NsrR-NO (nitrite-sensitive repressor binding nitric oxide) [50–52] and Lrp-Leu (leucine-responsive regulatory protein binding leucine) [53–55] whose naming schemes already indicate the metabolic link. The latter, namely, protein complex Cra (catabolite repressor activator) and protein monomer Lrp (leucine-responsive regulatory protein) form in the presence of appropriate metabolites, *i.e.*, fructose 1,6-bisphosphate/fructose 1-phosphate and leucine, complexes affecting their regulatory effect. The remaining hub is the metabolic-domain vertex representing guanosine 5'-diphosphate 3'-diphosphate (ppGpp). Besides its special domain-affiliation among the low intra-domain degree hubs, ppGpp acts as an important regulator of both, metabolism and transcriptional processes. More precisely, it regulates several enzyme activities as well as numerous transcription initiations by allosterically binding to RNA polymerase.

So far, we demonstrated that the protein interface of the *E. coli* network reconstruction acts as a bridging module between regulatory and metabolic domain enabling their interaction and communication. Therefore, we expect the betweenness centrality to directly highlight vertices from the interface. Indeed, ten out of the top-25-ranked (still including currency metabolites) vertices are from the interface (see Table I in [S1 Text](#), column 5), while overall the interface only accounts for about 18% of the vertices of the network. Especially, the already mentioned protein-compound complexes Crp-cAMP and DksA-ppGpp are among these compounds. In general, currency metabolites and trivial compounds (see above) as well as global regulators are among the central components with respect to betweenness. Apart from that, biochemical reactions building up and/or breaking down these metabolites and proteins as well as the other involved reactants pertain to the most betweenness-central components. Component association to functional systems allows to assess the systemic feature and by considering the corresponding network affiliation to depict the candidates for cross-systemic key elements. In this manner the network analysis allows us to detect the central role of Crp-cAMP, Lrp-Leu and ppGpp on purely topological grounds, as each component is the focus of such a functional system with high betweenness. Additionally among the top-ranked vertices with respect to betweenness centrality are five further cross-systemic components which are assigned to the protein interface, namely, phosphorylated PhoB (PhoB-P), Fur-Fe²⁺, and three outer membrane proteins (Omp), OmpC, OmpE and OmpF (Table I in [S1 Text](#)). The former two components are transcription factors and therefore acting in the gene regulatory domain, while at the same time they are protein complexes binding a metabolic small molecule depicting the connection to the metabolic processes. The latter three, the outer membrane porins, form hydrophilic channels, enabling non-specific diffusion of small molecules across the outer membrane [56–58]. In this role these proteins represent the most obvious connections of gene regulatory and metabolic domain—their encoding genes are highly regulated while the porins enable numerous metabolic transport reactions.

By focusing on the connecting domain of gene regulation and metabolism, the two centrality measures reinforce the key role of further cross-systemic elements. Considering the protein

interface-induced subgraph both centralities point out the vertices that top the list of the above-discussed downwards traversing paths (Table J in [S1 Text](#)). In more detail, both major systems contributing to the downwards traversing paths are represented each by three vertices, namely, PTS and RNR system ([Fig 7](#), panels A and B). Having a look at the intra-domain degree fraction, which put the focus on protein interface vertices as described above, additionally highlights a representative of the upwards traversing path system NtrBC ([Fig 7](#), panel C), as the second non-hub (Table K in [S1 Text](#)). This corroborates the predictions from the traversing paths and, thus, shows that our new topological measure reveals cross-systemic elements which otherwise only stand out under detailed scrutiny of a large amount of biological information.

Discussion

Here we present an integrative network covering metabolic processes as well as regulatory events of *E. coli* but, especially, the interaction between both systems. With more than 10,000 vertices, it comprises around two third of the metabolic processes currently integrated in metabolic reconstructions [6] and concerning regulatory events, the presented network incorporates more than 95% of the established transcription-related processes [10]. Both, metabolic and gene regulatory processes are integrated on a genome scale. This approach differs from the procedures, where one of the two provides the network basis which is then expanded by closely related processes in the other subsystem. The latter is the dominant approach, for example, in conventional metabolic reconstructions which solely involve the encoding genes indirectly. Hitherto, integration of transcriptomics data could only be achieved using the so-called gene-protein-reaction (GPR) associations. On the one hand, this procedure limits the applicable data set to metabolic genes and, on the other hand, it acts on the assumption that all expressed enzymes are present in their active form. Starting from the integrative *E. coli* network, integrating transcriptomics data is much more straightforward and, more importantly, the complete data set can be applied. In this way, multi-domain variants of the frequently employed network-based interpretation of ‘omics’ data [59–63] can be formulated and indirect and regulatory impacts on metabolism can be examined.

The novelty of the reconstruction, the connection of metabolism and gene regulation, allows us not only to investigate the separate systems but also to assess their interactions. The most relevant connecting links are proteins, on the one hand, those acting as enzymes and, on the other hand, metabolic transcription factors. The functional classification, together with the topological analysis, suggests a network division into three domains: metabolic domain, protein interface and regulatory domain. This partition was corroborated by different connectivity measures and reflects a biologically reliable categorization in two delimited modules linked by a bridging module.

The principal structural feature of the network model, the three-domain organization, is reminiscent of the ‘bow-tie’ architectures frequently discussed in the theory of complex systems, where an input and an output layer are connected via a (typically much smaller) intermediate network [64–66]. Such a bow-tie structure (or, rather, the presence of several nested bow-tie architectures) has for example been discussed for metabolic networks [67], where the diversity of inputs (nutrients) and outputs (biomass components) is much larger than the intermediate processing layer. It has been hypothesized that such a bow-tie organization is a prerequisite for the robust operation of a complex system [64, 65]. Here we observe a bow-tie organization in a system consisting of a rich ‘material flow’ system (metabolism) and a similarly rich ‘control’ system (gene regulation) connected via a protein interface.

As our topological assessment shows, the bridging character of the protein interface entails a bottleneck functionality. The analysis of the new topological measure, termed *traversing paths*, highlighted three major biological systems represented by 12 vertices forming more than 40% of these paths (comprising in total 1465 distinct vertices). These traversing path systems, namely phosphotransferase system (PTS), ribonucleotide reducing (RNR) and nitrogen regulation two-component signal transduction (NtrBC) system, are well-investigated ones with key biological relevance for *E. coli*'s metabolism as well as its gene regulation suggesting that a topologically prominent position points to an important biologically functional entity.

Further detection of cross-systemic key elements in the network was accomplished using additional topological measures. In particular, two centrality measures were studied to account for different aspects of importance in terms of direct linkage and participation in network flow. Apart from conspicuous components, such as trivial compounds, currency metabolites and global regulators, a group of seven hubs were revealed by degree centrality whose characteristic is a significant low intra-domain degree fraction what numerically reflects the bridging feature of the protein interface. As expected, these components are located in the interface except for one, the vertex representing guanosine 5'-diphosphate 3'-diphosphate (ppGpp) which is affiliated to the metabolic domain. On the other hand, the inspection of betweenness centrality highlights rather biological systems than single components and as such point to key components detected before in their functional context. Besides trivial compounds and currency metabolites, this includes Crp-cAMP (cyclic-AMP receptor protein binding cyclic-AMP), Lrp-Leu (leucine-responsive regulatory protein binding leucine) and ppGpp which stand out due to their intra-domain degree fraction as well as seven further components already revealed as hubs.

Intriguingly, the interface-specific key elements of the network could be corroborated by exactly these two centrality measures. The assessment of the interface-induced subgraph using both centralities emphasizes altogether eight vertices of the downwards traversing paths discussed above contributing to the two major systems PTS and RNR. Taking into account the intra-domain degree fraction points out a representative of the upwards traversing path system NtrBC. In conclusion, the importance of vertices revealed by the here presented traversing paths could be reinforced by well-established topological measures showing the predictive power of the new measure.

Eventually, the key elements of the integrative *E. coli* network according to both centralities illustrate the importance of the different domains and their combined consideration (Fig 8). Unsurprisingly, the majority of key elements are affiliated to the metabolic domain and represent trivial compounds and currency metabolites, e.g., H^+ , H_2O , ATP and $NAD(P)^+$. Moreover, predominantly cross-systemic components top this combined list of central elements. First of all, the vertices emphasized also by their low intra-domain degree fraction attract attention, namely, Crp-cAMP, Lrp-Leu and ppGpp. These vertices demonstrate the value of the integrative approach: Only when embedded in domain context their vertex importance emerged. In case of the former two components, additionally, the composition unveils the cross-systemic role, i.e., a transcriptional factor protein binding a metabolic small molecule affecting its regulatory activity. Likewise the two regulatory key elements, Fur- Fe^{2+} and PhoB-P, exhibit this conspicuous linkage to the metabolic domain illustrating their cross-systemic property. In other words, they belong to the so-called metabolic transcription factors and, thus, are related to the upwards interface. The opposite is the case for the three metabolic Omp (outer membrane porin) transporters that are among the key elements. While their metabolic linkage is more than obvious, the relation to the regulatory domain appears when the encoding genes are examined. These are highly regulated amongst others by the global

Vertex name	DC	BC	Property
Proton	1	1	□
H ₂ O	2	2	□
ATP	5	3	■
Phosphate (P)	4	4	■
Proton (periplasmic)	6	5	□
Crp-cAMP, transcriptional dual regulator	3	9	▲
ADP	10	6	■
outer membran porin F	7	24	▼
outer membran porin C	7	29	▼
H ₂ O (periplasmic)	12	31	□
outer membran porin E	9	34	▼
Fur-Fe ²⁺ , transcriptional dual regulator	25	21	▲
Pyrophosphate	18	30	■
NAD ⁺	13	35	■
Phosphate (periplasmic)	19	36	■
PhoB-P, transcriptional dual regulator	45	12	▲
Lrp-Leucine, transcriptional dual regulator	43	17	▲
Guanosine 5'-diphosphate 3'-diphosphate	35	26	▲
NADP ⁺	21	41	■
Glutamate	24	40	▲
Pyruvate	30	39	▼
Coenzyme A	23	48	■
CO ₂	32	41	■
NH ₄ ⁺	31	46	▲

Fig 8. Key elements of the integrative *E. coli* network. Key elements with respect to degree (DC) and betweenness centrality (BC) are listed according to their rank as well as their functional characteristic and cross-systemic property, respectively. Open squares denote trivial compounds and filled squares indicate currency metabolites. The colored arrows depict the cross-systemic contribution—downwards interface-related (▼), upwards interface-related (▲). The orange arrows emphasize the cross-systemic components with significant low intra-domain degree fraction and the golden ones point out elements indirectly linking to one of the major traversing paths systems.

<https://doi.org/10.1371/journal.pcbi.1006962.g008>

regulators Crp-cAMP, Fur-Fe²⁺, Lrp-Leu and PhoB-P. In this manner, the Omp's are classical representatives of proteins related to the downwards interface, even though they are not affiliated with it. The remaining key elements are three metabolic small molecules which are counter-intuitively also related to the interface and the cross-systemic elements detected by the traversing paths. While in case of pyruvate the connection to PTS is apparent at first glance (Fig 7, panel A), the link of glutamate and ammonium and the NtrBC system is less perceptible. The actual connecting element is glutamine which is the ligase product of glutamate and ammonium. It activates the (de)uridylylation of the regulatory protein PII which, in turn, inhibits NtrB autophosphorylation [36, 68]. Altogether, the links to the three major traversing path systems are certainly not the only important processes these elements are involved in but they reinforce their biologically central roles. Remarkably, these connecting elements show up when considering the entire network while to acknowledge their importance the interface-specific analysis is needed.

Beyond the detection of key elements, the integrative approach will allow to examine the interplay and distribution of short-term and long-term regulation in *E. coli*'s metabolism. While metabolic regulation of, for instance, enzyme activities occurs on a short time-scale, regulation of gene expression is a long-term control process. Both types of regulation have been incorporated in the network even though only on a qualitative level, *i.e.*, as activator or inhibitor. Like this, the different effective ranges in metabolism can be assessed and, thus, its covering by one or both regulation types where central metabolism is said to be highly controlled.

From the perspective of recent advances in network theory

With their balance of structural detail and functional simplicity, network models are capable of revealing organizational principles, which are hard to recognize on a smaller systemic scale (*e.g.*, by analyzing individual pathways) or in functionally richer system representations (*e.g.*, in dynamical models). One purpose of the network provided here is to enable work at the interface of statistical physics and systems biology, where the rich toolbox of complex network analysis is employed to identify functionally relevant non-random features of such biological networks.

The recent work of [69], for example, showed that network structure can reveal, whether an enzyme is susceptible rather to genetic knockdown or pharmacologic inhibition. While in the present study, the network measures do not distinguish between different kinds of vertices or links, the rich biological meta data concerning the different biological roles of the components could be translated into distinct vertex and edge classes. In our own investigation [29] we used this fact to study, in a further example of such an interdisciplinary effort, the balance of robustness and sensitivity in the interdependent network of gene regulation and metabolism, based on the reconstructed network provided here.

In general, we expect that our network reconstruction can serve as a relevant data resource for the application of methods from the analysis of multiplex [70] and other multilayer networks [20, 71]. Recently, there has been a growing interest in the properties of these systems, especially in the presence of explicit interdependencies between vertices [70, 72]. In contrast to monoplex networks interdependent networks can show a qualitatively different robustness against failures, *i.e.*, cascading failures leading to a sudden system breakdown at a critical initial attack size [73, 74]. The case of different vertex types (as opposed to different edge types) has been considered, for example, in the context of secure communication in a network where eavesdroppers control sets of vertices [75].

On a general level, analyzing statistics of paths with respect to the network's large-scale structure, like the domain-traversing paths used here, might prove useful for the evaluation of other networks that show (possibly more than one) interface-like features.

Concluding remarks

In summary, the analysis of network topology allows to determine key system components in the integrative *E. coli* network. In line with expectations, trivial compounds as well as currency metabolites showed up regardless of the measure that has been applied. In addition, further obvious components including several global regulators were identified. More striking is the detection of components and systems which solely emerge when analyzing specifically the interface. These hidden elements are associated to two of the biologically well-investigated functional subsystems, PTS and NtrBC. Both well-established and newly designed measures of the interface point out the same subsystems, and even the analysis of the entire network discloses components indirectly related to these hidden subsystems.

Apart from trivial and currency metabolites, every detected key element of the entire network contributes to some extent to the downwards and/or upwards interface. This unlooked-for cross-systemic property is reflected either in the complex composition, the intra-domain degree fraction, the proximity to key systems, and/or the interplay with regulatory and metabolic processes. The biological relevance of these components supports their detection and reinforces the predictive power of the novel traversing path measure. In general, we believe that the presented integrative *E. coli* network allows further investigations of the interplay of metabolism and gene regulation which will provide insights into cellular, system-wide responses.

Methods

The interconnected *E. coli* network is based on the EcoCyc database [22], release 20.0, which includes verified information of metabolic and regulatory processes (corresponds to RegulonDB 8.6 [10]) for *E. coli* K-12 substr. MG1655. The network is represented as a graph comprising four different types of vertices: (1) encoding genes, (2) protein monomers and complexes (including enzymes), (3) small compounds, and (4) (bio)chemical reactions (Table A in S1 Text). The protein vertices are further subclassified into protein monomers, protein-protein complexes, protein-compound complexes, and protein-RNA complexes. Regarding the edges, we distinguish three types: encoding and catalyzing associations, reaction connections to educts and products, and regulatory links to sources and targets (Table B in S1 Text).

Extraction of database information

First, relevant information of the database has been extracted and arranged (Algorithm 1 in Fig 9). For each regulatory process, the respective source and target were specified and converted to match one of the vertex types ('regulation.dat', file name of the EcoCyc-archive). To this end, the transcript units were separated into promoter, genes and terminator (if applicable), and the regulatory processes were multiplied per comprising gene. Moreover, each regulating RNA has been translated into its encoding gene to meet the vertex types. In case of the metabolic processes, the reaction educts and products as well as the catalyzing enzymes have been assembled and converted to match one of the vertex groups, the respective educt and product stoichiometry have been assigned and the reaction compartmentation and reversibility have been assessed ('reactions.dat'). Thereby, as cell compartments the periplasmic space, the inner membrane, and the cytosol have been taken into account and reversible reactions have been split up.

Second, vertex candidates have been validated ('reactions.dat', 'compounds.dat', 'proteins.dat', 'genes.dat', 'rnas.dat') and divided into *reaction*, *compound*, *protein monomer*, *protein-protein complex*, *protein-compound complex*, *protein-RNA complex*, and *gene*. In doing so, generic terms such as DIPEPTIDES have been substituted ('classes.dat') and double annotations, e.g., CPD-15709 and FRUCTOSE-6P have been decoded. Thereupon, the compositions and the encoding genes of the assembled proteins have been gathered and matched to the vertex groups and the respective logical operation and stoichiometry have been annotated ('protcplx.col'). Based on the validated vertex lists, the regulatory and metabolic processes have been updated whereby each process was removed with at least one unidentified vertex resulting in the final edge lists. Fig C in S1 Text provides a flowchart of this algorithmic procedure.

Network implementation

With the validated vertex and edge lists the graph has been assembled and its largest weakly connected component has been extracted. The three domain partition MD—PI—RD (Table 1 and A in S1 Text) as well as the two-domain partition are implemented as vertex properties

```

regprocs = AssembleRegulatoryProcesses()
    Extract information on regulatory processes ('regulation.dat')
    REG_type,REG_ID,REG_source,REG_target,REG_mode = ParseRegulation()
    for rT in REG_type do
        for rID in REG_ID[rT] do
            Convert regulation source and target to match one of the vertex groups
            if REG_source[iT][iID] == RNA then
                Translate RNA into corresponding genes
            if REG_target[iT][iID] == transcript unit then
                Split up transcription units into promoters, genes and terminators
                Translate promoters and terminators into corresponding genes
            s,t = Match2Vertex(REG_source[rT][rID],REG_target[rT][rID])
            regprocs.append(rT,rID,s,t,REG_mode[rT][rID])

metprocs = AssembleCompartmentedMetabolicProcesses()
    Extract information on metabolic processes ('reaction.dat')
    Rxn_ID,Rxn_enz,Rxn_dir,Rxn_l,Rxn_r,Rxn_loc = ParseReactions()
    for mID in Rxn_ID do
        Match catalyzing enzyme, educts and products to one of the vertex groups
        enz,left,right = Match2Vertex(Rxn_enz[mID],Rxn_l[mID],Rxn_r[mID])
        Annotate reaction compartmentation
        left_comp,right_comp = AnnotateCompartment(Rxn_loc[mID],left,right)
        Split up reversible reactions and reverse 'right-to-left' reactions
        if Rxn_dir == 'reversible' then
            metprocs.append(mID+'_f',enz,left_comp,right_comp)
            metprocs.append(mID+'_r',enz,right_comp,left_comp)
        else if Rxn_dir == 'right-to-left' then
            metprocs.append(mID,enz,right_comp,left_comp)
        else
            metprocs.append(mID,enz,left_comp,right_comp)

valReg,valMet,valVertices = ValidateProcessesVertices(regprocs,metprocs)
    Compile vertex candidates from regulation sources and targets as well as metabolic
    reactions, the corresponding educts and products, and encoding enzymes
    vertCands = AssembleCandidates(regprocs,metprocs)
    Assign vertex candidates to the seven types: reaction, compound, gene, protein
    monomer, protein-protein-complex, protein-compound-complex, protein-rna-complex
    rxns,cmps,gns,prts,ppc,pcc,prc = ComposeVertexLists(vertCands)
    Decode generic terms ('classes.dat') and double annotations
    curReg,curMet,curRxns,curCmps = Curation(regprocs,metprocs,rxns,cmps)
    Prune vertex lists regarding unmapped candidates
    valVertices = PruneVertices(curRxns,curCmps,gns,prts,ppc,pcc,prc)
    Prune regulatory and metabolic processes with respect to validated vertex lists
    valReg,valMet = PruneRegMetProcesses(curReg,curMet,valVertices)
    Export lists of validated vertices and links, i.e., processes

```

Fig 9. Algorithm 1. Extraction of database information (EcoCyc, release 20.0) on regulatory and metabolic processes.

<https://doi.org/10.1371/journal.pcbi.1006962.g009>

affiliation and *metabolic*. The initial categorization of both partitions is based on the vertex type 'reaction' which is denoted as purely metabolic and interface-related if all educts and products are compounds and proteins, respectively. Mixed educt and product types demand further clarification later on. Similarly, non-ambiguous vertices of type 'compound', 'protein' and 'gene' are affiliated based on the affiliation of their neighbor vertices. This means that, if the influential adjacent vertices have the same affiliation, the vertex will be assigned to the same or its assignment needs a detailed consideration. In this way, genes and proteins that are

Table 1. Comparison of vertex composition and the coverage to the model from [11] of the integrative *E. coli* network (Largest WCC), the underlying full graph and the EcoCyc database (release 20.0).

Vertices	Largest WCC		Full graph		Database**	
	EcoCyc	iMC1010*	EcoCyc	iMC1010*	EcoCyc	iMC1010*
Reaction	4693	569/ 767	7251	601/ 767	2617	717/ 767
Compound	2681	557/ 615	2785	558/ 615	2678	614/ 615
Gene	2545	971/1010	2801	981/1010	4506	1010/1010
Protein monomer	1917	771/ 817	2012	775/ 817	5708	815/ 817
Protein-protein complex	929		986			
Protein-compound complex	100		103			
Protein-RNA complex	3		4			
	12868	2868/3209	15942	2915/3209	15509	3151/3209

*accounted only for enzymatic reactions and unique metabolites (1076 and 762 in total);

** reactions in EcoCyc database are case insensitive for compartmentation and reversibility

<https://doi.org/10.1371/journal.pcbi.1006962.t001>

not involved in any regulatory process can be assigned to the metabolic domain (MD). Subsequently, deferred vertex affiliations are resolved iteratively based on their neighbor vertices affiliation until no further vertex affiliations can be assigned. The final ambiguous vertices, in total 13 of 12868, are assigned as interface vertices since they cannot be uniquely assigned to metabolic or regulatory domain. Algorithms 2A and 2B in Figs 10 and 11 show the detailed domain affiliation process of a vertex.

Moreover, the mapping to the *E. coli* model of [11] has been annotated which integrates the metabolic network iJR904 published by [5] and the transcription regulatory events related to the encoding genes of the catalyzing enzymes. To this end, genes, proteins, metabolites as well as biochemical reactions of the metabolic model have been mapped to the EcoCyc database (release 20.0), in a first step automatically based on their identifier and the resulting dictionaries have been manually curated. As the EcoCyc database does not account for compartmentation of compounds and reactions as well as for exchange reactions, unique metabolites and internal reactions have been considered resulting in a coverage of more than 93%. By additionally disregarding internal transport reactions a coverage of 96.5% can be achieved (Table 1).

Integrating the manually curated Covert dictionaries, each vertex has attributed (1) a unique identifier, according to the EcoCyc identifier but also indicating the compartment, (2) a unique type reference, (3) a unique assignment of the model components from [11], if applicable, and (4) the affiliations of the two- and three-domain partition. Furthermore, vertices of types *gene* and *reaction* have (5a) a name assigned, the Blattner ID and the EC number, if applicable. The remaining vertices have additionally (5b) a compartment assigned, where cytosol (c), extracellular space (e), periplasmic space (p), inner membrane (i), outer membrane (o) and membrane in general (m) were taken into account. Similarly, each edge of the network has the attribute (1) type, specifying the connected vertices, and the corresponding (2) stoichiometry, where zero is assigned if not applicable or ambiguous. For edges depicting regulatory processes the stoichiometry actually denotes the mode of regulation, namely activation (+, 1) inhibition (−, −1) or combined (0). These edges additionally have assigned (3) an identifier, according to the EcoCyc identifier and (4) a name, specifying the regulation type. All other edge types can be classified as either representing conjunct or disjunct links in the sense that all or solely one incoming link is required for functionality (Table B in S1 Text).

The fully annotated integrative reconstruction of *E. coli*'s metabolic and regulatory processes is provided as a graph representation in S1 File.

```

aff = AssignNonAmbiguousVertices()
Metabolic regulatory processes will be interpreted as metabolic links
MetRegs = [Regulation-of-Enzyme-Activity, Regulation-of-Reactions]
for v in Vtypes(network.vertices) == reaction do
  if Vtypes(Educts(v) ^ Products(v)) == compound then aff(v) = metabolic
  else if Vtypes(Educts(v) ^ Products(v)) == (compound ∨ protein) then
    aff(v) = interface
  else aff(v) = tba
for v in Vtypes(network.vertices) == compound do
  Compounds which participate in at least one reaction
  if # InvolvedRxns(v) > 0 then
    if # aff(InvolvedRxns(v)) == 1 then aff(v) = aff(InvolvedRxns(v))
    else if metabolic in aff(InvolvedRxns(v)) then aff(v) = metabolic
    else if interface in aff(InvolvedRxns(v)) then aff(v) = interface
    else aff(v) = ambiguous
  Compounds which adjacent vertices are either compounds or proteins
  else if Vtypes(Neighbors(v)) == (compound ∨ protein) then
    if regulation in Etypes(OutEdges(v)) then
      if REGtypes(OutEdges(v)) in MetRegs then aff(v) = metabolic
      else aff(v) = ambiguous
    else aff(v) = tba
for v in Vtypes(network.vertices) == protein do
  Proteins with enzymatic function
  if enzyme - reaction in Etypes(OutEdges(v)) then
    if Vtypes(OutNeighbors(v)) == reaction then
      if # aff(OutNeighbors(v)) == 1 then aff(v) =
        aff(OutNeighbors(v))
      else if metabolic in aff(OutNeighbors(v)) then aff(v) = metabolic
      else if interface in aff(OutNeighbors(v)) then aff(v) = interface
      else aff(v) = tba
    else if regulation in Etypes(OutEdges(v)) then
      if REGtypes(OutEdges(v)) in MetRegs then aff(v) = metabolic
      else aff(v) = regulatory
    else if # aff(InvolvedRxns(OutNeighbors(v))) == 1 then aff(v) =
      aff(InvolvedRxns(OutNeighbors(v)))
    else if metabolic in aff(InvolvedRxns(OutNeighbors(v))) then
      aff(v) = metabolic
    else if interface in aff(InvolvedRxns(OutNeighbors(v))) then
      aff(v) = interface
    else aff(v) = tba
  Proteins involved in metabolic reactions
  else if educt - reaction in Etypes(OutEdges(v)) then aff(v) = interface
  Proteins involved in regulatory processes
  else if regulation in Etypes(OutEdges(v)) then
    if REGtypes(OutEdges(v)) in MetRegs then aff(v) = metabolic
    else aff(v) = regulatory
  Proteins involved in protein complex formation
  else if protein - complex in Etypes(OutEdges(v)) then
    if reaction - product in Etypes(InEdges(v)) then aff(v) = interface
    else if regulation in Etypes(InEdges(v)) then
      if REGtypes(InEdges(v)) in MetRegs then aff(v) = metabolic
      else aff(v) = regulatory
    else aff(v) = interface
  else aff(v) = ambiguous

```

Fig 10. Algorithm 2A. Network affiliation compilation based on vertex type, and the vertex neighbors types and affiliations. Affiliation assignment for non-ambiguous reactions, compounds and proteins. Continued in Algorithm 2B in Fig 11.

<https://doi.org/10.1371/journal.pcbi.1006962.g010>

```

aff = AssignNonAmbiguousVertices() function resumption
  for v in Vtypes(network.vertices) == gene do
    if regulation in Etypes(OutEdges(v)) then aff(v) = regulatory
    else if regulation in Etypes(InEdges(v)) then aff(v) = regulatory
    else if regulatory in aff(OutNeighbors(v)) then aff(v) = regulatory
    else if metabolic in aff(OutNeighbors(v)) then aff(v) = metabolic
    else aff(v) = ambiguous
Assign affiliation for vertices formerly denoted as ambiguous
aff = AssignAmbiguousVertices(aff)
  additionalRun = True
  while additionalRun do
    additionalRun = False
    for v in aff(network.vertices) == ambiguous do
      if # aff(AllNeighbors(v)) == 1 then
        aff(v) = aff(AllNeighbors(v))
        additionalRun = True
      else if (regulatory ^ metabolic) in aff(AllNeighbors(v)) then
        aff(v) = ambiguous
      else if regulatory in aff(AllNeighbors(v)) then
        aff(v) = regulatory
        additionalRun = True
      else if metabolic in aff(AllNeighbors(v)) then
        aff(v) = metabolic
        additionalRun = True
      else aff(v) = ambiguous
  for v in aff(network.vertices) == ambiguous do
    aff(v) = interface

```

Fig 11. Algorithm 2B. Network affiliation compilation based on vertex type, and the vertex neighbors types and affiliations. Affiliation assignment for non-ambiguous genes and vertices assigned as *ambiguous*.

<https://doi.org/10.1371/journal.pcbi.1006962.g011>

Graph properties concerning intra- and inter-module connectivity

The following measures have been used in the assessment of the graph partitioning scheme.

Inter-module edge fraction c . Given the set of vertices with the domain label D , edges connecting these vertices to a vertex of a different label are considered *external*, while edges between vertices of the same label are *internal*. We call

$$c_D = \frac{\#(\text{external edges of } D)}{\#(\text{external} + \text{internal edges of } D)}$$

the inter-module edge fraction of domain D .

Network modularity M . denotes the degree to which a given partition divides the network in highly connected groups, modules, which are comparably sparsely connected among each other. Therefore, the intra-module links are counted against the total degree of the module vertices (Eq 1),

$$M = \sum_{j=1}^{N_M} \left(\frac{L(v_{M_j}, w_{M_j})}{L_G} - \left(\frac{\text{deg}(v_{M_j})}{2L_G} \right)^2 \right) \quad (1)$$

with N_M - # of modules, L_G - # of links of graph G , $L(v_M, w_N) = \sum_{v \in M} \sum_{w \in N} \text{link}(v, w)$,

$$\text{deg}(v_M) = \sum_{v \in M} k_v.$$

Here k_v is the degree of vertex v and $\text{link}(v, w)$ denotes an undirected edge between vertices v and w . Note that this formulation of modularity (taken from [30]) coincides with the definition from [76].

Domain-traversing paths

A *traversing path* connects the regulatory and the metabolic domains via the protein interface, specifically, a traversing path of length k is of the form

$$[(u, v_1), (v_1, v_2), \dots, (v_{k-1}, w)] \tag{2}$$

where the vertices u and w are from the regulatory and the metabolic domain (and vice versa) and the vertices v_i are distinct and part of the protein interface. Starting from the set of edges directly at the intersection of two domains iteratively the vertex successors of the interface domain as well as the final, first successor in the third domain have been determined (Algorithm 3 in Fig 12).

Vertex centrality

The key elements of the integrative *E. coli* network have been determined based on two graph properties.

```

input : graph G = {V,E},
       map aff : V → {regulatory, interface, metabolic}

downTP, upTP = TraversingPaths( G, aff )
  downTP, upTP = list()
  for (source, target) in E do
    if aff(target) == interface then
      if aff(source) == regulatory then
        | InterfaceSuccessors( G, [source, target], metabolic, downTP )
      else if aff(source) == metabolic then
        | InterfaceSuccessors( G, [source, target], regulatory, upTP )

recursively determine all successors
InterfaceSuccessors( G, vList, aimAff, sucList )
  for suc in outneighbours( vList[-1] ) do
    vL = list(vList)
    if suc not in vL then
      vL.append( suc )
      if aff(suc) == interface then InterfaceSuccessors( G, vL, aimAff,
        sucList )
      else if aff(suc) == aimAff then
        | sucList.append( vL )

```

Fig 12. Algorithm 3. Recursive algorithm for the determination of the, so-termed, domain-traversing paths from regulatory to metabolic domain and *vice versa* truly passing the interface domain.

<https://doi.org/10.1371/journal.pcbi.1006962.g012>

Degree Centrality DC . is a local centrality measure and denotes the total number of in- and out-going edges of a vertex, (Eq 3),

$$DC(v) = k_v = k_v^{\text{in}} + k_v^{\text{out}}. \quad (3)$$

Here, the vertices with a total degree greater than 50 are termed hubs (see the degree distribution in Fig D in S1 Text).

By additionally accounting for the domain boundaries, the intra-domain degree fraction ξ (also termed embeddedness [44]) has been defined as ratio of internal degree, within domain D , and total degree of a vertex, (Eq 4),

$$\xi_D(v) = \frac{k_v^{\text{int}}}{k_v} = \frac{1}{k_v} \sum_{w \in D} (A_{vw} + A_{wv}) \quad (4)$$

where A denotes the adjacency matrix of the graph.

Betweenness Centrality BC . describes the impact on the flux through the network, under the assumption that the transfer follows the shortest paths. In particular, it quantifies the fraction of shortest paths between all pairs of vertices which involve the designated vertex (Eq 5),

$$BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (5)$$

where σ_{st} is the number of all shortest-paths between the vertices s and t while $\sigma_{st}(v)$ yields the number of these paths that run through v [41].

Supporting information

S1 Text. Supplementary information. Compilation of additional material for the integrative network model of *E. coli*.

(PDF)

S1 File. Graph representation. The fully annotated integrative reconstruction of *E. coli*'s metabolic and regulatory processes is provided as a graph representation.

(XML)

Author Contributions

Conceptualization: Stefan Bornholdt, Marc-Thorsten Hütt.

Data curation: Anne Grimbs.

Formal analysis: Anne Grimbs, David F. Klosik.

Funding acquisition: Stefan Bornholdt, Marc-Thorsten Hütt.

Investigation: Anne Grimbs.

Methodology: Anne Grimbs, David F. Klosik.

Project administration: Stefan Bornholdt, Marc-Thorsten Hütt.

Supervision: Stefan Bornholdt, Marc-Thorsten Hütt.

Validation: David F. Klosik.

Visualization: Anne Grimbs, David F. Klosik.

Writing – original draft: Anne Grimbs, Marc-Thorsten Hütt.

Writing – review & editing: Anne Grimbs, David F. Klosik, Stefan Bornholdt, Marc-Thorsten Hütt.

References

1. Kochanowski K, Sauer U, Noor E. Posttranslational regulation of microbial metabolism. *Current Opinion in Microbiology*. 2015; 27:10–17. <https://doi.org/10.1016/j.mib.2015.05.007> PMID: 26048423
2. Escalante A, Salinas Cervantes A, Gosset G, Bolivar F. Current knowledge of the *Escherichia coli* phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. *Applied Microbiology and Biotechnology*. 2012; 94(6):1483–1494. <https://doi.org/10.1007/s00253-012-4101-5> PMID: 22573269
3. Goncalves E, Bucher J, Ryll A, Niklas J, Mauch K, Klamt S, et al. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol Biosyst*. 2013; 9(7):1576–1583. <https://doi.org/10.1039/c3mb25489e> PMID: 23525368
4. Edwards JS, Palsson B. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97(10):5528–5533. <https://doi.org/10.1073/pnas.97.10.5528> PMID: 10805808
5. Reed JL, Vo TD, Schilling CH, Palsson B. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology*. 2003; 4(9):R54. <https://doi.org/10.1186/gb-2003-4-9-r54> PMID: 12952533
6. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*. 2007; 3:121. <https://doi.org/10.1038/msb4100155> PMID: 17593909
7. Orth JD, Palsson B, Fleming RMT. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus*. 2010; 4(1). <https://doi.org/10.1128/ecosalplus.10.2.1> PMID: 26443778
8. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Molecular Systems Biology*. 2011; 7(1):535. <https://doi.org/10.1038/msb.2011.65> PMID: 21988831
9. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nature biotechnology*. 2017; 35(10):904. <https://doi.org/10.1038/nbt.3956> PMID: 29020004
10. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeda D, Muñiz-Rascado L, García-Sotelo JS, et al. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Research*. 2015; 44(D1):D133–D143. <https://doi.org/10.1093/nar/gkv1156> PMID: 26527724
11. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson B. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*. 2004; 429(6987):92–96. <https://doi.org/10.1038/nature02456> PMID: 15129285
12. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*. 2007; 3:101. <https://doi.org/10.1038/msb4100141> PMID: 17437026
13. Samal A, Jain S. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Systems Biology*. 2008; 2(1):21. <https://doi.org/10.1186/1752-0509-2-21> PMID: 18312613
14. Gianchandani EP, Joyce AR, Palsson B, Papin JA. Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. *PLoS Computational Biology*. 2009; 5(6):e1000403. <https://doi.org/10.1371/journal.pcbi.1000403> PMID: 19503608
15. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(41):17845–17850. <https://doi.org/10.1073/pnas.1005139107> PMID: 20876091
16. Imam S, Schäuble S, Brooks AN, Baliga NS, Price ND. Data-driven integration of genome-scale regulatory and metabolic network models. *Frontiers in Microbiology*. 2015; 6:409. <https://doi.org/10.3389/fmicb.2015.00409> PMID: 25999934
17. Vivek-Ananth R, Samal A. Advances in the integration of transcriptional regulatory information into genome-scale metabolic models. *Biosystems*. 2016; 147:1–10. <https://doi.org/10.1016/j.biosystems.2016.06.001> PMID: 27287878

18. Hao T, Wu D, Zhao L, Wang Q, Wang E, Sun J. The Genome-Scale Integrated Networks in Microorganisms. *Frontiers in Microbiology*. 2018; 9:296. <https://doi.org/10.3389/fmicb.2018.00296> PMID: 29527198
19. Kumar S, Mahajan S, Jain S. Feedbacks from the metabolic network to the genetic network reveal regulatory modules in *E. coli* and *B. subtilis*. *PLOS ONE*. 2018; 13(10):1–36. <https://doi.org/10.1371/journal.pone.0203311>
20. Radde NE, Hütt MT. The Physics behind Systems Biology. *EPJ Nonlinear Biomedical Physics*. 2016; 4(1):7–. <https://doi.org/10.1140/epjnbp/s40366-016-0034-8>
21. Galperin MY, Fernandez-Suarez XM, Rigden DJ. The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. *Nucleic Acids Research*. 2017; 45(D1):D1–D11. <https://doi.org/10.1093/nar/gkw1188> PMID: 28053160
22. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, et al. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*. 2012; 41(D1):D605–D612. <https://doi.org/10.1093/nar/gks1027> PMID: 23143106
23. Ideker T, Krogan NJ. Differential network biology. *Molecular Systems Biology*. 2012; 8:565. <https://doi.org/10.1038/msb.2011.99> PMID: 22252388
24. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEx, the Network Data Exchange. *Cell Systems*. 2015; 1(4):302–305. <https://doi.org/10.1016/j.cels.2015.10.001> PMID: 26594663
25. Ku Yu M, Kramer M, Dutkowski J, Srivas R, Licon K, Kreisberg JF, et al. Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems. *Cell Systems*. 2016; 2(2):77–88. <https://doi.org/10.1016/j.cels.2016.02.003>
26. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*. 2013; 9(1):693. <https://doi.org/10.1038/msb.2013.52> PMID: 24084808
27. Smallbone K. Standardized network reconstruction of *E. coli* metabolism. arXiv preprint arXiv:13042960. 2013;4.
28. Liu JK, O'Brien EJ, Lerman JA, Zengler K, Palsson B, Feist AM. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC Systems Biology*. 2014; 8(1):110. <https://doi.org/10.1186/s12918-014-0110-6> PMID: 25227965
29. Klosik DF, Grimbs A, Bornholdt S, Hütt MT. The interdependent network of gene regulation and metabolism is robust where it needs to be. *Nature Communications*. 2017; 8:534. <https://doi.org/10.1038/s41467-017-00587-4> PMID: 28912490
30. Guimera R, Nunes Amaral LA. Functional cartography of complex metabolic networks. *Nature*. 2005; 433(7028):895–900. <https://doi.org/10.1038/nature03288> PMID: 15729348
31. Deutscher J. The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology*. 2008; 11(2):87–93. <https://doi.org/10.1016/j.mib.2008.02.007> PMID: 18359269
32. Thelander L, Reichard P. Reduction of Ribonucleotides. *Annual Review of Biochemistry*. 1979; 48(1):133–158. <https://doi.org/10.1146/annurev.bi.48.070179.001025> PMID: 382982
33. Fontecave M, Nordlund P, Eklund H, Reichard P. The Redox Centers of Ribonucleotide Reductase of *Escherichia coli*. In: Nord FF, editor. *Advances in Enzymology and Related Areas of Molecular Biology*. vol. 65. Wiley-Blackwell; 1992. p. 147–183. Available from: <http://dx.doi.org/10.1002/9780470123119.ch4>.
34. Jordan A, Reichard P. Ribonucleotide reductases. *Annual Review of Biochemistry*. 1998; 67(1):71–98. <https://doi.org/10.1146/annurev.biochem.67.1.71> PMID: 9759483
35. Torrents E. Ribonucleotide reductases: essential enzymes for bacterial life. *Frontiers in Cellular and Infection Microbiology*. 2014; 4:52. <https://doi.org/10.3389/fcimb.2014.00052> PMID: 24809024
36. Jiang P, Ninfa AJ. Regulation of autophosphorylation of *Escherichia coli* nitrogen regulator II by the PII signal transduction protein. *J Bacteriol*. 1999; 181(6):1906–1911. PMID: 10074086
37. Reitzer L. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annual Review of Microbiology*. 2003; 57(1):155–176. <https://doi.org/10.1146/annurev.micro.57.030502.090820> PMID: 12730324
38. Brown DR, Barton G, Pan Z, Buck M, Wigneshweraraj S. Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nature Communications*. 2014; 5:4115. <https://doi.org/10.1038/ncomms5115> PMID: 24947454
39. Mathews CK. DNA precursor metabolism and genomic stability. *The FASEB Journal*. 2006; 20(9):1300–1314. <https://doi.org/10.1096/fj.06-5730rev> PMID: 16816105
40. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. 2010; 32(3):245–251. <http://dx.doi.org/10.1016/j.socnet.2010.03.006>
41. Newman M. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc.; 2010.

42. Kiley PJ, Beinert H. Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster. *FEMS Microbiology Reviews*. 1998; 22(5):341–352. <https://doi.org/10.1111/j.1574-6976.1998.tb00375.x> PMID: 9990723
43. Travers A, Muskhelishvili G. DNA supercoiling—a global transcriptional regulator for enterobacterial growth? *Nat Rev Micro*. 2005; 3(2):157–169. <https://doi.org/10.1038/nrmicro1088>
44. Fortunato S, Hric D. Community detection in networks: A user guide. *Physics Reports*. 2016; 659:1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
45. Kolb A, Busby S, Buc H, Garges S, Adhya S. Transcriptional regulation by cAMP and its receptor protein. *Annual Review of Biochemistry*. 1993; 62(1):749–797. <https://doi.org/10.1146/annurev.bi.62.070193.003533> PMID: 8394684
46. Fic E, Bonarek P, Gorecki A, Kedracka-Krok S, Mikolajczak J, Polit A, et al. cAMP Receptor Protein from *Escherichia coli* as a Model of Signal Transduction in Proteins—A Review. *Journal of Molecular Microbiology and Biotechnology*. 2009; 17(1):1–11. <https://doi.org/10.1159/000178014> PMID: 19033675
47. Magnusson LU, Farewell A, Nyström T. ppGpp: a global regulator in *Escherichia coli*. *Trends in Microbiology*. 2005; 13(5):236–242. <https://doi.org/10.1016/j.tim.2005.03.008> PMID: 15866041
48. Potrykus K, Cashel M. (p)ppGpp: still magical? *Annual Review of Microbiology*. 2008; 62(1):35–51. <https://doi.org/10.1146/annurev.micro.62.081307.162903> PMID: 18454629
49. Srivatsan A, Wang JD. Control of bacterial transcription, translation and replication by (p)ppGpp. *Current Opinion in Microbiology*. 2008; 11(2):100–105. <https://doi.org/10.1016/j.mib.2008.02.001> PMID: 18359660
50. Spiro S. Regulators of bacterial responses to nitric oxide. *FEMS Microbiology Reviews*. 2007; 31(2):193–211. <https://doi.org/10.1111/j.1574-6976.2006.00061.x> PMID: 17313521
51. Partridge JD, Bodenmiller DM, Humphrys MS, Spiro S. NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Molecular Microbiology*. 2009; 73(4):680–694. <https://doi.org/10.1111/j.1365-2958.2009.06799.x> PMID: 19656291
52. Tucker NP, Le Brun NE, Dixon R, Hutchings MI. There's NO stopping NsrR, a global regulator of the bacterial NO stress response. *Trends in Microbiology*. 2010; 18(4):149–156. <https://doi.org/10.1016/j.tim.2009.12.009> PMID: 20167493
53. Ernsting BR, Atkinson MR, Ninfa AJ, Matthews RG. Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli*. *Journal of Bacteriology*. 1992; 174(4):1109–1118. <https://doi.org/10.1128/jb.174.4.1109-1118.1992> PMID: 1346534
54. Calvo JM, Matthews RG. The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiological Reviews*. 1994; 58(3):466–490. PMID: 7968922
55. Brinkman AB, Ettema TJG, De Vos WM, Van Der Oost J. The Lrp family of transcriptional regulators. *Molecular Microbiology*. 2003; 48(2):287–294. <https://doi.org/10.1046/j.1365-2958.2003.03442.x> PMID: 12675791
56. Schulz GE. Bacterial porins: structure and function. *Current Opinion in Cell Biology*. 1993; 5(4):701–707. [https://doi.org/10.1016/0955-0674\(93\)90143-E](https://doi.org/10.1016/0955-0674(93)90143-E) PMID: 8257610
57. Jap BK, Walian PJ. Structure and functional mechanism of porins. *Physiological Reviews*. 1996; 76(4):1073–1088. <https://doi.org/10.1152/physrev.1996.76.4.1073> PMID: 8874494
58. Schirmer T. General and Specific Porins from Bacterial Outer Membranes. *Journal of Structural Biology*. 1998; 121(2):101–109. <https://doi.org/10.1006/jsbi.1997.3946> PMID: 9615433
59. Marr C, Geertz M, Hütt MT, Muskhelishvili G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol*. 2008; 2(1):18. <https://doi.org/10.1186/1752-0509-2-18> PMID: 18284674
60. Sonnenschein N, Geertz M, Muskhelishvili G, Hütt MT. Analog regulation of metabolic demand. *BMC Syst Biol*. 2011; 5(1):40. <https://doi.org/10.1186/1752-0509-5-40> PMID: 21406074
61. Sonnenschein N, Golib Dzib JF, Lesne A, Eilebrecht S, Boulkroun S, Zennaro MC, et al. A network perspective on metabolic inconsistency. *BMC Systems Biology*. 2012; 6(1):41. <https://doi.org/10.1186/1752-0509-6-41> PMID: 22583819
62. Knecht C, Fretter C, Rosenstiel P, Krawczak M, Hütt MT. Distinct metabolic network states manifest in the gene expression profiles of pediatric inflammatory bowel disease patients and controls. *Scientific Reports*. 2016; 6:32584. <https://doi.org/10.1038/srep32584> PMID: 27585741
63. Beber ME, Sobetzko P, Muskhelishvili G, Hütt MT. Interplay of digital and analog control in time-resolved gene expression profiles. *EPJ Nonlinear Biomedical Physics*. 2016; 4(1):8. <https://doi.org/10.1140/epjnbp/s40366-016-0035-7>

64. Kitano H. Biological robustness. *Nature Reviews Genetics*. 2004; 5:826–837. <https://doi.org/10.1038/nrg1471> PMID: 15520792
65. Kitano H. Towards a theory of biological robustness. *Molecular Systems Biology*. 2007; 3:137. <https://doi.org/10.1038/msb4100179> PMID: 17882156
66. Friedlander T, Mayo AE, Tlusty T, Alon U. Evolution of Bow-Tie Architectures in Biology. *PLoS Computational Biology*. 2014; 11(3):e1004055. <https://doi.org/10.1371/journal.pcbi.1004055>
67. Csete M, Doyle J. Bow ties, metabolism and disease. *Trends in Biotechnology*. 2004; 22(9):446–450. <https://doi.org/10.1016/j.tibtech.2004.07.007> PMID: 15331224
68. van Heeswijk WC, Westerhoff HV, Boogerd FC. Nitrogen assimilation in *Escherichia coli*: putting molecular data into a systems perspective. *Microbiology and Molecular Biology Reviews*. 2013; 77(4):628–695. <https://doi.org/10.1128/MMBR.00025-13> PMID: 24296575
69. Jensen KJ, Moyer CB, Janes KA. Network Architecture Predisposes an Enzyme to Either Pharmacologic or Genetic Targeting. *Cell systems*. 2016; 2(2):112–121. <https://doi.org/10.1016/j.cels.2016.01.012> PMID: 26942229
70. Radicchi F, Bianconi G. Redundant Interdependencies Boost the Robustness of Multiplex Networks. *Physical Review X*. 2017; 7:011013. <https://doi.org/10.1103/PhysRevX.7.011013>
71. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *Journal of Complex Networks*. 2014; 2(3):203–271. <https://doi.org/10.1093/comnet/cnu016>
72. Gao J, Buldyrev SV, Stanley HE, Havlin S. Networks formed from interdependent networks. *Nature Physics*. 2012; 8(1):40–48. <https://doi.org/10.1038/nphys2180>
73. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. *Nature*. 2010; 464(7291):1025–1028. <https://doi.org/10.1038/nature08932> PMID: 20393559
74. Son SW, Bizhani G, Christensen C, Grassberger P, Paczuski M. Percolation theory on interdependent networks based on epidemic spreading. *Europhysics Letters*. 2012; 97:16006. <https://doi.org/10.1209/0295-5075/97/16006>
75. Krause SM, Danziger MM, Zlatić V. Hidden Connectivity in Networks with Vulnerable Classes of Nodes. *Physical Review X*. 2016; 6:041022. <https://doi.org/10.1103/PhysRevX.6.041022>
76. Girvan M, Newman ME. Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002; 99(12):7821–7826. <https://doi.org/10.1073/pnas.122653799>