The modular biochemical reaction network structure of cellular translation

Bruno Cuevas-Zuviría [a,b] , Evrim Fer [a,c], Zachary R. Adam [a,d] and Betül Kaçar [a,*]

[a] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA
[b] Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid, Madrid, Spain
[c] Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI, USA
[d] Department of Geosciences, University of Wisconsin-Madison, Madison, WI, USA

*To whom correspondence should be addressed: bkacar@wisc.edu

Abstract

Translation is an essential attribute of all living cells. At the heart of cellular operation, it is a chemical information decoding process that begins with an input string of nucleotides and ends with the synthesis of a specific output string of peptides. The translation process is interconnected with gene expression, physiological regulation, transcription, and responses to signaling molecules, among other cellular functions. Foundational efforts have uncovered a wealth of knowledge about the mechanistic functions of and many interactions between components of translation, but the broader biochemical connections between translation, metabolism and polymer biosynthesis that enable translation to occur have not been comprehensively mapped. Here we present a multilayer graph of biochemical reactions describing the translation, polymer biosynthesis and metabolism networks of an *Escherichia coli* cell. Intriguingly, the compounds that compose these three layers are distinctly aggregated into three modes regardless of their layer categorization. Multimodal mass distributions are well-known in ecosystems, but this is the first such distribution reported at the biochemical level. The connectivity distributions of the translation and metabolic networks are each likely to be heavy-tailed, but the polymer biosynthesis network is not. A multimodal mass-connectivity distribution indicates that the translation and metabolism networks are each distinct, adaptive biochemical modules, and that the gaps between the modes reflect evolved responses to the functional use of metabolite, polypeptide and polynucleotide compounds. The chemical reaction network of cellular translation opens new avenues for exploring complex adaptive phenomena such as percolation and phase changes in biochemical contexts.

Introduction

Cellular life is a network of molecular components that are dynamic and adaptive. Life can only exist as a gradient of free energy, and metabolic networks can account for the uptake and excretion of substrates alongside the synthesis of key biomolecules (Oberhardt et al., 2009). Metabolic networks trace the routes that energy travels through cells, but genetic sequences and their associated regulatory interactions (Chubukov et al., 2014, 2012; Lewis, 2011) specify which pathways will be constructed, governed, and maintained in any given cell. While a great part of current research is focused on filling specific knowledge gaps at the molecular level, the study of the overall cellular network might provide a holistic view of life's properties.

At the cellular level, analyses of metabolic networks have inspired decades of research into complexity (Barabási et al., 1999; Jeong et al., 2000; Newman, 2004; Ravasz et al., 2002). These analyses have drawn connections between network attributes such as a heavy-tailed connectivity distributions and general complex behaviors such as *resilience*, *adaptability, and modularity*. Following initial studies of metabolic networks, there have been numerous descriptions of heavy-tailed networks at other biological levels such as ecological interactions, including theories of generative mechanisms that are relevant to prebiotic world (Albert and Barabasi, 2001; Newman, 2003a; Nghe et al., 2015; Shenhav et al., 2005; Wagner and Fell, 2001). These generative mechanisms lead to continuous, heavy-tailed distributions consistent with a power law as an asymptotic sampling limit is approached; discontinuous or multimodal distributions are not explicitly prescribed by these physical models.

While the cellular level is usually the fundamental starting point to describe life properties, life manifests at different nested scales (e.g., cell, organism, ecosystem). At the largest biological level (e.g., populations and ecosystems), systems show emergent traits as *multimodality* or even *discontinuities* in their properties. One of the most studied properties showing discontinuities is size: depending on the ecosystem, organisms spanning some size ranges will not exist. Some examples were found in the arboreal forest (Holling, 1992), birds (Fischer et al., 2008), fishes and plankton (Havlicek and Carpenter, 2001), and mammal (Lambert et al., 2006) groupings can exhibit multimodal or discontinuous size distributions. Ecological multimodal aggregations arise from a confluence of dynamic processes consisting of an overlapping combination of both positive feedback (e.g., overabundant energy resources available for consumption at one size or temporal scale but not at others) and negative feedback (e.g., predation pressures such as grazing that can make survival at specific scales extremely challenging, or few and highly variable available resources) (Stow et al., 2007). Discontinuities and modalities in biology also appear to be idiosyncratic; gaps or troughs in one ecosystem do not necessarily map to those found in

others, and persistent gaps can nevertheless vary and shift over time (Nash et al., 2014). It is not clear whether these network-level attributes of biology ought to extend downward to the biomolecular scale.

Considering these two scale dependent aspects of biological networks, we asked whether these topological features associated with complex high-level biological systems can be found at the cellular level. We hypothesized that these features, if they show up within the cell system, require the definition of processes lying in an underlying hierarchy, as we find in natural ecosystems (e.g., *trophic chains*). Therefore, we address this question at the cellular level by building a translation reaction network in *Escherichia coli (E. coli)*, which we combine with its metabolic and biosynthetic networks (**Figure 1**). Translation is a molecular information decoding process that begins with a ribonucleotide sequence as input and a peptide sequence as output, with remarkable robustness, accuracy and adaptability (Irving et al., 2021; Koonin and Novozhilov, 2016; Novozhilov et al., 2007; Prossliner et al., 2021; Starosta et al., 2014; Steinchen et al., 2020; Zaher and Green, 2009). The decoding process that converts triplets of RNA monomers into the attachment of a specific amino acid residue into a growing peptide sequence is remarkable in that it is the earliest (and possibly only) emergent system that processes discretized information and manifests semantic relationships entirely in chemical terms. The chemosynthetic nature of the TM (i.e., information processing via chemical reactions, as compared to electronic signals in the brain or in digital computers) enables its description as a chemical reaction network, which is amenable to systems-level analyses as has been done for metabolic networks (Fuente et al., 2013; Jeong et al., 2000).

Here, we present the translation reaction network that includes the core processes of translation (initiation, elongation, termination), detailed assembly steps required to replicate the ribosome, and well-documented features for the stress response of the TM to maintain translation functionality (**Figure 2**). To connect the translation and metabolic networks, we constructed a polymer biosynthesis network that explicitly links enzymatic components catalyzing the metabolic reactions with the mRNA reading and peptide synthesis steps of translation (**Figure 3**). Comprehensive mapping of the connections between translation and metabolism is necessary to investigate context-dependent feedbacks between the cellular environment, substrate/nutrient availability, and ribosomal protein activity.

Methods

Generation of the translation and polymer biosynthesis reaction databases

The purpose of this study is to investigate translation as a chemosynthetic process and to integrate this process with a chemical reaction network describing metabolism. To serve this purpose, the level of description for translation includes a chemical accounting of ribosome

assembly and replication, as well as adjacent active site (A, P and E) interactions, since each of these interactions has downstream effects on the synthesis of biopolymers that enable autocatalytic replication of the entire system (**Figure 2**). Alternative translation steps (e.g., initiation on a leaderless mRNA) and stress responses that provide robustness to the canonical translation process were also included when such steps were described with sufficient data from primary literature.

A list of *E. coli* translation components was obtained from the EcoCyc database (Keseler et al., 2021) (GO: 006412). Supplementary File 1 summarizes the components of translation used in this network, apart from the ions and metabolites (e.g., GTPs) required to operate. The reactions for tRNA aminoacylation by tRNA synthetases, rRNA maturation and modifications, ribosomal protein modifications and maturations and peptidyl tRNA hydrolysis were obtained from the EcoCyc (Keseler et al., 2021) database and were also confirmed by literature synthesis (**Supplementary File 1**). The reactions for ribosomal subunit assembly, initiation, elongation and termination steps and stress response were built for *E. coli* (**Supplementary File 2**). The generation of translation reactions focused on five processes: a) elongation, which is largely conserved across all organisms; b) initiation, where some variations have been described between and within domains of life; c) termination; d) translation machinery biosynthesis, which are the assembly steps for replicating a ribosome; and e) the response to amino acid starvation. The reactions were manually annotated. To describe the elongation process in detail, the reaction model accounts for each of the three ribosome active sites. This includes a large number of reactions involving each of the distinct combinations of amino acids at each position. The metabolic and translation component masses were retrieved from the KEGG API. The translation machinery subnetwork is attached as **Supplementary File 3**.

The polymer biosynthesis subnetwork was constructed to link the metabolic and translation subnetworks together by writing non-stoichiometric reactions that approximate the assembly of amino acids into the assembled enzymatic polypeptides. Each reaction within the metabolic subnetwork associated with a specific enzyme resulted in a new reaction placed within the polymer biosynthesis subnetwork. For reactions that involved multiple protein components, the logical relationships of the protein components that enabled specific reactions were automatically read and parsed to generate protein-protein interaction maps within the molecular biosynthesis subnetwork.

Metabolic reaction database

The metabolic subnetwork was obtained from (Feist et al., 2007), who reconstructed *E. coli* metabolism. We built the network from its chemical reactions, and we linked them to their corresponding enzymes. In some cases, those enzymes require the formation of

heterocomplexes described through complex-formation reactions in the biosynthetic network (see **Figure 1**). We obtained the information about each of these enzymes through the Kyoto Encyclopedia of Genes and Genomes (KEGG). The metabolic subnetwork is attached as **Supplementary File 4** and the biosynthesis subnetwork, and the final network are attached as **Supplementary Files 5**

Chemical reaction network integration

All chemical reactions from each of the three networks (translation, polymer biosynthesis and metabolism) were uniformly processed into an aggregate network data file for graphing and analysis (**Figures 1 and 3A**). The aggregate network data file forms a bipartite graph with one class of nodes representing chemical compounds, and another class of nodes representing logical chemical relationships (i.e., chemosynthesis and/or biosynthesis assembly processes) that describe generating biomolecules from underlying substrates. The aggregate network is composed of 1357 reactions for translation, 1455 reactions for polymer biosynthesis and 2085 reactions for cellular metabolism, each with varying proportions of intra- and internetwork connectivity. The complete network is provided in **Supplementary Files 5 and 6, respectively.**

Network visualization, processing, and analyses

We define the connectivity or degree of a compound based on the number of reactions involving it (either as reactant or product). The degree probability distribution of each network was analyzed using the power-law Python library (Alstott et al., 2014) to determine the best statistical model that described the connectivity data. To measure how nodes are relevant to keep the network cohesion between modules, we used the betweenness centrality metric, as implemented in Gephi. This metric consists of measuring the number of times that a node appears in the shortest path between two different nodes (Newman 2001).

The KEGG database provided the molecular structures and sequences required to compute the molecular weight of each of the compounds (proteins, RNAs, metabolites, etc.). In the case of complexes, we add the mass of their components, but we disregard their stoichiometry of assembly due to a general lack of quantification of these steps. We generated the molecular-weight histograms using two strategies: i) simple counting of the number of occurrences in the network, and ii) by considering both the number of occurrences and the degree of a compound in the network (*degree-weighted histograms*). This last metric weighs the relevance of instances that appear few times but are highly connected within the network.

The network embeddings are generated using the node2vec algorithm (Grover and Leskovec, 2016) as implemented in PyTorch Geometric (Fey and Lenssen, 2019). We use standard parameters to generate the embeddings. The resulting 128 dimensions of each of these

embeddings were converted into 2D embeddings by using the t-distributed Stochastic Neighborhood Embedding (t-SNE) technique. Global network visualization of the aggregate network was carried out in Gephi 0.9.2 (Bastian et al., 2009). The code employed to perform this analysis together with the data are provided in the article GitHub repository https://github.com/kacarlab/Translation-Network.

Results

Translation is a process described in textbooks in great detail, but there are still many unknowns about its function and modes of integration with other cellular systems. By reviewing the literature, we compiled the reactions of translation and generated a bipartite graph represented in **Figure 2**. This network includes the core of translation processes (initiation, elongation and termination) together with the assembly of the TM, the recycling of ribosomes and the response to amino acid starvation. A feature of this network is the high level of redundancy of some of the reactions, given that translation is based on parallel processes involving different tRNAs.

All chemical reactions for translation and polymer biosynthesis were then incorporated with that for metabolism. The chemical object names in each network were screened for consistent annotation, their molecular masses and object classifications were cross-referenced or inferred (i.e., proteins, RNAs, codons, amino acids, energy molecules, and relevant ions). A full map of the integrated network, with node sizes weighted by node degree and reactions color-coded by network layer category is displayed in **Figure 3A**. The overall network contains more than 8000 nodes and 24000 links. When nodes are placed using a graph-layout algorithm (Force-Atlas in this case), the translation layer arises as a separated group of nodes different from metabolism and biosynthesis. This apparent isolation is explained by the low number of connections joining the translation layer with the metabolic layer (Figure 3B), which is limited to the molecules providing energy (e.g., GTP) and building blocks (e.g., tRNAs). Due to this sparsity, we could hypothesize that these molecules and the reactions providing those might have a central role keeping the overall network together. **Figure 4** highlights the connections that maintain the cohesion between the metabolic and translation layers. We have clustered reactions and compounds involved in the same module (e.g., cofactor synthesis, elongation), and only represented the compounds that join the different modules. We colored the network using the betweenness centrality of each node (**see Methods**). Such a metric represents how often we visit a node when traversing the network among its different domains. The network layout shows the separation between the translation and metabolisms layers, which maintain cellular network cohesion through modules such as the tRNA charging placed between the networks. Together with

transport modules that provide the bulk of metabolic precursors in the heterotrophic *E. coli* metabolism, tRNA charging is one of the most central modules of the cell.

An important attribute of large networks is their degree distribution: the probability of finding nodes with a given number of connections. In bipartite reaction networks, as the ones considered in this work, the degree distribution of each compound represents the number of reactions in which it participates. Networks associated with complex systems usually show heterogeneous degree distributions, where a few nodes show a remarkable number of connections, while random networks are characterized by having their distributions peeking around a mean number of connections (Albert and Barabasi, 2001). We use the log-ratios of different statistical fits to determine the more likely degree distribution. For objects in metabolism and translation layers, their connectivities follow statistically probable heavy-tailed distributions (**Figure 5A**, **Table 1**), which is not the case of the polymer biosynthesis layer degree distribution ($p>0.05$). The only object in the polymer biosynthesis network that is heavily connected to other compounds is a generic node 'peptide' (**Figure 3A**), which represents intermediate peptide sequences that for various reasons have not completed the termination step and thus have no distinct chemosynthetic or functional impact on the network. Network maps of metabolism display intracellular and extracellular reactions as distinct from one another; we have labeled extracellular reactions as Transporters (**Figure 3A**). This distinction has been noted in prior coarse visualizations of *E. coli* metabolism (Kim et al., 2007). The overall degree distribution of the full network follows a heavy-tailed degree distribution with an exponent of 1.690 (**Figure S1**).

Although the network represents a coarse-grained representation of the cell processes and components, we can assume a certain relation between the degree of a compound and its relevance in a cell. For instance, a molecule such as ATP might only show up once in the network, but its large number of connections make it more relevant than a secondary metabolite. We use the degree of each compound to reweight the mass distribution of cellular compounds in the network. A multimodal distribution can be discerned in the panels of **Figure 5B** (by compound type) and **Figure 5C** (by layer) and in the supplementary **Figures S2A** and **S2B** (which show the same data without weighting compounds by their degree). Mass distribution clusters (modes) across all network layers occur at around 240, 40,000 and 2,000,000 atomic mass units (amu). The central regions of these modes contain the most highly connected compounds from the different layers. There are bins in the troughs without reported molecules, which indicates that the distribution may be so clustered around modes that they may also be discontinuous. All modes contain some objects from the TM network; enzymatic polymers that enable metabolism are almost entirely confined to the middle peak; and metabolic compounds compose the two smallest modes. The overlap between compounds from different layers in each mode is most likely attributable to compounds of similar composition (small molecules, polypeptides, and

polynucleotides respectively) regardless of the functions they perform or the layer they are assigned to within this network. **Figure 5D** represents a degree-weighted histogram of the times that each molecule (protein, RNA, metabolite) shows up in a reaction, either as a free entity or associated with other molecules. The resulting distribution also shows strong multimodality, with visible peaks associated with the RNA that are not visible in **Figure 5B** because of their complexation with proteins.

To understand how their components are connected within and outside each network, we applied Node2Vec non-supervised learning to generate multi-dimensional representations (embeddings) as vectors for each node in the network that reflect its neighborhood of connectivity. We then reduced these multi-dimensional vectors to two dimensions using a t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction clustering technique. The resulting embeddings provide a representation of which nodes share a common neighborhood by placing them closer in an unitless 2D plane.

The translation and the metabolic subnetworks represent two different domains that can be easily distinguished (**Figure S3**), while the biosynthesis nodes appear scattered across these two different domains without any distinct clustering. Within the translation network, we find two large clusters of nodes and many smaller groups. Within these smaller groups there are nodes representing each of the amino acids processing cycles and the tRNA-synthetase process. Although the process of amino acid processing cycle was labeled as metabolic (given that it was present in the set of metabolic reactions), in actual cells it belongs to an overlap zone between translation and metabolism. In contrast, we do not find any cluster in the metabolic network containing distinct components of the TM.

Discussion

Bioinformatic databases are constructed and annotated in such a way as to enable the high-throughput reconstruction of many biosynthetic steps of metabolism and gene regulation (Bar-Joseph et al., 2003), and also to reconcile these steps with sequences found within an organism's genome in an organized pipeline (gene → protein → function; e.g., (Becker and Palsson, 2008). Prior network descriptions and analyses of translation have focused on its dynamic attributes afforded by its autocatalytic motifs (Roy et al., 2021), genetic regulation of metabolic processes (Grimbs et al., 2019), the links between TM function and organism-level cellular dynamics (Thiele et al., 2012) and a mapping of genotype to phenotype useful for interpreting a wealth of -omics cellular data (Thiele et al., 2009). The chemosynthetic function of the TM and the topological attributes of the overall network did not figure significantly into these prior efforts.

The integration of translation and metabolism reactions for *E. coli* describes a naturally occurring chemical network that is hierarchical and spans at least six orders of magnitude of compound mass. Hierarchy is most clearly resolved when network connectivity per compound is plotted against compound mass (**Figure S4)**. The energy costs required to generate individual molecular components of the cell can be approximated through molecular masses (Hu et al., 2020). By inference, the energy costs associated with synthesizing and replicating the largest TM components outstrip those of enzymes and metabolites by at least 1-2 orders of magnitude. The energy cost per molecular component demonstrates the hierarchical importance of the TM in specifying chemical functions made possible by biopolymers within the cell. In this way, network hierarchy correlates the degree to which high-mass components cannot be easily or obviously substituted for any combination of functions carried out by lower-mass components, as their functions are unique and indispensably specific to cellular operation.

As cellular operation requires a hierarchy of different mass entities, it also involves modules with different connectivity patterns. The embedding analysis depicted in **Figure S3** reveals fundamentally new information describing cellular chemosynthesis: the metabolic and polymer biosynthesis reactions do not cluster apart from each other in any significant way, but the TM and the extracellular transport reactions cluster in distinct regions of the graph. This would indicate that protein biosynthesis is mostly shaped by the same structural drivers and stimuli of the intracellular metabolism module and is not likely to exhibit independent adaptive behavior.

Continuous, multimodal or discontinuous distributions

Based on prior observations of biological networks across the hierarchy of life, we articulated three distinct predictions for the distribution of compounds in the integrated translation network (see **Figure S5**). In a parallel to 'trophic hierarchy', one hypothesized driver of hierarchical connection in cellular biochemistry is the ability to distribute energy and maximize power in a dynamic way across independently functioning layers. Each distinct molecular layer is self-organizing and both feeds upon and into the others. Though each layer contains distinct connectivity peaks, the overall mass distribution is continuous. In an 'energy availability' distribution, the hierarchical structures of polymer biosynthesis and translation are built directly atop, and extensions of, energy pathways enabled by metabolism. Though the translation and polymer biosynthesis layers of the network may contain 'heavy tails', the bulk of compounds in each layer all coincide with a peak generated by metabolic compounds. In a distribution shaped by 'molecular distinction', the association of compounds with a given layer is less significant than the functions that the compounds carry out in the cellular environment. Multimodality or discontinuity would be an unsurprising feature of such a system, as distinct molecular properties would lead to sorting at the network level. Each of these predicted distributions carry basic

implications about the primordial relationship(s) between emergent metabolic and translation functions, and about identifying measurable network attributes that fundamentally distinguish abiotic and biotic systems. We predicted that energy availability would be the dominant organizing driver of cellular biochemistry, given that it is indispensable to the overall function of translation.

The observed multimodal (and perhaps discontinuous) mass-connectivity distribution of the network seems most consistent with the 'distinct molecules' prediction. The network's multimodal mass-connectivity distribution is the first such distribution reported for biological systems at the cellular level. The multimodal mass distribution of the integrated chemical reaction network is independently supported by chemical assays of living *E. coli* cells. Non-targeted metabolomic assays of small molecular mass compounds using mass spectrometry (TOF-MS) and different preparatory analytic weighting techniques consistently show a peak in metabolite compounds at approximately 150-350 amu, and a tapering of compounds approaching ~2000 amu, even after accounting for decreased detector sensitivity with increasing mass (Dwivedi et al., 2010; Fuhrer et al., 2011). This approximately aligns with the centroid and distribution range of the smallest mass mode. For polypeptides, proteomic assays conducted across taxa as diverse as *E. coli, S. cerevisiae* and *H. sapiens* consistently show a broad correlation between translated genomic length and *in vivo* enzyme length (Geiger et al., 2012). For *E. coli*, this specifically includes a median peak length of about 209 amino acids, which corresponds to a peak in the mass distribution of polypeptides of approximately 28,000 amu if the average amino acid mass is assumed to be 136 amu (the average molar mass of all 20 essential amino acids). This approximately aligns with the centroid of the second mass mode. A whole-cell assay of all *E. coli* enzymes, regardless of composition and function, show that highly connected objects central to both metabolism and translation (e.g., the ribosome, TCA cycle components, glycolysis and fermentation components, etc.) compose comparable protein mass fractions under a variety of growing conditions (Mori et al., 2021). The ribosome itself is consistently composed of 2:1 mass fractions of RNA to protein (Kostinski and Reuveni, 2020). In sum, independent raw cell molecular assays indicate sharply discretized compound mass modes that are attributable to the chemical properties of RNA polymers, proteinaceous polypeptides, and base metabolites, regardless of their cellular functions. Molecules that fall within the troughs between the modes are undoubtedly present in cells in some form, but would not seem to play essential roles in maintaining an *E. coli*'s physiology, activating chemosynthetic pathways, or generating adaptive responses to stimuli. Multimodal distributions present a unique challenge to theoretical studies of network generation, and may require elaboration using multiplexes that can account for multimodality (Bianconi, 2013).

Causes and evolutionary implications of biochemical modularity

An assessment of the possible drivers of multimodality and discontinuity in the chemical reaction networks of bacteria may open new ways of characterizing the evolvability of cellular systems, and for distinguishing them from complex abiotic systems. The mass distribution modes trace at least two modularized systems (translation and metabolism, **Figures 3A, 4 and S4**) that, by virtue of their chemical isolation from one another, are each able to independently adapt to conditions that the cell may encounter (Hintze and Adami, 2008). We use the term 'module' here to denote a subset of the nodes in a network that are densely connected while being sparsely connected with nodes in other modules (Hintze and Adami, 2008; Srivastava et al., 2014). Modules can exhibit hierarchical properties and function as key units of adaptive evolution because organismal fitness depends on their performance (Gyorgy and Vecchio, 2014; Hartwell et al., 1999; Qi and Ge, 2006; Ravasz et al., 2002; Venkataram et al., 2020).

The *E. coli* metabolic network is only able to exist via the functions afforded by polymeric peptides. Accounting for these polymeric peptides in the chemical reaction network introduces the first modal trough (**Figure 5B**, right panel) between the largest metabolites and the smallest polypeptides. This trough is also present in objects that compose the translation network (**Figure 5C**, left panel). From a purely physical perspective, there are no *a priori* reasons for a discontinuity of mass size between these compounds. A discontinuity may be understandable, though, in biofunctional terms if every resources for synthesis are limited (Hu et al., 2020). The synthesis of an extremely large metabolite necessitates the precise placement of assorted atoms and molecular groups, connected with different bond types. If there is a single placement error, or inadequate feedstock of starting materials, the resulting compound is unlikely to be functional. An expenditure of chemical resources to produce increasingly larger (and increasingly error-prone) metabolites has adverse effects on the persistence of the overall cellular system. The synthesis of a linear oligopeptide of comparable mass, by contrast, can make use of repetitive interconnections of a much smaller number of object types (i.e., amino acids), many of which also have similar chemical properties. A reduction of the number of building block types and of the kinds of bonds through which they are connected can increase production fidelity (and resultant functionality) of peptidic complexes compared to massive metabolites. The polypeptide assembly process cannot scale downward indefinitely, however, since there are natural limits to how small a chemically functional polypeptide may be. Oligopeptides must attain some minimal length to form structural folds, or to form a pocket in which specialized cofactor chemistry can occur that is isolated from the cellular environment (Goldman and Kacar, 2021). The tradeoffs between synthesizing extremely large metabolites or extremely small polypeptides to ensure overall functionality can introduce a natural trough in mass distribution between these two

chemical regimes that is attributable to both the constraints brought about by the chemical properties of their different building blocks and of the emergent functions they enable.

The complex label in **Figures 5B and C** masks the role of the individual components of large cellular machineries. **Figure 5D** shows the decomposition of those complexes into their constituent types of molecules: RNA, proteins, metabolites and ions. The trough between the second (polypeptide) and third (polynucleotide) modes appears to arise for entirely different reasons that are distinct to the TM. As with the first mode gap, there are no obvious physical or chemical reasons why there should not be an overlap between large polypeptide and small ribonucleotide polymer compounds, and the observed trough is counter to our predictions. The ribosome, however, has evolved to localize many of the TM's protein-protein, protein-ribozyme, and ribosomal interactions on itself, thereby spatially isolating the functions of translation from the surrounding metabolic and physiological activities in the cell (Hartwell et al., 1999). These functions are further isolated in time and space by regulatory mechanisms, such as via binding with the properly formed initiation factors (Benelli and Londei, 2009; Milón and Rodnina, 2012; Sharma and Woodson, 2019; Simonetti et al., 2008), by corresponding activated tRNA complexes (Schuette et al., 2009) and through spatial heterogeneity arising from complex physiological relationships with nucleoids (Bakshi et al., 2015). Our translation network description includes a rich array of subunit interactions, which we record as interactions involving an object with the mass of the ribosome. This creates a sharp, isolated peak in the distribution centered at the ribosome's mass. The ready clustering of the TM in the network topology (**Figure 3A**) and embedded vector and dimensionality reduction analysis (**Figure S3**) depict and corroborate the biochemical facets of translation's modularity, as do *in vitro* experimental systems that employ cell-free protein synthesis (Hartman et al., 2007; Niederholtmeyer et al., 2013).

There are numerous examples of systems that may be characterized as complex self-organizing, adaptive phenomena, but which lack the evolvability, emergent novelty and agency of living systems. Far-from-equilibrium physical systems such as the formation of galactic and planetary structures (Pérez-Mercader, 2002), earthquakes (Gutenberg and Richter, 1944), solar flares (Krucker and Benz, 1998) and sandpile or avalanche dynamics (Bak et al., 1988) all exhibit continuous scaling across many orders of magnitude, but few or no discontinuities. Discrete size classes observed in biology, on the other hand, are evidence of multiple dynamic regimes (Garmestani et al., 2009) in which signaling, and the perturbative effects of novelty can propagate and be controlled in directions both up and down scales of a hierarchy (Holling and Gunderson, 2002).

A multimodal biochemical distribution may be interpreted as indicating an evolved response of modularization to significant constraints placed on a complex adaptive chemical reaction

network (Hintze and Adami, 2008). In biochemical systems, energy flows through the entire hierarchy mostly in the direction from metabolites up to the translation components, but modal gaps and troughs arise due to a complicated reconciliation process between the basic building blocks that compose large objects in the system, energy availability for synthesis and optimization of functional traits that large objects can exhibit (Shekhtman et al., 2015). In conjunction with a heavy-tailed connectivity distribution, a multimodal mass distribution may be one of a select suite of indicators diagnostic of biological dynamics that can be observed across the entirety of life's various spatiotemporal scales.

Implications for prebiotic chemistry

Broader relationships between translation and the rest of the cell may also encode unique, systems-level relationships about the conditions in which translation arose. As one example, it is possible that some information-processing attributes of the TM are primordially rooted in complex feedbacks between ribonucleotides and metabolites, or from intra-network dynamics of the TM, as much as from the emergence of the ribosome's large and small subunit mechanical activities.

Generating hierarchically connected networks across so many orders of magnitude, with modules that each contain at least one autocatalytic cycle (Braakman and Smith, 2013; Preiner et al., 2019; Roy et al., 2021; Zubarev et al., 2015) presents a significant challenge to the field of prebiotic chemistry (Orgel, 2006; Shapiro, 2006; Smith et al., 2014). If the biochemical modality and modularity observed in this network are genuinely universal attributes of life, they may have manifested prior to the evolution of the cell and arisen due to general physical phenomena such as percolation (Sornette, 2000). The observed coupling between translation, protein biosynthesis and metabolic networks presents a new model system in which the network phenomenon of percolation may be studied in a chemical context (Schwartz et al., 2002), e.g., as an example of an emergent system in which multiple networks with distinct topologies coexist and are connected via common elements (Leicht and D'Souza, 2009) and discontinuities stem from mechanisms that lead to network (dis)assortativity (Newman, 2003b; Newman and Girvan, 2004). Theoretical analyses of such conditions show a possible connection between the number of distinct modes and the possibility of multiple discontinuous phase transitions enabled by multiplex networks (Kryven and Bianconi, 2019). Phase transitions have long been recognized as potentially important behavioral features that are correlated with complex adaptive systems (Paperin et al., 2011).

The relationships between complex adaptive systems and information processing attributes afforded by network topology are directly relevant to developing new theories of the origin of

translation. The decoding process that links nucleotide sequence identity to peptide sequence functionality is highly optimized for error tolerance through a property known as code degeneracy, which can be described in informatic and computational terms without explicit reference to the underlying biochemistry. It is unclear how an inherently informatic, network-level property such as code degeneracy would arise on the basis of tracing the individual origins of the ribosome's chemical components and interaction partners in isolation from this wider network (Gonzalez et al., 2019). Recent discoveries have shown that 'biological' attributes such as self-reproduction, adaptation and evolvability (which together characterize much of biological agency) are intrinsically linked to one another through coupling of Hopf and Turing instabilities within a small network of reactions with a cubic autocatalytic motif (Muñuzuri and Pérez-Mercader, 2022). These attributes are observable even in the absence of the intricacies of macromolecular mechanisms or a diversified biochemical metabolic network. Further studies of coupled metabolic and translation chemical reaction network properties may help to develop new hypotheses about the origin of translation as an information-processing modular network, rather than about how translation can arise due to the chemical relationships between amino acids and nucleotides or through the specific biochemical activities of TM components. Specifically, one may test whether the modularization of information processing centered on the ribosome in modern cells can be decentralized across a primordial TM network.

Conclusions

Life exists, in part, by cohesively channeling energy from abundant, small-mass objects into the synthesis of a specific array of large-mass objects. These larger objects afford functions and activities that smaller objects, for various reasons, cannot perform. The inherent tension between utilizing available energy in an efficient way and ensuring diverse functions and activities will be reliably carried out can lead to multimodal mass distributions. Multimodality in this sense may be diagnostically biological, as an evolved trait that is exhibited by adaptive, self-organizing and self-assembling populations that can exhibit and propagate novel responses to stimuli and which are subject to natural selection. It is a property that has been observed in many different ecological systems. Our study of a combined network of metabolism and translation shows that mass multimodality can arise in biochemistry, with distinct adaptive drivers for different thoughs between modes, even before physiological or gene regulatory mechanisms of cellular operation are considered. This extends multimodality down to the lowest levels at which biological population dynamics can exist: the chemistry within individual bacterial cells. Translation stands apart in the cell as a functional module that is as much informatic as it is chemosynthetic. Its modularity is achieved in part by a possible driver for biochemical modality that concentrates many enzyme and ribozyme component interactions on the ribosomal complex, which has the effect of chemically isolating the translation function from much of the

chemical activity going on elsewhere in the cell. Looking forward, it may be possible to use an integrated metabolism and translation network to study the emergence of the translation function as a network-level module in a way that is abstracted from the mechanistic biochemical activities carried out by the ribosome and its interaction partners.

Acknowledgements

This work is supported by the Wisconsin Alumni Research Foundation, the John Templeton Foundation Grant #61926 and the Universidad Politécnica de Madrid (UPM) Margarita Salas Fellowship, founded by the “Unión Europea-NextGeneration EU“ (code UP2021-035).

Conflict of Interest

The authors declare no conflicts of interest

Figures and Tables

Table 1. A tabulated comparison of power law and exponential distributions (log-likelihood ratio), significance estimates (p-value) and best-fit exponent values (k) for the cumulative connectivity distributions of the translation, biosynthesis and metabolism networks.

Layer	log-likelihood ratio	p-value	k
Translation	207.44	<0.001	1.421
Polymer biosynthesis	24.10	0.36	2.279
Metabolism	1484.49	<0.001	1.747
All-layers	2508.31	<0.001	1.690

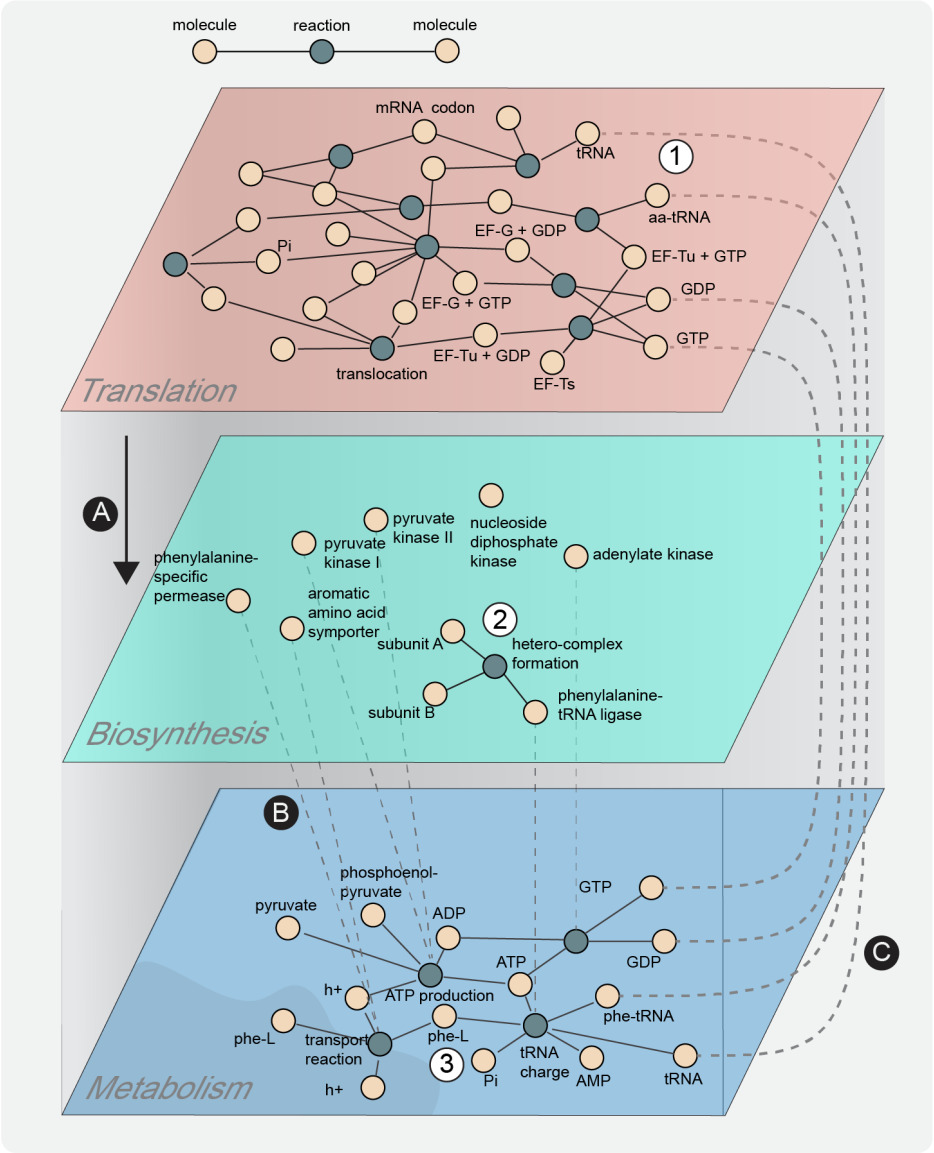

Figure 1. Schematic representation of the integrated cellular translation network. We built a translation network (1) that includes all the relevant steps (e.g. initiation, peptidic bond formation, translocation, termination). The translation network generates peptides that appear in a biosynthesis layer (A). While this layer shows many disconnected proteins that lack interactions with other proteins, some of them require the formation of hetero-multimers (B) to carry out their metabolic functions. Proteins in the biosynthesis layer act as enzymes or transporters enabling reactions (B) that generate energy and building blocks of translation such as ATP, GTP and charged tRNA from in a metabolism layer (3). Finally, the free energy and building blocks of metabolism will be employed as substrate inputs for translation (C). The integrated network spans the cellular chemosynthetic process, and does not explicitly include the genetic, regulatory or compartmentalization aspects of the cell.

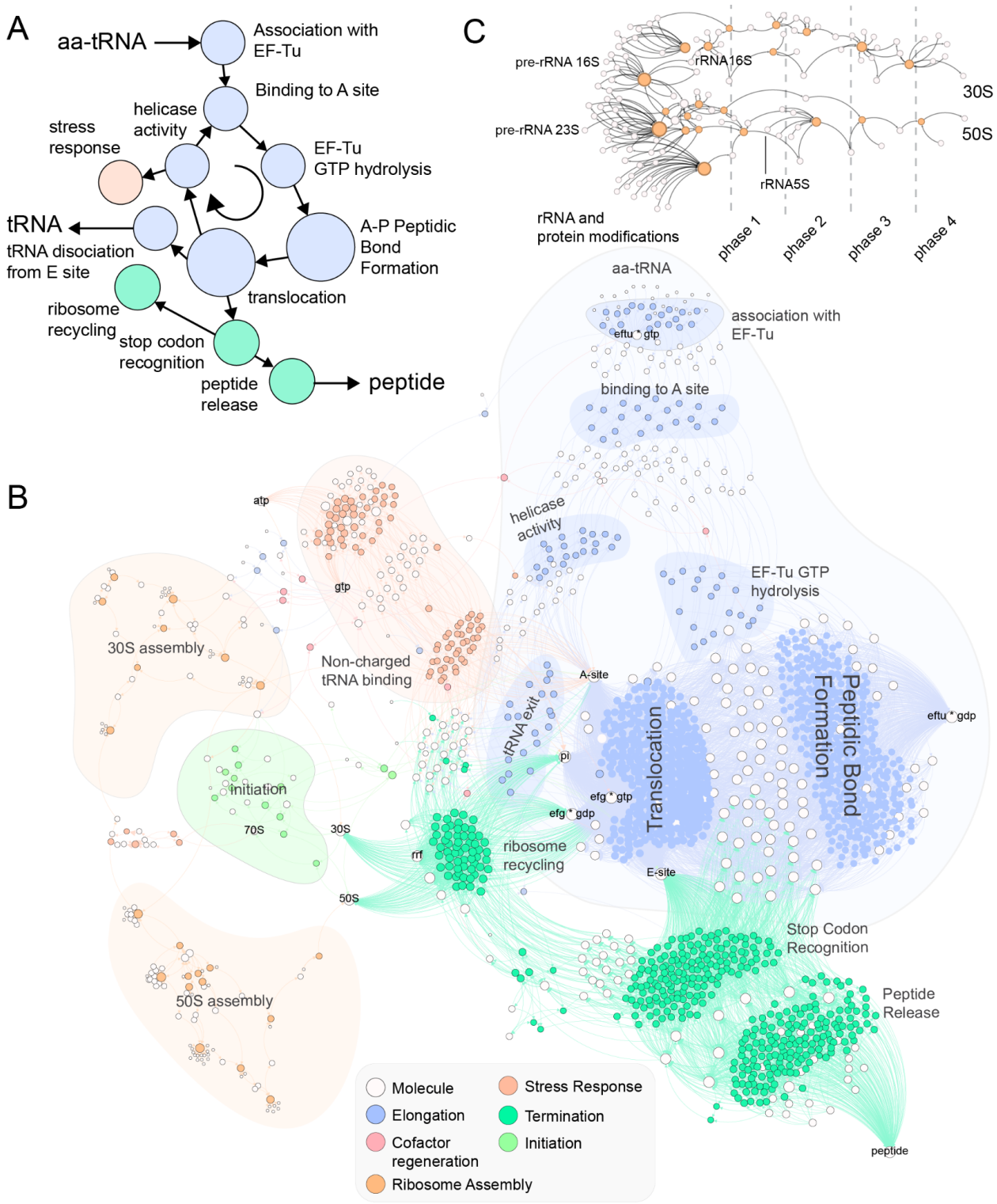

Figure 2. Overview of the translation portions of the cellular chemosynthesis reaction network database. **A.** Visualization of the complete translation chemical reaction network using Gephi

0.9.2. **B.** A depiction of the 'canonical' translation process encoded within the translation network.

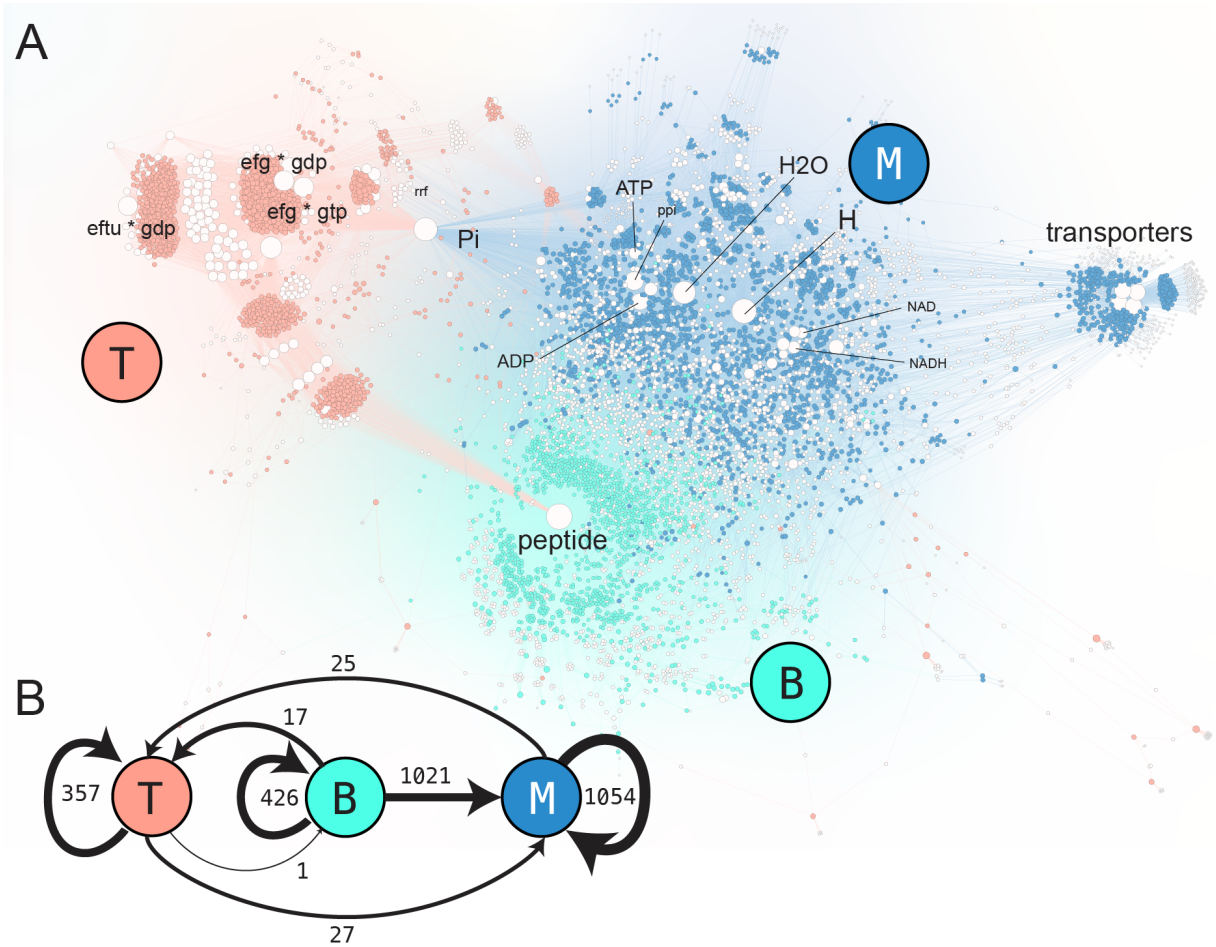

Figure 3. A. Full map with Translation (T), Polymer Biosynthesis (B) and Metabolic (M) networks, depicted with Gephi version 0.9.2 using a ForceAtlas2 algorithm**. B.** Visualization of the number of connections between the domains of the network.

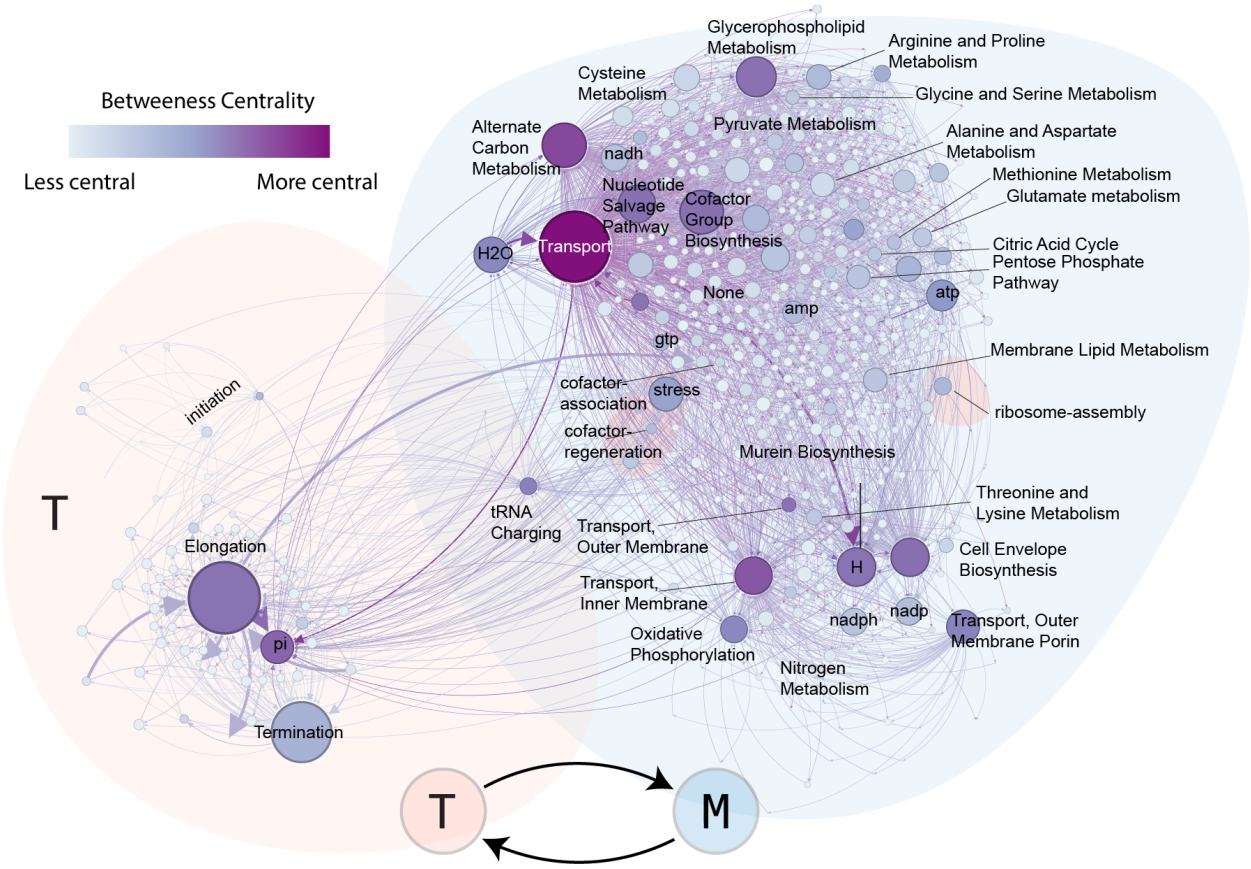

Figure 4. Reduced Network depicting the modules of translation and metabolism. The white-to-purple color scale represents the betweenness centrality of each node. The more visited is a node in the path between two random nodes, the higher its *betweenness-centrality*.

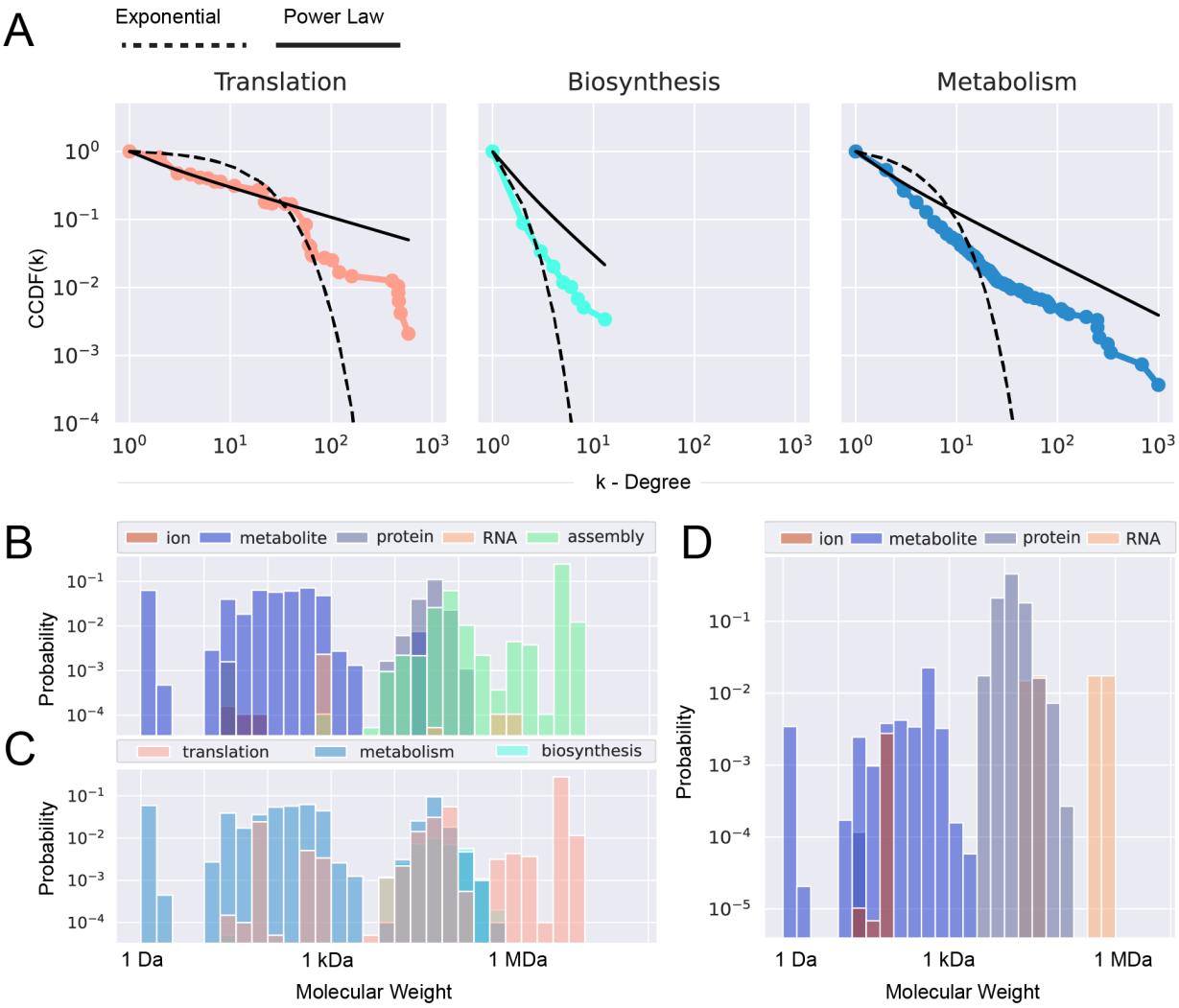

Figure 5. A. Cumulative degree distribution and distribution fit statistics for each network. Examples of the different fits (exponential and power law) tested using the power-law software tool. x-axis, node degree, y-axis, node degree *complementary cumulative density function* (CCDF). B. Degree-Weighted histogram of the molecular weights of the components of the full network, grouped by type. C. Degree-Weighted histogram of the molecular weights of the components of the full network, grouped by layer. D. Degree-Weighted histogram of the molecular weights of the molecules that constitute the complexes of the full-network, grouped by type.

Supplementary Figures

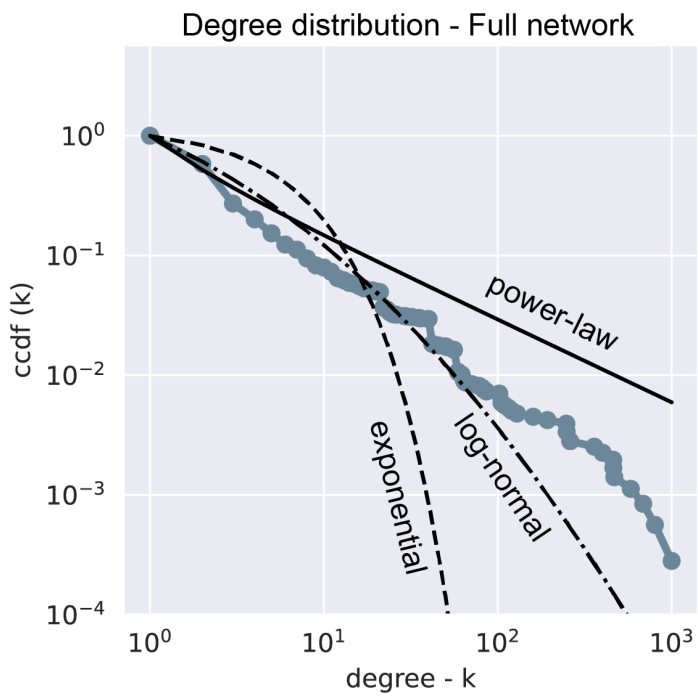

Figure S1: Complementary Cumulative Density Function of the observed degree distribution for the full network. Continuous line represents a power-law fit and discontinuous line represents an exponential fit.

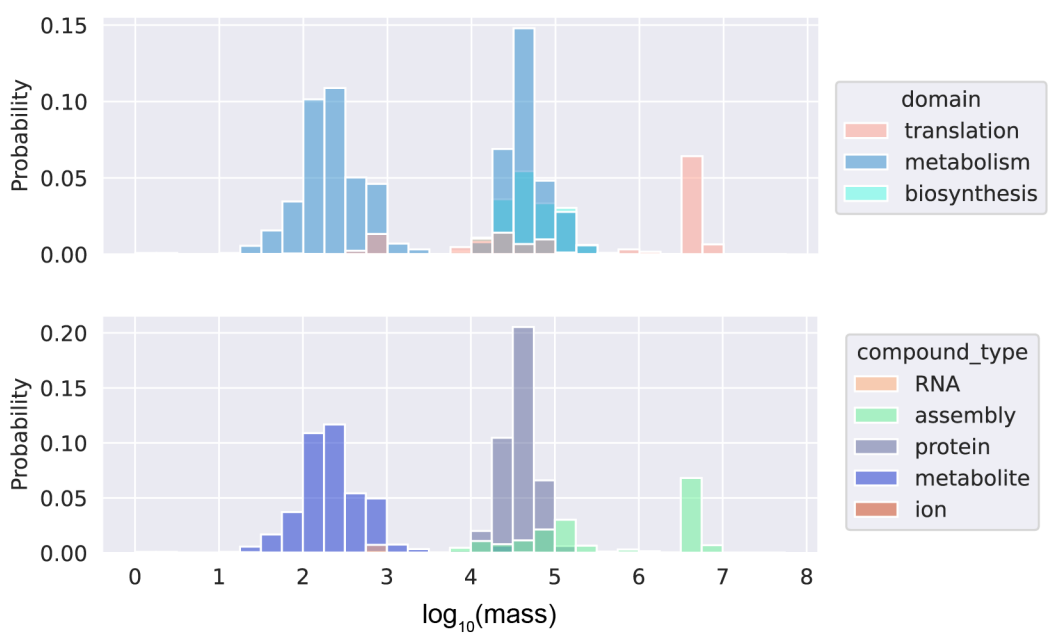

Figure S2: A. Histogram of the molecular weights of the components of the full network, grouped by type. B. Histogram of the molecular weights of the components of the full network, grouped by layer.

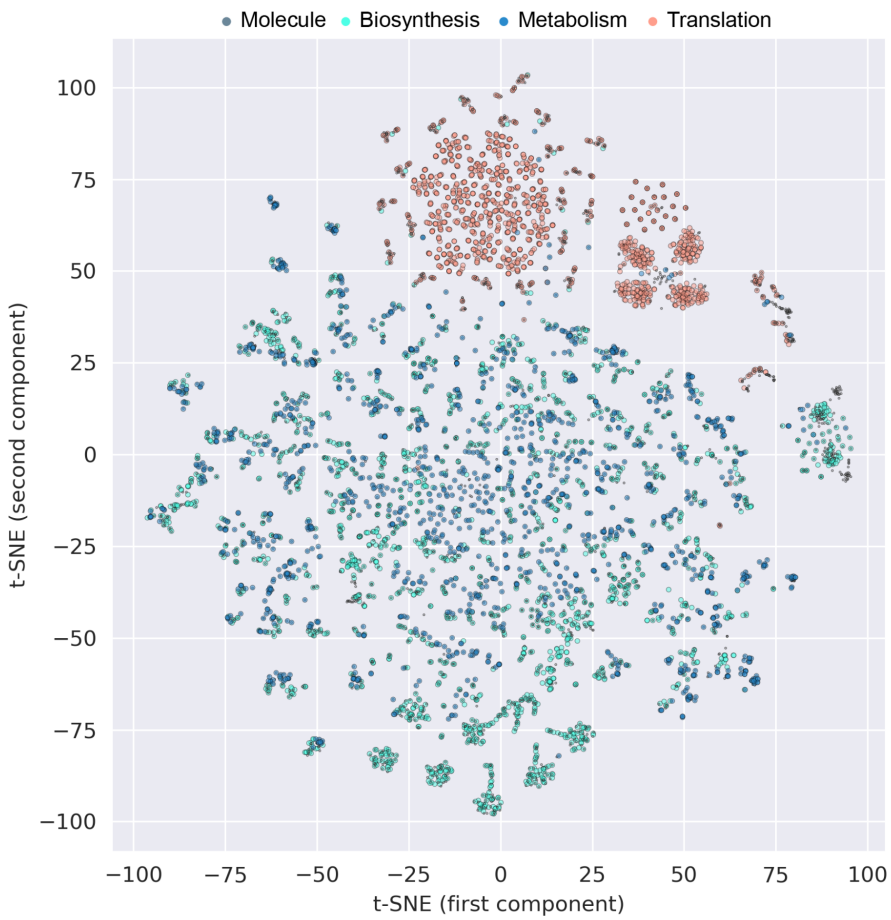

Figure S3: T-SNE clustering of node2vec embeddings of the full network. This representation allows to distinguish the hub structures of intracellular metabolism (upper grouping of blue), extracellular transport (upper right clusters of four in blue) and translation (lower left grouping of orange). Colors: dark blue, Metabolism; light blue, Biosynthesis; orange, Translation; gray, compounds.

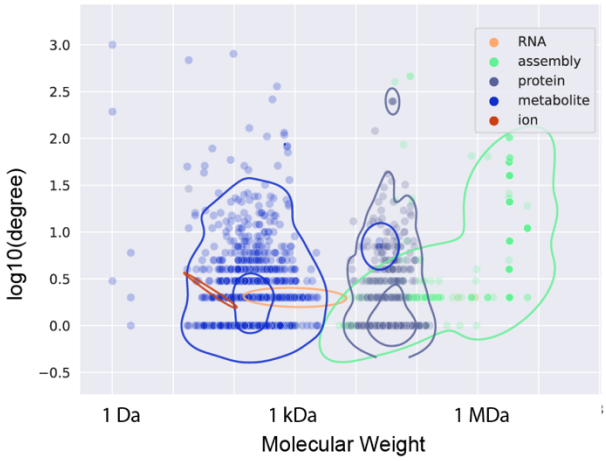

Figure S4: Raw (unbinned) scatter plot of connectivity degree versus molecular weight for each component of the full network.

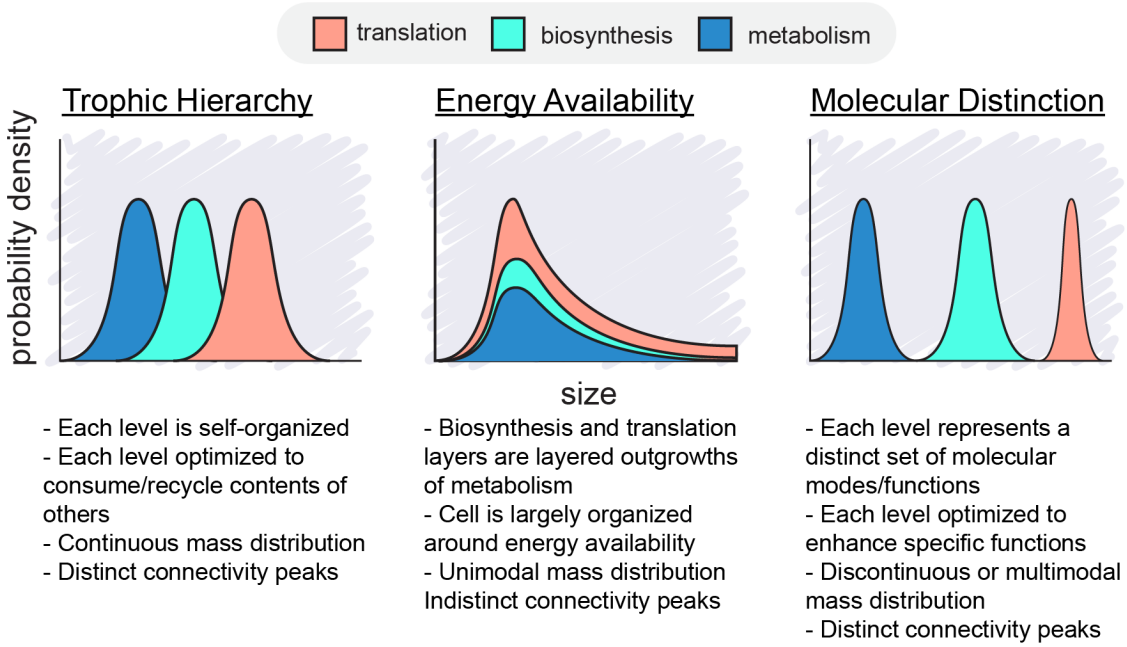

Figure S5: Explanation of the different modes of the size distribution.

References

Albert R, Barabasi A-L. 2001. Statistical mechanics of complex networks. *Arxiv*. doi:10.1103/revmodphys.74.47

Alstott J, Bullmore E, Plenz D. 2014. powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *Plos One* **9**:e**8**5**7**77. doi:10.1371/journal.pone.0085777

Bak P, Tang C, Wiesenfeld K. 1988. Self-organized criticality. *Phys Rev A* **38**:364–374. doi:10.1103/physreva.38.364

Bakshi S, Choi H, Weisshaar JC. **2**015. The spatial biology of transcription and translation in rapidly growing Escherichia coli. *Front Microbiol* **6**:636. doi:10.3389/fmicb.2015.00636

Barabási A-L, Albert R, Jeong H. 1999. Mean-field theory for scale-free random networks. *Phys Statistical Mech Appl* **272**:173–187. doi:10.1016/s0378-4371(99)00291-5

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK. 2003. Computational discovery of gene modules and regulatory networks. *Nat Biotechnol* **21**:1337–1342. doi:10.1038/nbt890

Bastian M, Heymann S, Jacomy M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc Int Aaai Conf Web Soc Media* **3**:361–362. doi:10.1609/icwsm.v3i1.13937

Becker SA, Palsson BO. 2008. Context-Specific Metabolic Networks Are Consistent with Experiments. *Plos Comput Biol* **4**:e**10**0008**2**. doi:10.1371/journal.pcbi.1000082

Benelli D, Londei P. 2009. Begin at the beginning: evolution of translational initiation. *Res Microbiol* **160**:493–**5**01. doi:10.1016/j.resmic.2009.06.003

Bianconi G. 2013. Statistical mechanics of multiplex networks: Entropy and overlap. *Phys Rev E* **87**:062806. doi:10.**11**03/physreve.87.062806

Braakman R, Smith E. 2013. The compositional and evolutionary logic of metabolism. *Phys Biol* **10**:01**100**1. doi:10.1088/1478-3975/10/1/011001

Chubukov V, Gerosa L, Kochanowski K, Sauer U. 2014. Coordination of microbial metabolism. *Nat Rev Microbiol* **12**:327–340. doi:10.1038/nrmicro3238

Chubukov V, Zuleta IA, Li H. 2012. Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc National Acad Sci* **109**:5127–5132. doi:10.1073/pnas.1114235109

Dwivedi P, Puzon G, Tam M, Langlais D, Jackson S, Kaplan K, Siems WF, Schultz AJ, Xun L, Woods A, Hill HH. 2010. Metabolic profiling of Escherichia coli by ion mobility-mass spectrometry with MALDI ion source. *J Mass Spectrom* **45**:1383–1393. doi:10.1002/jms.1850

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**:121.

doi:10.1038/msb4100**15**5
Fey M, Lenssen JE. 20**19**. Fast Graph Representation Learning with PyTorch Geometric. Arxiv. doi:10.48550/arxiv.1903.02428
Fischer J, Lindenmayer DB, Montague-Drake R. 2008. The role of landscape texture in conservation biogeography: a case study on birds in south-eastern Australia. *Divers Distrib* **14**:38–46. doi:10.1111/j.1472-4642.2007.00411.x
Fuente IMD la, Cortes JM, Pelta DA, Veguillas J. 2013. Attractor Metabolic Networks. Plos One 8:e58284. doi:10.1371/journal.pone.0058284
Fuhrer T, Heer D, Begemann B, Zamboni N. 2011. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal Chem* **83**:**70**74–7080. doi:10.1021/ac2012**67**k
Garmestani A, Allen C, Gunderson L. 2009. Panarchy: Discontinuities Reveal Similarities in the Dynamic System Structure of Ecological and Social Systems. *Ecol Soc* **14**. doi:10.5751/es-02744-140115
Geiger T, Wehner A, Schaab C, Cox J, Mann M. 20**12**. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins*. *Mol Cell Proteomics* **11**:M111.014050. doi:10.1074/mcp.m111.014050
Goldman AD, Kacar B. 2021. Cofactors are Remnants of Life's Origin and Early Evolution. *J Mol Evol* **89**:127–133. doi:10.1007/s00239-020-09988-4
Gonzalez DL, Giannerini S, Rosa R. 2019. On the origin of degeneracy in the genetic code. *Interface Focus* **9**:20190038. doi:10.1098/rsfs.2019.0038
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:e100**69**62. doi:10.1371/journal.pcbi.1006962
Grover A, Leskovec J. 2016. node2vec: Scalable Feature Learning for Networks. Arxiv.
Gutenberg B, Richter CF. 1944. Frequency of earthquakes in California*. *B Seismol Soc Am* **34**:185–188. doi:10.**17**85/bssa0340040185
Gyorgy A, Vecchio DD. 2014. Modular Composition of Gene Transcription Networks. Plos Comput Biol 10:e1003486. doi:10.1371/journal.pcbi.1003486
Hartman MCT, Josephson K, Lin C-W, Szostak JW. 2007. An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. Plos One 2:e972. doi:10.1371/journal.pone.0000972
Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999. From molecular to modular cell biology. *Nature* **402**:C47–C52. doi:10.1038/35011540
Havlicek TD, Carpenter SR. 2001. Pelagic species size distributions in lakes: Are they discontinuous? *Limnol Oceanogr* **46**:1021–1033. doi:10.4319/lo.2001.46.5.1021
Hintze A, Adami C. 2008. Evolution of Complex Modular Biological Networks. Plos Comput Biol 4:e23. doi:10.1371/journal.pcbi.0040023

Holling CS. 1992. Cross-Scale Morphology, Geometry, and Dynamics of Ecosystems. *Ecol Monogr* **62**:447–502. doi:10.2307/2937313
Holling CS, Gunderson LH. 2002. Resilience and adaptive cycles. *In: Panarchy: Understanding Transformations in Human and Natural Systems, 25-62*.
Hu X-P, Dourado H, Schubert P, Lercher MJ. 2020. The protein translation machinery is expressed for maximal efficiency in Escherichia coli. *Nat Commun* **11**:5260. doi:10.1038/s41467-020-18948-x
Irving SE, Choudhury NR, Corrigan RM. 2021. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol 19:256–271. doi:10.1038/s41579-020-00470-y
Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L. 2000. The large-scale organization of metabolic networks. Nature **407**:651–654. doi:10.1038/3503**66**27
Keseler IM, Gama-Castro S, Mackie A, Billington R, Bonavides-Martínez C, Caspi R, Kothari A, Krummenacker M, Midford PE, Muñiz-Rascado L, Ong WK, Paley S, Santos-Zavaleta A, Subhraveti P, Tierrafría VH, Wolfe AJ, Collado-Vides J, Paulsen IT, Karp PD. 2021. The EcoCyc Database in 2021. Front Microbiol 12:7**110**77. doi:10.3389/fmicb.2021.711077
Kim P-J, Lee D-Y, Kim TY, Lee KH, Jeong H, Lee SY, Park S. 2007. Metabolite essentiality elucidates robustness of Escherichia coli metabolism. Proc National Acad Sci **104**:13638–13642. doi:10.1073/pnas.0703262104
Koonin EV, Novozhilov AS. 2016. Origin and Evolution of the Universal Genetic Code. *Annu Rev Genet* **51**:1–18. doi:10.1146/annurev-genet-120116-024713
Kostinski S, Reuveni S. 2020. Ribosome Composition Maximizes Cellular Growth Rates in E. coli. *Phys Rev Lett* **125**:028103. doi:10.1103/physrevlett.125.028103
Krucker S, Benz AO. 1998. Energy Distribution of Heating Processes in the Quiet Solar Corona. *Astrophysical J Lett* **501**:L213–L216. doi:10.1086/311474
Kryven I, Bianconi G. 2019. Enhancing the robustness of a multiplex network leads to multiple discontinuous percolation transitions. Phys Rev E 100:020301. doi:10.1103/physreve.100.020301
Lambert TD, Malcolm JR, Zimmerman BL. 2006. Amazonian small mammal abundances in relation to habitat structure and resource abundance. *J Mammal* **87**:766–776. doi:10.1644/05-mamm-a-261r1.1
Leicht EA, D'Souza RM. 2009. Percolation on interacting networks. Arxiv. doi:10.48550/arxiv.0907.0894
Lewis M. 2011. A Tale of Two Repressors. *J Mol Biol* **409**:14–27. doi:10.1016/j.jmb.2011.02.023
Milón P, Rodnina MV. 2012. Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol* **47**:334–348. doi:10.3109/10409238.2012.678284
Mori M, Zhang Z, Banaei-Esfahani A, Lalanne J, Okano H, Collins BC, Schmidt A, Schubert OT, Lee D, Li G, Aebersold R, Hwa T, Ludwig C. 2021. From coarse to fine: the absolute Escherichia coli proteome under diverse growth conditions. Mol Syst Biol 17:e9536.

doi:10.15252/msb.20209536
Muñuzuri AP, Pérez-Mercader J. 2022. Unified representation of Life's basic properties by a 3-species Stochastic Cubic Autocatalytic Reaction-Diffusion system of equations. *Phys Life Rev* **41**:64–83. doi:10.1016/j.plrev.2022.03.003
Nash KL, Allen CR, Angeler DG, Barichievy C, Eason T, Garmestani AS, Graham NAJ, Granholm D, Knutson M, Nelson RJ, Nyström M, Stow CA, Sundstrom SM. 2014. Discontinuities, cross-scale patterns, and the organization of ecosystems. *Ecology* **95**:654–667. doi:10.1890/13-1315.1
Newman MEJ. 2004. Analysis of weighted networks. Phys Rev E 70:056131. doi:10.1103/physreve.70.056131
Newman MEJ. 2003a. The Structure and Function of Complex Networks. *Siam Rev* **45**:167–256. doi:10.1137/s003614450342480
Newman MEJ. 2003b. Mixing patterns in networks. Phys Rev E 67:026126. doi:10.1103/physreve.67.026126
Newman MEJ, Girvan M. 2004. Finding and evaluating community structure in networks. Phys Rev E 69:026113. doi:10.1103/physreve.69.026113
Nghe P, Hordijk W, Kauffman SA, Walker SI, Schmidt FJ, Kemble H, Yeates JAM, Lehman N. 2015. Prebiotic network evolution: six key parameters. *Mol Biosyst* **11**:3206–3217. doi:10.1039/c5mb00593k
Niederholtmeyer H, Stepanova V, Maerkl SJ. 2013. Implementation of cell-free biological networks at steady state. Proc National Acad Sci 110:15985–15990. doi:10.1073/pnas.1311166110
Novozhilov AS, Wolf YI, Koonin EV. 2007. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biol Direct* **2**:24. doi:10.1186/1745-6150-2-24
Oberhardt MA, Palsson BØ, Papin JA. 2009. Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320–320. doi:10.1038/msb.2009.77
Orgel L. 2006. In the beginning. Nature **439**:915–915. doi:10.1038/439915a
Paperin G, Green DG, Sadedin S. 2011. Dual-phase evolution in complex adaptive systems. *J Roy Soc Interface* **8**:609–629. doi:10.1098/rsif.2010.0719
Pérez-Mercader J. 2002. Astrobiology, The Quest for the Conditions of Life 337–360. doi:10.1007/978-3-642-59381-9_22
Preiner M, Xavier JC, Vieira A do N, Kleinermanns K, Allen JF, Martin WF. 2019. Catalysts, autocatalysis and the origin of metabolism. Interface Focus 9:20190072. doi:10.1098/rsfs.2019.0072
Prossliner T, Gerdes K, Sørensen MA, Winther KS. 2021. Hibernation factors directly block ribonucleases from entering the ribosome in response to starvation. *Nucleic Acids Res* **49**:gkab017-. doi:10.1093/nar/gkab017
Qi Y, Ge H. 2006. Modularity and Dynamics of Cellular Networks. Plos Comput Biol 2:e174.

doi:10.1371/journal.pcbi.0020174
Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. 2002. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **297**:1551–1555. doi:10.1126/science.1073374
Roy A, Goberman D, Pugatch R. 2021. A unifying autocatalytic network-based framework for bacterial growth laws. *P Natl Acad Sci Usa* **118**:e2107829118. doi:10.1073/pnas.2107829118
Schuette J, Murphy FV, Kelley AC, Weir JR, Giesebrecht J, Connell SR, Loerke J, Mielke T, Zhang W, Penczek PA, Ramakrishnan V, Spahn CMT. 2009. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *Embo J* **28**:755–765. doi:10.1038/emboj.2009.26
Schwartz N, Cohen R, ben-Avraham D, Barabási A-L, Havlin S. 2002. Percolation in directed scale-free networks. Phys Rev E 66:015104. doi:10.1103/physreve.66.015104
Shapiro R. 2006. Small Molecule Interactions were Central to the Origin of Life. *Q Rev Biology* **81**:105–126. doi:10.1086/506024
Sharma IM, Woodson SA. 2019. RbfA and IF3 couple ribosome biogenesis and translation initiation to increase stress tolerance. Nucleic Acids Res **48**:359–372. doi:10.1093/nar/gkz1065
Shekhtman LM, Shai S, Havlin S. 2015. Resilience of networks formed of interdependent modular networks. *New J Phys* **17**:123007. doi:10.1088/1367-2630/17/12/123007
Shenhav B, Solomon A, Lancet D, Kafri R. 2005. Transactions on Computational Systems Biology ILecture Notes in Computer Science. pp. 14–27. doi:10.1007/978-3-540-32126-2_2
Simonetti A, Marzi S, Jenner L, Myasnikov A, Romby P, Yusupova G, Klaholz BP, Yusupov M. 2008. A structural view of translation initiation in bacteria. *Cell Mol Life Sci* **66**:423. doi:10.1007/s00018-008-8416-4
Smith JI, Steel M, Hordijk W. 2014. Autocatalytic sets in a partitioned biochemical network. *J Syst Chem* **5**:2. doi:10.1186/1759-2208-5-2
Sornette D. 2000. Critical Phenomena in Natural Sciences, Chaos, Fractals, Self Organization and Disorder: Concepts and Tools. *Springer Series Syne* 239–256. doi:10.1007/978-3-662-04174-1_12
Srivastava A, Kumar S, Ramaswamy R. 2014. Two-layer modular analysis of gene and protein networks in breast cancer. *Bmc Syst Biol* **8**:81. doi:10.1186/1752-0509-8-81
Starosta AL, Lassak J, Jung K, Wilson DN. 2014. The bacterial translation stress response. *Fems Microbiol Rev* **38**:**117**2–1201. doi:10.1111/1574-6976.12083
Steinchen W, Zegarra V, Bange G. 2020. (p)ppGpp: Magic Modulators of Bacterial Physiology and Metabolism. Front Microbiol 11:2072. doi:10.3389/fmicb.2020.02072
Stow C, Allen C, Garmestani A. 2007. Evaluating Discontinuities in Complex Systems: Toward Quantitative Measures of Resilience. Ecol Soc 12. doi:10.5751/es-02055-120126
Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. 2012. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. Plos One 7:e45635. doi:10.1371/journal.pone.0045635
Thiele I, Jamshidi N, Fleming RMT, Palsson BØ. 2009. Genome-Scale Reconstruction of

Escherichia coli's Transcriptional and Translational Machinery: A Knowledge Base, Its Mathematical Formulation, and Its Functional Characterization. Plos Comput Biol 5:e1000312. doi:10.1371/journal.pcbi.1000312
Venkataram S, Monasky R, Sikaroodi SH, Kryazhimskiy S, Kacar B. 2020. Evolutionary stalling and a limit on the power of natural selection to improve a cellular module. Proc National Acad Sci 117:18582–18590. doi:10.1073/pnas.1921881117
Wagner A, Fell DA. 2001. The small world inside large metabolic networks. *Proc Royal Soc Lond Ser B Biological Sci* **268**:1803–1810. doi:10.1098/rspb.2001.1711
Zaher HS, Green R. 2009. Quality control by the ribosome following peptide bond formation. Nature **457**:161–166. doi:10.1038/nature07582
Zubarev DY, Rappoport D, Aspuru-Guzik A. 2015. Uncertainty of Prebiotic Scenarios: The Case of the Non-Enzymatic Reverse Tricarboxylic Acid Cycle. *Sci Rep-uk* **5**:8009. doi:10.1038/srep08009