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 53706
[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 53706
[d] Department of Geosciences, University of Wisconsin-Madison, Madison, WI, USA 53706

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

Abstract

Translation is a foundational 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. However, the underlying 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. Contrary to our predictions of a continuous mass distribution, 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.

Keywords: Translation, metabolism, modularity, complex systems, adaptability

Significance Statement

One challenge in tracing the origin of life is understanding the emergence and coupling of information processing and metabolism, which often involves debating which function preceded the other. These functions can be articulated as a unified network of interacting objects that include RNAs, proteins and metabolites. We mapped all translation factors, their interactions and substrates as biochemical reactions and connected them to their corresponding metabolic pathways. The resulting network exhibits a multimodal mass distribution, and the metabolic and translation layers each likely have a heavy-tailed connectivity distribution, indicating modular coupling between the two. At the ecological level, this combination of traits is associated with both stability and adaptability. This work establishes a basis for exploring properties observed throughout biological organization that may be requisite features of prebiotic chemical networks.

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 (1). Metabolic networks trace the routes that energy travels through cells, but genetic sequences and their associated regulatory interactions (2–5) specify which pathways will be constructed, governed, and maintained in any given cell.

At the nexus between the metabolic and genetic, there is translation. Translation is a chemical information decoding process that begins with an input string of ribonucleotides and ends with a specific output string of peptides. The decoding process between polynucleotides and polypeptides is robust (6, 7), with an extraordinary capability for error-correction (8) and stress response (9–12). 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, network-level system that processes discretized information and manifests semantic relationships entirely in chemical terms. Reconstructing the chemical origins of translation is, in some respects, akin to recapitulating the very origins of living systems.

The mechanical process of translation relies heavily upon a unique RNA polymer called the ribosome, but the ribosome can only function following its precise assembly and interaction with a coterie of supporting polypeptide enzymes and polynucleotide ribozymes. Translation is

therefore a network-level phenomenon that coexists with the cell's metabolic network, carrying out informatic processes via chemosynthetic reactions. The network of reactions that makes translation possible may be collectively referred to as the translation machinery (TM). The evolution of the TM represents a fundamental innovation in the origin of life, and explaining its emergence is concomitantly a major challenge in the field of prebiotic chemistry (13).

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 analyses as has been done for metabolic networks (14, 15). Translation is commonly described as hierarchically related to metabolism, but hierarchies can exist in different arrangements and topologies, each with potentially unique physical or evolutionary implications. It is possible that the 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. Mapping the connections between translation and metabolism is necessary to investigate context-dependent feedbacks between the cellular environment, substrate/nutrient availability, and ribosomal protein activity (16, 17). Here we generate a comprehensive biochemical reaction database for translation in *Escherichia coli (E. coli)* to assess whether topological features associated with complex adaptive systems are detectable in the resulting graph (**Figure 1**). The network includes 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, a polymer biosynthesis network was constructed; this network explicitly links enzymatic components catalyzing the metabolic reactions with the mRNA reading and peptide synthesis steps of translation (**Figure 3**).

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/leaderless translation steps 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 (18) (GO: 006412). The list of translation term members included ribosomal proteins (22 SSU proteins, 33 LSU proteins), rRNA operons (7 for 5S rRNA operon, 6 for 16S rRNA operon, 6 for 23S rRNA operon), 3 initiator factors, 6 elongation factors (including their copy genes), 4 release factors, 8 ribosome maturation factors, 2 peptide hydrolases and deformylases, and 8 putative translation proteins. In addition to the enzymes, 21 tRNAs (including initiator tRNA), mRNA codons, 20 amino acids, energy molecules (ATP and GTP), phosphate, acetyl CoA and CoA, water, and ions (Mg^{+2}, Zn^{+2}, Mn^{+2}, 4Fe-4S) were added to the list. 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 (18) 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) stress response. 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 calculated using the information provided by 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 modified from Feist et al. (19) (**Supplementary File 4**). Its reactions were expanded to explicitly include all referenced enzymes that make each of the metabolic reactions possible. Such enzymes were provided as KEGG gene identifiers, used to link reactions with the properties of their enzymes through the KEGG API. The metabolic subnetwork is attached as **Supplementary File 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 1A, 1B 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 1433 reactions for translation, 1455 reactions for polymer biosynthesis and 2085 reactions for cellular metabolism, each with varying proportions of intra- and internetwork connectivity. The number of objects in each level can be reduced to a simplified depiction of number of objects interacting within and between the different levels (**Figure 1C**). The level of description for each network includes an accounting of key objects and their interactions required for precise biosynthesis of macromolecules (**Figure 1D**). The biosynthesis subnetwork and the final network are attached as **Supplementary Files 6 and 7,** respectively**.**

Network visualization, processing and analyses

Global network analysis and visualization of the aggregate network was carried out in Gephi 0.9.2 (20). The cumulative probability distribution profile of the connectivities of each of the chemical compounds in the data file were generated and analyzed using the power-law Python library (21) to determine the best statistical model that described the connectivity data. The node2vec algorithm (22), as implemented in PyTorch-Geometric (23) was used to generate the aggregate-network node embeddings. The dimensions of the vectors were reduced into a 2-dimensional space by applying the t-Distributed Stochastic Neighbor Embedding (t-SNE) method. Jupyter notebooks detailing the network processing and analysis are provided at Zenodo: 10.5281/zenodo.7502221.

Results

All chemical reactions for translation and polymer biosynthesis were used to construct bipartite graphs, which were then incorporated with that for metabolism (**Figure 1**)**.** 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), and an integrated bipartite graph was generated to depict and analyze all interactions between objects (**Figure 1**).

A full map of the integrated chemosynthesis network, with nodes weighted by object mass and reactions color-coded by subnetwork category is displayed in **Figure 3A**. For objects in

metabolism and translation, their connectivities form statistically probable heavy-tailed distributions (**Figure 3B**, **Table 1**). The polymer biosynthesis connectivity distribution p-value is greater than 0.05, which indicates that it is statistically indistinguishable between exponentially truncated or heavy-tailed possibilities (**Table 1**). 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 (24).

A multimodal mass distribution of cellular compounds in the network can be discerned in the panels of **Figure 3C**. Mass distribution clusters (modes) across all network layers occur at around 200, 30,000 and 2,000,000 atomic mass units (amu). The central regions of these modes contain the most highly connected compounds from the different layers. 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.

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. t-SNE is a visualization and clustering analysis method that can be used to reduce a complex set of relationships into a low-dimensional representation of adjacent objects. The two-dimensional axes are mathematical constructs; they do not necessarily correspond to physical or chemical characteristics of the chemical reaction network nodes. These techniques prioritize grouping by neighborhood connectivity over functional role, which is a known trait of Node2Vec embeddings. The resulting embeddings allow retrieval of community structures within a network.

In the Node2Vec mapping, the translation and the metabolic subnetworks represent two different domains that can be easily distinguished (**Figure 4**), while the biosynthesis nodes appear scattered across these two different domains without any distinct clustering. The cluster in the upper right portion of the metabolism network describes a suite of membrane transport

reactions, and the smaller metabolic clusters also form around a shared set of reactants with similar composition (e.g., **Figure 4**). 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 (**Figure 4**). Although this process (i.e., 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 (25), and also to reconcile these steps with sequences found within an organism's genome in an organized pipeline (gene → protein → function; e.g., (26)). Prior network descriptions and analyses of translation have focused on its dynamic attributes afforded by its autocatalytic motifs (27), genetic regulation of metabolic processes (28), the links between TM function and organism-level cellular dynamics (29) and a mapping of genotype to phenotype useful for interpreting a wealth of -omics cellular data (30). The chemosynthetic function of the TM and the topological attributes of the overall network did not figure significantly into these prior efforts. At the same time, network analyses of metabolic chemosynthesis have been instrumental in implicating network topology attributes such as a continuous, heterogeneous (heavy-tailed) connectivity distributions with complex systems behavior (14, 31–33).

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 3C**). The energy costs required to generate individual molecular components of the cell can be approximated through molecular masses (34). 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 4** 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. The clustering of the membrane transport reactions may be interpreted as indicating that maintenance of boundary conditions across the cell wall is a chemically distinct (and perhaps indispensable) function for which there is no counterpart within either intracellular redox metabolism or the TM (35–37).

Continuous versus multimodal groupings

The network's multimodal mass-connectivity distribution is the first such distribution reported for biological systems at the cellular level. Combinations of heavy-tailed aggregation, multimodality and discontinuity have been observed at the highest (ecological) levels of biological organization, in such diverse systems as arboreal forest (38), bird (39) fish and plankton (40) and mammal (41) groupings. Ecological multimodal aggregations arise from a confluence of dynamic processes operating at specific temporal periodicities and spatial scales (42, 43). Each aggregation can be so strongly influenced or controlled by small scale processes, which are distinct in speed or size, that the broader distribution of ecosystem resources is affected accordingly (44, 45). These processes form a complicated landscape, and can consist 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) (46). 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 (47).

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 (48, 49). 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 (50). 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 (51). The ribosome itself is consistently composed of 2:1 mass fractions of RNA to protein (52). 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.

There are numerous descriptions of scale-free networks of biological chemical systems, including theories of generative mechanisms that are relevant to prebiotic chemistry scenarios (33, 53–56). These 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 prescribed. Multimodal distributions present a unique challenge to theoretical studies of network generation, and may require elaboration using multiplexes that can account for multimodality (57).

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 and 4**) that, by virtue of their chemical isolation from one another, are each able to independently adapt to conditions that the cell may encounter (58). 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 (58, 59). Modules can exhibit hierarchical properties and function as key units of adaptive evolution because organismal fitness depends on their performance (32, 60–63).

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 3C**, right panel) between the largest metabolites and the smallest polypeptides. This trough is also present in objects that compose the translation network (**Figure 3C**, 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 (34). 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 (64). 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 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 (61). These functions are further isolated in time and space by regulatory mechanisms, such as via binding with the properly formed initiation factors (65–68), by corresponding activated tRNA complexes (69) and through spatial heterogeneity arising from complex physiological relationships with nucleoids (70). 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 4**) depict and corroborate the biochemical facets of translation's modularity, as do *in vitro* experimental systems that employ cell-free protein synthesis (71, 72).

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 (73), earthquakes (74), solar flares (75) and sandpile or avalanche dynamics (76) 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 (77) in which signaling, and the perturbative effects of novelty can propagate and be controlled in directions both up and down scales of a hierarchy (78).

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 (58). 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 (79). 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

Generating hierarchically connected networks across so many orders of magnitude, with modules that each contain at least one autocatalytic cycle (27, 80–82) presents a significant challenge to the field of prebiotic chemistry (83–85). 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 (86). The 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 (87), e.g., as an example of an emergent system in which multiple networks with distinct topologies coexist and are connected via common elements (88) and discontinuities stem from mechanisms that lead to network assortativity (89, 90). 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 (91). Phase transitions have long been recognized as potentially important behavioral features that are correlated with complex adaptive systems (92).

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 (93). 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 an elegant coupling of Hopf and Turing instabilities within a small network of reactions with a cubic autocatalytic motif (94). 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.

Chemical Reaction Network	log-likelihood ratio	p-value	k
Translation	247.44	<0.001	1.434
Polymer biosynthesis	24.10	0.36	2.279
Metabolism	1484.49	<0.001	1.747

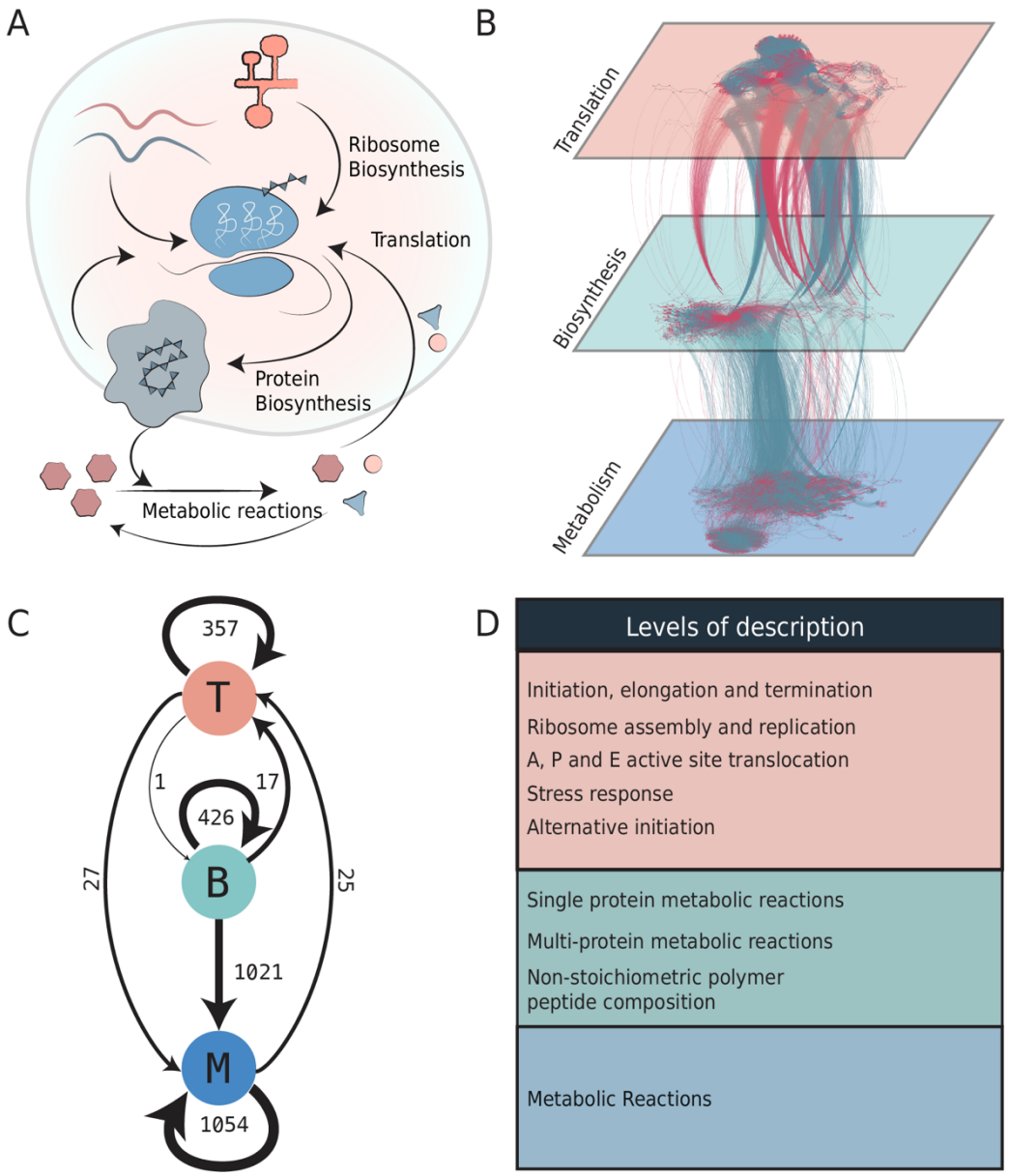

Figure 1. Overview of the level of systemic detail incorporated into the cellular chemosynthesis reaction network database. **A**. Schematic overview of translation network reconstruction including the underlying biochemical reactions. The metabolic network was obtained from Feist et al. (19) and was expanded to include translation and protein biosynthesis chemical reactions. **B**. Visualization and interconnectivity of each of the three chemosynthesis layers that enable translation to occur. **C**. An object connectivity diagram depicting the total number of distinct objects and their participation in reactions that either connect within a layer (circular arrows) or across layers (curved and straight arrows). **D.** Levels of description of chemical reactions for each network layer.

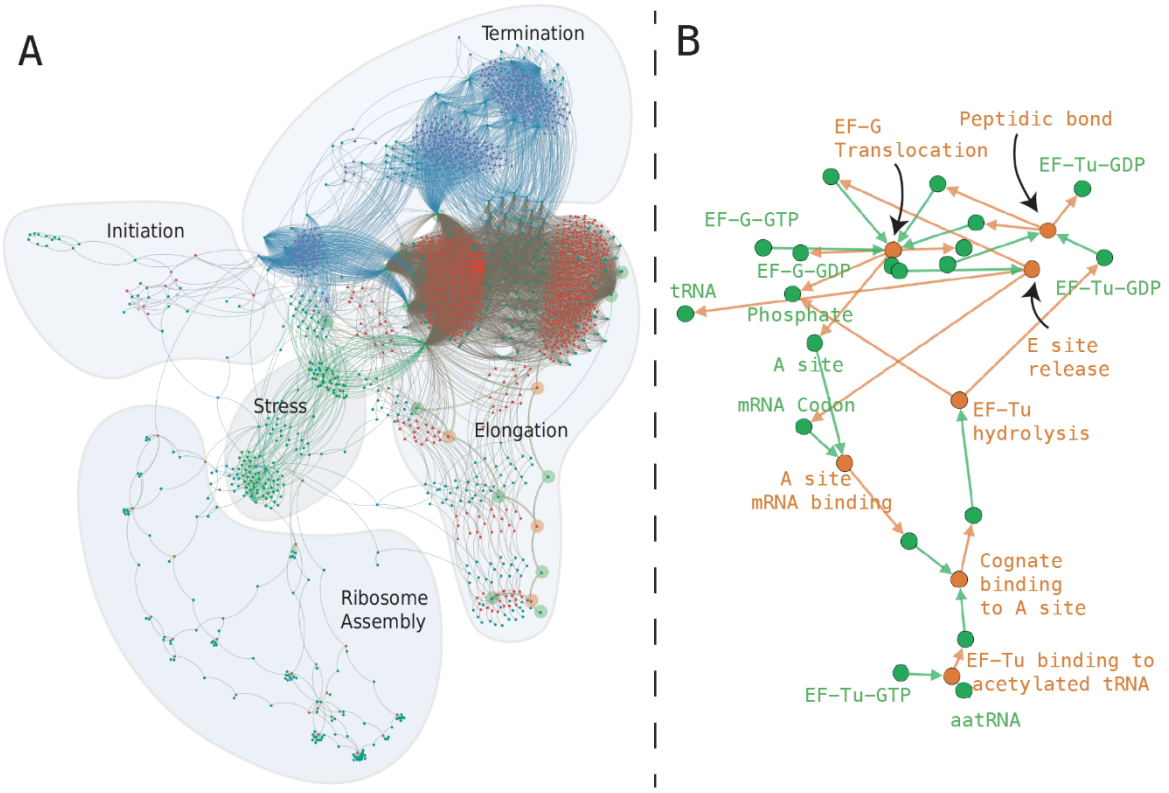

Figure 2. Overview of the translation portions of the cellular chemosynthesis reaction network database. **A.** Visualization of the complete translation chemical reaction network using Gephi 0.9.2. **B.** A depiction of the 'canonical' elongation process encoded within the translation network.

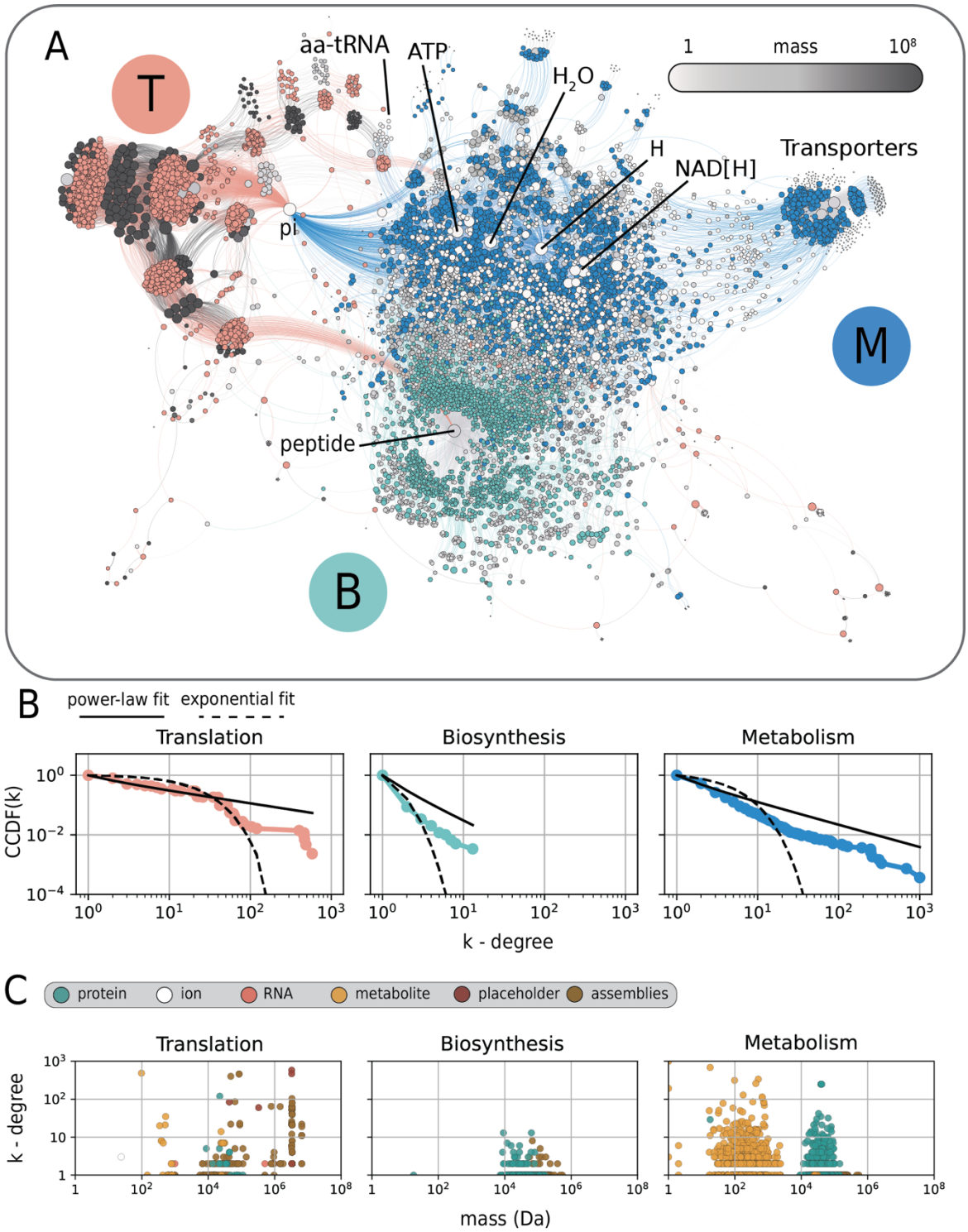

Figure 3. A. Full map with Translation (T), Polymer Biosynthesis (B) and Metabolic (M) networks, depicted with Gephi version 0.9.2 using a ForceAtlas2, scale 20, value 500 gravity algorithm. **B.** Cumulative degree distribution and distribution fit statistics for each network. Examples of the different 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) **C.** A depiction of

the mass-connectivity distribution for all objects in the full network, color coded according to their chemical categories. x-axis, node mass; y-axis, node degree.

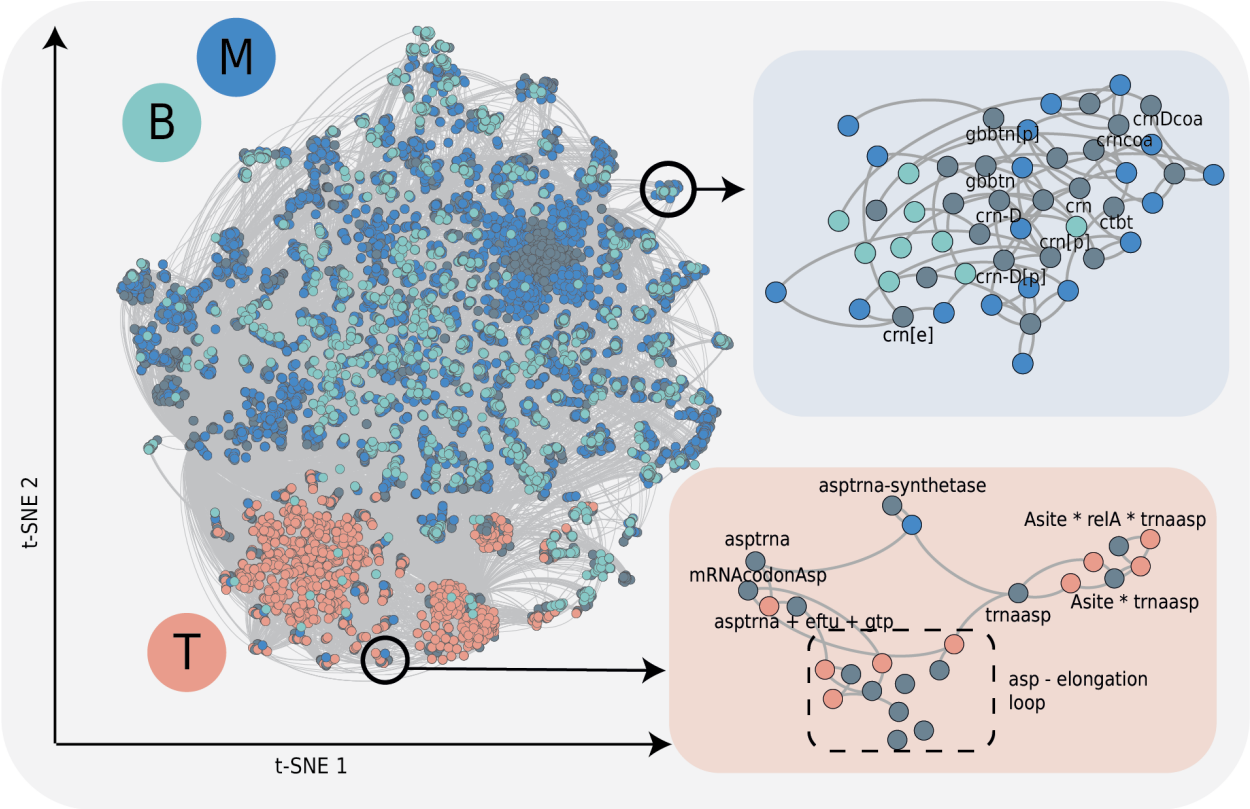

Figure 4. A clustering analysis using node2vec embeddings and the t-SNE values in two dimensions clearly distinguishes the hub structures of intracellular metabolism (upper grouping of blue), extracellular transport (upper right clusters of four in blue) and translation (lower left grouping of orange). Colors: dark blue, Metabolism; light blue, Biosynthesis; orange, Translation; gray, compounds.

References

1. M. A. Oberhardt, B. Ø. Palsson, J. A. Papin, Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5, 320–320 (2009).

2. M. Lewis, A Tale of Two Repressors. *J Mol Biol* 409, 14–27 (2011).

3. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318–356 (1961).

4. V. Chubukov, L. Gerosa, K. Kochanowski, U. Sauer, Coordination of microbial metabolism. *Nat Rev Microbiol* 12, 327–340 (2014).

5. V. Chubukov, I. A. Zuleta, H. Li, Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc National Acad Sci* 109, 5127–5132 (2012).

6. E. V. Koonin, A. S. Novozhilov, Origin and Evolution of the Universal Genetic Code. *Annu Rev Genet* 51, 1–18 (2016).

7. A. S. Novozhilov, Y. I. Wolf, E. V. Koonin, 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 (2007).

8. H. S. Zaher, R. Green, Quality control by the ribosome following peptide bond formation. *Nature* 457, 161–166 (2009).

9. W. Steinchen, V. Zegarra, G. Bange, (p)ppGpp: Magic Modulators of Bacterial Physiology and Metabolism. *Front Microbiol* 11, 2072 (2020).

10. T. Prossliner, K. Gerdes, M. A. Sørensen, K. S. Winther, Hibernation factors directly block ribonucleases from entering the ribosome in response to starvation. *Nucleic Acids Res* 49, gkab017- (2021).

11. A. L. Starosta, J. Lassak, K. Jung, D. N. Wilson, The bacterial translation stress response. *Fems Microbiol Rev* 38, 1172–1201 (2014).

12. S. E. Irving, N. R. Choudhury, R. M. Corrigan, The stringent response and physiological roles of (pp)pGpp in bacteria. *Nat Rev Microbiol* 19, 256–271 (2021).

13. Y. I. Wolf, E. V. Koonin, On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct* 2, 14–14 (2007).

14. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks. *Nature* 407, 651–654 (2000).

15. I. M. D. la Fuente, J. M. Cortes, D. A. Pelta, J. Veguillas, Attractor Metabolic Networks. *Plos One* 8, e58284 (2013).

16. A. Yadav, *et al.*, The Modular Adaptive Ribosome. *Plos One* 11, e0166021 (2016).

17. J. W. Whittaker, Cell-free protein synthesis: the state of the art. *Biotechnol Lett* 35, 143–152 (2013).

18. I. M. Keseler, *et al.*, The EcoCyc Database in 2021. *Front Microbiol* 12, 711077 (2021).

19. A. M. Feist, *et al.*, A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3, 121 (2007).

20. M. Bastian, S. Heymann, M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc Int Aaai Conf Web Soc Media* 3, 361–362 (2009).

21. J. Alstott, E. Bullmore, D. Plenz, powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. *Plos One* 9, e85777 (2014).

22. A. Grover, J. Leskovec, node2vec: Scalable Feature Learning for Networks. *Arxiv* (2016).

23. M. Fey, J. E. Lenssen, Fast Graph Representation Learning with PyTorch Geometric. *Arxiv* (2019) https:/doi.org/10.48550/arxiv.1903.02428.

24. P.-J. Kim, *et al.*, Metabolite essentiality elucidates robustness of Escherichia coli metabolism. *Proc National Acad Sci* 104, 13638–13642 (2007).

25. Z. Bar-Joseph, *et al.*, Computational discovery of gene modules and regulatory networks. *Nat Biotechnol* 21, 1337–1342 (2003).

26. S. A. Becker, B. O. Palsson, Context-Specific Metabolic Networks Are Consistent with Experiments. *Plos Comput Biol* 4, e1000082 (2008).

27. A. Roy, D. Goberman, R. Pugatch, A unifying autocatalytic network-based framework for bacterial growth laws. *P Natl Acad Sci Usa* 118, e2107829118 (2021).

28. A. Grimbs, D. F. Klosik, S. Bornholdt, M.-T. Hütt, A system-wide network reconstruction of gene regulation and metabolism in Escherichia coli. *Plos Comput Biol* 15, e1006962 (2019).

29. I. Thiele, *et al.*, Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. *Plos One* 7, e45635 (2012).

30. I. Thiele, N. Jamshidi, R. M. T. Fleming, B. Ø. Palsson, 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 (2009).

31. A.-L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks. *Phys Statistical Mech Appl* 272, 173–187 (1999).

32. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297, 1551–1555 (2002).

33. M. E. J. Newman, The Structure and Function of Complex Networks. *Siam Rev* 45, 167–256 (2003).

34. X.-P. Hu, H. Dourado, P. Schubert, M. J. Lercher, The protein translation machinery is expressed for maximal efficiency in Escherichia coli. *Nat Commun* 11, 5260 (2020).

35. D. Deamer, J. P. Dworkin, S. A. Sandford, M. P. Bernstein, L. J. Allamandola, The First Cell Membranes. *Astrobiology* 2, 371–381 (2002).

36. D. W. Deamer, J. Oro, Role of lipids in prebiotic structures. *Biosystems* 12, 167–175 (1980).

37. S. E. Maurer, D. W. Deamer, J. M. Boncella, P.-A. Monnard, Chemical Evolution of Amphiphiles: Glycerol Monoacyl Derivatives Stabilize Plausible Prebiotic Membranes. *Astrobiology* 9, 979–987 (2009).

38. C. S. Holling, Cross-Scale Morphology, Geometry, and Dynamics of Ecosystems. *Ecol Monogr* 62, 447–502 (1992).

39. J. Fischer, D. B. Lindenmayer, R. Montague-Drake, The role of landscape texture in conservation biogeography: a case study on birds in south-eastern Australia. *Divers Distrib* 14, 38–46 (2008).

40. T. D. Havlicek, S. R. Carpenter, Pelagic species size distributions in lakes: Are they discontinuous? *Limnol Oceanogr* 46, 1021–1033 (2001).

41. T. D. Lambert, J. R. Malcolm, B. L. Zimmerman, Amazonian small mammal abundances in relation to habitat structure and resource abundance. *J Mammal* 87, 766–776 (2006).

42. C. S. Holling, Understanding the Complexity of Economic, Ecological, and Social Systems. *Ecosystems* 4, 390–405 (2001).

43. R. V. O'Neill, D. L. Deangelis, J. B. Waide, T. F. H. Allen, A Hierarchical Concept of Ecosystems. (MPB-23), Volume 23 (2021) https:/doi.org/10.1354/books/unregistered/9780691236605.

44. J. Kolasa, Ecological Systems in Hierarchical Perspecitive: Breaks in Community Structure and Other Consequences. *Ecology* 70, 36–47 (1989).

45. T. F. H. Allen, T. B. Starr, *Hierarchy: Perspectives for Ecological Complexity* (University of Chicago Press, 2017) https:/doi.org/10.7208/chicago/9780226489711.001.0001.

46. C. Stow, C. Allen, A. Garmestani, Evaluating Discontinuities in Complex Systems: Toward Quantitative Measures of Resilience. *Ecol Soc* 12 (2007).

47. K. L. Nash, *et al.*, Discontinuities, cross-scale patterns, and the organization of ecosystems. *Ecology* 95, 654–667 (2014).

48. T. Fuhrer, D. Heer, B. Begemann, N. Zamboni, High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal Chem* 83, 7074–7080 (2011).

49. P. Dwivedi, *et al.*, Metabolic profiling of Escherichia coli by ion mobility-mass spectrometry with MALDI ion source. *J. Mass Spectrom.* 45, 1383–1393 (2010).

50. T. Geiger, A. Wehner, C. Schaab, J. Cox, M. Mann, Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins*. *Mol Cell Proteomics* 11, M111.014050 (2012).

51. M. Mori, *et al.*, From coarse to fine: the absolute Escherichia coli proteome under diverse growth conditions. *Mol Syst Biol* 17, e9536 (2021).

52. S. Kostinski, S. Reuveni, Ribosome Composition Maximizes Cellular Growth Rates in E. coli. *Phys Rev Lett* 125, 028103 (2020).

53. P. Nghe, *et al.*, Prebiotic network evolution: six key parameters. *Mol Biosyst* 11, 3206–3217 (2015).

54. A. Wagner, D. A. Fell, The small world inside large metabolic networks. *Proc Royal Soc Lond Ser B Biological Sci* 268, 1803–1810 (2001).

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

56. B. Shenhav, A. Solomon, D. Lancet, R. Kafri, “Transactions on Computational Systems Biology I” in Lecture Notes in Computer Science., (2005), pp. 14–27.

57. G. Bianconi, Statistical mechanics of multiplex networks: Entropy and overlap. *Phys Rev E* 87, 062806 (2013).

58. A. Hintze, C. Adami, Evolution of Complex Modular Biological Networks. *Plos Comput Biol* 4, e23 (2008).

59. A. Srivastava, S. Kumar, R. Ramaswamy, Two-layer modular analysis of gene and protein networks in breast cancer. *Bmc Syst Biol* 8, 81 (2014).

60. S. Venkataram, R. Monasky, S. H. Sikaroodi, S. Kryazhimskiy, B. Kacar, Evolutionary stalling and a limit on the power of natural selection to improve a cellular module. *Proc National Acad Sci* 117, 18582–18590 (2020).

61. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, From molecular to modular cell biology. *Nature* 402, C47–C52 (1999).

62. Y. Qi, H. Ge, Modularity and Dynamics of Cellular Networks. *Plos Comput Biol* 2, e174 (2006).

63. A. Gyorgy, D. D. Vecchio, Modular Composition of Gene Transcription Networks. *Plos Comput Biol* 10, e1003486 (2014).

64. A. D. Goldman, B. Kacar, Cofactors are Remnants of Life's Origin and Early Evolution. *J Mol Evol* 89, 127–133 (2021).

65. A. Simonetti, *et al.*, A structural view of translation initiation in bacteria. *Cell Mol Life Sci* 66, 423 (2008).

66. I. M. Sharma, S. A. Woodson, RbfA and IF3 couple ribosome biogenesis and translation initiation to increase stress tolerance. *Nucleic Acids Res* 48, 359–372 (2019).

67. P. Milón, M. V. Rodnina, Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol* 47, 334–348 (2012).

68. D. Benelli, P. Londei, Begin at the beginning: evolution of translational initiation. *Res Microbiol* 160, 493–501 (2009).

69. J. Schuette, *et al.*, GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *Embo J* 28, 755–765 (2009).

70. S. Bakshi, H. Choi, J. C. Weisshaar, The spatial biology of transcription and translation in rapidly growing Escherichia coli. *Front Microbiol* 6, 636 (2015).

71. M. C. T. Hartman, K. Josephson, C.-W. Lin, J. W. Szostak, An Expanded Set of Amino Acid Analogs for the Ribosomal Translation of Unnatural Peptides. *Plos One* 2, e972 (2007).

72. H. Niederholtmeyer, V. Stepanova, S. J. Maerkl, Implementation of cell-free biological networks at steady state. *Proc National Acad Sci* 110, 15985–15990 (2013).

73. J. Pérez-Mercader, Astrobiology, The Quest for the Conditions of Life. 337–360 (2002).

74. B. Gutenberg, C. F. Richter, Frequency of earthquakes in California*. *B Seismol Soc Am* 34, 185–188 (1944).

75. S. Krucker, A. O. Benz, Energy Distribution of Heating Processes in the Quiet Solar Corona. *Astrophysical J Lett* 501, L213–L216 (1998).

76. P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality. *Phys Rev A* 38, 364–374 (1988).

77. A. Garmestani, C. Allen, L. Gunderson, Panarchy: Discontinuities Reveal Similarities in the Dynamic System Structure of Ecological and Social Systems. *Ecol Soc* 14 (2009).

78. C. S. Holling, L. H. Gunderson, Resilience and adaptive cycles. *In: Panarchy: Understanding Transformations in Human and Natural Systems, 25-62* (2002).

79. L. M. Shekhtman, S. Shai, S. Havlin, Resilience of networks formed of interdependent modular networks. *New J Phys* 17, 123007 (2015).

80. D. Y. Zubarev, D. Rappoport, A. Aspuru-Guzik, Uncertainty of Prebiotic Scenarios: The Case of the Non-Enzymatic Reverse Tricarboxylic Acid Cycle. *Sci Rep-uk* 5, 8009 (2015).

81. R. Braakman, E. Smith, The compositional and evolutionary logic of metabolism. *Phys Biol* 10, 011001 (2013).

82. M. Preiner, *et al.*, Catalysts, autocatalysis and the origin of metabolism. *Interface Focus* 9, 20190072 (2019).

83. J. I. Smith, M. Steel, W. Hordijk, Autocatalytic sets in a partitioned biochemical network. *J Syst Chem* 5, 2 (2014).

84. L. Orgel, In the beginning. *Nature* 439, 915–915 (2006).

85. R. Shapiro, Small Molecule Interactions were Central to the Origin of Life. *Q Rev Biology* 81, 105–126 (2006).

86. D. Sornette, Critical Phenomena in Natural Sciences, Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. *Springer Series Syne*, 239–256 (2000).

87. N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, S. Havlin, Percolation in directed scale-free networks. *Phys Rev E* 66, 015104 (2002).

88. E. A. Leicht, R. M. D'Souza, Percolation on interacting networks. *Arxiv* (2009) https:/doi.org/10.48550/arxiv.0907.0894.

89. M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks. *Phys Rev E* 69, 026113 (2004).

90. M. E. J. Newman, Mixing patterns in networks. *Phys Rev E* 67, 026126 (2003).

91. I. Kryven, G. Bianconi, Enhancing the robustness of a multiplex network leads to multiple discontinuous percolation transitions. *Phys Rev E* 100, 020301 (2019).

92. G. Paperin, D. G. Green, S. Sadedin, Dual-phase evolution in complex adaptive systems. *J Roy Soc Interface* 8, 609–629 (2011).

93. D. L. Gonzalez, S. Giannerini, R. Rosa, On the origin of degeneracy in the genetic code. *Interface Focus* 9, 20190038 (2019).

94. A. P. Muñuzuri, J. Pérez-Mercader, 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 (2022).