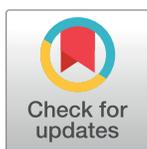RESEARCH ARTICLE

Rules, hypergraphs, and probabilities: The three-level analysis of chemical reaction systems and other stochastic stoichiometric population processes

Eric Smith[1,2,3]*, Harrison B. Smith[1,4], Jakob Lykke Andersen[5]

1 Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan, **2** Santa Fe Institute, Santa Fe, New Mexico, United States of America, **3** Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia, United States of America, **4** Blue Marble Space Institute of Science, Seattle, Washington, United States of America, **5** Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

* desmithelsi.jp

Abstract

We consider problems in the functional analysis and evolution of combinatorial chemical reaction networks as rule-based, or three-level systems. The first level consists of rules, realized here as graph-grammar representations of reaction mechanisms. The second level consists of stoichiometric networks of molecules and reactions, modeled as hypergraphs. At the third level is the stochastic population process on molecule counts, solved for dynamics of population trajectories or probability distributions. Earlier levels in the hierarchy generate later levels combinatorially, and as a result constraints imposed in earlier and smaller layers can propagate to impose order in the architecture or dynamics in later and larger layers. We develop general methods to study rule algebras, emphasizing system consequences of symmetry; decomposition methods of flows on hypergraphs including the stoichiometric counterpart to Kirchhoff's current decomposition and work/dissipation relations studied by Wachtel *et al.*; and the large-deviation theory for currents in a stoichiometric stochastic population process, deriving additive decompositions of the large-deviation function that relate a certain Kirchhoff flow decomposition to the extended Pythagorean theorem from information geometry. The latter result allows us to assign a natural probabilistic cost to topological changes in a reaction network of the kind produced by selection for catalyst-substrate specificity. We develop as an example a model of biological sugar-phosphate chemistry from a rule system published by Andersen *et al.* It is one of the most potentially combinatorial reaction systems used by biochemistry, yet one in which two ancient, widespread and nearly unique pathways have evolved in the Calvin-Benson cycle and the Pentose Phosphate pathway, which are additionally nearly reverses of one another. We propose a probabilistic accounting in which physiological costs can be traded off against the fitness advantages that select them, and which suggests criteria under which these pathways may be optimal.

Author summary

The dynamics of chemical reaction systems, and their change under evolution, are complex because constraints on what is possible, and how hard it is for evolution to discover or maintain any given function, arise variously from chemical mechanisms, network structure, and the stochastic dynamics generated by these on thermodynamic landscapes. Well-developed methods from algebra, topology, and probability exist to study each of these levels, and have been employed extensively within levels. We study here the inter-level relations that propagate constraint and order upward and downward from mechanisms to population-level dynamics, for both the function and the evolution of chemical reaction networks and more general systems that transform their members in sets rather than individually. We show that some important universal properties in living systems, such as the minimality of the reaction sequence in the Calvin cycle, can be derived directly as consequences of chemical mechanism, while others, such as its pruning from a combinatorially large set of alternatives, probably reflect relations between network topology and a probability measure related to dissipation, which we derive. We are interested in the information that specifies full systems, as it is supplied within levels and by way of their mutual relations.

1 Introduction: From rules and micro-statistics to macro-phenomena in complex combinatorial systems

The three-level architecture of many combinatorial and complex stochastic systems

Systems that may be defined by quite complicated and heterogeneous components and interactions may nonetheless be tractable to analysis if their state-spaces and event-spaces do not grow too large to organize and navigate [1, 2]. At a different extreme, systems with very large state- and event-spaces produced by combinatorial composition may be impractical to represent explicitly yet nonetheless remain tractable if these spaces have high symmetry and not-too-much structural diversification [3–5]. It is at the interaction of these two forms of "large-ness"—where heterogeneous mechanisms and interactions act combinatorially to generate large state- and event-spaces, that we encounter many frontiers of complexity [6].

One class of complex systems for which a broad approach has come forth are the so-called *rule-based systems* [7–13]. For many of these, the phenomena of interest are realized as stochastic processes, in which case we also refer to them as "three-level" systems. The first level consists of "rules" that abstract the classification and transformation of objects of some kind. Formal models of rules, together with prescriptions for applying them to particular objects while preserving the contexts in which they act [11, 12, 14], enable the inductive construction of object and event types by recursive rule application, which then furnish a second level of description. For stochastic phenomena, the models of literal objects and events, together with choices about the possible states in which they can occur, can take on the interpretation of generators of stochastic processes [15] producing trajectories and probability dynamics in the state spaces that constitute a third level of description.

In this paper we demonstrate a program of analysis for the phenomena generated by a three-level system, with an emphasis on the interaction of heterogeneity in the generating structures and combinatorial interactions among them, both from rules to graphs and from graphs to probability dynamics. We are interested in the way specification of the system—by

us as model-builders or by constraints imposed physically or emerging from selection in nature—may be introduced at one level in the hierarchy and then propagate across levels to create order in the output at other levels. This propagation may occur by restricting combinatoriality, in which case we may be able to derive its consequences by direct arguments, or it may guide combinatorial expansion that drives emergence and robustness of collective phenomena through laws of large numbers. We will exhibit cases of each, and study how jointly they enable the design and control of function in complex combinatorial systems.

The stochastic stoichiometric population processes among the three-level systems

For simplicity of approach, our treatment will apply to what we term *stochastic stoichiometric population processes* (SSPPs): discrete-state stochastic processes in which objects are counted as individuals grouped into types, with states corresponding to their joint counts in populations, and in which each elementary event removes some set of individuals from a population and replaces it with some other set. The SSPPs are an expressive abstraction in that they are known to include phenomena with computationally complex search and optimization problems [16], yet they are defined from a small collection of primitives [17, 18] and admit numerous topological [19] and other [13, 18, 20] modes of analysis.

The SSPPs are well known and developed as models of chemical reaction networks (CRNs) [21–27], but we refer to them in general form to emphasize that they have much wider applications including Darwinian populations [28] and genetic lifecycles [29, 30]. To many of these the narrow and restrictive assumptions about dynamics that are natural for elementary CRNs do not apply [31].

The generative relation of rules in the form of graph grammars to stoichiometric chemical networks has been studied extensively by Andersen *et al.* [11, 14, 32–35]. Separately, the generative relation between the topology of stoichiometric systems and their physical thermodynamics and more general large-deviation behavior has been studied by numerous authors [21, 22, 24–27, 31, 36–43]. Therefore we will focus here on phenomena that draw jointly on properties at all three levels, and on the combinatorics both of rules and of stochastic events. (A somewhat different study of thermochemical characteristics together with molecular combinatorics is that of [44], who make a computational library of small linear-chain molecules and explore the formation free energy and reaction opportunities for all distinct concatenations of oxidation states of the carbon centers, finding that biological metabolites are enriched among more stable and more soluble compounds relative to a uniformly sampled library).

What questions arise for dynamics that is both combinatorial and complex?

The interaction of system structure with combinatorics is familiar from the relation of internal energy landscapes and entropy in equilibrium thermodynamics [4, 45–47]. With suitable generalizations to non-equilibrium contexts, the same kinds of relations can be derived between the stoichiometric graph and its large-deviation statistics [31, 41, 48, 49]. For rule-based systems, a second interplay of generating structure with combinatoriality arises between the rule algebras and the graph with its associated spaces of states and transitions [9, 11, 12].

We will apply our general constructions to an example problem from the biochemistry of sugar-phosphates. These compounds stand out among metabolites for the very large list of molecules and reactions that *could be* generated from a small set of generating mechanisms [34], because those can act on most molecules, in some cases at multiple positions. In biochemistry this rule redundancy has enabled the assembly of networks that recombine sugars

without material loss and with modest dissipation, and within these networks, pathways have been selected that are both very specific and highly conserved.

In terrestrial evolution we find the emergence of a single, ancient, universal pathway for sugar re-arrangement in the anabolic Calvin-Benson-Bassham cycle [50] through substrate specificity of catalysts, from a formally-infinite chemical network that could diffusely perform the same conversion. Moreover, the Calvin-cycle pathway is nearly reversed in direction to form a second ancient, nearly-universal pathway in the catabolic Pentose-Phosphate pathway [51–56]. We want to account both for how the two functions of catalysts—realizing reaction mechanisms and specifically restricting substrates [57, 58]—relate in the process of pathway evolution, and also why these particular pathways are the outcomes. Our example is motivated in part by its reversibility and potential for near-equilibrium kinetics in biologically pertinent regimes. This simplification permits us to separate out the quantitative impacts of most kinetic features for a separate and secondary analysis, which can be approached in future work with methods of *Metabolic Control Analysis* [59, 60].

Three questions we wish to answer are: 1) Which aspects of pathway specificity derive from rule properties and which from ancillary network restriction? 2) How are macroscopic transport capabilities and force/flux response that might be targets of selection attributable to lower-level and distributed catalyst specificities that can be modified to select them? and 3) What cost functions might be assignable that bridge the functional properties of the network as a transport mechanism and the informational attributes of that network as a product of selection?

Technical and general-method results

A variety of results derived below address the foregoing biologically-motivated questions, but support the analysis of SSPPs as three-level systems more generally:

For rules we study the presence of discrete symmetries and resulting conserved quantities that constrain all conversions that can be performed by composition of rule actions. We show that certain optimality properties such as shortest path length and a topological characteristic related to minimization of dissipation can be computed from the rule algebra alone without the intermediate step of explicit construction of stoichiometric networks through network expansion and subsequent exhaustive enumeration of integer-flow solutions on the resulting networks.

For stoichiometric networks we study the generalization of Kirchhoff's current decomposition laws from electric circuit theory [61]. The systematic construction of bases of null flows of stoichiometry has been shown before to be useful for analysis of dissipation [37]. In our graph-grammar model it will enable a complete construction of the null space in terms of a fixed class of basis types for networks of indefinite size.

We situate our analysis of work and dissipation in a somewhat more general study of nesting hierarchies for graphs than is done by Wachtel *et al.* [27], because our concern is less with defining and computing transduction and efficiency (already fully performed by them), than with deriving all concepts within the representational abstraction of the hypergraph. Where Wachtel *et al.* invoke several *ad hoc* criteria specific to chemistry (and only for "elementary" reactions with all species kept explicit)—such as their class of admissible *chemical transformations* defined in terms of element conservation, and dissipation derived inherently in terms of energy and the local detailed balance commonly assumed in stochastic thermodynamics [62, 63]—we obtain all notions of system/environment relations and their jointly performed transformations solely from stoichiometry and graph embedding, and entropy decompositions from large-deviation functions. We do this to emphasize that the SSPPs are a much wider class

to which these abstractions can be applied than to CRNs with energy conservation [31], and that even within chemistry a variety of different coarse-grainings [64] may entail different conservation laws and functional forms for likelihood within a standard modeling framework.

In the dynamics of probability under the master equation, we re-present the work-dissipation identities from [22, 27], and derive a certain representation for the large-deviation function for currents known from [41] (and proceeding from a slightly different derivation, in [42, 43]), in a suitable steady-state limit of an integral expression for log-likelihood of general non-equilibrium macro-trajectories. In particular we show that the LDF possesses an additive decomposition in the form of the extended Pythagorean theorem of information geometry [65, 66], for current solutions under the mass-action rate law in hierarchies of nested graphs.

1.1 Organization of the presentation

Sec. 2 introduces SSPPs as a model class and their topological representation by hypergraphs. The stoichiometric counterparts to Kirchhoff current laws are derived for steady states, along with related identities for chemical work delivery and dissipation noted in [27]. Definitions of system/environment decomposition for graphs, and reducibility or irreducibility of flows on graphs, are given in terms of graph topology and stoichiometry alone. These will serve as a basis to define boundary conditions for driven networks and to formally distinguish work transduction through stoichiometric and non-stoichiometric coupling (termed "tight" and "loose" coupling in [27]).

Sec. 3 introduces the generative relationship from a rule algebra representing chemical reaction mechanisms by graph grammars [11, 14] to the set of chemical species and the hypergraph of reactions that defines the SSPP. Here we introduce a model of sugar-phosphate chemistry from [34], and a thermochemical landscape produced from modern group-contribution methods [67]. From these we can cast the emergence of a universal Calvin cycle and Pentose-Phosphate pathway as a problem of selection from a combinatorially large network faced historically by biological evolution.

We frame the problem of chemical conversion on networks in terms of integer linear programming, and exhaustively enumerate all stoichiometrically independent solutions to the Calvin cycle conversion on the computer-generated graph. We then show how rules can be used to decompose the space of null flows for graphs of any size, and how conservation laws at the rule level propagate to constrain solutions at any order, leading to proofs of solution minimality without exhaustive enumeration.

Sec. 4 derives the large-deviation theory for concentrations and currents on general SSPPs, and applies it to flows in the null space from the example of Sec. 3. We derive an integral form for time-dependent large-deviation functions (LDFs) from their associated Hamilton-Jacobi theory [68–70], and compute the deviation rate function for currents from the stochastic-process generator. We then show how the information geometry induced by the Kullback-Leibler divergences of mass-action currents on a graph and any of its subgraphs leads to an additive decomposition by the extended Pythagorean theorem, following the same partition used to decompose Kirchhoff components of null flows in Sec. 2.

We use the simplifying limit of linear response about a thermodynamic equilibrium to show how a measure of network resistance to chemical conversions can be defined for stoichiometric systems, and how this is dominated by contributions from graph topology shown in Sec. 3, for realistic thermochemical landscapes governing the actual Calvin cycle and Pentose Phosphate pathway. Finally we argue that the LDF for currents with its Pythagorean decomposition is a natural cost function for the contribution of physiology to the problem of natural selection for catalyst specificity in biological populations also modeled as SSPPs.

2 The stochastic stoichiometric population processes as a model class

Our modeling abstraction will be the stochastic stoichiometric population processes (SSPPs). In a *population process*, entities are population members that come in P (finite or countable) types, and population states are counts of their members by type, corresponding to lattice points in the non-negative orthant of $\mathbb{Z}^P$. Change events are *stoichiometric*, meaning that in each elementary event some set of individuals, specified by type-counts, is removed from the population and some other set is added to produce the next state of the population. To declare a stochastic realization of a process we add to the state space and stoichiometry a specification of the sampling rule and rate parameters for the activation of transitions.

2.1 Declaring a model (as far as the generator)

Our specification of a model is equivalent to standard constructions for chemical reactions [22], replicating populations evolving under selection [28], or lifecycle models with explicit genetic stoichiometry [29]. This section serves just to fix notation and definitions. Further detail enabling solutions for quantities introduced here is provided in S1 File A 1.

Population members are termed *species*, and indexed $p \in 1, \ldots, P$, and the species themselves are given formal labels $X \equiv [X_p]$, which we arrange as a column vector. States are column vectors of non-negative integer counts $n \equiv [n_p]$. Probabilities are written as column vectors $\rho \equiv [\rho_n]$ over the state index n, with $\Sigma_n \rho_n = 1$.

Probabilities evolve under under a continuous-time, discrete state master equation [71]

$$\frac{\partial \rho_n}{\partial t} = \sum_{n'} \mathbb{T}_{nn'} \rho_{n'}, \tag{1}$$

with *transition matrix* $\mathbb{T} \equiv [\mathbb{T}_{nn'}]$. $\mathbb{T}$ is a stochastic matrix, meaning $1^T \mathbb{T} \equiv 0$.

We refer to the transition types generically as *reactions*, and represent each reaction as a directed link between a pair of *complexes*, following Horn and Jackson [17] and Feinberg [18]. Complexes are *multisets* of species, (meaning that a complex may contain multiple members of the same species [11]), and are denoted with an index *i*. The type membership in complexes is represented in a matrix Y with columns $y^i \equiv [y_p^i]$, giving the count of each type p in complex *i*, known as the *stoichiometric coefficients*.

A reaction from complex *i* to complex *j* is represented by a *schema* in the standard chemical notation $y^j \cdot X \xrightarrow{k_{ji}} y^i \cdot X$, which we write in terms of formal vector products between stoichiometric vectors and species labels, annotated with its rate constant k_{ji}. To simplify notation we suppose that there is at most one reaction in either direction between a unique pair of complexes, so that unidirectional reactions can be labeled with ordered pairs *ji*.

A collection of reaction schemata specify a stoichiometric model in the form of a *hypergraph* [19]—a set of nodes with a set of hyper-edges in which each hyper-edge is an (ordered or unordered) pair of sets or nodes—that we label $\mathcal{G}$ and write simply as the set of its reactions:

$$\mathcal{G} \equiv \left\{ y^j \cdot X \xrightarrow{k_{ji}} y^i \cdot X \right\}_{ji}. \tag{2}$$

In our treatment of open SSPPs—processes in which members enter or leave populations from some environment—all allowed state changes must follow from explicitly declared stoichiometry. Thus a system plus its environment are given a joint representation by a graph, of which the system component is a proper subgraph $\mathcal{G}' \subset \mathcal{G}$.

To compactly represent transition matrices, rather than writing out the matrix element for each state transition explicitly, we write the elements as shift operators acting on state indices. $\mathbb{T}$ then becomes a sum over the the number of ordered complex pairs ji. As a notation for this representation, we use the exponential form $e^{y^i \cdot \partial/\partial \mathrm{n}^T}$ for the shift operator $\mathrm{n} \rightarrow \mathrm{n} + y^i$ acting to the right.

To trigger events, complexes are sampled independently and proportionally from the population without replacement. The number of possible samples of complex i from a population in state n defines a complex *activity* of the form

$$\mathrm{n}^{\underline{y^i}} \equiv \prod_p \mathrm{n}_p^{\underline{y_p^i}} \equiv \prod_p \frac{\mathrm{n}_p!}{(\mathrm{n}_p - y_p^i)!}, \tag{3}$$

where we use the underbar notation for the (component-wise) falling factorial [23]. The *half-reaction rate* for reaction ji is a product of the activity (3) of the input complex i with the rate constant k_{ji} labeling the schema. With these choices, the transition matrix on the graph (2) takes the form:

$$\mathbb{T} = \sum_{ji} k_{ji} \left(e^{(y^i - y^j) \cdot \partial/\partial \mathrm{n}^T} - 1 \right) \mathrm{n}^{\underline{y^i}}. \tag{4}$$

Forms for the case of detailed balance. For networks with rate constants that admit a detailed-balance equilibrium, the half-reaction rate constants can be expressed in terms of pairs of potentials, one defined for species and the other for reactions. We denote one-particle chemical potentials for species with a vector $\mu_\circ \equiv [\mu_\circ^p]$, from which a chemical potential for any complex i is constructed as $\mu_\circ^i \equiv \mu_\circ^T y^i$. For reactions we introduce a vector $\mu^\ddagger \equiv [\mu_{\langle ji \rangle}^\ddagger]$ indexed by *unordered* pairs $\langle ji \rangle$, which may be thought of as "one-complex" chemical potentials at the transition state. In terms of these, the half-reaction rate constants are given by

$$k_{ji} = e^{-\left(\mu_{\langle ji \rangle}^\ddagger - \mu_\circ^i \right)/k_B T}. \tag{5}$$

First moments, the rate law, and the stoichiometric matrix. We denote expectations in the distribution ρ_n with angle brackets, as in

$$\langle \mathrm{n} \rangle \equiv \sum_\mathrm{n} \rho_\mathrm{n} \mathrm{n} \tag{6}$$

for the first moment. The deterministic equation for the time dependence of $\langle \mathrm{n} \rangle$ is called the *rate law*. It is expressed in terms of the expected fluxes

$$v_{(ji)} \equiv k_{ji} \langle \mathrm{n}^{\underline{y^i}} \rangle - k_{ij} \langle \mathrm{n}^{\underline{y^j}} \rangle, \tag{7}$$

where we use the index with parentheses (ji) to indicate that each bidirectional reaction only needs to be counted once, but it requires an (arbitrary) assignment of directionality to assign a sign convention, indicated by the index order ji.

We denote the *stoichiometric matrix* between currents and species as $\mathbb{S} \equiv [\mathbb{S}_{p\,(ji)}]$ with the column at reaction index (ji) given by

$$\mathbb{S}_{(ji)} = y^j - y^i. \tag{8}$$

The time-dependence of the first moment (6) is written in matrix form in terms of currents as

$$\frac{d}{dt}\langle n \rangle = \mathbb{S}\nu. \tag{9}$$

For non-equilibrium steady states in sub-graphs driven by sources and sinks of species, in cases where we do not write the graph for the driving environment explicitly, we will sometimes replace Eq (9) with a steady-state relation

$$J^{\text{ext}} = \mathbb{S}\nu, \tag{10}$$

defining the sign convention for external source currents J^{ext} as positive for flow out of the system or the graph.

Mean-field approximation. The *mean-field approximation* to the rate law (7) replaces averages of factorial moments with powers of the first moment:

$$\nu_{(ji)} \sim k_{ji}\langle \mathrm{n} \rangle^{y^i} - k_{ij}\langle \mathrm{n} \rangle^{y^j}. \tag{11}$$

In Eq (11) we introduce the notation $\sim$ for approximations defined by asymptotic leading-exponential dependence. These (including the mean-field approximation) are not generally regulated small-parameter expansions in the sense of perturbation series, but are often arrived at by saddle-point approximation. We use $\approx$ to denote small-parameter approximations such as leading-order Taylor's series expansions.

We show in S1 File A 2 how the source Eq (10) with the mean-field form for currents (11) is inverted within the stoichiometric subspace of a graph for a profile $\langle n \rangle$, which is needed to compute chemical potentials μ and the current solutions ν that appear in expressions for dissipation below.

If we denote by $\underline{n}$ the value of $\langle \mathrm{n} \rangle$ at any thermodynamic *equilibrium* (entailing the assumption of detailed balance), then species chemical potentials corresponding to the mean-field approximation take the forms for the mass-action rate law,

$$
\begin{aligned}
\underline{\mu}^p &\sim \mu_\circ^p + \log \underline{n}_p \\
\mu^p &\sim \underline{\mu}^p + \log\left(\frac{\langle \mathrm{n}_p \rangle}{\underline{n}_p}\right),
\end{aligned}
\tag{12}
$$

respectively at and away from equilibrium. The first line of Eq (12) sets $\underline{\mu}^i = \underline{\mu}^j$ for any reaction ji in Eq (11); implying $\underline{\mu}\mathbb{S} = 0$. We introduce the notation

$$\underline{J}^{\ddagger}_{\langle ji \rangle} \equiv k_{ji}\underline{n}^{y^i} = k_{ij}\underline{n}^{y^j} \tag{13}$$

for the equal and opposite half-reaction currents at such an equilibrium.

Dissipation by a current. The general expression for dissipation of chemical potential to heat, by a vector of currents ν in the presence of chemical potentials μ, is

$$\dot{Q} = -\mu \mathbb{S}\nu \tag{14}$$

Taking the linear-order expansion about equilibrium of $\log(\langle n \rangle / \underline{n})$, in the expression (12) for μ and (11) for ν, gives the quadratic order approximation of Eq (14):

$$\dot{Q}(\nu) \approx \nu^T \operatorname{diag}\left[\frac{1}{\underline{J}^{\ddagger}}\right] \nu. \tag{15}$$

An expression for $\dot{Q}(\nu)$ directly in terms of J^{ext} from Eq (10) is given in Eq. (A24) of S1 File A 2.

A special case of the quadratic order dissipation (15) that fully separates the contributions of ν from those of the thermochemical background is that in which by suitable choice of units we may take $\underline{J}^{\ddagger} \equiv 1$. We term this the "topological" dissipation:

$$\dot{Q}_{\text{top}}(\nu) \quad \equiv \nu^T \nu = \sum_{(ji)} \nu_{(ji)}^2. \tag{16}$$

$\dot{Q}_{\text{top}}$ is directly connected to network topology for those currents in which the topology dictates the relative magnitudes of all reaction fluxes. The idealization $\underline{J}^{\ddagger} \equiv 1$ can be reasonably approximated in thermochemical equilibria at empirically calibrated reaction free energies for our example system of sugar-phosphate chemistry, at concentration profiles relevant to physiological values.

Linear-response impedance of a network to a through-flow.   In the linear-response regime, a natural measure of impedance of a network to a source current $-J^{\text{ext}}$ can be defined as a generalization of the expression $R = V/I = P/I^2$ [61] for electrical resistance (in which V is voltage across a resistor, I current through it, and P dissipated power within it). Using Eq (10), the ratio of dissipated power to a natural quadratic scalar measure for throughput becomes

$$R \equiv \frac{\dot{Q}(\nu)}{J^{\text{ext} T} J^{\text{ext}}} = \frac{J^{\text{ext} T}(\mathbb{S} \operatorname{diag}[\underline{J}^{\ddagger}] \mathbb{S}^T) J^{\text{ext}}}{J^{\text{ext} T} J^{\text{ext}}} \tag{17}$$

The impedance (17) is particular to the ratio of components in ν produced by the mass-action solution, which depend in general on all three of: the conversion J^{ext} performed, the topology of the network, and the thermochemical background represented by $\underline{J}^{\ddagger}$. R is independent, however, of the overall scale factor given by $J^{\text{ext} T} J^{\text{ext}}$, and in that sense is a function of the structural properties of the response of the network to driving, but not of the magnitude of the dissipation itself.

Non-linear extensions of the quadratic dissipation function.   The quadratic form (15) is (up to differing factors of 2) a linear-response limit of two different non-linear expressions at large flow-rate. One is the general dissipation rate (14), familiar from its long use in non-equilibrium thermodynamics [63, 72–75], and the other is a large-deviation rate function for currents [41–43]. Through these the cost function (17) per squared-unit transport can likewise be generalized in two ways with different probabilistic interpretations.

2.2 The generator as the middle level in 3-level rule-based systems

The foregoing construction applies to any stochastic stoichiometric population process, providing, for example, a model of populations of literal molecular species and literal reactions among them. For many phenomena, however, the literal reaction network is neither the most compact nor the most complete reflection of our knowledge of the process at work. Chemical reactions are grouped into equivalence classes in terms of common reaction *mechanisms* [76], so named because they perform specific conversions of chemical bonds among conserved atomic centers, which together are generally only fragments of whole molecules.

Here we will be interested in stoichiometric systems that are not merely declared *ad hoc* but are themselves generated by specified mechanisms. The generalizations from the particular chemical notion of reaction mechanism will be termed *rules*, and the approach to creating state spaces, networks, and dynamics from rules is called *rule-based* modeling [10].

We will also term phenomena described in this way **three-level systems**, for the hierarchical levels with generative relations between them: those of *rules*, stochastic process *generators* specified as in Eq (4) from the topology of the hypergraph, and *ensemble dynamics* on the state space or its bundle of histories. Fig 1 shows the general relation among levels in a three-level system, illustrating with the case of CRNs. In addition to the chemical interpretations of each level and the generative relations between them, we note the main mathematical objects that formalize the levels, and the parts of overall solutions that each contributes.

For chemistry a natural intermediate level of abstraction, known as a *graph grammar* [14] represents molecules as ordinary graphs in which atoms are (typed) nodes and bonds are (typed) links. Reaction mechanisms correspond to rewriting rules for graph fragments, which retain atomic centers and reconfigure bonds. The mathematical relation bridging the (lower) level of rules and the (higher) level of stoichiometric reaction networks is the structure-preserving embedding, of the graph fragment altered under the reaction mechanism within
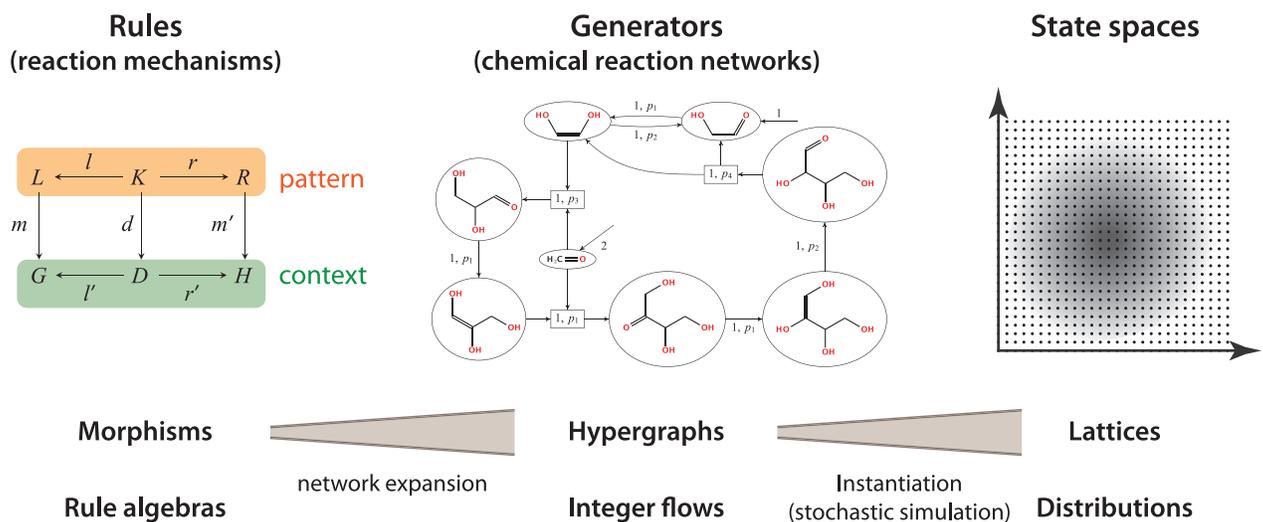
Fig 1. The three levels (upper labels) in a Stochastic Stoichiometric Population Process realized as a rule-based system, with chemical reactions as an example. Rules correspond to reaction mechanisms, in which a *context K* comprising the active atoms can support two bond configurations that we generically term patterns: a reactant pattern *L* is converted to a product pattern *R* by the reaction. The reactant pattern is embedded as a subgraph in one or more literal molecules *G* by a map *m*, and the conversion maps *l* and *r* on patterns are used to generate embeddings *d* and *m′*, and conversions *l′* and *r′* to new literal molecules *H*, so that the remainders of the literal molecules outside the reacting bonds are "carried along" by the mechanism in a structure-preserving way. The second level formed by the action of rules on molecules is, in our treatment, the **generator** of a stochastic process in the form of a chemical reaction network (CRN) connecting literal molecule types by literal reaction types. The third level is a **state space** in which collections of the molecules evolve stochastically under the generator. Middle labels give the mathematical structures that express each level. For rules, they are morphisms from category theory; the commutative diagram is known as a *double pushout*. For reaction networks, the representation is called a *multi-hypergraph* (because the inputs and outputs may have multiple copies of the same molecule type, and so are termed "multisets"). For state spaces in a population process, the states are points in a lattice where the coordinates count molecule copy-numbers. Each level is connected to the next by a (generally) one-to-many generative relation: mechanisms generate CRNs (both molecules and reactions) through *network expansion*, in which the same rule may be instantiated in many different reactions. Sets of transitions from the CRN as a generator are embedded in the state space as paths of population states; the same reaction sequence may have indefinitely or infinitely many images through states with different numbers of molecules. The bottom labels give result-types at each level. For rules, they consist of the algebra of dependencies for activation of a rule on patterns created or eliminated by other rules. For reaction networks, they may be integer-flow solutions to a conversion problem. For state spaces, they are stochastically evolving population states, or distributions over states and their transitions that may evolve deterministically under a master equation. Left and middle panels are reproduced from [11], Fig 6.6 and Fig 12.3 respectively, and other terms used here are explained at length in that dissertation.

the full graph of a molecule. The parts of the input molecules not altered by the reaction mechanism constitute a context for the atomic centers acted on by the rule, which is propagated through the rewrite to produce new output molecules. Structure-preserving embedding is realized in a formal rule-based language as a double-pushout from category theory [11, 14].

By recursive application of all rules from some set to a starting set of seed molecules, a stoichiometric graph can be formed, in which each reaction is an image of the generating rule. Chemical species have no such simple mapping, as each species (except starting inputs) is created through the joint action of the rule as a creator of bond patterns, and embedding of the reacting centers in molecular contexts. The notion of an algebra of rules arises because both the embedding step and the rewriting step may create or destroy instances of graph patterns to which other rules apply, so that the application of rules can generally be non-commutative [77].

Each level-crossing in a three-level system produces a one-to-many map, which may map finite domains to either finite or indefinitely large ranges. Thus each rule in a finite set of mechanisms may be realized in indefinitely many literal reactions in the generated hypergraph, and and each reaction in a hypergraph may govern indefinitely many explicit transitions in the lattice of states.

2.3 Features from stoichiometry: Conserved quantities, null flows, and the stoichiometric subspace

From the stoichiometric matrix $\mathbb{S}$ defined in Eq (8), three important subspaces are defined that govern the conservation and dynamics of matter, currents, and probability, under any stochastic model and at any rate parameters and particle content. Here and below, we will use the term **flow** to refer to any assignment of values to a vector of fluxes $v \equiv [v_{(ji)}]$. The three subspaces of the stoichiometry, together with graph partitioning and hierarchical nesting, will serve as a basis for our treatment of through-flows, linear decomposition of flows and definitions of flow reducibility or irreducibility, and transduction of chemical work in the rest of the presentation.

The set of flow vectors v satisfying

$$
\begin{aligned}
\mathbb{S}\,v &= 0 \\
\Rightarrow \quad \mu\,\mathbb{S}\,v &= 0 \quad ; \ \forall \mu,
\end{aligned}
\tag{18}
$$

are called *null flows* of $\mathbb{S}$. In the second line of Eq (18), μ may be any row vector of species chemical potentials. The null flows are thus non-dissipating.

The set of (row) vectors $c \equiv [c_p]^T$ on species satisfying

$$
\begin{aligned}
c\,\mathbb{S} &= 0 \\
\Rightarrow \quad \frac{d}{dt}\langle cn \rangle &= 0
\end{aligned}
\tag{19}
$$

for $\langle n \rangle$ evolving under Eq (9), giving the component c_p an interpretation as the measure of a *conserved quantity* c of the stoichiometry ascribed to species p.

The image of $\mathbb{S}$, denoted $\mathrm{im}\,\mathbb{S}$, is called the *stoichiometric subspace*, also called the *stoichiometric compatibility class*. We denote its dimension by $s \equiv \dim \mathrm{im}\,\mathbb{S}$, which is the rank of $\mathbb{S}$. It follows that $s + \dim(\ker \mathbb{S}^T)^T = |\{p\}|$, the number of species, and that $s + \dim \ker \mathbb{S} = |\{\langle ji \rangle \mid k_{ji} \text{ or } k_{ij} \neq 0\}|$, the number of reactions.

Any method for obtaining a linearly-independent basis for the space $\ker \mathbb{S}$ gives the stoichiometric generalization of Kirchhoff's laws [61] for decomposition of currents in a closed electric circuit. Because the elements of $\mathbb{S}$ are integers, such a basis can always be found (for

finite networks) in integer components. In Sec. 3.3 we will consider a particular decomposition making use of symmetries among generative rules.

2.4 Graphs and their subgraphs

2.4.1 Conventions and notation for nested graphs and flows. A small number of conventions for defining graph, subgraph, boundary, and complement, summarized next, will enable us to use graph topology together with the rank/nullity decomposition of stoichiometric matrices to specify the admissible source terms to which flows in a graph respond, to define transduction through a graph, and to separate the cases coupled by stoichiometry from those coupled by interior potentials. Our definitions here minimize formalism where possible: see S1 File A 4 a for more technical definitions.

Graph, subgraph, complement, and stoichiometry. From the specification (2) of a graph $\mathcal{G}$ in terms of the set of its reactions, we specify a **subgraph** $\mathcal{G}' \subset \mathcal{G}$ from a subset of the reactions in $\mathcal{G}$. The **complement** to a subgraph $\mathcal{G}' \subset \mathcal{G}$, denoted $(\mathcal{G}')^\perp \subset \mathcal{G}$, comprises the reactions in $\mathcal{G}$ and not in $\mathcal{G}'$. We denote the restriction of a stoichiometric matrix $\mathbb{S}$ on $\mathcal{G}$ to both the species and the reactions in a subgraph $\mathcal{G}' \subset \mathcal{G}$ by $\mathbb{S}' \equiv \mathbb{S}|_{\mathcal{G}'}$.

Boundary and interior of a subgraph. From these conventions for specifying graphs and subgraphs, the natural definition of the **boundary** of a subgraph $\mathcal{G}' \subset \mathcal{G}$, denoted $\partial_\mathcal{G} \mathcal{G}'$, is the set of species with nonzero stoichiometry in both some complex from $\mathcal{G}'$ and some complex from $(\mathcal{G}')^\perp$. Any species with nonzero stoichiometry in complexes from $\mathcal{G}'$ but not its complement is considered interior to $\mathcal{G}'$.

Supporting graphs and boundaries of flows in a graph. A special class of subgraphs of a graph $\mathcal{G}$ are those that contain only the active reactions in one or another flow on $\mathcal{G}$. For a flow v, we will call the **support** of v, or **supporting graph** of v, denoted $\text{supp} v \subset \mathcal{G}$, the subgraph of $\mathcal{G}$ consisting of the reactions on which v is nonzero. Denoting by $J(v) \equiv \mathbb{S}v$ the vector of net currents to all species nodes resulting from flow v, we define the **boundary** of v in $\mathcal{G}$ as $\partial_\mathcal{G} v \equiv \{p \mid J_p(v) \neq 0\}$.

Conservative flow through a subgraph. Flows that are non-null within $\mathcal{G}'$, but are the restriction of null flows in $\mathcal{G}$, will take the place in the constructions below of what [22, 27] term *emergent cycles*. We call these **conservative flows** through $\mathcal{G}'$. Any such flow v may be written as a sum of its restrictions to $\mathcal{G}'$ and the complement $(\mathcal{G}')^\perp$ as $v = v|_{\mathcal{G}'} + v|_{(\mathcal{G}')^\perp}$. By construction it does not change concentrations of species within either $\mathcal{G}'$ or $(\mathcal{G}')^\perp$, meaning that all fluxes are balanced within either subgraph except on its boundary nodes, and the fluxes to those nodes from the subgraph and the complement are likewise balanced.

Feasible transformations from the complement of a subgraph. If a flow v is null in $\mathcal{G}$, its boundary $\partial_\mathcal{G} v \equiv \varnothing$. If it is a conservative flow through a subgraph $\mathcal{G}' \subset \mathcal{G}$, the currents from the two projections $J(v|_{\mathcal{G}'}) = -J\left(v|_{(\mathcal{G}')^\perp}\right)$ will not generally be zero. (Note that the boundaries of the two sub-flows are therefore contained in the boundary of $\mathcal{G}'$: $\partial_\mathcal{G} v|_{\mathcal{G}'} = \partial_\mathcal{G} v|_{(\mathcal{G}')^\perp} \subseteq \partial_\mathcal{G} \mathcal{G}'$). We will term any such source current $J\left(v|_{(\mathcal{G}')^\perp}\right)$ to $\partial_\mathcal{G} \mathcal{G}'$ a **feasible transformation** carried out by $(\mathcal{G}')^\perp$. The feasible transformations by this stoichiometric definition take the place of what [27] call the *chemical transformations*, which the authors define only in terms of chemical element conservation.

2.4.2 Definitions of flow reducibility and irreducibility. We will call a flow v **irreducible** if $\dim \ker \mathbb{S}|_{\text{supp} v} = 0$: removal of any reaction from supp v results in a graph that cannot support the net conversion (if any) performed by v. Note that since stoichiometry is discrete, the

flux components in an irreducible flow are all integer multiples of some common denominator. It then follows as well from Eq (11) that all species degrees of freedom in the stoichiometric space are determined as functions of the fluxes in v and the values of any conserved quantities of the stoichiometry.

If $\dim \ker \mathbb{S}|_{\text{supp } v} > 0$ then v is **reducible**. For any reaction (ji) in the support of a null flow in $\ker \mathbb{S}|_{\text{supp } v}$, the magnitude of that null flow may be chosen to set the flux through (ji) to zero, resulting in a flow on supp $v - (ji)$. In this way the supporting graphs for one or more irreducible flows may be extracted from the supporting graph of a reducible flow. We will call any sequence of such removals of reactions from the supporting graph of a reducible flow v to reach the supporting graph of an irreducible flow a **reduction** of v.

2.4.3 Dimension of a reaction in a null flow. Wachtel *et al.* [27] term the transduction of chemical work between sources and sinks through direct coupling by the stoichiometry "tight coupling", and the alternative, coupling mediated by the buildup and consumption of chemical potential at internal reactions but not constrained to occur at fixed flow ratios, "loose coupling". They provide examples but do not systematize the conditions leading to either case.

To identify those conditions, for any reaction $(ji) \in$ supp v, define the **dimension** of (ji) within v to be $\dim_v(ji) \equiv \dim \ker \mathbb{S}|_{\text{supp } v} - \dim \ker \mathbb{S}|_{\text{supp } v-(ji)}$. $\dim_v(ji)$ counts the number of independent null flows on supp v that cannot be activated without passing through (ji). Transduction of chemical work between two components of a null flow v by non-stoichiometric coupling through a reaction (ji) will be possible only if $\dim_v(ji) > 1$.

2.4.4 Work-dissipation identity for a flow through a subgraph. Chemical work is defined so that the equality between the chemical work delivered from an external source to a graph, and the dissipation by internal reactions within the graph, is an accounting identity. To simplify notation in this section, we let $\mathcal{G}$ stand for a graph through which a conservative flow v passes, and do not introduce an explicit notation for the complement. We suppose that the net conversion $J(v) \equiv \mathbb{S}v$ is feasible in $\mathcal{G}$ and in the (implicit) complement. Chemical work delivered to or extracted from the whole network by fluxes in v are transferred through species in $\partial_{\mathcal{G}} v$.

For μ any vector of chemical potentials on the species in $\mathcal{G}$, and the fact that a source current balancing the flux from v must be $-\mathbb{S}v$, it follows from Eq (14) that

$$\dot{W}_{\partial_{\mathcal{G}} v}(-\mathbb{S}v) = \dot{Q}_{\mathcal{G}}(v). \tag{20}$$

The chemical work delivered to $\mathcal{G}$ by an external source balancing v equals heat dissipated by v in the reactions within $\mathcal{G}$.

Chemical work transduced by a single reaction. Suppose v is a null flow in $\mathcal{G}$, and that $(ji) \in$ supp v is a reaction with $\dim_v(ji) = d$. When a non-equilibrium chemical potential μ is imposed on the species in $\mathcal{G}$ (by any combination of sources and sinks, which need not even be in supp v), we wish to understand the role of the null flow in redistributing chemical work and dissipation around supp v, and of the reaction (ji) in transducing chemical work among species in $\partial_{\mathcal{G}}(ji)$ (the collection of species in the complexes j and i).

By construction, the components $v|_{(ji)^\perp}$ of v outside the reaction (ji) perform a feasible transformation. The chemical work delivered by $v|_{(ji)^\perp}$ to the species in $\partial_{\mathcal{G}}(ji)$, $\dot{W}\left(J\left(v|_{(ji)^\perp}\right)\right) = (\mu^i - \mu^j)v|_{(ji)} = \dot{Q}\left(v|_{(ji)}\right)$, the dissipation within the reaction (ji) due to v.

Suppose that this complement flow can be written as a sum of two flows $v|_{(ji)^\perp} = v_1 + v_2$, both with boundaries $\partial_{\mathcal{G}}(v_1) \subseteq \partial_{\mathcal{G}}(ji)$ and $\partial_{\mathcal{G}}(v_2) \subseteq \partial_{\mathcal{G}}(ji)$; that is: each component v_1 or v_2 is conservative on supp v except where it deposits or withdraws current from species in complexes i or j. Because this decomposition is done in the stoichiometric subspace of $(ji)^\perp$, v_1 and

v_2 are individually also feasible conversions. Since the delivered work $\dot{W}\left(J\left(v|_{(ji)^\perp}\right)\right)$ in linear in $v|_{(ji)^\perp}$ given μ, we have the decomposition $\dot{W}(\mathbb{S}v_1) + \dot{W}(\mathbb{S}v_2) = \dot{Q}\left(v|_{(ji)}\right)$. Work is transduced through reaction (ji) if one of $\dot{W}(\mathbb{S}v_1)$ and $\dot{W}(\mathbb{S}v_2)$ is positive and the other negative.

Chemical transduction efficiency. Suppose that work is transduced, and that $\dot{W}(\mathbb{S}v_1)$ is positive. Wachtel *et al.* [27] define the *transduction efficiency* from v_1 to v_2 through reaction (ji) in the background μ as

$$\eta(v \mid \mu) \quad \equiv \frac{-\dot{W}(\mathbb{S}v_2)}{\dot{W}(\mathbb{S}v_1)} = 1 - \frac{\dot{Q}\left(v|_{(ji)}\right)}{\dot{W}(\mathbb{S}v_1)}. \tag{21}$$

The ordinary application of Eq (21), and the only one considered in [27], is to mass-action flows in which v is the entire current through reaction (ji). In addition to that case, we will consider the role of null cycles in redistributing chemical work and dissipation in cases where there may be other flows through the network, such that $v|_{(ji)}$ is not the entire current through reaction (ji), and the chemical potential drop $(\mu^i - \mu^j)$ across reaction (ji) receives other contributions.

2.4.5 Stoichiometric versus potential coupling defined through flow reduction. For a reaction (ji) with $\dim_v(ji) = d$, let $\{v_1, \ldots, v_d\}$ be the elements in a basis for $\ker\mathbb{S}|_{\mathrm{supp}\, v}$ that satisfy $(ji) \in \mathrm{supp}\, v_k$ (each basis element passes through (ji)). There will then be a set of d real numbers $\{\alpha^1, \ldots, \alpha^d\}$ for which $\Sigma_k \alpha^k v_k = v$ and so in particular $\Sigma_k \alpha^k v_k|_{(ji)} = v|_{(ji)}$.

Again using linearity of $\dot{W}$ given μ, we may decompose the work-dissipation identity (20) in terms of this basis as

$$\sum_{k=1}^{d} \dot{W}\left(\alpha^k \mathbb{S}v_k|_{(ji)^\perp}\right) \quad = \dot{Q}\left(v|_{(ji)}\right). \tag{22}$$

The work term in Eq (22) for the restriction of each basis element v_k to the complement of (ji) in $\mathcal{G}$, evaluates simply to $\dot{W}\left(\alpha^k \mathbb{S}v_k|_{(ji)^\perp}\right) = \alpha^k(\mu^i - \mu^j)v_k|_{(ji)}$.

If $d = 1$, the case of stoichiometric coupling, each of $\dot{Q}$ and $\dot{W}$ in Eq (22) contains a single term. The overall scale of v may be factored out of both expressions, and the ratio (21)— however v is divided into two feasible transformations outside reaction (ji) – will then depend only on the inner products of the vector μ of chemical potentials with parameters from the stoichiometry.

For $d > 1$, in addition to decompositions of the form available when $d = 1$, it is also possible to consider cases in which external v_1 and v_2 are the restrictions of any partial sums of terms in Eq (22) having respectively positive and negative signs, thus writing

$$v_\sigma \quad = \sum_{k=1}^{d} \alpha_\sigma^k v_k|_{(ji)^\perp} \ ; \qquad \sigma \in \{1, 2\}. \tag{23}$$

This is Wachtel *et al.*'s [27] "loose", or non-stoichiometric, coupling. Independent nullity of v_1 and v_2 requires that $\partial_\mathcal{G} v_1 = \partial_\mathcal{G} v_2 = \partial_\mathcal{G}(ji)$, and the projection of their two completions on (ji) gives

$$\sum_{k=1}^{d} (\alpha_1^k + \alpha_2^k)v_k|_{(ji)} \quad = v_{(ji)}. \tag{24}$$

Then the chemical-potential drop $(\mu^i - \mu^j)$ factors out of both work terms in the ratio (21), leaving the efficiency in the form

$$\eta(\nu \mid \mu) \quad \equiv \frac{-\sum_{k=1}^{d} \alpha_2^k \nu_k|_{(ji)}}{\sum_{k=1}^{d} \alpha_1^k \nu_k|_{(ji)}}, \tag{25}$$

independent of the potential landscape μ given the expansions (23) for ν_1 and ν_2.

Remark: complementary regimes of maximum efficiency. For stoichiometrically coupled reactions at $d = 1$, maximum efficiency (21) is attained in the limit of vanishing kinetic barrier $\mu_j - \mu_i \to 0$, relative to the potential drops that determine $\dot{W}(\mathbb{S}\nu_1)$. In the alternative case of fully non-stoichiometrically-coupled transduction at $d > 1$, the efficiency (25) is maximized in the limit of arbitrarily high reaction barrier, where $\nu_{(ji)} \to 0$ in Eq (24) and chemical work is transduced "around" the focal reaction from ν_1 to ν_2.

We will return in Sec. 4.4 to give examples of each of these cases in the context of larger networks. There we will also show how redistribution of chemical potential around networks is responsible for minimization of overall dissipation by mass-action flows. Some further respects in which efficiency measures on chemical networks are more complex and diverse than simple thermal efficiencies are noted in S1 File A 6.

3 Sample combinatorially-generated topology: Sugar-phosphate chemistry from five rules

Here we illustrate the propagation of constraints and symmetries across a three-level system, from its generating rules to the thermodynamic phenomenology captured in its large-deviation functions, in a biochemical network model that is of independent interest as a putative instance of evolutionary combinatorial optimization. The example is drawn from biological sugar-phosphate chemistry, developed as a graph-grammar model in [34]. The uniform stoichiometry $(CH_2O)_n$ of sugars, which requires hydroxyl or carbonyl groups at every carbon center (or bridging oxygen in cyclic forms), leads to a high degree of potential combinatorial complexity in networks generated from very few mechanisms [33, 78].

Sugar chemistry in metabolism is distinctive for both retaining part of this combinatorics and excluding other large parts, suggesting questions about the criteria and complexity of search and optimization problems evidently solved by selection. At the same time, all core biological sugar metabolism is sugar-phosphate metabolism [79], with the positioning and reactions of phosphate groups providing crucial free energy modulation and protection for some carbon centers, so that sugar-phosphate metabolism remains much simpler than unrestricted sugar chemistry [80, 81].

In this section we present the generating rules for a model and a small, computationally-generated reaction network that can be extended to unlimited size by induction on the length of the largest carbon chains. The reaction network contains the two universal biochemical pathways in the Calvin-Benson-Bassham (CBB) cycle [50] and the Pentose-Phosphate Pathway (PPP), which perform the same sugar-phosphate recombination process in reversed orders. We present an exhaustive and ordered enumeration of all integer flows in the network performing this conversion, generated as integer linear programming solutions in the graph-grammar modeling system MØD [11, 14], which grows combinatorially in the network size, but can likewise be systematically extended to arbitrary networks.

We will emphasize the role of discrete *symmetries* respected by the generating rules, and the ways in which these, together with a net conversion such as that of CBB/PPP, constrain the forms of all possible pathways performing the conversion. In particular, the biological Calvin

cycle can be shown to be one of two uniquely minimal solutions under these rules, by a direct graphical construction that bypasses a need for exhaustive enumeration.

We also use symmetries to construct a basis for the null space of the stoichiometry in this system, which we show decomposes into a classification of null cycles that remains complete in networks of any size from these rules. The null cycles presented here will be used later in examples of the information-geometric decomposition of large deviation functions and the assignment of costs to reduction sequences from network flows to irreducible pathways as defined in Sec. 2.4.2, a reduction plausibly performed by natural selection for progressively more substrate-specific enzymes. Finally we will use the null basis elements in examples of both stoichiometric and non-stoichiometric coupling of chemical work that is performed by the process of equilibration to arrive at Onsager's minimum-dissipation solutions [72, 73] for flow through a network.

3.1 System construction from rule declaration and network expansion

3.1.1 A biologically motivated throughput problem. Biological sugar-phosphate metabolism is found in the core of both anabolic and catabolic chemistry. Anabolically, it is the sub-network in the Calvin cycle that prepares ribulose-5-phosphate to receive CO_2 driven by hydrolysis of the main backbone in the enzyme RuBisCO, forming two molecules of 3-phosphoglycerate, which are reduced to glyceraldehyde-3-phosphate to initiate two repeats of the pathway, rendering it autocatalytic. In anabolism, ribulose-5-phosphate is re-arranged to glyceraldehyde-3-phosphate to enter one of the glycolytic pathways of energy metabolism.

The common feature of both pathways is the lossless rearrangement of (CH_2O) groups between chains of two lengths (3 and 5) that have no common divisor. As attested in the fact that both are ancient, very widespread, and near-reverses of each other, they can also operate near equilibrium at physiological substrate concentrations [55] except in the steps of phosphorylation or phosphate hydrolysis (from different cofactors and kinetically controlled), a property that is important for the thermodynamic efficiency of the Calvin cycle with the unusual mechanism of action of RuBisCO [82], and for the yield of chemical work in the Pentose-Phosphate Pathway.

The five reaction mechanisms that support biological sugar-phosphate chemistry are listed in Table 1. Each appears in the rule set from [34] as a pair of bidirectional reactions, which are listed in the table together. (A sixth rule from [34], the phosphorylating cleavage reaction performed by phosphoketolase, is not used as it is not a feature of the sugar-phosphate re-arrangement system responsible for the Calvin-Benson cycle and the standard Pentose Phosphate pathway. It has, however, been found to enable a novel non-oxidative glycolytic pathway in yeast [83]). The list includes one isomerization (aldose-ketose conversion), one condensation (aldolase), one hydrolysis (phosphohydrolase), and two

Table 1. The five rules that generate biological sugar-phosphate conversions, from [34]. The rule name and a brief annotation for its action are shown.

Name	Description
Aldolase	Aldol addition / retro-aldol cleavage
Aldose-Ketose	Aldehyde-ketone interconversion
Phosphohydrolase	Phosphate hydrolysis
Transaldolase	Move C_3 end
Transketolase	Move C_2 end
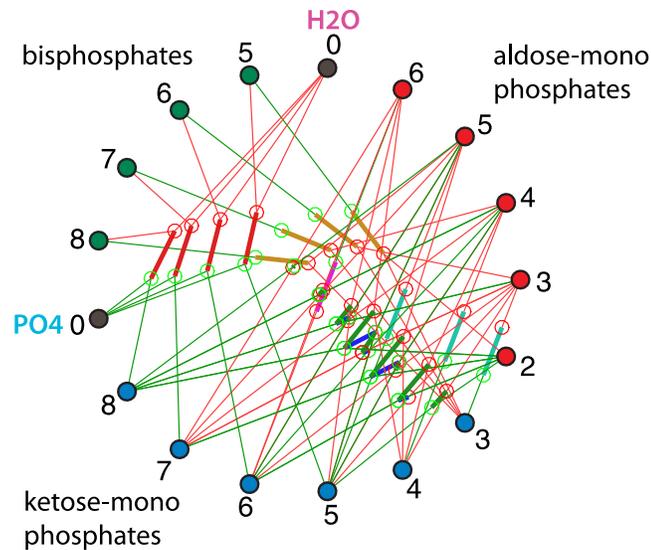
Fig 2. The diagram layout to be used. Filled circles are species in the hypergraph notation following Feinberg [18]. Aldose (red) and ketose (blue) monophosphates, and bisphosphates (green), are arranged in the series shown, and the carbon number of the sugar is indicated in the outer ring. All edges generated for these compounds by network expansion from the rules in Table 1 are shown in doubly-bipartite graph form. Open circles are complexes, heavy lines are reactions, and light lines indicate stoichiometry.

https://doi.org/10.1371/journal.pcsy.0000022.g002

recombining reactions (transaldolase and transketolase) that transfer, respectively, the C_3 or C_2 ends from ketose onto aldose sugars.

The rule system is conservative of (CH_2O) groups and redox-neutral. In the example developed here we omit keto-enol tautomerization, so carbonyl migration within chains and the resulting branched-chain synthesis are not included in the model. (Their effects may be seen in [33]).

A joint program of network expansion with molecule synthesis was carried out using the graph-grammar system MØD with the rules in Table 1, and glyceraldehyde-3-phosphate and water as starting compounds. The resulting network, extended to sugar size C_8, includes 17 molecules and 28 reactions and is shown in Fig 2. Sugar-phosphate nodes are labeled with the carbon-chain length, and are colored and grouped into aldose-monophosphates (red), ketose-monophosphates (blue), and bisphosphates (green). Water and orthophosphate (0-carbon species) are the remaining two nodes in the network (grey).

Because the aldehyde groups on which aldol addition acts are created only at the ends of sugars, all compounds are simple linear sugar-phosphates. The rules in this model do not distinguish stereochemistry, meaning that stereoisomers are treated collectively in each node. These simplifications make it possible to refer to all species with an index notation, by symbols A_n, K_n, or B_n, respectively for aldose, ketose, or bis-phosphorylated sugars of n carbons.

Reactions (hyper-edges) are presented using doubly-bipartite simple graphs (two kinds of nodes and two kinds of links) in which each graphical element corresponds to an element in the analysis of [17, 18], and to a term in the rate equation. Filled circles are species, open circles are complexes; heavy lines are reaction edges, and thin lines show the stoichiometry of the complexes by connecting each complex to the species it comprises. The reference direction for each reaction edge is indicated, in this case, by assigning different colors (green and red) respectively to its input and output complex nodes.

3.2 Integer-flow solutions and their supporting graphs

The reaction schema that defines the universal conversion problem for the PPP and the CBB cycle is

$$5 \, \text{GAP} + 2 \, \text{H}_2\text{O} \rightleftharpoons 3 \, \text{Ru5P} + 2 \, \text{H}_3\text{PO}_4. \tag{26}$$
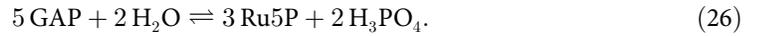
The forward direction characterizes CBB, and the reverse direction PPP.

An integer linear program (ILP) solver was used to enumerate the first 763 solutions $\{f_i\}_{i=1}^{763}$, to $J^{\text{ext}} = \mathbb{S}f_i$ for J^{ext} the net conversion (26). The solutions are exhaustively enumerated in increasing order of the number of reactions used (equal to the number of edges in the supporting graph), which we designate $\#_{\text{reactions}}$, and within each $\#_{\text{reactions}}$, in increasing order of the sum of (absolute) magnitudes of the currents through all reactions. The CBB conversion appears in this list as f_{13}, and canonical PPP as f_1. The two flow solutions are shown on their supporting graphs, with input and output source currents, in Fig 3. The graph annotation over reactions shows the number of times each is activated in the corresponding solution f_i.

Fig 4 shows the values of $\#_{\text{reactions}}$ and $\dot{Q}_{\text{top}}$ from Eq (16) at $v = f_i$ for the 763 ILP solutions. $\#_{\text{reactions}}$ suggests the genomic complexity required to realize a particular flow, and $\dot{Q}_{\text{top}}$ suggests (as we will show later with thermochemical data) the relative free energy costs of different flows to perform the same conversion. Both are plausible as factors in the selection of the biologically attested pathways.

3.2.1 Prelude and fugue structure of all solutions to the conversion problem entailed by the rule algebra. The combinatorics of exhaustively-enumerated integer flow solutions is readily systematized because constraints from phosphate number, aldose number, and backbone length on the application of the rules from Table 1 impose certain common architectures on all solutions.

The first architecture we note is a decomposition of the overall conversion (26) into what we term preludes and fugues. All solutions possess (up to addition of null cycles) a composite reaction that we term a *prelude* of the form 2 AlKe ⊕ PHL˚AL, where AlKe is the triose-
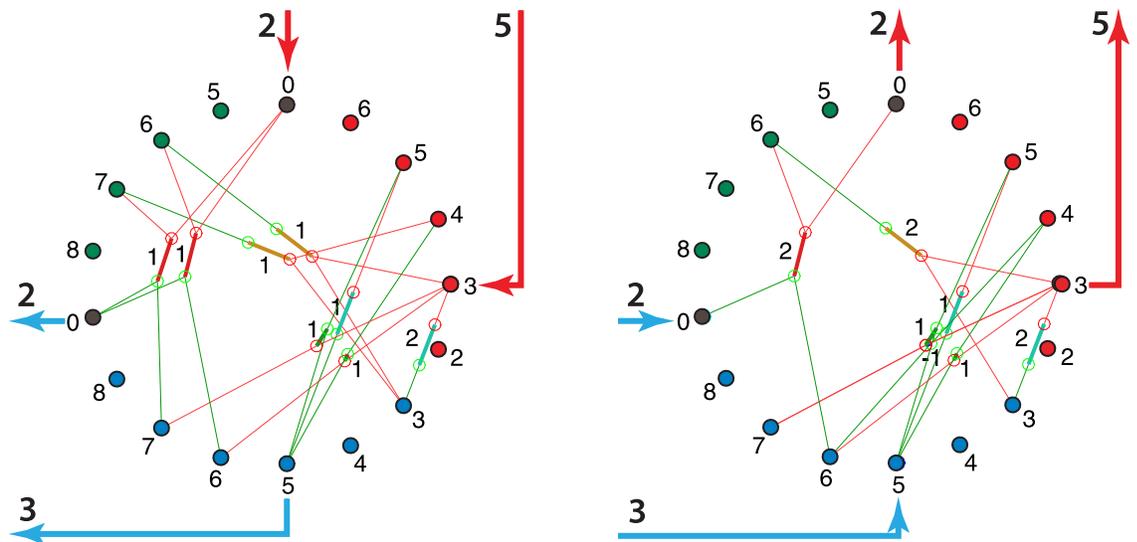
Fig 3. Left panel: the sugar re-arrangement part of the Calvin-Benson cycle as an integer-flow solution v with net currents $J^{\text{ext}} = \mathbb{S}v$ from the environment to the network. Right panel: one version of the canonical Pentose-Phosphate Pathway as an integer flow solution with the equal and opposite J^{ext}.
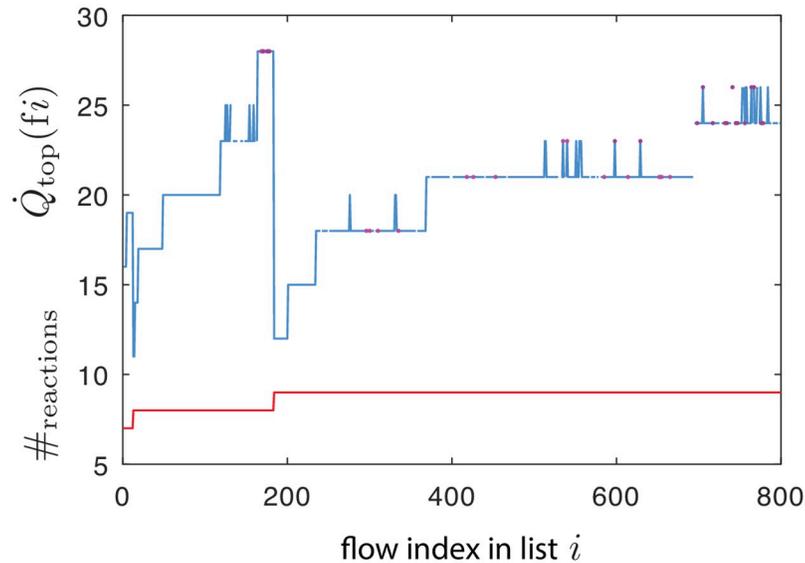
Fig 4. 763 solutions f$_i$ to the throughput problem, showing the number of reactions in the supporting graph f$_i$ **(red) and the topological contribution to dissipation $\dot{Q}_{top}$ (blue and symbols) from Eq (16) evaluated for f$_i$.** 19 integer flows (light magenta dots) produce a graph f$_i$ that appears again as the supporting graph of a second solution (dark magenta dots).

isomerase (TIM) reaction $A_3 \leftrightarrows K_3$ between GAP and DHAP. We use the direct-sum notation $\oplus$ to refer to subsets of reactions that act independently on their respective inputs, and composition $\circ$ to indicate that the reaction PHL takes as its input the output from reaction AL, the two forming together an overall conversion. The aldol condensation may act on any aldose sugar, so the preludes as a class have the schema

$$2\,GAP + 2\,A_n + 2\,H_2O \rightleftharpoons 2\,K_{n+3} + 2\,H_3PO_4. \tag{27}$$
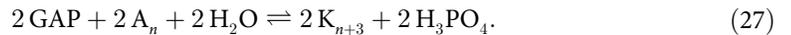
Four "pure" preludes are possible, with n ranging from 2 to 5, which convert two molecules of the same aldose phosphate, plus DHAP, to two molecules of the same bisphosphate. (These are shown in graphical form of Fig 2 of S1 File B 1). Six other "mixed" preludes may be formed, which add 1/2 of two instances of the schema (27) with two different aldose lengths n_1 and n_2. We see that PPP in the right panel of Fig 3 uses a pure prelude with $n = 3$, whereas CBB in the left panel uses a mixed $n = 3/n = 4$ prelude.

If, from the overall reaction schema (26) we extract the prelude schema (27), the remainder is a combination of TAL, TKL, and AlKe/KeAl edges that we term a *fugue* for its typically cyclic form, with the schema

$$3\,GAP + 2\,K_{n+3} \rightleftharpoons 3\,Ru5P + 2\,A_n. \tag{28}$$
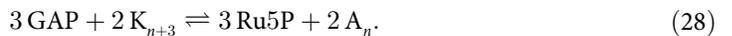
(For mixed preludes, corresponding mixed fugues are the complements).

In the case that $n = 3$ then $A_n \leftrightarrow GAP$ and we may cancel products in the schema (28) to yield

$$GAP + 2\,K_6 \rightleftharpoons 3\,Ru5P.$$

If $n = 2$ then $K_{n+3} \leftrightarrow Ru5P$ and we may cancel reactants in schema (28) to yield

$$3\,GAP \rightleftharpoons Ru5P + 2\,A_2.$$

The remaining diversity of integer flow solutions results from adding integer multiples of null flows to any instance of this prelude-fugue backbone. Some null cycles are supported entirely within fugues, while others may span preludes and fugues. We turn now to a decomposition of the null space making use of rule symmetries.

3.3 The null spaces

We can show (see S1 File B for more detail) that the null space $\ker\mathbb{S}$ for the graph of Fig 2 can be expanded in a basis of elementary flows from only four categories, and moreover that this expansion extends inductively to be complete for generalizations of our example network to any maximal carbon size.

Two of the classes are similar, and differ only in their construction either from TAL or from TKL reactions. They are a category of null flows that we term *braided cycles* for the way they pass a sugar-phosphate end-group along backbones in a "twisted" network topology. (See Fig 7 in S1 File B 2 for elaboration).

All such loops can be reduced to a basis of elementary loops of length three, which we therefore term **trefoils**. We label a trefoil with the carbon lengths of the three backbones that exchange the end-group. Fig 5 shows the "234 TKL trefoil" as an example, first in the graph context from Fig 2, and then emphasizing the symmetry and twist of the reaction connectivity.

A third category (graphed in Fig 3 of S1 File B 1 is closely related to the TAL and TKL trefoils, by the relation that the difference of two preludes (27) beginning with A_n and A_m, is twice the TAL reaction (see Table 2)

$$A_n + K_{m+3} \rightleftharpoons A_m + K_{n+3}. \tag{29}$$
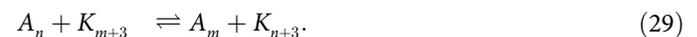
Therefore subtracting the TAL edge (29) from 1/2 the prelude differences results in a TAL-like trefoil in which one of the "backbones" is the liberated ortho-phosphate (a formal place-
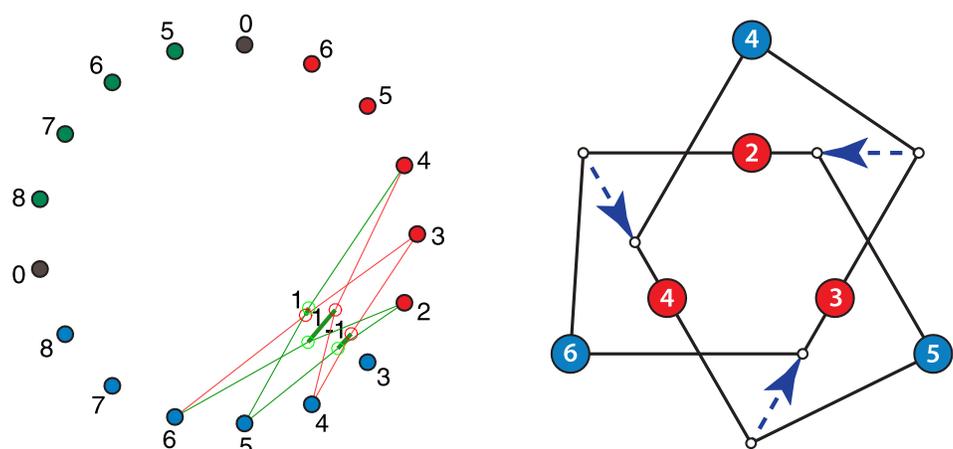
Fig 5. The 234 TKL trefoil shown in two representations. Right-hand side: as a projection from the network from Fig 2. Left-hand side: with the 3-fold symmetry that conserves carbon number and aldose-plus-ketose number exhibited, as well as the Möbius-band topology that makes this a non-degenerate transport cycle.

https://doi.org/10.1371/journal.pcsy.0000022.g005

Table 2. Rule symbols paired with the conversions the rules make on 1-complexes or 2-complexes of sugar-mono-phosphates. Legend: AlKe (aldose-ketose); KeAl (ketose-aldose); TAL (transaldolase); TKL (transketolase); AL (aldol-ase); PHL (phosphohydrolase). In the fifth line, TAL ∘ TKL indicates function composition: the output of a TKL reaction is input to the TAL reaction. In the sixth line, AlKe ⊕ KeAl indicates parallel application, AlKe on input A_n and KeAl on input K_{n+1}.

| reaction	schema
AlKe/KeAl	$A_n \leftrightharpoons K_n$
TAL	$A_n + K_m \leftrightharpoons k_{n+3} + A_{m-3}$
TKL	$A_n + K_m \leftrightharpoons k_{n+2} + A_{m-2}$
AL	$A_n + K_3 \leftrightharpoons B_{n+3}$
PHL	$B_n + H_2O \leftrightharpoons K_n + H_3PO_4$
TAL ∘ TKL	$A_{n+1} + K_n \leftrightharpoons A_n + K_{n+1}$
AlKe ⊕ KeAl	$A_n + K_n + 1 \leftrightharpoons K_n + A_{n+1}$

https://doi.org/10.1371/journal.pcsy.0000022.t002

holder for a "zero-length aldose-phosphate A_0"; see Eq. (B3) in S1 File B 1 for further justification).

The fourth category of null flows follows from the observation that the reaction composition TAL ∘ TKL performs the net conversion $A_{n+1} + K_n \leftrightharpoons A_n + K_{n+1}$ which can be reversed by the direct sum of AlKe ⊕ KeAl reactions at adjacent carbon number. Fig 6 shows a resulting $n = 4/n = 5$ AlKe null cycle as an example.

3.4 Null cycles and nested graphs from the supporting graph of a reducible flow

To further illustrate the use of the null flows in the previous section to produce nested graph sequences, we begin with a reaction network shown in Fig 7, obtained from the union of the supporting graphs of three irreducible flows: f_{14}, f_{18}, and f_{193}. All three use the same mixed, $n = 2/n = 3$ prelude, so null cycles arise only in the fugue sub-network of the union graph.

The graph of Fig 7 supports three independent null cycles: two TKL trefoils and the AlKe null flow from Fig 6. These are shown in the context of the fugue sub-graph (re-arranged to
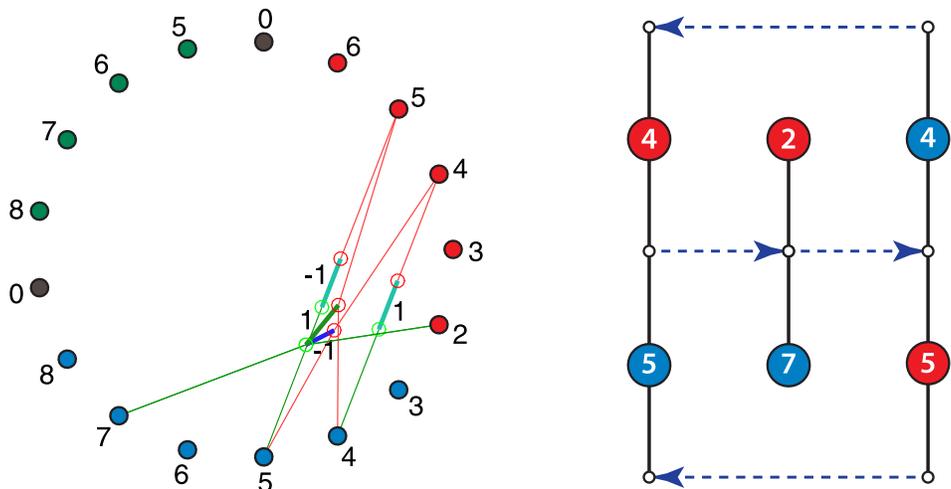
Fig 6. An AlKe null cycle shown two ways. Right-hand side: as a projection from the network from Fig 2. Left-hand side: showing how the TKL ∘ TAL composite reaction balances the direct sum (⊕) of two otherwise-independently feasible AlKe and KeAl reactions.

https://doi.org/10.1371/journal.pcsy.0000022.g006
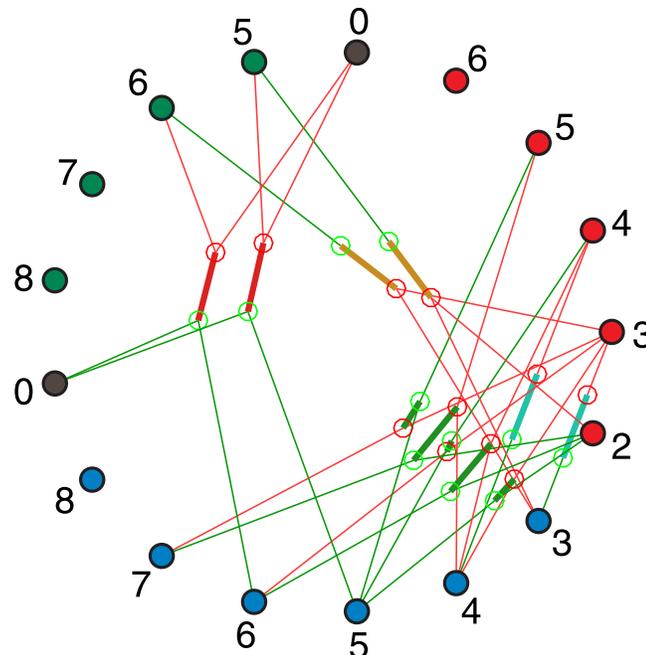
Fig 7. A network formed from the union of the supporting graphs for three irreducible flows f_{14}, f_{18} and and f_{193}.
A subgraph with 8 edges hosts f_{14} uniquely. It is the first lattice diagram in Fig 9 below. The union of that graph with
e98 (the lattice edge from (A_2, K_6) to $(A_4, K_4,)$) hosts f_{535} and f_{629} with one additional TKL trefoil of backbones 234.
The further union with e23 (the lattice edge from (A_3, K_7) to (A_5, K_5)) and e97 (the lattice edge from (A_2, K_7) to (A_5, K_4))) adds a second TKL trefoil with backbones 235 and common edge e96 with the first trefoil. Two firings of the TIM
reaction and single firings of each AL/PHL sequence are fixed by the topology for the conversion (26), so the prelude
on this graph is independent of the background. The 5 TKL edges and the remaining C_4 AlKe edge constitute the
fugue. Like the preludes, the AlKe reaction is topologically constrained to fire one time. The only two degrees of
freedom responsive to the kinetics are the circulations in the two trefoils, illustrated below in Fig 14.

emphasize symmetries) in Fig 8. The 234 TKL trefoil from Fig 5 appears, coupled to a 235 TKL
trefoil through the reaction edge e96 (numbering given to edges in the automated MØD net-
work expansion).

We will return in Sec. 4.4 to use these null flows to study chemical work transduction and
the redistribution of chemical potential in mass-action solutions for reducible versus irreduc-
ible flows. We note here that the the graph of Fig 6 supports only one (the AlKe) null cycle,
and the dimension of each of its reactions is likewise 1. Therefore transduction through this
null flow is always stoichiometrically coupled. In contrast, the supporting graph for the two
TKL trefoils (middle panel in Fig 8) has one reaction of dimension 2 (e96), where non-stoi-
chiometrically-coupled transduction can occur.

3.5 Deriving constraints on flows from rule properties using lattice representations of reactions and flows

Up to now we have used the rule-level of representation only to generate networks, and then
used enumeration to derive bases for the Kirchhoff flow decompositions on those networks,
without further direct appeal to the rules. We show next how properties of rules can be used to
directly derive constraints on flow solutions without the intermediary step of enumeration.

We will derive the minimum value $\dot{Q}_{\text{top}} = 11$ in Fig 4 and the reason exactly two flow solu-
tions are possible from this rule set that achieve that value. The rule properties we will use are
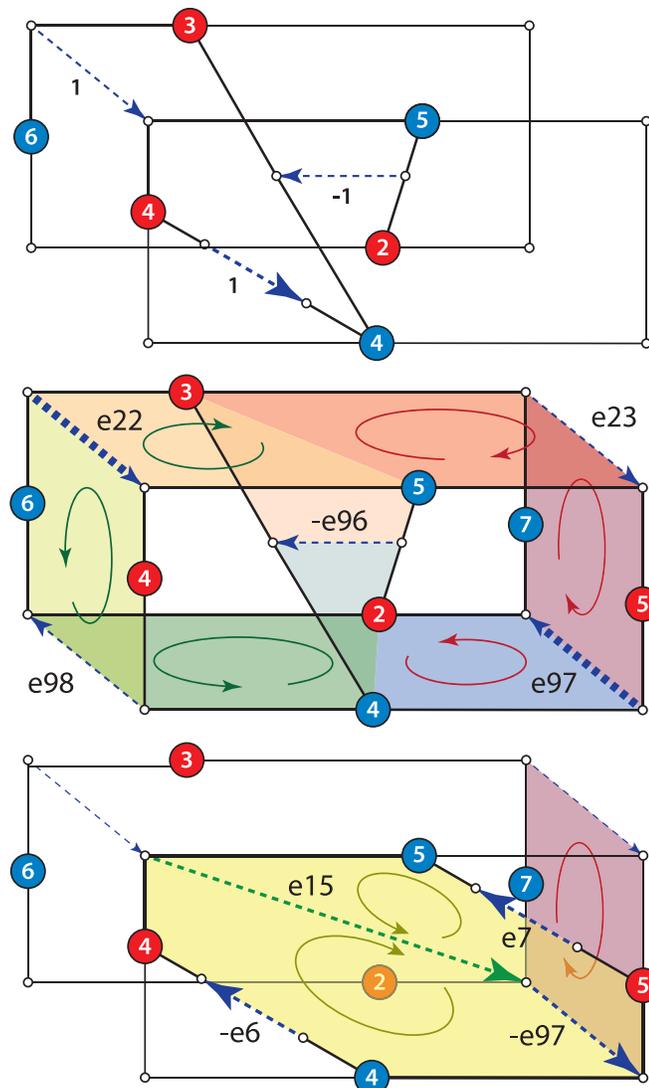
Fig 8. Top panel shows the three reactions in f_{14} that are not part of the prelude creating F6P from GAP + DHAP and Ru5P from GLP and DHAP. Middle panel expands the set of species and reactions to correspond to the supporting graph in Fig 7. A TKL trefoil with aldose backbone lengths 2, 3, 4 covers the faces on the left-hand side of the cube, shown with green circulation arrows. A second TKL trefoil with aldose backbone lengths 2, 3, 5 covers the faces on the right-hand side of the cube, shown with red circulation arrows. The two trefoils are non-stoichiometrically coupled through any potential drop that arises on e96. Arrows are shown in the directions that make trefoils null. The sense of edges in the MØD listing, relative to the drawn arrows, is indicated with ± signs. All currents and potentials will be similarly signed relative to drawn arrow directions. Bottom panel shows the AlKe null cycle of Fig 6 overlaid on the 235 TKL trefoil, with which it shares edge 97. The pair of dark-gold circulations shows the stoichiometrically coupled reaction directions.

discrete symmetries and associated conservation laws, which propagate upward from the rule level through graph generation to emerge as constraints on flow solutions. We will introduce a lattice representation for reactions as an alternative to the hypergraph, on which solutions to the conversion problem $J = \mathbb{S}v$ become simple closed curves on the lattice in place of stoichiometric flows.

3.5.1 Rules as lattice moves. Table 2 shows each of the rules used to generate our network as a mapping of complexes that consist of single species or pairs of species. The 2-species
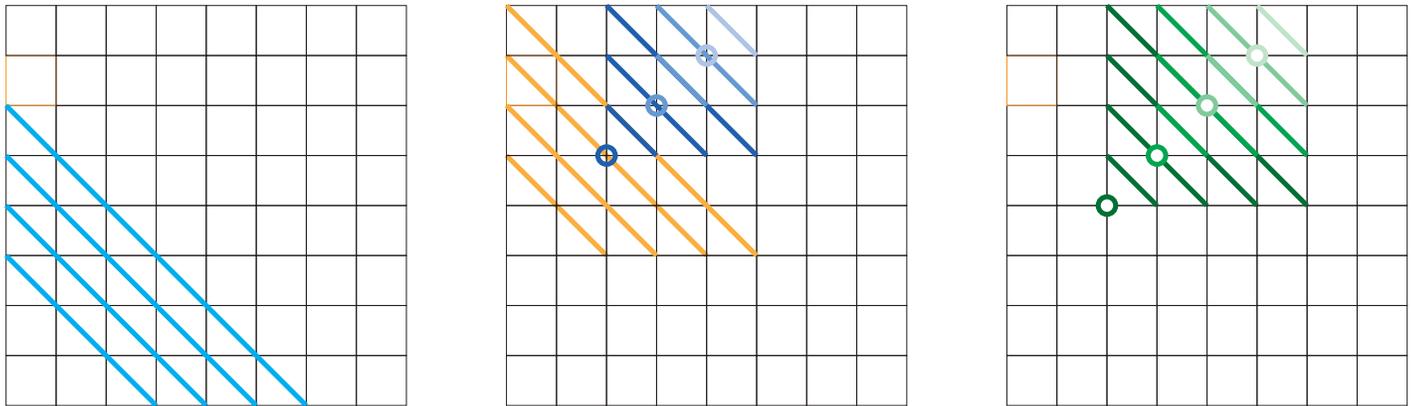
Fig 9. Lattice diagrams for all edges (plus the composite PHL ∘ AL) in null flows. Left panel: AlKe edges (cyan). Center panel: TAL edges (blues) and PHL ∘ AL (orange). Right panel, TKL edges (greens). To aid visibility for edges that overlap, TAL and TKL edges are grouped into "tiers" anchored at fixed points (circles), with darkness distinguishing between tiers. TAL plus PHL ∘ AL and TKL both have 10 distinct edges.

https://doi.org/10.1371/journal.pcsy.0000022.g009

complexes all consist of one aldose and one ketose sugar-monophosphate. We may plot any complex (A_n, K_m) in an integer lattice at coordinates (n, m). 1-complexes occupy the axes with either n or m equal to zero. If we formally treat orthophosphate net of its eliminated water (see Eq. (B3) of S1 File B 1) as the "zero-carbon" aldose-phosphate A_0, the output of the rule composition PHL ∘ AL, a ketose monophosphate with eliminated orthophosphate, may be regarded interchangeably as a 1-complex or the 2-complex with A_0. The maps of Table 2 become links between lattice points, as shown in Fig 9.

It will be important later for proving pathway properties to note the way the rules in Table 2 act on the difference of aldose and ketose counts. Since TAL, TKL, and PHL ∘ AL have both input and output 2-complexes with one aldose and one ketose, the difference between their counts is zero and is left unchanged by the map. Only AlKe changes aldose-minus-ketose count to its opposite. It follows that no flow omitting AlKe edges can accomplish the net conversion (26), which may be written using the formal "zero-carbon aldose" A_0 for phosphate as $5 A_3 \leftrightharpoons 2 A_0 + 3 K_5$, and so requires a net AlKe conversion of 3.

In order to plot solutions as simple closed curves, we need rules to disassemble and assemble complexes. We do this by shifting between 2-complexes and 1-complexes by removal or addition of species. These become vertical or horizontal moves in the lattice diagram. Table 3 gives names to three such lattice maps, which include the input and output of species from J^{ext}, and also the allocation of DHAP (output by the TIM reaction $A_3 \rightharpoonup K_3$ from GAP) in all flows to 2-complexes. Inclusion of GAP (A_3) in a complex is a positive horizontal shift in the lattice, while input of DHAP (k_3) or removal of Ru5P (K_5) are respectively upward and downward vertical shifts.

Table 3. Three "reactions" that interconvert 1-complexes and 2-complexes through the input or output of species in the net conversion schema (26), and to simplify later plots, the species DHAP which is always produced from GAP by AlKe at flux 2.

reaction	lattice map
GAP_in	$(K_n) \rightharpoonup (A_3, K_n)$
DHAP_in	$(A_n) \rightharpoonup (A_n, K_3)$
Ru5P_out	$(A_n K_5) \rightharpoonup (A_n)$

https://doi.org/10.1371/journal.pcsy.0000022.t003

3.5.2 Mapping hyperflows to simple cycles in lattice graphs. Taking the maps in Tables 2 and 3 as the set of available elementary links, we show in Fig 10 the irreducible flow solutions f_{14}, f_{13}, and f_{194}, represented as lattice graphs. Each flow solution is a simple closed curve in the lattice, because it converts all incorporated inputs to eliminated outputs. Moreover, this series of solutions is *minimal* in the sense of using the fewest reactions solving the schema (26) from a seed A_n, entailed by the conservation laws of these rules.

Details of the construction, and some further properties of flow solutions, are given in S1 File D 1. Here we note two particular features of the series in Fig 10.

Lossless carbon shuffling by an algorithm. f_{14}, f_{13}, and f_{194} are the first three members of an infinite series of solutions with a common structure. An overall reaction sequence, shown in the fourth panel of Fig 10, is applied successively to a pair of aldoses A_n and A_{n+1}, taking in one DHAP and then one GAP and eliminating one Ru5P, to increment the count n by 1. After two repetitions, a single AlKe edge followed by one GAP input and one Ru5P elimination returns the compound A_n starting the cycle. The aldose size n serves as a *counting register* incremented by 1 in a twice-executed loop and then decremented by 2 in a reset sequence.

These solutions use no TAL edges and only one AlKe edge each (beyond the flux 2 in the TIM reaction, common to all solutions), thus avoiding duplicate use of any edge. Because none of these operations passes repeatedly through the same edge, the three reactions in the increment loop contribute +3 to $\dot{Q}_{\text{top}}$ in Eq (16), and two reactions in the reset sequence contribute +2 to $\dot{Q}_{\text{top}}$, summing to +8 for all loops in the third panel of Fig 10 at $n \geq 4$. A remaining +4 is contributed to $\dot{Q}_{\text{top}}$ from the twice-used TIM reaction, shown as a simple closed curve through the origin in the final panel of Fig 10. Thus $\dot{Q}_{\text{top}} = 12$ for all solutions in the third panel for $n \geq 4$. (As a corollary, we see why solutions with mixed preludes can reach lower values of $\dot{Q}_{\text{top}}$ than those with "pure" preludes that use two aldol condensations of the
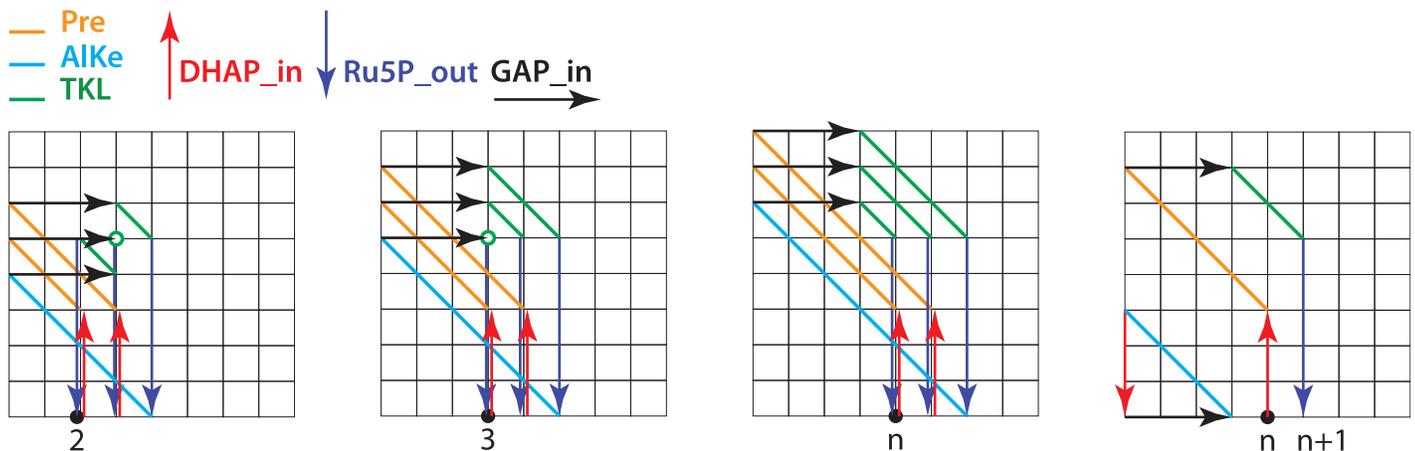
Fig 10. Lattice diagrams for a sequence of complexes that carry one autocatalytic A_n backbone through the sequence of its conversions in one full turn of the CBB cycle. To simplify the diagrams, the inputs are taken to be $3A_3 + 2K_3$ (corresponding to 3 GAP + 2 DHAP). First panel, starting with A_2, is f_{14} from Fig 8 of S1 File C 2. Second panel, starting with A_3 is f_{13} from the left panel of Fig 3. Third panel, starting with A_4 and representative of A_n for any $n \geq 4$, is f_{194}. The links in each path are to be read in the order that makes a closed cycle. The starting aldose is a 1-complex A_n, on the lower axis (black dot). Additions of DHAP are red up arrows, forming the 2-complex input to PHL ∘ AL edges (orange), producing 1-complexes (ketoses) on the vertical axis. Additions of GAP are black right arrows, forming the 2-complex inputs to TKL edges (green). Extractions of Ru5P are blue down arrows, converting 2-complexes back to single aldoses on the horizontal axis, from which new complexes are created by DHAP addition. The sub-flow Ru5P_out ∘ TKL ∘ GAP_in ∘ PHL ∘ AL ∘ DHAP_in, shown in the fourth panel, forms one turn of an "algorithmic" loop incrementing the value n for the starting aldose 1-complex A_n. AlKe conversions (cyan), followed by Ru5P_out ∘ TKL ∘ GAP_in in each of the first three panels, reset two iterations of this algorithmic loop to its starting complex. Also shown in the fourth panel is a 3-cycle of $A_3 \rightarrow K_3$ which, added to the previous graphs, would permit a starting configuration of $5A_3 + 0K_3$, at the cost of slightly greater complexity in plots.

https://doi.org/10.1371/journal.pcsy.0000022.g010

same aldose input. Potentially greater genomic complexity to encode two specific enzymes is exchanged for lower dissipation in the operation of the cycle).

Exactly two minimal and self-seeding solutions. All loops of the kind in Fig 10 at $n \geq 4$ are *exclusively autocatalytic* by the terminology of [84] (see [64, 85–87] for related treatments with other nomenclatures).: they require that one of the aldose or ketose species where the loop contacts the lattice boundaries be provided, in order for the flow to be realizable.

The sole exception to the value $\dot{Q}_{\text{top}} = 12$ for the infinite series in Fig 10 arises for the flows f_{13} and f_{14} shown in the first two panels of the figure. If we consider the full series in descending order of the starting A_n, as n passes through values 3 and 2, exactly one of the two lower the TKL edges passes through the fixed-point complex (A_3, K_5)—there is no reaction to perform – reducing $\dot{Q}_{\text{top}}$ to the value 11. The third TKL edge can never be made to pass through this fixed point, because no C_1 aldose phosphate exists as a realization of the formal label A_1.

As a corollary we see that the two flows f_{13} and f_{14} with $\dot{Q}_{\text{top}} = 11$ in Fig 4 are also the only two in the series from Fig 10 that are not essentially autocatalytic. Any curve passing through the TKL fixed-point complex (A_3, K_5) (GAP_in, then Ru5P_out) can be "re-routed" through the origin (Ru5P_out, then GAP_in); in the language of *siphons* [85], they are "self-priming".

3.6 A thermochemical landscape

To assign relative dissipations or pathway resistances in the regime of linear response near equilibrium, we must identify a thermochemical landscape from which to derive values of the bidirectional equilibrium transition-state currents $\underline{J}^{\ddagger}$ introduced in Eq (13). As a first step, free energies of formation were obtained from the platform eQuilibrator 3.0 [67, 88], which uses sophisticated modern group-contribution algorithms to assign these from molecule descriptors such as SMILES or INCHI representations, which are calibrated against database values from KEGG [89–91].

The left-hand panel in Fig 11 shows carbon concentrations in equilibrium distributions computed for $\Delta G'^0$ values from eQuilibrator for both non-stereochemical and stereochemical
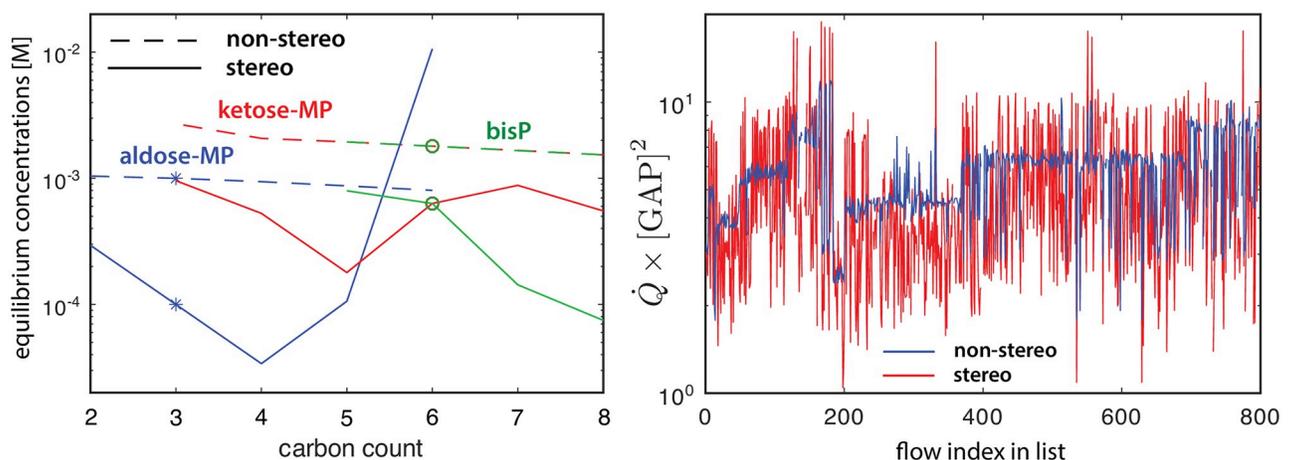
Fig 11. Left-hand panel: an equilibrium concentration profile from both non-stereochemical and stereochemical formation free energies, with hydrolyzing potential [H3PO4]/[H2O] set to ensure [F16BP]/[F6P] $\equiv 1$ (dark green circles), and a geometric decay per carbon determined by [GAP] to make concentrations within a group roughly independent of chain length. Right-hand panel: the values $\dot{Q}$ from Eq (15) at the thermodynamic landscape in the upper panel, showing the qualitative similarity to $\dot{Q}_{\text{top}}$ retained but also the sensitivity to stereochemical corrections.

SMILES representations of the molecules. (Currently MØD does not retain stereochemistry in automated network expansions; therefore in automated applications of networks to arbitrary size, only the non-stereochemical estimates can be used consistently). For non-stereochemical SMILES it is possible to choose conserved quantities that give nearly-uniform concentrations for sugar-phosphates of a common type (aldose/ketose/bisphosphate), qualitatively resembling physiological conditions [92, 93]. Comparable but somewhat less uniform equilibria exist for stereochemical SMILES.

To assign kinetic parameters we adopt a limiting case of "barrierless" kinetics, in which the half-reaction rate constant from the complex with lower activity is set to a constant value over all reactions that we take to be 1 in appropriate rate units. The half-reaction rate constant from the complex with higher activities is then the equilibrium constant. This assumption is equivalent biologically to the assumption that all enzymes have evolved turnover rates up to the diffusion limit for substrates to the enzymes.

(For context, we note that factoring out independent kinetic parameters from formation free energies is the complementary limit to the differential analysis of distributed control by these parameters in *Metabolic Control Analysis* (MCA) [59]. It is most appropriate for near-equilibrium reactions where elasticities tend to be large and control parameters small—so for the sugar chemistry in this model but not for phosphorylation and dephosphorylation which in metabolism are mainly kinetically controlled. As kinetic constraints will in almost any case reflect the interaction of molecular mechanisms [94, 95] with selection, useful application of MCA will generally require dedicated and *ad hoc* treatments).

From these values the near-equilibrium dissipations (15) at unit conversion from schema (26) are shown in the right-hand panel of Fig 11. Semiquantitative agreement of $\dot{Q}(\nu)$ with $\dot{Q}_{\text{top}}$ from Eq (16) is seen for the non-stereochemical landscape, as expected from its uniform distribution. The qualitative dependence on network topology remains visible as a background for the stereochemical landscape, but variation with details of flow topology becomes comparable to $\dot{Q}_{\text{top}}$ differences.

From the resulting thermodynamic and kinetic landscape, we compute steady-state solutions under the mass-action rate law (7, 9) in examples of network dissipation and chemical work transduction below.

4 The information associated with large deviations at single instants and through time

We have so far shown how topology can *constrain* properties of flows through different graphs that all perform the same net conversion. Separately, from exhaustive enumeration of flows and their supporting graphs, we have shown how dissipation responds to flow topology and thermochemical context. However, such methods offer no direct comparison between mass-action flows on different graphs, nor do they give us a principled relation between modifying a graph by eliminating reactions (*e.g.* by the evolution of specificity within a family of catalysts for the same reaction mechanism), and the performance and cost of networks thus modified.

Here we will derive such relations using *Large-Deviation Functions* (LDFs) and the geometries on spaces of potentials and flows that the LDFs naturally induce [42, 43, 65]. LDFs arise in driven networks as a group of *information divergences* [30, 31]. They include as well-known special cases the thermodynamic potentials of equilibrium systems [46, 48], but they generalize much more widely: to currents in driven systems [41–43] or even systems lacking reaction reversibility [31], from ensembles of states to ensembles of histories [49, 96], and to multi-level or multi-scale population processes including biological evolutionary processes [28–30].

We will provide here only a brief derivation and summary of the most-essential concepts needed to compute and interpret information divergences, and the particular forms for currents in driven systems that we will use to understand steady-state flows on graphs. Didactic treatments are widely available [28, 31, 49, 97–102], and we provide a more detailed and explanatory treatment in S1 File F.

We introduce basic ideas for distributions over states at a single instant of time, where the crucial relations of divergences to biased sampling and log-likelihood as a measure of information require the least construction to explain. We then generalize to driven currents, not as a parallel independent application as in [42, 43], but as a simplifying case within the full generalization from distributions over states to those over histories.

Our main result will be that mass-action flows on nested subgraphs furnish an additive decomposition of log-likelihood, within a single, nested partition function for biased sampling from equilibrium reference distributions. Through this decomposition we assign specific and principled cost measures to changes of graph topology by reaction elimination. We use these results also to understand the role of work transduction to arrive at mass-action flows as solutions that are both minimum-dissipation and minimum-LDF conditioned on topology and through-flow.

4.1 Generating functions and Legendre duality for fluctuations at a single time

For a distribution ρ_n at a single time, its cumulant-generating function (CGF) $\psi(\theta)$ for the number vector n is defined as

$$e^{\psi(\theta)} \equiv \sum_n e^{\theta n} \rho_n. \tag{30}$$

Here $\theta \equiv [\theta_p]$ is a (row) vector indexed by the species index p of n, and θn is the Euclidean inner product.

The gradient of the CGF is the expectation of n in the distribution sampled with an exponential bias function $e^{\theta n}$:

$$\frac{\partial \psi}{\partial \theta} = n(\theta) \equiv \sum_n n \, \rho_n^{(\theta)}. \tag{31}$$

$\psi(\theta)$ is convex in θ, and therefore $n(\theta)$ is invertible to a function $\theta(n)$. We denote the Legendre transform with inverted argument n by

$$S(n) \equiv \left. \left\{ \theta n - \psi(\theta) \right\} \right|_{\theta(n)}. \tag{32}$$

By construction the gradient of Eq (32) gives the inverse function

$$\frac{\partial S}{\partial n} = \theta(n). \tag{33}$$

Properties and interpretations of $S(n)$ as an information divergence are reviewed in S1 File F 1 a. We note here only that, for ρ_n having support at population counts $n_p \gg 1$, and suitably approximated from the CGF by saddle-point methods in the *Gärtner-Ellis theorem* [46], $S(n)$ takes on the interpretation of the *Large-Deviation Function* (LDF) for fluctuations of n, and ρ_n

is approximated to leading exponential order as

$$\rho_n \sim e^{-S(n)}. \tag{34}$$

In Eq (34) $S(n)$, as an approximation to $-\log \rho_n$, is an instance of a *Hartley information* [103], establishing a basic relation between divergence measures and the log-likelihood meaning for information. In the same approximation, Eq (33) shows that the gradient of log-likelihood at any n is the sampling bias vector θ needed to set $n(\theta) = n$.

It follows further from Eq (33) and inverse relation (31) giving $n(\theta)$ from ψ, that $S(n)$ can be written as an integral in the positive-definite Hessian of ψ, as

$$
\begin{aligned}
S(n) &= \int_{n_0}^{n} \theta \, dn' \\
&= \int_0^{\theta} \theta' \left. \frac{\partial^2 \psi}{\partial \theta \partial \theta^T} \right|_{\theta'} d\theta'^{T},
\end{aligned}
\tag{35}
$$

where we denote by $n_0 \equiv \Sigma_n \, n\rho_n$, the mean without biasing. We will reference the form (35) where we return below to its counterpart in a log-likelihood *rate* for systems characterized by a steady-state current deviation.

4.1.1 Hamilton-Jacobi equation and integral form for the one-time potential. For a distribution ρ evolving in time under Eq (1), the LDF $S(n)$ can be shown [31, 70] (see S1 File F 2 for details) to evolve under an equation known as the *Hamilton-Jacobi equation* from dynamical systems [1]:

$$\frac{\partial S}{\partial t} = \mathcal{L}(\theta(n), n). \tag{36}$$

The function $\mathcal{L}$ of Legendre-dual variables θ and n, for a generator of the form (4), will become [31, 39, 40]

$$\mathcal{L}(\theta, n) = \sum_{ji} k_{ji} \left(1 - e^{\theta(y^j - y^i)} \right) n^{y^i}. \tag{37}$$

The existence of a Hamilton-Jacobi equation for S (and thereby also for its Legendre transform ψ) gives an integral expression for the CGF of the form

$$\psi_T(\theta_T) = \int_0^T dt \left\{ \dot{\theta} n - \mathcal{L}(\theta, n) \right\} + \psi_0(\theta_0), \tag{38}$$

in which the integral is evaluated along time-dependent functions $\theta(t)$, $n(t)$ that are solutions to dynamical equations obtained from vanishing of the first variational derivative:

$$
\begin{aligned}
\frac{dn}{dt} + \frac{\partial \mathcal{L}}{\partial \theta} &= 0 \\
-\frac{d\theta}{dt} + \frac{\partial \mathcal{L}}{\partial n} &= 0 \\
n_0 &= \frac{\partial \psi_o}{\partial \theta_0}.
\end{aligned}
\tag{39}
$$

The gradient condition (31) is satisfied by the integral (39), and under Legendre transform the corresponding integral representation for the LDF becomes

$$
\begin{aligned}
S_T(n_T) \quad &= \theta_T n_T - \psi_T(\theta_T) \\
&= \int_0^T dt \{ \theta \dot{n} + \mathcal{L}(\theta, n) \} + S_0(n_0),
\end{aligned}
\tag{40}
$$

which may be checked to satisfy Eqs (33) and (36). The integral form (40) is known as *Hamilton's Principal Function* or the *action* functional [1] which motivates our designation by S.

Whereas the original construction (35) for $S(n)$ assigns a log-likelihood only to a single-time deviation, the identification of this likelihood with $S_T(n_T)$ in Eq (40) attaches to that single-time condition, through the stationarity equations (39) for which it provides a final-time boundary condition, a least-improbable history to have arrived at value n_T under conditions of free evolution of ρ under $\mathcal{L}$.

Remark: distinct roles of the CGF and LDF. We have given the two integral forms (38) and (40) for ψ and $\mathcal{L}$ respectively, despite their elementary relation through the Legendre transform, because they play distinct roles in the construction of Hamilton's Principle function as the LDF. The variational equations (39) are derived from the role of ψ as a CGF under biased sampling of n, while it is the evaluation of $S_T(n_T)$ on any such trajectory solution that assigns a value to the LDF. When both integrals are constructed for free evolution with only a final-time condition n_T, the variational equations from either integral are the same. We will turn next, however, to the more general problem of evaluating log-likelihood for *trajectories* that deviate from those of free evolution, and for currents as well as configurations. In the solution of that problem, the variational equations from ψ and from S will differ, but their roles in assigning large-deviation values will remain as we have described.

We note for later use that the θ-gradient of $\mathcal{L}$ appearing in Eq (39) for dn/dt evaluates to

$$
-\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{(ji)} (y^j - y^i) v_{ji} = \mathbb{S} v,
\tag{41}
$$

where $v_{ji} = e^{\theta(y^j - y^i)} k_{ji} n^{y^i} - e^{-\theta(y^j - y^i)} k_{ij} n^{y^j}$ is a current of mean-field form (11) with n substituting for $\langle n \rangle$, under half-reaction rate constants biased by the gradient of the log-likelihood of ρ at the point through which n passes.

4.2 Conditioning along the course of trajectories: cumulant-generating and large-deviation functionals

To compute log-likelihoods for trajectories, rather than bias sampling at a single final time, we bias the event probabilities under which systems evolve. On a non-equilibrium process we can construct generating functionals and large-deviation functionals both for counts n as before and also for reaction fluxes v. To bias sampling for reaction events we modify the Liouville representation (37) for the generator to a form derived in [41]:

$$
\tilde{\mathcal{L}}_\eta(\theta, n) \equiv \sum_{ji} k_{ji} \left(1 - e^{\theta(y^j - y^i) + \eta_{ji}} \right) n^{y^i},
\tag{42}
$$

in which $\{\eta_{ji}\}$ are independent parameters (potentially time-dependent) on all unidirectional reactions. (Our Eq (42) corresponds to the negative of the Hamiltonian derived as Eq. (37) of [41]. The construction is similar to the potential in Eq. (12) of [43], though the parametrization of those families deals directly with currents as single-time values, rather than embedding them in a context of generating functionals for extended-time histories of concentrations and

currents, as in the approach here). To simplify the treatment here, we only consider biasing parameters that are antisymmetric on bidirectional reactions, meaning that $\eta_{ji} \equiv -\eta_{ij}$.

In the presence of continuous-time biasing of event probabilities, the stationarity conditions that replace Eq (39) become

$$
\begin{aligned}
\frac{dn}{dt} + \frac{\partial \tilde{\mathcal{L}}_\eta}{\partial \theta} &= 0 \\
-\frac{d\theta}{dt} + \frac{\partial \tilde{\mathcal{L}}_\eta}{\partial n} &= 0 \\
n_0 &= \frac{\partial \psi_o}{\partial \theta_0}.
\end{aligned}
\tag{43}
$$

$\partial \tilde{\mathcal{L}}_\eta / \partial \theta$ continues to satisfy a relation of the form (41), but in place of the reaction fluxes under free evolution, under the generator $\tilde{\mathcal{L}}_\eta$ these become

$$
v_{ji} = -\frac{\partial \tilde{\mathcal{L}}_\eta}{\partial \eta_{ji}} = e^{\theta(y^j - y^i) + \eta_{ji}} k_{ji} n^{y^i} - e^{-\theta(y^j - y^i) - \eta_{ji}} k_{ij} n^{y^j}.
\tag{44}
$$

Under continuous-time biasing, v is the Legendre dual variable to η, depending not only point-wise in time, but on the full functional form of $\eta(t)$ through the variational Eq (43). A functional counterpart to the single-time construction, given in S1 File G, shows that the trajectory large-deviation function*al*, now with a trajectory $v(t)$ as its functional argument (which we indicate with a square bracket), can again be written in the integral form of Hamilton's principal function as

$$
S_T(n_T; v] = \int_0^T dt \left\{ (\theta \mathbb{S} + \eta) v + \tilde{\mathcal{L}}_\eta(\theta, n) \right\} + S_0(n_0).
\tag{45}
$$

The functional form (45) applies for arbitrary time-dependent η and v. Explicit time-derivative terms that were present in the integral (40) for n_T alone have been absorbed into the expression $(\theta \mathbb{S} + \eta) v$ through the variational equations (43).

The first variational derivative of S_T, a single-time variation with respect to n_T and an extended-time functional variation with respect to v, can be shown to evaluate to

$$
\frac{\partial S_T(n_T; v]}{\partial n_T} = \theta_T ; \quad \frac{\delta S_T(n_T; v]}{\delta v_t} = \eta_t,
\tag{46}
$$

establishing that S_T is a potential for the dual variables to n_T and v.

Steady-state currents in detailed-balance networks. Our biochemical examples concern steady-state flows in networks with detailed-balance rate constants (microscopic reversibility). The restriction of Eq (45) to the case of time-independent sources η that create deviations through sequences of steady states in such networks, which we designate by $\underset{\text{SS,DB}}{\rightarrow}$, is developed in S1 File G 3, and evaluates as a steady-state limit of Eq. (G44) to

$$
\begin{aligned}
S_T[n, v] &\underset{\text{SS,DB}}{\rightarrow} \int_0^T dt \int_0^v \eta(v') dv' + S_0(n_0) \\
&= \int_0^T dt \int_0^{\eta(v)} \eta' \left. \frac{-\partial^2 \tilde{\mathcal{L}}_\eta}{\partial \eta \, \partial \eta^T} \right|_{\eta'} d\eta'^T + S_0(n_0).
\end{aligned}
\tag{47}
$$

In the second line of Eq (47) we have used the relation (44) to write the *integrand* of $S_T[n, v]$ as an integral $d\eta'$ in the Hessian of $\tilde{\mathcal{L}}_\eta$, paralleling the construction for $S(n)$ in terms of the Hessian of the CGF ψ in Eq. (35). Thus we give a direct interpretation of the large-deviation functional for persistent currents: just as $S(n)$ is a negative log-likelihood and Hartley entropy for deviations of states from their most-likely value, the negative log-likelihood for a trajectory that maintains a current away from the mass-action value under the free generator accumulates additively (the unlikelihood accumulates multiplicatively) at a rate given by the integrand in Eq (47).

We will use the form (47) to assign log-likelihood costs to any profile of currents on a graph, first as deviations of *any* steady-state current from the equilibrium, and then as deviations of one steady-state current from another. Of particular interest will be the deviation of a more-restricted mass-action solution on a subgraph $\mathcal{G}' \subset \mathcal{G}$ from the less-restricted mass-action solution on $\mathcal{G}$.

Flows induced by external sources. S1 File G 4 shows how external sources may be added to a cumulant-generating functional to produce through-flows in a graph. The result is a modification of the stationarity equations (43), but the form of the large-deviation functional is unchanged from Eq (45) which reduces to Eq (47) in the steady state on detailed-balance networks. Moreover, if a mass-action solution v to $J^{\text{ext}} = \mathbb{S}v$ induces a steady-state concentration profile n on some graph, no bias parameter η is needed and the θ values are those that produce n as a deviation from the corresponding equilibrium $\underline{n}$, as computed in Sec. 4.1.

4.3 An additive decomposition of large-deviation functions for currents from the geometry of dissipation

The most general state for a flow v on a graph can be produced with a combination of external sources J^{ext} and internal biasing weights η. We show in S1 File G 6 that for any such flow, the integrand in Eq (47) satisfies a decomposition known as the *extended Pythagorean theorem* in Information Geometry [65, 104]:

$$
\begin{aligned}
\tilde{\mathcal{L}}_{\theta\mathbb{S}+\eta} - (\theta\mathbb{S}+\eta)\frac{\partial\tilde{\mathcal{L}}}{\partial\eta}\bigg|_{\theta\mathbb{S}+\eta} &= \left[\tilde{\mathcal{L}}_{\theta\mathbb{S}} - \tilde{\mathcal{L}}_0\right] - \theta\frac{\partial\tilde{\mathcal{L}}}{\partial\theta}\bigg|_{\theta\mathbb{S}} \\
&\quad + \left[\tilde{\mathcal{L}}_{\theta\mathbb{S}+\eta} - \tilde{\mathcal{L}}_{\theta\mathbb{S}}\right] - \eta\frac{\partial\tilde{\mathcal{L}}}{\partial\eta}\bigg|_{\theta\mathbb{S}+\eta}.
\end{aligned}
\tag{48}
$$

The left-hand side of Eq (48) behaves like the squared length of the hypotenuse of a right triangle, and the two terms on the right-hand side like the squares of the base and rise in Euclidean geometry. Making use of the Hessian integral form in the second line of Eq (47), we regard the left-hand integral as occurring over a path from no current ($\eta' = 0$) to the the arbitrary flow v at $\eta' = \theta\mathbb{S} + \eta$. That integration contour can be broken into two legs—from 0 to $\theta\mathbb{S}$ and from $\theta\mathbb{S}$ to $\theta\mathbb{S} + \eta$—in such a way that these two legs are orthogonal under the natural geometry induced by the Hessian of $\tilde{\mathcal{L}}_\eta$.

The first line on the right-hand side of Eq (48), the integral from 0 to $\theta\mathbb{S}$, corresponds to a contour of mass-action solutions built up in response to increments $dJ^{\text{ext}} = \mathbb{S}\, dv_{\text{MA}}$ leading from zero to the final value $J^{\text{ext}} = \mathbb{S}v_{\text{MA}}$ at $\theta\mathbb{S}$. The second line of Eq (48), the integral from $\theta\mathbb{S}$ to $\theta\mathbb{S} + \eta$, involves only addition of null flows that integrate to $v - v_{\text{MA}}$. This integral is performed keeping fixed $J^{\text{ext}} = \mathbb{S}v$. Because null flows result in no dissipation against any chemical potential (as we noted in Sec. 2.4.4), the possible cross-terms in writing the left-hand side of

Eq (48) as two summands vanish, and only the remaining terms on the right-hand side are nonzero.

The individual orthogonal legs of these large-deviation integrands can then be written as the integrals of a quadratic form, as we did in Eq. (35) for the single-time case, along the contours for θ and η identified above:

$$
\begin{aligned}
\left[\tilde{\mathcal{L}}_{\theta\mathbb{S}} - \tilde{\mathcal{L}}_0\right] - \theta\frac{\partial\tilde{\mathcal{L}}}{\partial\theta}\bigg|_{\theta\mathbb{S}} &= \int_0^{J^{\text{ext}}} \theta \, dJ^{\text{ext}\prime} \\
&= \int_0^\theta \theta' \frac{dJ^{\text{ext}}}{d\theta'} d\theta' \\
\left[\tilde{\mathcal{L}}_{\theta\mathbb{S}+\eta} - \tilde{\mathcal{L}}_{\theta\mathbb{S}}\right] - \eta\frac{\partial\tilde{\mathcal{L}}}{\partial\eta}\bigg|_{\theta\mathbb{S}+\eta} &= \int_0^v \eta \, dv' \\
&= \int_0^\eta \eta' \frac{dv}{d\eta'} d\eta'.
\end{aligned}
\tag{49}
$$

Corollary: Because both terms in the Pythagorean decomposition (49) are non-negative, we identify the source $\theta\mathbb{S}$ and its associated flow v_{MA} as the minimum-large-deviation solution on the constraint surface $J^{\text{ext}} = \mathbb{S}v$ for a given graph.

4.3.1 Triangle inequalities between graphs and subgraphs. We may apply the Pythagorean decomposition to the problem of cost functions for nested subgraphs created by sequential removal of reactions. Consider a series of nested subgraphs $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{G}$ where the conversion J^{ext} is feasible in all three. Mass-action flow solutions will exist in both subgraphs, and we can write each in the form of the upper equation of Eq (49) with respect to its own subgraph. A mass-action solution in $\mathcal{G}_2$ also exists as a *flow* solution to $J^{\text{ext}} = \mathbb{S}v$ in the larger graph $\mathcal{G}_1$, but there it will generally differ from the mass-action solution in $\mathcal{G}_1$ by some null flow in $\mathcal{G}_1$.

The extended Pythagorean theorem (49) for the two integration paths to reach the same final current profile then requires that

$$
\int_0^{\theta_2(J^{\text{ext}})} \theta_2' \frac{dJ^{\text{ext}}}{d\theta_2'} d\theta_2' = \int_0^{\theta_1(J^{\text{ext}})} \theta_1' \frac{dJ^{\text{ext}}}{d\theta_1'} d\theta_1' + \int_0^\eta \eta' \frac{dv}{d\eta'} d\eta'
\tag{50}
$$

in which the two integrals on the left-hand side are performed in the subgraph $\mathcal{G}_1$ and the integral on the right-hand side is performed in the subgraph $\mathcal{G}_2$.

The second integral on the right-hand side of Eq (50) is the log-likelihood cost associated with removing reactions from $\mathcal{G}_1$ to restrict flows to $\mathcal{G}_2$. By continuing this procedure we may assign costs to all edge removals in the reduction sequence from Sec. 2.4.2, from any reducible flow to any irreducible flow within the supporting graph of the reducible flow.

4.3.2 Example from the sugar-phosphate network. In the regime of linear response, minimum large-deviation equates with minimum dissipation, given by the quadratic form (15). The extended Pythagorean theorem then reduces to the ordinary Euclidean Pythagorean theorem.

We illustrate the decomposition for the sugar-phosphate example with two transects connecting irreducible flows interpolated by single null flows. The integer solutions f_{14}, f_{193}, and f_{184} are shown in Table 4, together with the mass-action solutions on pair-wise union graphs, in the non-stereochemical thermodynamic landscape from Fig 11. Here $\mathcal{G}_1$ is the union of supporting graphs for either pair of irreducible flows, and $\mathcal{G}_2$ may be the supporting graph for

Table 4. Flow properties of the Pythagorean theorem illustrated in Fig 12. Edges in the graph of Fig 2 are listed in the first column, and fluxes v at through current $J^{\text{ext}} = \mathbb{S}v$ are given in columns for each flow solution. $\dot{Q}$ in the final row is multiplied by $[\text{GAP}]^2$ in the thermochemical background where the force-flux relation is solved. The trefoil current v° (the only null flow in the union graph) equals the flux in edge e23. At the minimum-dissipation solution $v^{\circ}_{\min} = 0.2956$. The relation (50) is fulfilled with the dissipation in the final row $\dot{Q}(v^{\circ}) \approx \dot{Q}_{14\cup193} + 1.6709 \times (v^{\circ} - v^{\circ}_{\min})^2$.

edge	f_{14}	$f_{14\cup193}$	f_{193}	$f_{193\cup184}$	f_{184}
e5	2	2	2	2	2
e1	1	1	1	1	1
e69	1	1	1	1	1
e40	1	1	1	1	1
e182	1	1	1	1	1
e22	1	1	1	1	1
e96	1	0.7044	0	0	0
e23	0	0.2956	1	1	1
e97	0	0.2956	1	0.5000	0
e6	1	1	1	0.5000	0
e7	0	0	0	0.5000	1
e15	0	0	0	0.5000	1
$\dot{Q}$	1.7796	1.6336	2.4626	2.1739	2.4627

either endpoint. The two integrals $\int d\theta'_{1,2}$ in Eq (50) give $\dot{Q}/2$ respectively in $\mathcal{G}_1$ and in either choice for $\mathcal{G}_2$. Fig 12 plots $\sqrt{\dot{Q}}$ for the mass-action flow $f_{14\cup193}$ and for the two irreducible flows, which lie one unit apart in the coefficient of the null flow, which is the 235 TKL trefoil shown in the second panel of Fig 8.
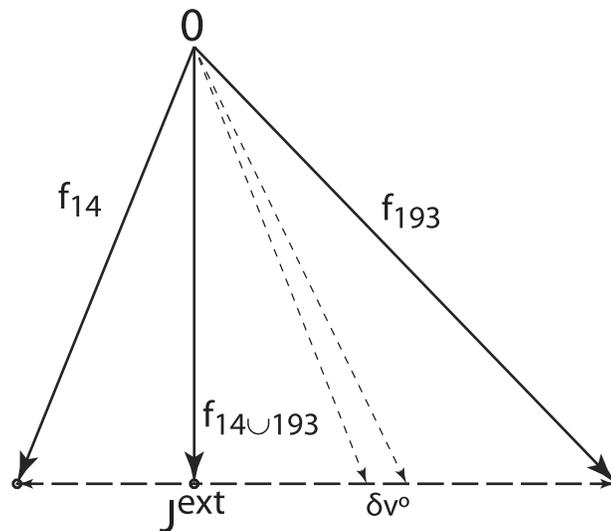
Fig 12. The Pythagorean theorem (50) in the linear-response regime where it is the ordinary Euclidean theorem. Two irreducible flows f_{14} and f_{193} differ by a TKL trefoil current v° with magnitude 1. In the hierarchy $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{G}, \mathcal{G}$ is the full graph from Fig 2, and $\mathcal{G}_1$ is the union of the supporting graphs for f_{14} and f_{193}. $\mathcal{G}_2$ may be the supporting graph for either f_{14} or f_{193}. Vertical solid arrow is square root of the integral $\int d\theta'_1$ in Eq (50) from zero to $\theta_1(J^{\text{ext}})$ in $\mathcal{G}_1$. Hypotenuse solid arrows are square roots of integrals $\int d\theta'_2$ in Eq (50) in either graph $\mathcal{G}_2$. Horizontal dashed arrows are square roots of integrals $\int d\eta'$ in either direction of v° in Eq (50). Details of the final flow parameters and the functional dependence of $\dot{Q}$ on v° are given in Table 4.

4.4 Transduction of chemical work

Next we use the solution series from Table 4 to illustrate the non-stoichiometric or stoichiometric transduction of chemical work defined in Sec. 2.4, as null flows allocate current along parallel pathways, and in doing so redistribute chemical potential to the minimum-dissipation values for mass-action flows. A series of mass-action solutions for current and chemical potential are computed for the non-stoichiometric thermochemical landscape from Fig 11.

Minimum transduction loss through stoichiometric coupling by a low-impedance graph. Fig 13 shows the AlKe null cycle from Fig 6 within the fugue component of the union supporting graph from f_{193} and f_{184}. The fluxes and potential drops are indicated for the solution $f_{193 \cup 184}$. The figure shows that $\{\Delta\mu_{15}, \Delta\mu_{97}\} \approx 500 \times \{\Delta\mu_6, \Delta\mu_7\}$. The two are highly unequal because TAL is a bimolecular reaction whereas AlKe is monomolecular; bimolecular reactions are suppressed by an additional power $\sim [GAP]$ in complex activities compared to monomolecular reactions in this concentration profile. The effect of low-impedance AlKe edges is to hold the net potential drop (here $\Delta\mu_{15} - \Delta\mu_{97}$) near zero, though $\Delta\mu_{15}$ and $\Delta\mu_{97}$ separately may remain large.

We may view this solution as an instance of transduction of work and thus chemical potential between the TAL edge e15 and the TKL edge e97, respectively taken as input and output boundaries from the full graph of Fig 6, by the subgraph $\mathcal{G}' \equiv \{e6, e7\}$. For all flows on this supporting graph, the prelude requires regeneration of the central complex $A_2 + K_7$ in Fig 6 at rate 1. In either irreducible flow this flux passes entirely through one of the bimolecular reactions. Null cycle activation distributes flux almost equally between e15 and e97, transducing work and redistributing chemical potential from the boundary of one reaction to the boundary of the other, in the process reducing the potential across both boundary pairs by 1/2. As noted in Sec. 2.4.5, because the AlKe edges have very low impedance, transduction through this $\mathcal{G}'$ incurs almost no parasitic loss to redistribute potential.

Transduction through non-stoichiometric coupling. Within the same fugue-subgraph, we may evaluate an instance of Eq. (25) for the 235 trefoil driven non-stoichiometrically through reaction e96 in Fig 14. As shown in the figure, the total current going through the complex-pair with potential drop $\Delta\mu_{96}$ is $v_{96} + v_{23} = v_6 - v_{98} = 1.1447$. The first of these
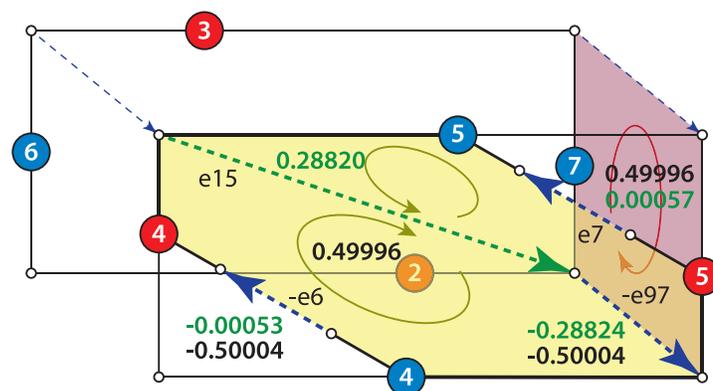
Fig 13. The chemical potentials in the solution on the union of supports for f_{193} and f_{184}, across all edges of the AlKe null cycle. Current through the null cycle is shown in black; potential drops in green. The potential drop across the AlKe edges is $\sim [GAP] \sim 10^{-3}$ smaller because they are first-order in organics, whereas the TAL and TKL edges are second-order. Therefore almost-all chemical work delivered by the current $v_{23} = 1$ (dark red circle) to the boundary of e97, which is saved from dissipation in e97 by the flow around the AlKe null cycle, is transduced stoichiometrically to the boundary of e15, where it is dissipated.

https://doi.org/10.1371/journal.pcsy.0000022.g013
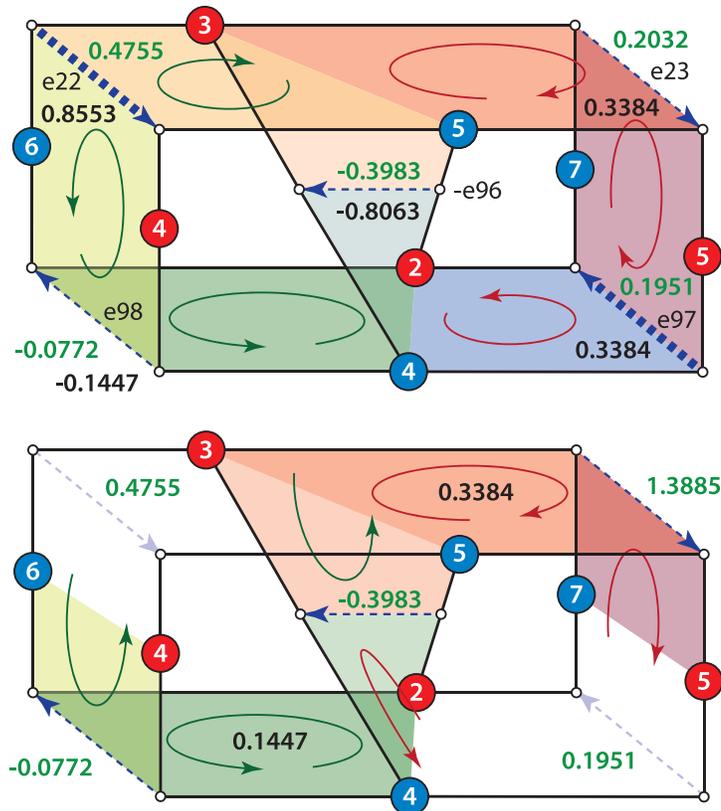
Fig 14. Two linked TKL-trefoils that appear in the union of the supporting graphs for two flows (f_{193} and f_{535}).
The Möbius boundary of each trefoil is formed by tracing a continuous path of solid (stoichiometric) links as they pass through species and complexes. A set of chemical-potential drops (green lettering) and currents (black lettering) for a $\dot{Q}$-minimizing solution are shown for each reaction. Top panel shows two basis elements for null flows on the complete supporting graph: the 234 trefoil (green shades) on the left and the 235 trefoil (red shades) on the right. The period-2 backbone cycles shown in Fig 7 of S1 File B 2 appear as simple circulations in their respective faces of the cube. Bottom panel shows "environment" reactions $K_6 + A_3 \rightharpoonup A_4 + K_5$ and $A_5 + K_4 \rightharpoonup K_7 + A_2$ removed as explicit sources of dissipation. A red cycle with current $j_2 = 0.3384$ fully accounts for the current supplied to the complexes bounding the "output" conversion $A_5 + K_4 \rightharpoonup K_7 + A_2$. A green cycle with current $j_1 = 0.1447$ flows through the subgraph, alongside a supply of the boundary complexes $K_6 + A_3 \rightharpoonup A_4 + K_5$ at rate 0.8553 across the potential 0.4755 from the complement to this subgraph in the complete graph (making the total current flow between these two complexes the topologically-constrained value of 1). Other sources (TAL edge and preludes) also couple to the boundary complexes, and serve to maintain their chemical potentials.

expressions is the sum of the two currents driven by $\Delta\mu_{96}$: one through e96 (the transducing reaction) and the other through the 235 TKL trefoil (the driven subgraph). The second expression is the driving supply current from an environment which is responsible for building up $\Delta\mu_{96}$, in this case accounted as supply of K_4 through the 234 TKL trefoil and the AlKe edge e6 shown in Table 4, because K_4 has no other sources or sinks. Eq. (25) for transduction efficiency then gives

$$\eta_{235}(v \mid \Delta\mu_{e96}) \quad = \frac{v_{23}}{v_{96} + v_{23}} = \frac{0.3384}{1.1447} \tag{51}$$

Transduction through more complex subgraphs. We may use the Kirchoff decomposition to analyze more complex cases, such as the contribution solely from the two TKL null flows in

the upper panel of Fig 14, to transduction through the subgraph $\mathcal{G}'$ shown in the lower panel, in a potential landscape constructed jointly by those null flows and other through-flows. These are the null flows added to f_{14} in extending its supporting graph to include those of f_{535} and then f_{193}. (We form the boundary of $\mathcal{G}'$ by excising edges e22 and e97 for simplicity, so that the input and output complexes contain no common species).

To isolate this transduction, we pro-rate the output current v_{23} between a null-flow and a through-flow contribution. The fraction of the 235 TKL current v_{23} accounted against the null flow $-v_{98}$ around the 234 TKL trefoil would be $-v_{98} \times v_{23}/(v_{96} + v_{23})$. The net efficiency measured as the ratio of that prorated output current times $\Delta\mu_{97}$ to the input chemical work $-v_{98}\Delta\mu_{22}$ is then

$$\eta_{\mathcal{G}'}(v \mid \mu) \quad = \frac{v_{23}}{v_{96} + v_{23}} \frac{\Delta\mu_{97}}{\Delta\mu_{22}} = \frac{0.3384}{1.1447} \frac{0.1951}{0.4755} \tag{52}$$

Whereas the potential $\Delta\mu_{96}$ cancels in the single-reaction transduction efficiency (51), the potential ratios across the graph $\mathcal{G}'$ will not likewise all cancel and thus will be sensitive to the effects of other flows through $\mathcal{G}$.

Transduction reduces whole-graph impedance by alleviating bottlenecks. Table 5 shows potential drops and fluxes on key reactions, along with whole-network dissipation $\dot{Q}$, for both thermochemical backgrounds from Fig 11, to illustrate the role of the transducing edge e96 as a bottleneck to whole-graph through-flow. A nested graph sequence from $f_{535\cup193}$ to f_{14} is shown, with the third column corresponding to the annotation in Fig 14.

While addition of both trefoils to f_{14} contributes to reducing $\dot{Q}$ in Table 5, they do so through different impacts on chemical potential and current profiles. When e98 is added, the potential drop $\Delta\mu_{22}$ is decreased as the 234 trefoil routes a part of the current v_{22} that is 1 in f_{14}, around edge e22 and through $-v_{98}$. The result is an *increase* in both the potential drop across, and the current through e96, and hence of the rate of chemical work delivered to its input complexes.

In contrast, when e23 and e97 from f_{193} are added *parallel to* e96, activating the downstream 235 trefoil, the potential $\Delta\mu_{96}$ and current v_{96} are both diminished, as e96 has been alleviated as

Table 5. Currents v and potential drops $\Delta\mu$ on the edges in TKL trefoils for the masks of f_{14}, f_{535}, and $f_{535} \cup f_{193}$. f_{535} adds one (234) TKL trefoil to f_{14} by adding edge e98, and f_{193} adds a second (235) TKL trefoil by adding edges e23 and e97. The final columns, $f_{535} \cup f_{193}$, provide the labeling for the transduction in Fig 14 in either potential. (eQNS) labels the thermochemical background for group contribution from non-stereochemical SMILES, and (eQS) labels the background with stereochemistry, approximating KEGG data values. The relation $v_{22} - v_{98} = 1 = v_6$ in all cases is the topologically constrained AlKe velocity E4P $\rightharpoonup$ Eu4P. Chemical-potential cycles must cancel around trefoils; thus $\Delta\mu_{23} + \Delta\mu_{97} - \Delta\mu_{96} = 0$ and $\Delta\mu_{22} + \Delta\mu_{98} - \Delta\mu_{96} = 0$.

	$f_{14}^{(eQNS)}$	$f_{535}^{(eQNS)}$	$f_{535\cup193}^{(eQNS)}$	$f_{14}^{(eQS)}$	$f_{535}^{(eQS)}$	$f_{535\cup193}^{(eQS)}$
v_{22}	1.0000	0.9609	0.8553	1.0000	0.3919	0.3645
$\Delta\mu_{22}$	0.5559	0.5341	0.4755	1.6440	0.6443	0.5992
v_{98}		-0.0391	-0.1447		-0.6081	-0.6355
$\Delta\mu_{98}$		-0.0209	-0.0772		-0.3395	-0.3548
v_{96}	1.0000	1.0391	0.8063	1.0000	1.6081	1.2891
$\Delta\mu_{96}$	0.4939	0.5133	0.3983	0.1896	0.3049	0.2444
v_{23}			0.3384			0.3465
$\Delta\mu_{23}$			0.2032			0.1824
v_{97}			0.3384			0.3465
$\Delta\mu_{97}$			0.1951			0.0620
$\dot{Q}$	1.7796	1.7772	1.6035	1.9743	1.0898	0.9842

https://doi.org/10.1371/journal.pcsy.0000022.t005

a bottleneck. The net impedance of the left-hand subgraph from the input complexes of e96 is reduced, so flux v_{98} through the 234 trefoil is further increased, while the potential drop $\Delta\mu_{22}$ and current v_{22} are further decreased.

4.5 The effective potentials in geometric coordinates

The orthogonality between null flows and the reaction-potential drops $\theta\mathbb{S}$ that result from species potentials θ is a special case of a more general orthogonality relation on all flow increments δv through the potential drops that are their duals under the Hessian of $\tilde{\mathcal{L}}_\eta$. In this section we study the eigenvectors of this Hessian, which serves as a metric on a geometric space of flow increments, to understand why the reductions to some integer flows f_i and not others enable significant network pruning while minimizing impact on network resistance from the removal of parallel flow paths.

4.5.1 The Hessian as a metric tensor, and the large-deviation integrand in geometric form.   Let $\mathcal{G}$ be a graph with stoichiometric matrix $\mathbb{S}$ and let $\{f_\mu\}$ be a basis for $\ker\mathbb{S}$. Each $f_\mu \equiv [f_{\mu(ji)}]$ is a column vector on the reaction index (ji), and we can expand any null flow

$$v \;\; = f_\mu \alpha^\mu. \tag{53}$$

In Eq (53) and later we adopt the Einstein summation convention for vector or tensor contraction, in which paired indices, with one index raised and one lowered, are summed.

For any such basis we introduce a dual basis $\{\bar{f}^\mu\}$ of projection operators, which are row vectors $\bar{f}^\mu \equiv \left[\bar{f}^\mu_{(ji)}\right]^T$ on the same reaction indices, defined so that a representation of the identity matrix on $\ker\mathbb{S}$ can be expanded as

$$\mathbb{I} = \sum_\mu f_\mu \tilde{f}^\mu, \tag{54}$$

in which we retain the explicit sum for the outer product (54) because this is not an index contraction. Eq (54) is equivalent to requiring

$$\tilde{f}^\mu f_v = \delta^\mu_v, \tag{55}$$

(the Kronecker δ symbol that equals 1 if $\mu = v$ and zero otherwise).

The row vector $\theta\mathbb{S} + v$ of potential drops across reactions can be expanded in the dual basis as

$$\theta\mathbb{S} + \eta \equiv h_\mu \tilde{f}^\mu. \tag{56}$$

The integrand appearing in the large-deviation functional (45) is then written in terms of current coordinates in the basis $\{\mu\}$ as

$$(\theta\mathbb{S} + \eta)v + \tilde{\mathcal{L}}_\eta(\theta, n) = h\alpha + \tilde{\mathcal{L}}(h, n). \tag{57}$$

We show in S1 File H 1 that the inverse of the Hessian of $-\tilde{\mathcal{L}}$ in Eq (47) defines a metric for inner products of current in the coordinates α as

$$g_{\mu v} \equiv f_\mu^T \left(\left. \frac{-\partial^2 \tilde{\mathcal{L}}_\eta}{\partial\eta\,\partial\eta^T} \right|_\circ \right)^{-1} f_v. \tag{58}$$

The metric transforms differentials of the Legendre-dual coordinates α and h according to the relation

$$dh_\mu^T = g_{\mu\nu}(\alpha)d\alpha^\nu. \tag{59}$$

We can then write the steady-state, detailed-balance large-deviation integrand from Eq (47) explicitly in terms of the coordinate basis $\{\mu\}$ and the inverse metric g^{-1} as

$$
\begin{aligned}
S_T[n, \nu] \quad &\underset{\text{SS,DB}}{\rightarrow} \int_0^T dt \int_0^{\alpha(\nu)} h(\alpha')d\alpha' + S_0(n_0) \\
&= \int_0^T dt \int_0^{h(\nu)} h'g^{-1}(h')dh'^T + S_0(n_0).
\end{aligned}
\tag{60}
$$

4.5.2 Flows through a sub-graph, in the linear-response regime. Following the program outlined in Sec. 2.4.1, we suppose that the null space on the total graph $\mathcal{G}$ defines the feasible conversions in steady-state through any subgraph. For an application to the sugar-phosphate example, the through-flow of interest is J^{ext} of schema (26), and the subgraph $\mathcal{G}'$ of interest may be the whole network of Fig 2, or the supporting graph of some smaller collection of flows.

To compare through-flows of $\mathcal{G}'$, we need only those basis elements of $\ker\mathbb{S}$ on $\mathcal{G}$ for which

$$\mathbb{S}|_{\mathcal{G}'}f_\mu = J^{\text{ext}} ; \quad \forall\mu. \tag{61}$$

For any pair of distinct indices μ and ν, $f_\mu - f_\nu$ will then be a null flow in $\mathcal{G}'$. It follows that any solution $\nu = f\alpha$ with $\mathbb{S}|_{\mathcal{G}'}\nu = J^{\text{ext}}$ must then have $1^T\alpha = 1$.

In the linear-response regime, the large-deviation integrand (57) is evaluated only to quadratic order in sources, and becomes, as we show in Eq. (H7) of S1 File H 2

$$h\alpha + \tilde{\mathcal{L}}(h, n) \approx \frac{1}{2}\alpha^T g_\circ \alpha \tag{62}$$

where $g_\circ$ denotes the metric in the un-driven (equilibrium) background.

The minimizer of the quadratic form (62) on the constraint surface $1^T\alpha = 1$, which as we have shown is the mass-action flow, will be the point of tangency of level sets of $\dot{Q}$ to that surface, as shown in Fig 15. Some algebra shows that this requires $h^T = 1/(1^Tg_\circ^{-1}1)$. The coordinates α at this solution, which we denote by $\alpha_\parallel$, are then given by

$$\alpha = \frac{g_\circ^{-1}1}{1^Tg_\circ^{-1}1} \equiv \alpha_\parallel. \tag{63}$$

We will refer to the dissipation $\dot{Q}$ net of the un-driven steady state as the *excess dissipation*, $\dot{Q}_{\text{ex}} \equiv h\alpha$. On the mass-action solution (63), it is given by

$$\dot{Q}_{\text{ex}}(\alpha_\parallel) = \alpha_\parallel^T g_\circ \alpha_\parallel = \frac{1}{1^Tg_\circ^{-1}1} = \min_{1^T\alpha=1}\dot{Q}_{\text{ex}}(\alpha). \tag{64}$$
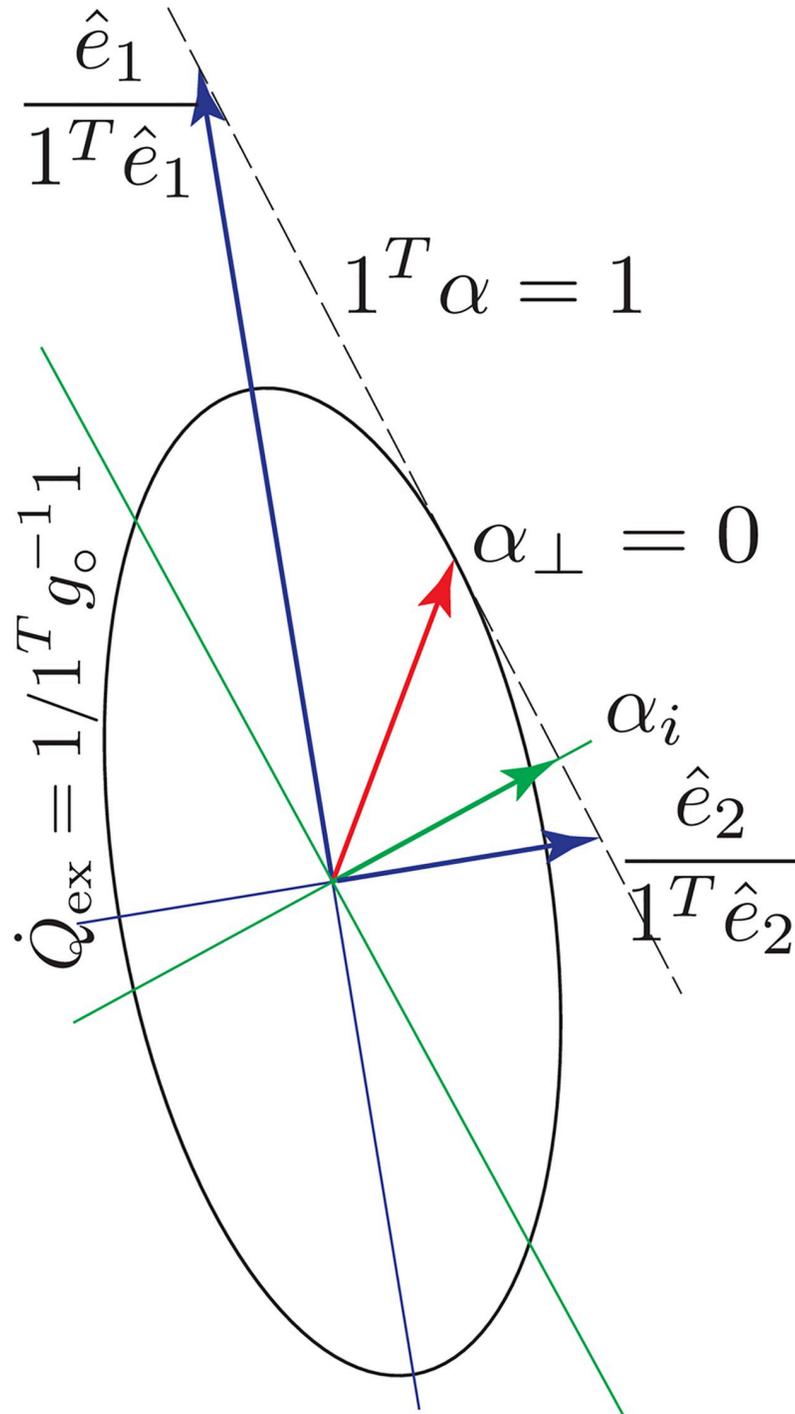
Fig 15. The stationarity condition (62), written in the dual coordinates of Eq (65). Ellipse is a surface of constant excess dissipation $\dot{Q}_{ex} = (\alpha - \alpha_\circ)^T g_\circ (\alpha - \alpha_\circ) = 1/(1^T g_\circ^{-1} 1)$ from Eq (66) below. This is the minimum of $\dot{Q}_{ex}$ on the surface $1^T \alpha = 1$ of fixed flux J^{ext}. $\hat{e}_1$ and $\hat{e}_2$ are the eigenvectors of $g_\circ$, normalized here to the $1^T \alpha = 1$ surface. α_i is the coordinate vector of some other integer flow solution f_i, restricted to support on a subgraph of the graph defining $g_\cdot$.

If a bias parameter η induces further null flows in $\mathcal{G}'$, we designate these as $\alpha_\perp$. Then the total current and the value h^T from Eq. (62) become

$$
\begin{aligned}
h^T &= g_\circ (\alpha_\parallel + \alpha_\perp) \\
\alpha &= \alpha_\parallel + \alpha_\perp,
\end{aligned}
\tag{65}
$$

in which $1^T \alpha_\perp = 0$. The integrand (62) and its relation to the excess dissipation then evaluate to

$$
h\alpha + \tilde{\mathcal{L}}(h, n) \approx \frac{1}{2} \left(\alpha_\parallel^T g_\circ \alpha_\parallel + \alpha_\perp^T g_\circ \alpha_\perp \right) = \frac{1}{2} \dot{Q}_{\mathrm{ex}}(\alpha).
\tag{66}
$$

In terms of Eq (66), The steady-state LDF (60) accordingly becomes

$$
S_T[n, v] \underset{\mathrm{SS,LR}}{\to} \frac{1}{2} \int_0^T dt \, \dot{Q}_{\mathrm{ex}} + S_0(n_0).
\tag{67}
$$

Example. For the transect example shown in Fig 12, with nested supporting graphs $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{G}$, the mass-action solution in $\mathcal{G}_1$ defines $\alpha_\parallel$, and the mass-action in $\mathcal{G}_2$ defines $\alpha = \alpha_\parallel + \alpha_\perp$. Applying the Pythagorean decomposition (50) to the resistance measure (17) gives

$$
R_{\mathcal{G}_2} - R_{\mathcal{G}_1} = \frac{\alpha_\perp^T g_\circ \alpha_\perp}{J^{\mathrm{ext}\,T} J^{\mathrm{ext}}}.
\tag{68}
$$

Fig (15) shows the isocontour of $\dot{Q}_{\mathrm{ex}}(\alpha)$ tangent to the constraint surface $1^T \alpha = 1$ at $\alpha_\perp = 0$. Other flows, such as an integer solution f_i with coordinates α_i are indicated, along with the eigenvectors of $g_\circ$, labeled $\hat{e}_{1,2}$ normalized to the $1^T \alpha = 1$ surface, which we consider next.

4.5.3 Eigenvalues and their relation to low-impedance flows. The metric eigenvectors—principle axes of the ellipsoid in Fig 15—provide another way to assess the divergence of mass-action flows on complex networks from their reductions to one or another irreducible pathway flow. In the lower panel of Fig 15 in S1 File E 1, the values of $\dot{Q}$ on the integer flow solutions from the ILP list are shown sorted. Interspersed with these are the dissipation rates of flows proportional to the eigenvectors of the metric, normalized for each case $g_\circ \alpha_q = \lambda_q \alpha_q$ to $1^T \alpha_q = 1$, ensuring that the resulting flow satisfies $J^{\mathrm{ext}} = \mathbb{S} f \alpha_q$ for the conversion (26).

f_{13}, the canonical Calvin-Benson cycle from the left panel of Fig 3, is among the lowest-impedance irreducible flows in the list, with $\dot{Q}_{f_{13}}[\mathrm{GAP}]^2 \approx 1.9244$, which may be compared to values for $\dot{Q}_{f_i}[\mathrm{GAP}]^2$ for other low-impedance flows in Table 5, and to $\dot{Q}_{\mathrm{tot}}[\mathrm{GAP}]^2 \approx 1.0332$ for the whole network.

By construction the eigenvectors are orthogonal, and from Eq (66) they contribute additively to $\dot{Q}_{\mathrm{ex}}(\alpha)$. If one of the metric eigenvectors is closely aligned with the whole-network mass-action flow, then flows contributing to transverse eigenvectors may be eliminated with little increase in dissipation on the reduced network. If one or more integer flows are are close to one of these eigenvectors, then network pruning to isolate that subset of flows can eliminate transverse directions without significantly increasing the dissipation relative to the unrestricted network.

Fig 16 shows that, for the non-stereochemical background in Fig 11, f_{13} is well approximated by linear combinations of only a few eigenvectors of $g_\circ$ with the lowest dissipations. The eigenvectors are indexed in increasing order of the dissipation produced by each as a pure flow. Two levels of approximation are exhibited in the figure, showing that the eigenvectors of $g_\circ$ that most-nearly approximate the dissipation of the whole network jointly account for most of the flow in f_{13}. Thus selection isolating f_{13} by removing reactions from the full network of
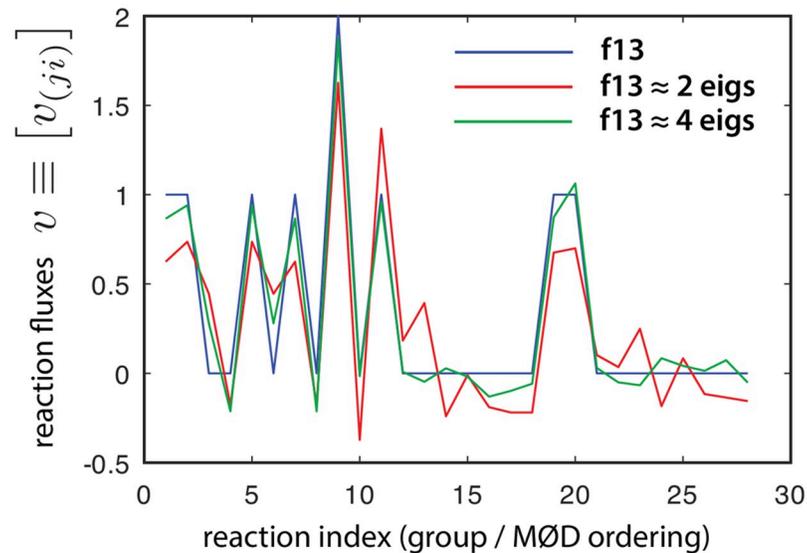
Fig 16. The integer flow solution (blue) for the canonical Calvin cycle (irreducible f_{13}) from the left panel of Fig 3, expanded in a basis of the eigenvectors of $g_{\cdot}$. The 28 reactions are listed in the order of Fig 11 in S1 File E 1. The velocity of each reaction is plotted as the ordinate. Red shows the contribution from eigenvectors 1 and 2; green shows the contribution from eigenvectors 1, 2, 3, and 10, respectively the two-largest, and four-largest coefficients in absolute magnitude in the eigenbasis expansion.

https://doi.org/10.1371/journal.pcsy.0000022.g016

Fig 2 can be performed with minimal increase in dissipative cost to perform the conversion (26).

4.6 The biological questions motivating a study of information

4.6.1 The information distinguishing a pathway within a network derived from rules. We chose sugar-phosphate chemistry to illustrate principles and strategy of rule-based modeling because it provides two combinatorial diversifications to be understood: from few rules come indefinitely large and networked stoichiometric graphs; and from the usual many-particle limits of thermodynamics the reactions in these graphs generate macroscopic state and transport behaviors. A highly combinatorial reaction network from few chemical mechanisms isolates an interesting problem in the evolution of pathway specificity and control, which is likely present for most biological pathways but nowhere else so clearly isolated as a combinatorial problem.

Many enzyme families are organized [57, 58] around a retained mechanism and catalytic site under strongly conservative selection, with sub-families diverging in substrate specificity or other modes of subfunctionalization. For reactions that form networks by many serendipitous combinations [105, 106], the asymmetry of conservation between mechanism and substrate discrimination in enzyme families suggests that the earliest primitive enzymes, by opening reaction mechanisms but not yet selecting precise substrates, would have brought large combinatorial networks into existence. Only through later selection for specificity—in part made possible by the productivity of earlier diffuse networks—would reactions be eliminated until only precise pathways remained.

Selection is understood as imparting adaptive information in populations, at some cost in organisms and the embodied resources to form them. What is increasingly appreciated is that a natural measure of the information imparted can be the same as the cost function measured in logarithmic units of population growth [107–110].

Here we wish to identify lower bounds on the information required to evolve primitive enzymes, and the degenerate networks they produce, into specific enzymes catalyzing networks assembled from precise pathways. This information will not generally be localized in single reactions, as it depends on substrate availability which is jointly determined across the network. An absolute lower bound on the information in specificity should not be rooted in the molecular biology of particular catalysts, though the latter could impose tighter bounds requiring more information than the abstract lower limit. We will propose here that a lower bound is given from the information divergence S_T of Eq (45) and its decompositions (48, 49), which assign a likelihood cost to the sampling bias needed within the underlying permissive network to produce the same current pattern as the one from removal of reactions by enzyme specificity.

4.6.2 The information divergence as a direct measure of evolutionary cost. Our proposal to use the LDF for currents as a cost measure to tie physiological events to selective events is driven by two considerations: First, relative entropies [111–113] and a variety of functions either derived from [114, 115] or related to [116] them are understood as measures of both information and cost [30, 107, 109, 110, 113] through selection in population processes. Second, the events within physiology form a nested partition function with the population-level events of reproduction or death through which selection acts in the form of differential rates.

The overall LDF for such a nested process may be decomposable (for example by a chain rule for relative entropies) into summands of which the network divergence S_T of Eq (45) characterizes the likelihood for events within physiology (chemical conversions) by alternate pathways, while another summand characterizes likelihoods through Malthusian selection events that alter population composition [28–30] by replacing more open with more restrictive networks to produce altered current routes through constraints. The robust population trajectory must be one that is jointly maximal in the product of these contributions to probability (inevitably along with many others), giving the implication that *ceteris paribus*, the improbability-cost of forcing current through a restricted subnetwork must not accumulate to a larger value than whatever probability gain is achieved at the population level, as reflected in selection to maintain the more evolved and constrained network.

We have not aimed in this paper to construct a model with both selection and physiology, which must be a separate project. We aim in this section only to justify large-deviation log-likelihood as a common denomination in which costs-per-event in chemical kinetics and in biological lifecycles can be quantitatively compared.

Remark. A naïve story to bolster the intuition for this argument may be made from the relation (67) of the LDF and the integrated dissipation in the linear-response regime: network restriction increases the chemical work dissipated per unit of some chemical conversion performed, such as PPP catabolism or Calvin-cycle anabolism, which is essential to an organism's viability. Excess work lost to dissipation should appear as a fitness cost in some other essential function, favoring lower-dissipation networks. For selection to maintain the more specific but more dissipative network, it must yield some other advantage—elimination of substrate loss to diffusion or side-reactions, toxicity, *etc.*—which favors the specific phenotype even at a higher throughput cost in dissipated work.

5 Conclusions: Consequences and future directions

Our purpose in this paper has been to introduce rule-based modeling for chemistry in a framework of relations that we call "three-level systems", and to illustrate the interactive and

complementary use of rules, stoichiometric graphs, and probability dynamics to recognize and characterize the loci of causation in such systems.

The main emphasis in the three-level framework is that each level has a generative relation to the one following, with the consequence that constraints, patterns, or information imposed at smaller simpler levels can propagate through the generative relations to govern patterns and dynamics at later levels even when those are combinatorially larger than their progenitors. Scale-independence of patterns is well understood to underlie the concept of macrostates [31, 46, 48], and the origin of thermodynamics from stochastic processes. Here we anchor the possible CRN generators of macrostates within lower-level algebras of rules.

Our main technical contributions were a program for graph and flow reduction that gave a general formulation of chemical work transduction and chemical-potential redistribution, and the use of the current-large-deviation rate to assign probabilistic costs to topological modifications of networks. We used the natural Hessian geometry [42, 43] induced by dissipation to derive an additive decomposition of large-deviation rates for currents, and gave the interpretation of the cost in terms of biased sampling [117].

Evolutionary context has motivated our choice of model systems and questions: These include understanding how topology governs decomposition of both flows and large-deviation rates, and how rules dictate shortest paths and paths of least resistance.

Evolutionary control naturally crosses levels in real molecular systems, from reaction mechanisms to substrate specificities to kinetic regulation, and the three-level framework suggests how these can be coordinated under selection: global complexity or minimality of reaction currents can be governed at the mechanism level and derived directly from rule properties without the insertion of costly and complex steps such as exhaustive enumeration and search of large combinatorial spaces, which selection may have limited capability to solve [118].

We introduced a lattice-graph representation of pathways as simple closed curves, directly expressing rule symmetries. By this approach we could show that the Calvin cycle is one of the two unique shortest paths for the conversion (26). The absence of thermochemical data on 2-phosphoglycolate may suggest that this compound is not readily formed or is unstable, thereby eliminating from possibility the only other comparably short path to the Calvin cycle.

Finally, we have suggested that large-deviation probabilities—precisely because they nest in the manner of partition functions for multilevel systems, and because they can be additively decomposed on nested networks—provide a natural measure of evolutionary cost from events at the physiological level, which must be at least compensated by selective advantage expressed as a large-deviation probability [30], to explain the persistence of evolved specificity. A denomination of all rate effects, from physiological to organism-population and ecological scales, in common units of large-deviation probability, is meant to provide robust lower bounds on constraints of thermodynamics on fitness, without regard to particular molecular and other mechanisms that may connect the two in nature.

Supporting information

S1 File. Supporting figures, data, and derivations.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Eric Smith, Harrison B. Smith, Jakob Lykke Andersen.

Formal analysis: Eric Smith.

Funding acquisition: Eric Smith, Harrison B. Smith.

Investigation: Eric Smith, Jakob Lykke Andersen.

Methodology: Eric Smith, Harrison B. Smith, Jakob Lykke Andersen.

Software: Harrison B. Smith, Jakob Lykke Andersen.

Supervision: Eric Smith.

Writing – original draft: Eric Smith.

Writing – review & editing: Harrison B. Smith, Jakob Lykke Andersen.

References

1. Goldstein H, Poole CP, Safko JL. Classical Mechanics. 3rd ed. New York: Addison Wesley; 2001.

2. Abelson H, Sussman GJ, Sussman J. Structure and Interpretation of Computer Programs. 2nd ed. Cambridge, MA: MIT Press; 1996.

3. Boltzmann L. Populäre Schriften. Leipzig: J. A. Barth; 1905.

4. Fermi E. Thermodynamics. New York: Dover; 1956.

5. Mézard M, Parisi G, Sourlas N, Toulouse G, Virasoro M. Nature of the spin-glass phase. Phys Rev Lett. 1984; 52:1156–1159. https://doi.org/10.1103/PhysRevLett.52.1156

6. Oono Y. The Nonlinear World: Conceptual Analysis and Phenomenology. New York: Springer; 2013.

7. Fontana W, Wagner G, Buss LW. Beyond Digital Naturalism. Artificial Life. 1994; 1:211–227. https://doi.org/10.1162/artl.1993.1.1_2.211

8. Fontana W, Buss LW. The barrier of objects: From dynamical systems to bounded organizations. In: Casti J, Karlqvist A, editors. Boundaries and Barriers. New York: Addison-Wesley; 1996. p. 56–116.

9. Danos V, Feret J, Fontana W, Harmer R, Krivine J. Rule-based modelling, symmetries, refinements. Formal methods in systems biology: lecture notes in computer science. 2008; 5054:103–122.

10. Harmer R, Danos V, Feret J, Krivine J, Fontana W. Intrinsic information carriers in combinatorial dynamical systems. Chaos. 2010; 20:037108. https://doi.org/10.1063/1.3491100 PMID: 20887074

11. Andersen JL. Analysis of Generative Chemistries. Denmark: Syddansk Universitet. Det Naturvidens-kabelige Fakultet; 2015.

12. Behr N, Danos V, Garnier I. Stochastic mechanics of graph rewriting. In: Shankar N, editor. Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, 2016. New York: ACM Press; 2016. p. 46–55.

13. Behr N, Krivine J. Rewriting Theory for the Life Sciences: A Unifying Theory of CTMC Semantics. In: Gadducci F, Kehrer T, editors. Graph Transformation, 13th International Conference, ICGT 2020, Proceedings, volume 12150 of Theoretical Computer Science and General Issues. Switzerland AG: Springer International Publishing; 2020. p. 185–202.

14. Andersen JL, Flamm C, Merkle D, Stadler PF. An intermediate level of abstraction for computational systems chemistry. Phil Trans R Soc A. 2017; 375:20160354. https://doi.org/10.1098/rsta.2016.0354 PMID: 29133452

15. Behr N, Danos V, Garnier I. Combinatorial Conversion and Moment Bisimulation for Stochastic Rewriting Systems. Logical Methods in Computer Science. 2020; 16:1–45.

16. Andersen JL, Flamm C, Merkle D, Stadler PF. Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete. J Sys Chem. 2012; 3:1. https://doi.org/10.1186/1759-2208-3-1

17. Horn FJM, Jackson R. General mass action kinetics. Arch Rat Mech Anal. 1972; 47:81–116. https://doi.org/10.1007/BF00251225

18. Feinberg M. Lectures on chemical reaction networks; 1979. lecture notes.

19. Berge C. Graphs and Hypergraphs. Rev. ed. Amsterdam: North-Holland; 1973.

20. Anderson DF, Craciun G, Kurtz TG. Product-form stationary distributions for deficiency zero chemical reaction networks. Bull Math Bio. 2010; 72:1947–1970. https://doi.org/10.1007/s11538-010-9517-4 PMID: 20306147

21. Qian H, Beard DA. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. Biophys Chem. 2005; 114:213–220. https://doi.org/10.1016/j.bpc.2004.12.001 PMID: 15829355

22. Polettini M, Esposito M. Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws. J Chem Phys. 2014; 141:024117. https://doi.org/10.1063/1.4886396 PMID: 25028009

23. Baez JC, Fong B. Quantum techniques for studying equilibrium in reaction networks. J Compl Netw. 2014; 3:22–34. https://doi.org/10.1093/comnet/cnu013

24. Ge H, Qian H. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: A mathematical theory. Phys Rev E. 2016; 94:052150. https://doi.org/10.1103/PhysRevE.94.052150 PMID: 27967115

25. Ge H, Qian H. Nonequilibrium thermodynamic formalism of nonlinear chemical reaction systems with Waage-Guldberg's law of mass action. Chem Phys. 2016; 472:241–248. https://doi.org/10.1016/j.chemphys.2016.03.026

26. Ge H, Qian H. Mathematical Formalism of Nonequilibrium Thermodynamics for Nonlinear Chemical Reaction Systems with General Rate Law. J Stat Phys. 2017; 166:190–209. https://doi.org/10.1007/s10955-016-1678-6

27. Wachtel A, Rao R, Esposito M. Free-Energy Transduction in Chemical Reaction Networks: from Enzymes to Metabolism. J Chem Phys. 2022; 157:024109. https://doi.org/10.1063/5.0091035 PMID: 35840395

28. Smith E, Krishnamurthy S. Symmetry and Collective Fluctuations in Evolutionary Games. Bristol: IOP Press; 2015.

29. Smith E. Beyond fitness: The nature of selection acting through the constructive steps of lifecycles. Evolution. 2023; 77:1967–1986. https://doi.org/10.1093/evolut/qpad068 PMID: 37161529

30. Smith E. Beyond fitness: The information imparted in population states by selection throughout life-cycles. Theor Popul Biol. 2024; 157:86–117. https://doi.org/10.1016/j.tpb.2024.04.003 PMID: 38615922

31. Smith E. Intrinsic and extrinsic thermodynamics for stochastic population processes with multi-level large-deviation structure. Entropy. 2020; 22:1137. https://doi.org/10.3390/e22101137 PMID: 33286906

32. Andersen JL, Flamm C, Merkle D, Stadler PF. Inferring chemical reaction patterns using rule composition in graph grammars. J Sys Chem. 2013; 4:4:1–14.

33. Andersen JL, Flamm C, Merkle D, Stadler PF. Generic strategies for chemical space exploration. Int J Comput Biol Drug Des. 2014; 7:225–258. https://doi.org/10.1504/IJCBDD.2014.061649 PMID: 24878732

34. Andersen JL, Flamm C, Merkle D, Stadler PF. Chemical Transformation Motifs—Modelling Pathways as Integer Hyperflows. IEEE/ACM Trans Comp Biol Bioinf. 2019; 16:510–523. https://doi.org/10.1109/TCBB.2017.2781724 PMID: 29990045

35. Andersen JL, Flamm C, Merkle D, Stadler PF. Chemical Transformation Motifs-Modelling Pathways as Integer Hyperflows. IEEE/ACM Trans Comp Biol Bioinf. 2019; 16:510–523. https://doi.org/10.1109/TCBB.2017.2781724 PMID: 29990045

36. Hill TL. Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism. New York: Academic Press; 1977.

37. Schnakenberg J. Network theory of microscopic and macroscopic behavior of master equation systems. Rev Mod Phys. 1976; 48:571–585. https://doi.org/10.1103/RevModPhys.48.571

38. Huang S, Li F, Zhou JX, Qian H. Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics: differentiation in living matters. J R Soc Interface. 2017; 14:20170097. https://doi.org/10.1098/rsif.2017.0097 PMID: 28490602

39. Krishnamurthy S, Smith E. Solving moment hierarchies for chemical reaction networks. J Phys A: Math Theor. 2017; 50:425002. https://doi.org/10.1088/1751-8121/aa89d0 PMID: 29333197

40. Smith E, Krishnamurthy S. Flows, scaling, and the control of moment hierarchies for stochastic chemical reaction networks. Phys Rev E. 2017; 96:062102. https://doi.org/10.1103/PhysRevE.96.062102 PMID: 29335680

41. Lazarescu A, Cossetto T, Falasco G, Esposito M. Large deviations and dynamical phase transitions in stochastic chemical networks. J Chem Phys. 2019; 151:064117. https://doi.org/10.1063/1.5111110

42. Kobayashi TJ, Loutchko D, Kamimura A, Sughiyama Y. Kinetic derivation of the Hessian geometric structure in chemical reaction networks. Phys Rev Res. 2022; 4:033066. https://doi.org/10.1103/PhysRevResearch.4.033208

43. Kobayashi TJ, Loutchko D, Kamimura A, Sughiyama Y. Hessian geometry of nonequilibrium chemical reaction networks and entropy production decompositions. Phys Rev Res. 2022; 4:033208. https://doi.org/10.1103/PhysRevResearch.4.033066

44. Jinich A, Sanchez-Lengeling B, Ren H, Goldford JE, Noor E, Sanders JN, et al. A thermodynamic atlas of carbon redox chemical space. Proc Nat Acad Sci USA. 2020; 117:32910–32918. https://doi.org/10.1073/pnas.2005642117 PMID: 33376214

45. Carnot S. Reflections on the Motive Power of Fire. Mendoza E. ed. New York: Dover; 1960.

46. Ellis RS. Entropy, Large Deviations, and Statistical Mechanics. New York: Springer-Verlag; 1985.

47. Huang K. Statistical Mechanics. New York: Wiley; 1987.

48. Touchette H. The large deviation approach to statistical mechanics. Phys Rep. 2009; 478:1–69. https://doi.org/10.1016/j.physrep.2009.05.002

49. Smith E. Large-deviation principles, stochastic effective actions, path entropies, and the structure and meaning of thermodynamic descriptions. Rep Prog Phys. 2011; 74:046601. https://doi.org/10.1088/0034-4885/74/4/046601

50. Bassham JA, Benson AA, Kay LD, Harris AZ, Wilson AT, Calvin M. The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. J Am Chem Soc. 1954; 76:1760–1770. https://doi.org/10.1021/ja01636a012

51. Horecker BL. Pathways of carbohydrate metabolism and their physiological significance. J Chem Educ. 1965; 42:244. https://doi.org/10.1021/ed042p244 PMID: 14328677

52. Metzler DE. Biochemistry: The Chemical Reactions of Living Cells. 2nd ed. New York: Academic Press; 2003.

53. Meléndez-Hevia Enrique and Isidoro Angel. The game of the pentose phosphate cycle. J Theor Biol. 1985; 117(2):251–263. https://doi.org/10.1016/S0022-5193(85)80220-4

54. Meléndez-Hevia Enrique and Waddell Thomas G and Montero Francisco. Optimization of Metabolism: The Evolution of Metabolic Pathways Toward Simplicity Through the Game of the Pentose Phosphate Cycle. J Theor Biol. 1994; 166(2):201–220. https://doi.org/10.1006/jtbi.1994.1018

55. Clasquin MF, Melamud E, Singer A, Gooding JR, Xu X, Dong A, et al. Riboneogenesis in Yeast. Cell. 2011; 145:969–980. https://doi.org/10.1016/j.cell.2011.05.022 PMID: 21663798

56. Stincone A, Prigione A, Cramer T, Wamelink MMC, Campbell K, Cheung E, et al. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. Biol Rev. 2015; 90:927–963. https://doi.org/10.1111/brv.12140 PMID: 25243985

57. Khersonsky O, Tawfik DS. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. Annu Rev Biochem. 2010; 79:471–505. https://doi.org/10.1146/annurev-biochem-030409-143718 PMID: 20235827

58. Khersonsky O, Malitsky S, Rogachev I, Tawfik DS. Role of Chemistry versus Substrate Binding in Recruiting Promiscuous Enzyme Functions. Biochem. 2011; 50:2683–2690. https://doi.org/10.1021/bi101763c PMID: 21332126

59. Sauro HM. Systems Biology: An Introduction to Metabolic Control Analysis. Ambrosius Publishing; 2018.

60. King E, Holzer J, North JA, Cannon WR. An approach to learn regulation to maximize growth and entropy production rates in metabolism. Front Syst Biol. 2023; 3:981866. https://doi.org/10.3389/fsysb.2023.981866

61. Purcell EM, Morin DI. Electricity and Magnetism. 3rd ed. Cambridge MA: Cambridge U. Press; 2013.

62. Esposito M, Van den Broeck C. Three Detailed Fluctuation Theorems. Phys Rev Lett. 2010; 104:090601. https://doi.org/10.1103/PhysRevLett.104.090601 PMID: 20366974

63. Seifert U. Stochastic thermodynamics, fluctuation theorems, and molecular machines. Rep Prog Phys. 2012; 75:126001. https://doi.org/10.1088/0034-4885/75/12/126001 PMID: 23168354

64. Blokhuis A, Lacoste D, Nghe P. Autocatalysis in Chemical Networks: Unifications and Extensions. Proc Nat Acad Sci USA. 2020; 117:25230–25236.

65. Amari SI. Information Geometry and its Applications. Springer Japan: Appl. Math. Sci. vol. 194; 2001.

66. Ay N, Jost J, Lê HV, Schwachhöfer L. Information Geometry. Cham, Switzerland: Schwinger International; 2017.

67. Beber ME, Gollub MG, Mozaffari D, Shebek KM, Flamholz AI, Milo R, et al. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. Nucl Acids Res. 2021; 50:D603–D609. https://doi.org/10.1093/nar/gkab1106

68. Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G, Landim C. Macroscopic fluctuation theory for stationary non equilibrium states. J Stat Phys. 2002; 107:635–675. https://doi.org/10.1023/A:1014525911391

69. Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G, Landim C. Large deviations of the empirical current in interacting particle systems. Theory Probab Appl. 07; 51:2–27. https://doi.org/10.1137/S0040585X97982256

70. Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G, Landim C. Towards a nonequilibrium thermodynamics: a self-contained macroscopic description of driven diffusive systems. J Stat Phys. 2009; 135:857–872. https://doi.org/10.1007/s10955-008-9670-4

71. van Kampen NG. Stochastic Processes in Physics and Chemistry. 3rd ed. Amsterdam: Elsevier; 2007.

72. Onsager L. Reciprocal Relations in Irreversible Processes. I. Phys Rev. 1931; 37:405–426. https://doi.org/10.1103/PhysRev.37.405

73. Onsager L. Reciprocal Relations in Irreversible Processes. II. Phys Rev. 1931; 38:2265–2279. https://doi.org/10.1103/PhysRev.38.2265

74. Glansdorff P, Prigogine I. Thermodynamic Theory of Structure, Stability, and Fluctuations. New York: Wiley; 1971.

75. Kondepudi D, Prigogine I. Modern Thermodynamics: From Heat Engines to Dissipative Structures. New York: Wiley; 1998.

76. March J. Advanced Organic Chemistry. New York: McGraw Hill; 1977.

77. Forbes AG, Burks A, Lee K, Li X, Boutillier P, Krivine J, et al. Dynamic Influence Networks for Rule-based Models. IEEE Trans Vis Comput Graph. 2018; 24:184–194. https://doi.org/10.1109/TVCG.2017.2745280 PMID: 28866584

78. Ricardo A, Carrigan MA, Olcott AN, Benner SA. Borate minerals stabilize ribose. Science. 2004; 303:196. https://doi.org/10.1126/science.1092464 PMID: 14716004

79. Braakman R, Smith E. The compositional and evolutionary logic of metabolism. Phys Biol. 2013; 10:011001. https://doi.org/10.1088/1478-3975/10/1/011001 PMID: 23234798

80. Weber AL. Sugars as the optimal biosynthetic carbon substrate of aqueous life throughout the Universe. Orig Life Evol Biosphere. 2000; 30:33–43. https://doi.org/10.1023/A:1006627406047 PMID: 10836263

81. Weber AL. Sugar model of the origin of life: Catalysis by amines and amino acid products. Orig Life Evol Biosphere. 2001; 31:71–86.

82. Tcherkez GGB, Farquhar GD, Andrews JT. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylasesmay be nearly perfectly optimized. Proc Nat Acad Sci USA. 2006; 103:7246–7251. https://doi.org/10.1073/pnas.0600605103 PMID: 16641091

83. Hellgren J, Godina A, Nielsen J, Siewers V. Promiscuous phosphoketolase and metabolic rewiring enables novel non-oxidative glycolysis in yeast for high-yield production of acetyl-CoA derived products. Metab Eng. 2020; 62:150–160. https://doi.org/10.1016/j.ymben.2020.09.003 PMID: 32911054

84. Gagrani P, Blanco V, Smith E, Baum D. Polyhedral geometry and combinatorics of an autocatalytic ecosystem. J Math Chem. 2023; 62:1012–1078. https://doi.org/10.1007/s10910-024-01576-x

85. Deshpande A, Gopalkrishnan M. Autocatalysis in Reaction Networks. Bull Math Biol. 2014; 76:2570–2595. https://doi.org/10.1007/s11538-014-0024-x PMID: 25245394

86. Andersen JL, Flamm C, Merkle D, Stadler PF. Defining Autocatalysis in Chemical Reaction Networks. J Sys Chem. 2020; 8:121–133.

87. Peng Z, Linderoth J, Baum DA. The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. PLoS Comp Biol. 2022; 18:e1010498. https://doi.org/10.1371/journal.pcbi.1010498 PMID: 36084149

88. Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, Milo R. An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. Bioinformatics. 2012; 28:2037–2044. https://doi.org/10.1093/bioinformatics/bts317 PMID: 22645166

89. Kanehisa M. The KEGG database. Novartis Found Symp. 2002; 247:91–101. https://doi.org/10.1002/0470857897.ch8 PMID: 12539951

90. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucl Acids Res. 2006; 34:D354–D357. https://doi.org/10.1093/nar/gkj102 PMID: 16381885

91. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucl Acids Res. 2014; 42:D199–D205. https://doi.org/10.1093/nar/gkt1076 PMID: 24214961

92. Gumaa KA, McLean P. The Pentose Phosphate Pathway of Glucose Metabolism. Biochem J. 1969; 115:1009–1029.

93. Pettersson G, Ryde-Pettersson U. A mathematical model of the Calvin photosynthesis cycle. Eur J Biochem. 1988; 175:661–672. https://doi.org/10.1111/j.1432-1033.1988.tb14242.x PMID: 3137030

94. Singh S, B Sunoj R. Molecular Machine Learning for Chemical Catalysis: Prospects and Challenges. Acc Chem Res. 2023; 56:402–412. https://doi.org/10.1021/acs.accounts.2c00801 PMID: 36715248

95. Yu H, Deng H, He J, Keasling JD, Luo X. UniKP: a unified framework for the prediction of enzyme kinetic parameters. Nat Commun. 2023; 14:8211. https://doi.org/10.1038/s41467-023-44113-1 PMID: 38081905

96. Jaynes ET. The minimum entropy production principle. Annu Rev Phys Chem. 1980; 31:579–601. https://doi.org/10.1146/annurev.pc.31.100180.003051

97. Doi M. Second quantization representation for classical many-particle system. J Phys A. 1976; 9:1465–1478. https://doi.org/10.1088/0305-4470/9/9/008

98. Doi M. Stochastic theory of diffusion-controlled reaction. J Phys A. 1976; 9:1479–1495. https://doi.org/10.1088/0305-4470/9/9/009

99. Peliti L. Path-integral approach to birth-death processes on a lattice. J Physique. 1985; 46:1469–1483. https://doi.org/10.1051/jphys:019850046090146900

100. Peliti L. Renormalization of fluctuation effects in $A + A \to A$ reaction. J Phys A. 1986; 19:L365–L367. https://doi.org/10.1088/0305-4470/19/6/012

101. Eyink GL. Action principle in nonequilibrium statistical dynamics. Phys Rev E. 1996; 54:3419–3435. https://doi.org/10.1103/PhysRevE.54.3419 PMID: 9965486

102. Kamenev A. Keldysh and Doi-Peliti Techniques for out-of-equilibrium Systems. In: Lerner IV, Althsuler BL, Fal'ko VI, Giamarchi T, editors. Strongly Correlated Fermions and Bosons in Low-Dimensional Disordered Systems. Heidelberg: Springer-Verlag; 2002. p. 313–340.

103. Hartley RVL. Transmission of information. Bell system technical journal. 1928;July:535–563. https://doi.org/10.1002/j.1538-7305.1928.tb01236.x

104. Chentsov NN. Nonsymmetrical distance between probability distributions, entropy and the theorem of pythagoras. Math Notes Acad Sci USSR. 1968; 4:686–691.

105. Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD. Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5′-phosphate synthesis. Mol Sys Biol. 2010; 6:436:1–13. https://doi.org/10.1038/msb.2010.88

106. Copley SD. The physical basis and practical consequences of biological promiscuity. Phys Biol. 2020; 17:051001. https://doi.org/10.1088/1478-3975/ab8697 PMID: 32244231

107. Kimura M. Natural selection as the process of accumulating genetic information in adaptive evolution. Genet Res, Camb. 1961; 2:127–140. https://doi.org/10.1017/S0016672300000616

108. Watkins C. Selective Breeding Analysed as a Communication Channel: Channel Capacity as a Fundamental Limit on Adaptive Complexity. IEEE. 2008;2008 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing.

109. Rivoire O, Leibler S. The Value of Information for Populations in Varying Environments. J Stat Phys. 2010; 142:1124–1166. https://doi.org/10.1007/s10955-011-0166-2

110. McGee RS, Kosterlitz O, Kaznatcheev A, Kerr B, Bergstrom CT. The cost of information acquisition by natural selection. bioRxiv. 2022;.

111. Iwasa Y. Free fitness that always increases in evolution. J Theor Biol. 1988; 135:265–281. https://doi.org/10.1016/S0022-5193(88)80243-1 PMID: 3256719

112. Sella G, Hirsch AE. The application of statistical physics to evolutionary biology. Proc Nat Acad Sci USA. 2005; 102:9541–9546. https://doi.org/10.1073/pnas.0501865102 PMID: 15980155

113. Hledík M, Barton N, Tkačik G. Accumulation and maintenance of information in evolution. Proc Nat Acad Sci USA. 2022; 119:e2123152119. https://doi.org/10.1073/pnas.2123152119 PMID: 36037343

114. Frieden BR. Science from Fisher Information: A Unification. Cambridge, UK: Cambridge U. Press; 2004.

115. Frank SA. Natural selection maximizes Fisher information. J Evol Biol. 2009; 22:231–244. https://doi.org/10.1111/j.1420-9101.2008.01647.x PMID: 19032501

116. Mustonen V, Lässig M. Fitness flux and ubiquity of adaptive evolution. Proc Nat Acad Sci USA. 2010; 107:4248–4253. https://doi.org/10.1073/pnas.0907953107 PMID: 20145113

117. Smith E. The information geometry of 2-field functional integrals. Info Geo. 2022; 5:427–492. https://doi.org/10.1007/s41884-022-00071-z

118. Valiant L. Probably Approximately Correct. New York: Basic Books; 2014.