Universal slow dynamics of chemical reaction networks

Masanari Shimada,* Pegah Behrad,* and Eric De Giuli
Department of Physics, Toronto Metropolitan University, Toronto, Ontario, Canada M5B 2K3

(Received 24 August 2023; revised 7 January 2024; accepted 4 March 2024; published 3 April 2024)

Understanding the emergent behavior of chemical reaction networks (CRNs) is a fundamental aspect of biology and its origin from inanimate matter. A closed CRN monotonically tends to thermal equilibrium, but when it is opened to external reservoirs, a range of behaviors is possible, including transition to a new equilibrium state, a nonequilibrium state, or indefinite growth. This study shows that slowly driven CRNs are governed by the conserved quantities of the closed system, which are generally far fewer in number than the species. Considering both deterministic and stochastic dynamics, a universal slow-dynamics equation is derived with singular perturbation methods and is shown to be thermodynamically consistent. The slow dynamics is highly robust against microscopic details of the network, which may be unknown in practical situations. In particular, nonequilibrium states of realistic large CRNs can be sought without knowledge of bulk reaction rates. The framework is successfully tested against a suite of networks of increasing complexity and argued to be relevant in the treatment of open CRNs as chemical machines.

DOI: 10.1103/PhysRevE.109.044105

I. INTRODUCTION

The goal of theory for complex systems is often to reduce the number of degrees of freedom from a large intractable number down to something manageable, whose dynamics can then be understood intuitively. Ideally, such a reduction should be principled, mathematically well controlled, and lead to a description in terms of universal effective variables. This challenge is especially acute in the biosciences where dizzying complexity is the norm. We address it for chemical reaction networks (CRNs), which provide the substrate for biochemistry and hence biology.

Model reduction for CRNs has a long history [1–3]. Since CRNs often contain a wide range of timescales, many reduction methods exploit this by quasiequilibrium or quasi-steady-state approximations [4]. However, existing theories employ different reductions for each particular CRN, thus not leading to any universal description. This may be sufficient for detailed analysis of a particular system, but makes cross-system analysis difficult and hinders unification of diverse phenomena. Moreover, these approaches work at the level of the rate equations, ignoring stochastic effects known to be important in biochemistry [5]. They may also fail to respect thermodynamic constraints [6].

Here we take a different route, grounded in hydrodynamics, a branch of condensed-matter physics [7]. The starting point for hydrodynamics is the observation that molecular timescales, on the order of picoseconds, are minuscule compared to macroscopic forcing timescales. Thus most degrees of freedom relax very rapidly to a state of local thermodynamic equilibrium. However, over macroscopic distances, forcing conditions can differ, thus leading to different local equilibria. In such a hydrodynamic limit in which forcing is slow in time and gradual in space, only a subset of degrees of freedom, dictated by symmetries and conservation laws, are important. Indeed, conserved quantities are precisely those whose densities need to be tracked, while the other degrees of freedom relax quickly and can be neglected.[1] Conventional continuum theories for fluids, elastic solids, liquid crystals, and others are all of this type, differing only in the assumed symmetries [7].

This vantage is natural for chemical reaction networks, since individual reactions conserve the number of each element, leading to a large number of conservation laws. While previous and ongoing work on CRNs in the mathematical literature has also exploited conservation laws [8–10], such approaches do not incorporate thermodynamic constraints, nor do they yield a universal description. Instead we combine notions from hydrodynamics with stochastic thermodynamics [11–16], building thermodynamic consistency in from the beginning. This is useful to quantify CRNs as chemical machines, as will be discussed later [15,16].

In particular, we consider slowly driven well-mixed physical CRNs and show that they are governed by conserved quantities, similar to hydrodynamic theories. We derive a universal slow-dynamics equation [(25) below] that can be applied to generic slowly driven CRNs. For a CRN with N species and L conserved quantities, the reduced theory involves only L variables, which is generally the number of elements, far fewer than the number of species. We work at the large-deviation level of the particle-number distribution, thus incorporating the leading stochastic effects.

*These authors contributed equally to this work.

[1]If a continuous symmetry is broken, then one also needs to track the density of its associated elastic variable, like the displacement field in an elastic solid. This phenomenon will play no role here.

This article is organized as follows. First, we define our CRN, emphasizing the role of microscopic reversibility. Then we analyze the rate equation in the slowly driven setting, showing that a naive perturbation expansion breaks down. This is cured by a singular perturbation theory, which leads to the slow-dynamics equation. We then extend the theory to include stochastic effects and test our theory with numerical simulations, showing its broad utility. Finally, we show how the theory can be extended to initial states that are far from equilibrium.

Species are indexed with $i, j, \ldots$ while reactions are indexed with $\alpha, \beta, \ldots$. We use vector notation whenever possible. For example, stoichiometric coefficients $p_{i\alpha}$ and $q_{i\alpha}$ are also written as $\vec{p}_\alpha$ and $\vec{q}_\alpha$. All contractions are explicitly indicated by dots. In CRNs, many functions appear that act componentwise on different species. We write $[\vec{f}(\vec{n})\vec{g}(\vec{n})\cdots]$ for the vector in species space whose components are $f_j(\vec{n})g_j(\vec{n})\cdots$. For example, $[\vec{n}^{\mathrm{eq}}\, e^{\vec{\ell}}]$ has components $n_j^{\mathrm{eq}}\, e^{\ell_j}$, etc. When such expressions are considered as diagonal matrices, we double the brackets, i.e., $[[\vec{f}]]$ is the matrix with elements $\delta_{ij}f_i$. To sum or multiply over all species we write $\sum[\vec{f}]$ and $\prod[\vec{f}]$, respectively. We also apply this notation to componentwise vectors over reactions.

II. SLOWLY DRIVEN CHEMICAL REACTION NETWORKS

We define a physical CRN as follows. We have N species X_i, interacting with M reactions α, split into the core and the boundary interactions. We write a general core reaction as

$$\sum_i p_{i\alpha}X_i \rightleftharpoons \sum_i q_{i\alpha}X_i, \tag{1}$$

where $p_{i\alpha}$ and $q_{i\alpha}$ are the stoichiometric coefficients for the molecular species as reactant and product, respectively.

The number of moles of all species is collected in a vector $\vec{n}$. Quantum mechanics requires that if a reaction α occurs with rate k_α^+, then its corresponding backward reaction must occur, with rate k_α^-. As a condition for existence of thermal equilibrium, these rates are not independent but constrained in ratio to satisfy

$$\frac{k_\alpha^+(\vec{n})}{k_\alpha^-(\vec{n})} = e^{-(\Delta G)_\alpha/RT}, \tag{2}$$

where $(\Delta G)_\alpha = (\vec{q}_\alpha - \vec{p}_\alpha)\cdot\vec{G}(\vec{n})$ is the difference in molar Gibbs free energy between products and reactants. Equation (2) is known as local detailed balance [13,15,17,18] or microscopic reversibility [19]. Importantly, it does not require or imply that the CRN be in thermal equilibrium or close to it. In a physical CRN, we require that all reactions in the core of the system satisfy (2). For the boundary interactions, which force the system, we consider intake and degradation pairs

$$X_i^{\mathscr{C}} \rightleftharpoons X_i, \tag{3}$$

with rates ϵr_i^+ and ϵr_i^-, respectively, where ϵ is a dimensionless constant. The $\mathscr{C}$ decoration denotes a chemostat. We explain later how the boundary interactions can be generalized.

The CRN is slowly driven when $\epsilon \ll 1$, meaning that all reservoir interactions are slow compared to internal reactions. This condition is quite natural. Indeed, bulk reaction rates are proportional to the volume of the system, Ω, while boundary rates are proportional to the surface area, $\partial\Omega$. Their ratio $L^* \equiv \Omega/\partial\Omega$ is a length, on the order of the linear dimension of the system. To obtain the dimensionless parameter ϵ, this must be compared with a microscopic length. For example, in the case of passive diffusion across a membrane with similar concentrations on both sides of the membrane, the microscopic length is $\delta \approx D/k_0 d$, where d is the membrane thickness and D is the diffusion constant.[2] Here ϵ can be defined as $\epsilon = \delta/L^*$, which for typical values $d \sim 10$ nm and $D \sim 10^{-9}$ m^2/s gives $\epsilon \sim 10^{-14}$ m$/L^*$, which is small even for microscopic systems. This example furthermore illustrates that the slow-driving condition can be avoided if systems have an anomalously large surface area (as in mitochondria), if diffusion is active (as in the sodium potassium pump), if large concentration differences are held across the membrane, or if important bulk reactions are significantly slower than k_0.

As a consequence of slow driving, after an initial relaxation the system will be close to a thermal equilibrium state of the core CRN. However, over long timescales it can have a nontrivial dynamics near evolving equilibria, just as a fluid that is stirred or poured will transition through a series of near equilibria, described in that case by the Navier-Stokes equations. Our main result is an evolution equation for the slow degrees of freedom, which, as we show, correspond to conserved quantities, in precise analogy with hydrodynamics. For a generic large CRN, the conserved quantities are the number of each element (H, C, O, etc.) and number of free electrons, if ions are present.

The local detailed balance condition (2) can also be derived microscopically. Indeed, modern transition rate theory [20] predicts from first principles

$$k_\alpha^+(\vec{n}) = ke^{-(\delta G)_\alpha(\vec{n})/RT}, \tag{4}$$

where $(\delta G)_\alpha(\vec{n}) = G_{A_\alpha} - \sum_i p_{i\alpha}G_i(\vec{n})$ is the difference in molar Gibbs free energy between the activated complex and the reactants, and $k = \Omega c^\circ k_0$ in terms of $k_0 = 1/2\pi\beta\hbar \approx 6\times 10^{12}$ Hz at 300 K. Here Ω is the system volume, assumed to be dominated by the solvent, and c° is standard concentration 1 mol/l. Equation (4) holds for both forward and backward reactions, *mutatis mutandis*, so (2) is satisfied.

For ideal dilute solutions, the free energies (chemical potentials) for each species take the form

$$G_i(\vec{n}) = \mu_i^\circ + RT\log(n_i/c^\circ\Omega) = RT\log\left(n_i/n_i^{\mathrm{eq}}\right), \tag{5}$$

where μ_i° is the chemical potential in standard conditions and $n_i^{\mathrm{eq}} = \Omega c^\circ e^{-\mu_i^\circ/RT}$. The forward reaction rates are thus proportional to $\prod_i(n_i/c^\circ\Omega)^{p_{i\alpha}}$, which is the law of mass action.

[2]From Fick's law, the diffusion flux at each point is $D\partial\phi/\partial x \approx D\Delta\phi/d$, where ϕ is the concentration, and the derivative is approximated by a finite difference, with d the membrane thickness. Integrating over the surface gives a factor $\partial\Omega$ so that $\epsilon r c^\circ \sim D\partial\Omega c^\circ/d$, where we assume $\Delta\phi \sim c^\circ$. Then $\epsilon \sim (D\partial\Omega c^\circ/d)/k_0 c^\circ\Omega$.

The flux of reaction α is

$$J_\alpha(\vec{n}) = k_\alpha^+(\vec{n}) - k_\alpha^-(\vec{n}) = k e^{-G_{A_\alpha}/RT}\left(\prod[(\vec{n}/\vec{n}^{\rm eq})^{\vec{p}_\alpha}] - \prod[(\vec{n}/\vec{n}^{\rm eq})^{\vec{q}_\alpha}]\right). \quad (6)$$

We can write the rates of reservoir interactions in the form

$$r_i^+ = r_i z_i^{\mathscr{C}}, \quad r_i^-(n_i) = r_i n_i/\Omega, \quad (7)$$

where $z_i^{\mathscr{C}}$ is the molar concentration of species i in its reservoir; in general, this can be time dependent. The numbers r_i have units of rate per mole times volume, while the factors $z_i^{\mathscr{C}}$ and n_i account for the law of mass action.

Note that to precisely distinguish energy from entropy, and hence to unambiguously identify heat flows, requires a microscopic Hamiltonian [11]. However, by comparing (2) to our rate parametrization for reservoir interactions, we can write

$$\frac{r_i^+}{r_i^-(n_i)} = \frac{\epsilon r_i z_i^{\mathscr{C}}}{\epsilon r_i n_i/\Omega} = e^{-[G_i(\vec{n})-W_i]/RT}, \quad (8)$$

where W_i is the work done by the reservoir in one intake reaction. This identifies $W_i = \mu_i^{c^\circ} + RT\log(z_i^{\mathscr{C}}/c^\circ)$, which is just the chemical potential of species i at the reservoir concentration, as expected.

To quantify how far a CRN is from equilibrium, we measure the entropy production rate

$$T\dot{S} = TR\sum\left((\vec{k}^+ - \vec{k}^-)\log\frac{\vec{k}^+}{\vec{k}^-}\right) \geqslant 0. \quad (9)$$

We define the $N \times M$ stoichiometric matrix $\mathbb{S} = [\mathbb{S}^0\ \mathbb{S}^{\mathscr{C}}]$, where $\mathbb{S}^0_{i\alpha} = q_{i\alpha} - p_{i\alpha}$ are the stoichiometric coefficients for the core reactions and $\mathbb{S}^{\mathscr{C}}_{i\alpha} = +1$ when α corresponds to a reservoir of species i and 0 otherwise. Using local detailed balance, the entropy production can be rewritten as

$$T\dot{S} = -\sum[(\vec{k}^+ - \vec{k}^-)(\mathbb{S}^T\cdot\vec{G}(\vec{n}) - (\mathbb{S}^{\mathscr{C}})^T\cdot\vec{W})] = -\vec{G}(\vec{n})\cdot\mathbb{S}\cdot(\vec{k}^+ - \vec{k}^-) + \vec{W}\cdot(\vec{r}^{\,+} - \vec{r}^{\,-}), \quad (10)$$

which will be useful below.

Note that in our setup, each species present in a reservoir has two concentrations: its dynamical concentration in the system, n_j/Ω, and its concentration in the reservoir, denoted by $z_j^{\mathscr{C}}$. These only become equal, in general, when the reservoir rate $r_j \to \infty$ (although we will see other situations below where they equilibrate). Thus having a fully chemostatted species, as often considered in the literature, is a strong-driving limit, since the corresponding species must be added or removed faster than any reaction rate in the system to maintain its constant concentration. Our setup more naturally respects real-world constraints.

Our choice of unimolecular reservoirs is simply for convenience. The theory trivially extends to the case where the reservoir supplies a complex Y_j, say, written as $\sum_i \tilde{p}_{ji}X_i$ in terms of species, where $\tilde{p}_{ji}$ are positive stoichiometric coefficients. We simply replace n_j/Ω in the external flux for the reaction with $\prod_i (n_i/\Omega)^{\tilde{p}_{ji}}$.

Besides slow forcing by reservoirs, one may also consider a subset of slow internal reactions. These will be discussed below.

III. DETERMINISTIC ANALYSIS

To illustrate our approach, we first consider the rate equations for our model; later we will generalize our results to include stochastic effects. The rate equations are

$$\partial_t\vec{n} = \mathbb{S}\cdot\vec{J}(\vec{n}), \quad (11)$$

where $\vec{J}$ is the vector of reaction fluxes. Separating the reactions into core and boundary, this becomes

$$\partial_t\vec{n} = \mathbb{S}^0\cdot\vec{J}^0(\vec{n}) + \epsilon[\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}/\Omega)], \quad (12)$$

where $\vec{J}^0(\vec{n})$ is the reaction flux vector for core reactions and $r_j = 0$ if there is no reservoir for species j. Equation (12) suggests a perturbative solution in $\epsilon : \vec{n} = \vec{n}^0 + \epsilon\vec{n}^1 + \cdots$. At leading order $\partial_t\vec{n}^0 = \mathbb{S}^0\cdot\vec{J}^0(\vec{n}^0)$, which describes a closed system. The system monotonically tends to thermal equilibrium, described by $\vec{J}^0(\vec{n}^0) = 0$. The general steady-state solution is

$$\vec{n}^0 = [\vec{n}^{\rm eq}\, e^{\vec{\ell}}] = \Omega c^\circ[e^{-\vec{\mu}^\circ/RT}e^{\vec{\ell}}], \quad (13)$$

where we must have $(\mathbb{S}^0)^T\cdot\vec{\ell} = 0$. Such vectors $\vec{\ell}$ are the conserved quantities of the closed CRN, called moieties [21]. Suppose there are L independent moieties. Let ζ be an $N \times L$ matrix whose columns give a basis of conserved quantities. Then we can write

$$\vec{\ell} = \zeta\cdot\vec{\eta}, \quad (14)$$

where $\vec{\eta}$ is a vector in moiety space.

To give this a physical interpretation, let $\vec{Y} = (e^-, \mathrm{H}, \mathrm{C}, \mathrm{O}, \ldots)$ be a vector of elements that appear in the CRN and write each species as an abstract sum of elements

$$X_j = \sum_e \tilde{\zeta}_{je}Y_e, \quad (15)$$

defining the atomic matrix $\tilde{\zeta}$. The condition for conservation of element e at reaction α is

$$0 = \sum_j q_{j\alpha}\tilde{\zeta}_{je} - \sum_j p_{j\alpha}\tilde{\zeta}_{je} = [(\mathbb{S}^0)^T\cdot\tilde{\zeta}]_{\alpha e}, \quad (16)$$

showing that $\tilde{\tilde{\zeta}}_e$ is a conserved quantity. In small CRNs, the moieties can differ from the elements. For example, if carbon and oxygen only appear in the CRN in multiples of CO, then their individual concentrations, while both conserved, are degenerate. However, for large CRNs where we expect our theory to be of interest, there are often no conserved quantities besides the elements.

Comparing (14) with (13), we see that nonzero $\vec{\ell}$ is equivalent to shifting chemical potentials by

$$\vec{\mu}^\circ \to \vec{\mu}^\circ - RT\zeta\cdot\vec{\eta}. \quad (17)$$

Therefore, $RT\vec{\eta}$ corresponds to a shift in the chemical potential of moieties. (In our rate parametrization a full transformation would require also shifting activation energies by $G_{A_\alpha} \to G_{A_\alpha} - RT\vec{p}_\alpha\cdot\zeta\cdot\vec{\eta}$.) It acts as a tilt in the free-energy landscape, which fixes the number of each moiety as

$$\vec{y} = \zeta^T\cdot\vec{n} = \zeta^T\cdot[\vec{n}^{\rm eq}\, e^{\zeta\cdot\vec{\eta}}] + O(\epsilon). \quad (18)$$

Although this gives L equations in L unknowns, they cannot be explicitly solved for $\vec{\eta}$.

At the next order we have

$$\partial_t \vec{n}^1 = (\mathbb{S}^0) \cdot (\partial \vec{J}^0/\partial \vec{n})|_{\vec{n}^0} \cdot \vec{n}^1 + [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}/\Omega)]|_{\vec{n}^0}. \quad (19)$$

If we multiply by ζ^T we get

$$\partial_t \zeta^T \cdot \vec{n}^1 = \zeta^T \cdot [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}^0/\Omega)], \quad (20)$$

which expresses moiety balance. This equation can be directly integrated:

$$\zeta^T \cdot \vec{n}^1(t) = \int_0^t dt' \zeta^T \cdot [\vec{r}(t')(\vec{z}^{\,\mathscr{C}}(t') - \vec{n}^0(t')/\Omega)].$$

Now $\vec{n}^0(t)$ relaxes monotonically to some thermal equilibrium state $\vec{n}^0(\infty)$. For simplicity we assume that the reservoirs are independent of time. Then we can write

$$\zeta^T \cdot \vec{n}^1(t) = t\zeta^T \cdot [\vec{r}\,(\vec{z}^{\,\mathscr{C}} - \vec{n}^0(\infty)/\Omega)] + \int_0^t dt' \zeta^T \cdot [\vec{r}\,(\vec{n}^0(\infty) - \vec{n}^0(t'))/\Omega].$$

The integrand in the second term goes to zero at large time, so this term will be sublinear in t at large time. Moreover, there is no opportunity for special cancellations with $\vec{z}^{\,\mathscr{C}}$ in the first term, since $\vec{n}^0$ is completely independent of the reservoirs. Thus, asymptotically,

$$\zeta^T \cdot \vec{n}^1(t) \sim t\zeta^T \cdot [\vec{r}\,(\vec{z}^{\,\mathscr{C}} - \vec{n}^0(\infty)/\Omega)], \quad (21)$$

which diverges at large t. After a time $t \sim 1/\epsilon$ we will have $\epsilon \vec{n}^1 \sim \vec{n}^0$ and the perturbation series breaks down. Thus the slow-driving limit $\epsilon \to 0$ is singular. This is not specific to our boundary conditions or simplifying assumptions, but is completely generic.

This mathematical singularity has a simple physical interpretation: When the system is open to reservoirs, elements can be exchanged with the environment. Over a long timescale $t \sim 1/\epsilon$, the relevant thermal equilibrium state can be completely different from its initial value. Fortunately, this suggests a cure to the long-time divergence: We need to consider a multiple-scale asymptotic analysis [22,23]. We introduce the slow time $\tau = t\epsilon$, so-named because $\tau \sim O(1)$ when $t \sim 1/\epsilon$, and replace the time derivative by

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \epsilon \frac{\partial}{\partial \tau}.$$

The leading-order solution remains the same, except that the coefficients $\vec{\eta}$ get promoted to functions of the slow time, capturing their evolution on long timescales. The $O(\epsilon)$ equation becomes

$$\partial_\tau \vec{n}^0 + \partial_t \vec{n}^1 = (\mathbb{S}^0) \cdot (\partial \vec{J}^0/\partial \vec{n})|_{\vec{n}^0} \cdot \vec{n}^1 + [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}^0/\Omega)].$$

Multiplying by ζ^T, we have

$$\zeta^T \cdot \partial_\tau \vec{n}^0 + \zeta^T \cdot \partial_t \vec{n}^1 = \zeta^T \cdot [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}^0/\Omega)]. \quad (22)$$

At this order, the divergences are cured if we impose

$$\zeta^T \cdot \partial_\tau \vec{n}^0 = \zeta^T \cdot [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}^0/\Omega)], \quad (23)$$

which are the slow-dynamics equations. As shown below, these same equations will remain valid for stochastic dynamics. We have L equations in L degrees of freedom (DOF). More explicitly,

$$\sum_{j,e'} \zeta_{je} n_j^{\mathrm{eq}} \exp\left(\sum_e \zeta_{je}\eta_e\right) \zeta_{je'} \partial_\tau \eta_{e'} = \sum_j \zeta_{je} r_j \left[z_j^{\mathscr{C}} - \frac{n_j^{\mathrm{eq}}}{\Omega} \exp\left(\sum_e \zeta_{je}\eta_e\right)\right]. \quad (24)$$

Defining the matrix $M(\vec{\eta}) = \zeta^T \cdot [[\vec{n}^{\mathrm{eq}}\, e^{\zeta \cdot \vec{\eta}}]] \cdot \zeta$ and the external flux $\vec{J}^{\mathscr{C}}(\vec{\eta}) = [\vec{r}(\vec{z}^{\,\mathscr{C}} - \vec{n}^{\mathrm{eq}}\, e^{\zeta \cdot \vec{\eta}}/\Omega)]$, we can write this as

$$M \cdot \partial_\tau \vec{\eta} = \zeta^T \cdot \vec{J}^{\mathscr{C}}, \quad (25)$$

which is our main result. Equation (25) is a strongly nonlinear system of equations governing the evolution of near-equilibrium states in a slowly driven CRN. The core CRN can be completely arbitrary, as long as it is detailed balanced and closed; open cases will be treated below. The reservoirs can be forced arbitrarily on the slow timescale, that is, $\vec{r}$ and $\vec{z}^{\,\mathscr{C}}$ can be arbitrary functions of τ. In particular, one can consider discontinuous stepwise forcings if needed.

Note that $\zeta^T \cdot \vec{n}$ is simply the number of moles of each moiety; thus from (23) the slow-dynamics equation is simply the conservation law for moieties, which under slowly driven conditions gives a closed set of equations. Remarkably, the number of degrees of freedom is reduced from N down to L. Moreover, for the usual case where moieties correspond to elements, these slow DOF are not arbitrary but are easily interpretable and universal across different CRNs.

If some moieties appear in the CRN but not in any reservoirs, then their concentrations are clearly conserved at their initial values. In this case the corresponding entries of the slow-dynamics equations can be immediately solved, leading to a further reduction in the number of evolving DOF.

Equation (25) is also remarkably universal in form: It does not depend on any activation energies in the core, and steady states are also independent of core chemical potentials. It can thus be applied to poorly characterized CRNs where only the stoichiometry, chemical potentials, and reservoir interactions are known.

The slow-dynamics equation describes the dissipative dynamics through near-equilibrium states. In the Supplemental Material [24] we show that at leading order the dissipation depends only on the slow dynamics and not also on $\vec{n}^1$ as might naively be expected. In particular, it takes a simple form

$$T\dot{S} = \vec{W} \cdot (\vec{r}^{\,+} - \vec{r}^{\,-}) + O(\epsilon^2), \quad (26)$$

which can be evaluated on the solution of (25).

It is important to emphasize that although the system is always close to a thermal equilibrium state of the closed CRN and although $\vec{n}^0$ takes precisely the form of a thermal equilibrium state, the system is nevertheless open, exchanging matter and energy with its surroundings. Therefore, general results for detailed-balanced systems do not apply to $\vec{n}^0$, once $\vec{\eta}$ is allowed to vary on the slow timescale.

There is a fundamental relationship between the number of reservoirs and internal properties of the CRN [17]. Here we adapt the arguments of [17] to our setup, to be used later in analysis of nonequilibrium steady states. We define a cycle as a vector $\vec{c}$ in reaction space in the kernel of $\mathbb{S}$: $\mathbb{S} \cdot \vec{c} = 0$.

This is a combination of reaction fluxes that do not affect the concentration of any species. Explicitly, this condition can be written $0 = \sum_{\alpha\in\text{core}} S^0_{j\alpha} c_\alpha + c_j \delta_{j\in\mathscr{C}}$, where $\delta_{j\in\mathscr{C}}$ is 1 if there is a reservoir of species j and 0 otherwise. Let C be the number of linearly independent cycles. A subset of cycles will be spanned by the core reactions alone, with dimension C_0; the remainder will necessarily involve a reservoir, giving an additional $C_{\mathscr{C}}$.

By the rank-nullity theorem of linear algebra, we have $N - L_r = M - C$, where (in full generality) L_r is the number of conserved moieties in the open system. We can write $L_r = L - L_{\text{br}}$, where L is the number of conserved moieties in the closed system and L_{br} is the number of conservation laws that are broken by the reservoirs. Likewise we can write $M = M_0 + M_{\mathscr{C}}$, where M_0 is the number of reactions in the core and $M_{\mathscr{C}}$ is the number of reservoirs.

In the closed system, we have $N - L = M_0 - C_0$. It follows then that [17,18]

$$M_{\mathscr{C}} = L_{\text{br}} + C_{\mathscr{C}}, \tag{27}$$

where the number of reservoirs equals the number of broken conservation laws plus the number of cycles involving the reservoirs; the latter are called emergent cycles and correspond to effective reactions performed by the CRN. Existence of an emergent cycle is a necessary condition for a CRN to reach a nonequilibrium steady state (NESS). Moreover, many biologically important NESSs have only a few emergent cycles [15].

As universal and general as Eq. (25) is, it relies on the assumption of slow driving. General CRNs may be bistable; the dynamics can end up in different equilibria, depending upon initial conditions. However, model systems for this phenomenon all correspond to strongly driven CRNs, typically with many chemostats [25–27]. Thus it is not obvious *a priori* whether such phenomena are captured by (25). In fact, as shown in Supplemental Material Eq. (25) has no unstable equilibria if the kernel of ζ is trivial, which is expected when $N \gg L$. Since an unstable equilibrium must exist in between stable equilibria, this implies that the slow dynamics does not have multiple equilibria. Instead, there is one equilibrium, which can vary over the τ timescale as external conditions evolve.

Finally, let us discuss the trivial extension when a subset of internal reactions are slow, of relative order ϵ compared to the bulk. We simply add the corresponding projected reaction flux to the right-hand sides of Eqs. (23)–(25). If these reactions respect all conserved quantities, then they disappear identically at leading order. If instead some conserved quantities are broken, then they will act in a similar way to external fluxes.

IV. STOCHASTIC ANALYSIS

We now extend our results to include stochastic effects; in a first reading, this section can be omitted. We begin from the Doi-Peliti path-integral formulation [28,29], which is an exact rewriting of the chemical master equation for the full counting statistics $\mathbb{P}(\vec{n}, t)$. For a self-contained review see [30]. The CRN is specified by the quasi-Hamiltonian (or Liouvillian) $H = H_0 + \epsilon H_{\mathscr{C}}$, with

$$H_0(\vec{n}, \vec{\nu}) = \sum_\alpha k_\alpha \left\{ (e^{\vec{\nu}\cdot(\vec{q}_\alpha - \vec{p}_\alpha)} - 1) \prod \left[\left(\frac{\vec{n}}{\vec{n}^{\text{eq}}} \right)^{\vec{p}_\alpha} \right] + (e^{-\vec{\nu}\cdot(\vec{q}_\alpha - \vec{p}_\alpha)} - 1) \prod \left[\left(\frac{\vec{n}}{\vec{n}^{\text{eq}}} \right)^{\vec{q}_\alpha} \right] \right\}, \tag{28}$$

$$H_{\mathscr{C}}(\vec{n}, \vec{\nu}) = \sum [\vec{r}(e^{-\vec{\nu}} - 1)\vec{n}/\Omega + \vec{r}(e^{\vec{\nu}} - 1)\vec{z}^{\,\mathscr{C}}]. \tag{29}$$

The $\vec{\nu}$ variables act as a per-species bias. As explained in the Supplemental Material [24], in the macroscopic regime where particle numbers are large the leading behavior of $\rho \equiv \log \mathbb{P}$ satisfies a Hamilton-Jacobi equation [31–34]

$$\frac{\partial \rho(\vec{n}, t)}{\partial t} = H(\vec{n}, -\vec{\nabla}_{\vec{n}} \rho(\vec{n}, t)). \tag{30}$$

Equation (30) goes beyond the Gaussian approximation as it includes rare trajectories between different attractors, if they exist. In the slowly driven case, we require that $\Omega c^\circ \epsilon \gg 1$, which ensures that the slow driving is more relevant than finite-size fluctuations from the bulk of the system.

Let $\{\vec{\ell}_x\}_{x=1}^L$ be a basis of the left kernel (or cokernel) of the stoichiometric matrix $\mathscr{K} = \text{coker}\,\mathbb{S}$, i.e., this system has L conserved quantities. We solve (30) under the initial condition

$$\rho(\vec{n}, t = 0) = \sum \left(\vec{n} \log \vec{\lambda}^{(0)} - \vec{\lambda}^{(0)} - \vec{n} \log \frac{\vec{n}}{e} \right), \tag{31}$$

$$\vec{\lambda}^{(0)} = \left[\vec{n}^{\text{eq}}\, e^{\vec{\ell}^{(0)}} \right], \tag{32}$$

where $\vec{\ell}^{(0)}$ is an arbitrary vector in $\mathscr{K}$. This is a large deviation function of a Poisson distribution in the limit $n_i \to \infty$,

$$\mathbb{P}(\vec{n}, 0) = \prod \left(\frac{(\vec{\lambda}^{(0)})^{\vec{n}} e^{-\vec{\lambda}^{(0)}}}{(\vec{n}/e)^{\vec{n}}} \right) \sim \prod \left(\frac{(\vec{\lambda}^{(0)})^{\vec{n}} e^{-\vec{\lambda}^{(0)}}}{\vec{n}!} \right)$$

with mean $\vec{\lambda}^{(0)}$.

Introducing the two time variables t and $\tau = \epsilon t$ as in the preceding section, Eq. (30) is rewritten as

$$\left(\frac{\partial}{\partial t} + \epsilon \frac{\partial}{\partial \tau} \right) \rho(\vec{n}, t, \tau) = H(\vec{n}, -\vec{\nabla}_{\vec{n}} \rho(\vec{n}, t, \tau)),$$

with

$$\rho(\vec{n}, t = 0, \tau = 0) = \sum_i \left(n_i \log \lambda_i^{(0)} - \lambda_i^{(0)} - n_i \log \frac{n_i}{e} \right).$$

We expand the solution of this equation as $\rho(\vec{n}, t, \tau) = \rho_0(\vec{n}, t, \tau) + \epsilon \rho_1(\vec{n}, t, \tau) + \cdots$. The leading equation is

$$\frac{\partial}{\partial t} \rho_0(\vec{n}, t, \tau) = H_0(\vec{n}, -\vec{\nabla}_{\vec{n}} \rho_0(\vec{n}, t, \tau)),$$

which is already solved by the initial condition. We cannot determine the τ dependence of the vector $\vec{\ell}(\tau) \in \mathscr{K}$ at this order.

At the next order we have

$$\frac{\partial}{\partial t} \rho_1 + \frac{\partial}{\partial \tau} \rho_0 = -\vec{\nabla}_n \rho_1 \cdot \vec{\nabla}_\nu H_0(\vec{n}, -\vec{\nabla}_n \rho_0) + H_{\mathscr{C}}(\vec{n}, -\vec{\nabla}_n \rho_0),$$

which can be written

$$\frac{\partial}{\partial \rho_1} t + \vec{\nabla}_n \rho_1 \cdot \vec{\nabla}_\nu H_0(\vec{n}, -\vec{\nabla}_n \rho_0) = \sum [(1 - \vec{n}/\vec{\lambda})(\partial_\tau \vec{\lambda} - \vec{J}^{\mathscr{C}})]. \tag{33}$$

Now we note that, deterministically, the system is always close to some equilibrium state, i.e., $|\vec{n} - \vec{n}^{\rm eq}| \sim \epsilon$, where $\vec{n}^{\rm eq}$ is of the form (13). Further deviations are exponentially suppressed in probability when $\Omega c^\circ \epsilon \gg 1$, as assumed. Consider a general equilibrium $\vec{n} = [\vec{n}^{\rm eq}\, e^{\zeta \cdot \vec{\eta}'}]$ where $\vec{\eta}\,' \neq \vec{\eta}$. On any such state, we have $\vec{\nabla}_\nu H_0 = 0$, so this equation can be directly integrated

$$\rho_1|_{\vec{n}=[\vec{n}^{\rm eq}\, e^{\zeta \cdot \vec{\eta}'}]} = \int^t dt' \sum [(1 - e^{\zeta \cdot (\vec{\eta}\,' - \vec{\eta})})(\partial_\tau \vec{\lambda} - \vec{J}^{\mathscr{C}})], \tag{34}$$

which may lead to secular divergences. We demand that for all nearby equilibria $|\vec{\eta}\,' - \vec{\eta}\,| \ll 1$, the right-hand side vanishes. We thus expand $e^{\zeta \cdot (\vec{\eta}\,' - \vec{\eta}\,)} \approx 1 + \zeta \cdot (\vec{\eta}\,' - \vec{\eta}\,)$ and impose

$$0 = \zeta^T \cdot (\partial_\tau \vec{\lambda} - \vec{J}^{\mathscr{C}}), \tag{35}$$

which is equivalent to (23), with $\vec{n}^{(0)}$ replaced by $\vec{\lambda}$. We thus recover the slow-dynamics equation in the stochastic approach, as the leading equation necessary to prevent long-time divergences in the singular perturbation expansion.

In this approximation, the particle distribution remains Poissonian at leading order, with mean $\vec{\lambda}$ that corresponds to $\vec{n}^0$ in the rate equations. Note that the first correction to Poissonian distributions is given by the solution to (33). Since this is a linear partial differential equation for ρ_1, it can be solved by the method of characteristics. This solution will depend on a trajectory of the closed system.

It is clear that (35) only prevents the leading divergences, and there are not enough DOF in $\vec{\lambda}(\tau)$ to prevent further ones; this implies that the full distribution must be non-Poissonian. To go beyond (35), it is easiest to use a cumulant generating function representation, as discussed in [24]. This analysis shows that, once (35) is solved, all higher-order divergences are tamed by solving a series of linear tensorial ordinary differential equations (ODEs). These ODEs all involve the same matrix M that appears in (25), indicating its central role for slow dynamics in CRNs.

V. NUMERICAL VALIDATION

We illustrate our theory with a series of models of increasing complexity. In realistic CRNs at room temperature, the reaction rates span a wide range of scales. This presents challenges both for numerical simulations and for the basis of our theory, which requires a timescale separation between the bulk and the reservoir interactions. Nevertheless, at high enough temperature such a separation can be found and the theory applied. Initially we consider models for which the logarithmic reaction rates span a modest range, corresponding to physical systems at high temperature. For such models the slow-driving condition is easily specified. Later we will consider CRNs with a wider range of rates.

A. ABC model

The first model, which we dub the ABC model, has three internal reactions

$$A \overset{k_1}{\rightleftharpoons} B, \tag{36}$$

$$A + B \overset{k_2}{\rightleftharpoons} 2B, \tag{37}$$

$$B \overset{k_3}{\rightleftharpoons} C \tag{38}$$

and two reservoirs

$$A \overset{r_A}{\rightleftharpoons} A^{\mathscr{C}}, \tag{39}$$

$$C \overset{r_C}{\rightleftharpoons} C^{\mathscr{C}}. \tag{40}$$

It is stoichiometrically trivial, but simple enough that the slow-dynamics equation can be analytically solved, as detailed in [24]. This solution, giving $\vec{n}^0(\tau)$, is compared with illustrative numerical results from the full rate equations, shown in Fig. 1(a). The leading-order analytical solution $\vec{n}^0(\tau)$ differs from numerical results by an amount of order ϵ, as expected.

The slow dynamics of the ABC model can also be solved at the stochastic level. As shown in [24], the probability remains Poissonian with computable mean. The solution agrees with direct numerical simulation of the master equation with the Gillespie algorithm [35], as shown in Figs. 1(b) and 1(c).

B. Methane combustion

We now consider a version of methane combustion (see Sec. VIII):

$$CH_4 + 3CO_2 \overset{k_1}{\rightleftharpoons} 2H_2O + 4CO, \tag{41}$$

$$O_2 + 2CO \overset{k_2}{\rightleftharpoons} 2CO_2, \tag{42}$$

$$H_2 + CO_2 \overset{k_3}{\rightleftharpoons} H_2O + CO, \tag{43}$$

$$H_2O \overset{r_3}{\rightleftharpoons} H_2O^{\mathscr{C}}, \tag{44}$$

$$O_2 \overset{r_5}{\rightleftharpoons} O_2^{\mathscr{C}}, \tag{45}$$

$$H_2 \overset{r_6}{\rightleftharpoons} H_2^{\mathscr{C}}. \tag{46}$$

Although small, this CRN has features typical of large physical networks: The stoichiometric analysis (see the Supplemental Material [24]) shows that there are three conserved quantities, corresponding to the concentrations of C, H, and O, as expected. We consider it in the high-temperature limit $T \to \infty$ where all bulk rates are equal $k_i \to 1$ (in appropriate units), and we furthermore set $n_i^{\rm eq} = 1$, $r_i = 1$, and $n_i^{\mathscr{C}} = 2$ for simplicity. Example numerical time evolutions are shown in Fig. 2(a) (colored lines) and compared with the result from the slow-dynamics equation (black solid lines). At $\epsilon = 0.01$ the results are indistinguishable, while even at $\epsilon = 0.3$ (dashed lines) the slow-dynamics result captures all qualitative features of the dynamics and provides a quantitative approximation with relative errors smaller than ϵ.

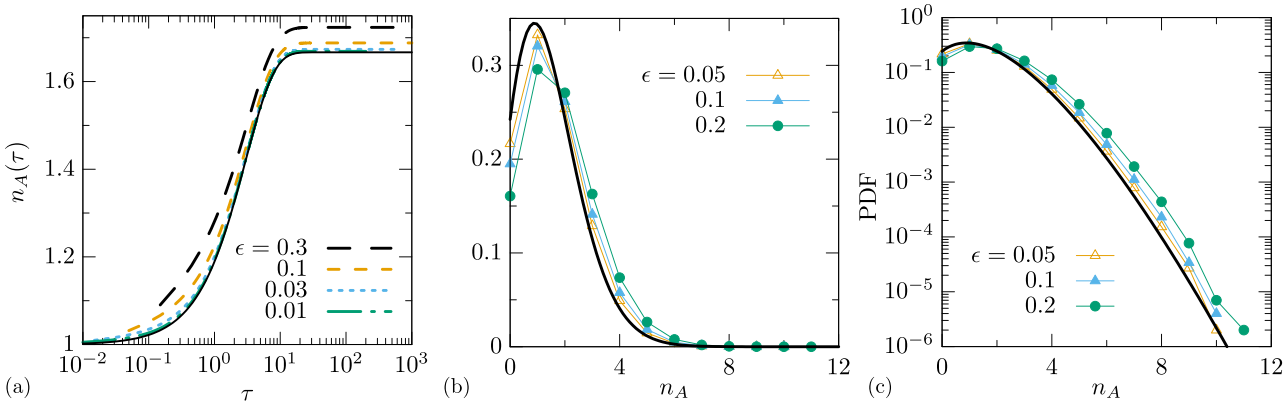

FIG. 1. The slow-dynamics equation tracks solutions to the full rate equations, even through nonequilibrium processes. (a) The ABC model with numerical solutions (dashed lines) and the slow-dynamics equation (solid line) for a range of ϵ with the parameters $k_1 = k_2 = k_3 = r_A = r_C = 1$, $n_A(t=0) = n_A^{\rm eq} = 1$, $n_B(t=0) = n_B^{\rm eq} = 2$, $n_C(t=0) = n_C^{\rm eq} = 3$, $n_A^{\mathscr{C}} = 2$, and $n_C^{\mathscr{C}} = 4$. The ABC model is solved at the full stochastic level, showing the probability distribution function (PDF) of species A on (b) linear and (c) logarithmic axes with the parameters $r_A = 1$, $r_C = 1/3$, $n_A(t=0) = n_B(t=0) = n_C(t=0) = 0$, $n_A^{\rm eq}/\Omega = 1$, $n_B^{\rm eq}/\Omega = 2$, $n_C^{\rm eq}/\Omega = 3$, $z_A^{\mathscr{C}} = 2$, $z_C^{\mathscr{C}} = 4$, and $\Omega = 3$.

The entropy production rate is shown in Fig. 2(b) (orange lines). As predicted by our analysis, the entropy production is well captured by the contribution from slow dynamics (black solid line). Thus, even though the system is always close to some thermal equilibrium state, it is nevertheless out of equilibrium and constantly producing entropy.

Moreover, for this choice of reservoirs, the final state of the system is a nonequilibrium steady state, with small but finite entropy production. It can be found by looking for steady states of the slow dynamics. This gives two equations for $e^{\eta_{\rm H}}$ and $e^{\eta_{\rm O}}$, which do not depend at all on the bulk dynamics. These are easily reduced to a cubic equation for $e^{\eta_{\rm O}}$, with strong positivity constraints on the coefficients, since rates and concentrations cannot be negative. This situation, that is, reduction to a polynomial in an activity e^{η_e}, is typical. An explicit example will be given below.

C. Broad spectrum of reaction rates

Here we consider an early Earth CRN modeling a submarine hydrothermal system containing formaldehyde, ammonia, and water, of interest for the origin of life as submarine hydrothermal systems are a potential source of abiotic amino acids [36,37]. Using Reaction Mechanism Generator (see Sec. VIII), we obtain species that can be obtained from the initial pool, along with the corresponding reactions, including their rates; the full list of 13 species and 40 reactions is in [24]. The histogram of reaction rates is shown in Fig. 3(a); it spans 30 orders of magnitude. This CRN conserves the concentrations of C, H, O, and N, so its slow dynamics is governed by only 4 DOF, a large reduction from the initial 13 DOF.

We solve the rate equations with reservoirs of CH_2O, H, CH_4O, and $C_2H_6O_2$ at $T = 1000$ K. For this choice of reservoirs, N is still conserved, as is the difference of C and O concentrations (since all reservoirs have an equal number of C and O atoms). Thus only two nontrivial DOF are needed to understand its slow dynamics. As shown in Fig. 3, despite the range of reaction rates spanning 30 orders of magnitude and initial concentrations spanning more than five orders of magnitude, the slow-dynamics equation quantitatively predicts the evolution of all species and the entropy production.

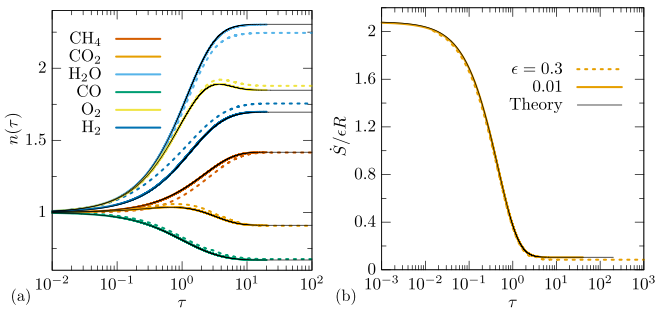

FIG. 2. Methane combustion model for (a) numerical solutions with $\epsilon = 0.3$ (colored dashed lines) and $\epsilon = 0.01$ (colored solid lines) and compared to the result from the slow-dynamics equation (black solid line lines). From top to bottom on the right side, concentrations are shown for H_2O, O_2, H_2, CH_4, CO_2, and CO. (b) The entropy production rate is extremely well captured by the slow dynamics, even at $\epsilon = 0.3$. These results use the parameters $k_i \to 1$, $n_i^{\rm eq} = 1$, $r_i = 1$, and $n_i^{\mathscr{C}} = 2$.

D. Long-timescale dynamics

The slow-dynamics equation is not limited to a description of mild transients between equilibrium states; the theory also applies when the concentrations are dynamic over very long timescales or when a steady state is not reached due to varying reservoir parameters. With the early Earth CRN, if coupled to reservoirs of CH_2O, NH_3, H, CHO, and H_2N, then the system has a very slow dynamics. Figure 4 shows the dynamics at both $T = \infty$ [Fig. 4(a)] and $T = 1000$ K [Fig. 4(b)]. In both cases, solutions of Eq. (25) track solutions to the rate equations over many orders of magnitude in τ. Figure 4(a) shows convergence to a steady state; in Fig. 4(b) a steady state has not yet been reached, but the dynamics appears to be approaching a nonequilibrium steady state of Eq. (25), as shown in Fig. 4(c). These examples show that the final state of the CRN need not be close at all to the initial state for Eq. (25) to capture the dynamics.

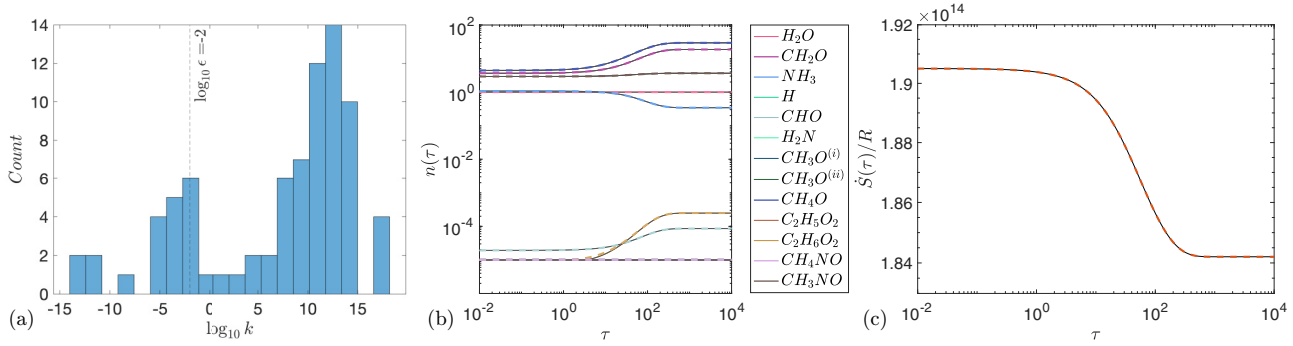

FIG. 3. The slow-dynamics equation can be applied to CRNs with a broad range of reaction rates. (a) Histogram of reaction rates (including both forward and reverse reactions), along with the chosen value of ϵ in simulations. (b) Numerical integration of the rate equations (colored dashed lines), along with the prediction from slow dynamics (thin black solid lines). (c) The entropy production (orange dashed line) is well predicted by the slow-dynamics contribution (thin black solid line).

E. Autotrophic core

As a final example, we consider a very large reaction network of 404 reactions and 375 species, proposed in [38] as a minimal CRN from which to construct the amino acids and nucleic acid monomers necessary for life, along with cofactors needed for their synthesis, from primitive building blocks, namely, H_2, CO_2, and NH_3; the authors have in mind an aqueous environment like a serpentizing hydrothermal vent [37]. Here we show how our theory can be used with such a CRN. More details on the CRN appear in [24].

Consider first reservoirs of H_2O, H_2, CO_2, and NH_3, as suggested in [38]. These reservoirs break conservation of H, O, N, and C and do not create any emergent cycles. Therefore, the system cannot evolve to a NESS, but must relax eventually to an equilibrium state. The slow dynamics is four dimensional and easily solved numerically. Since the left-hand side of the slow-dynamics equation involves the M matrix, which has contributions from the equilibrium concentrations of all species, one needs to know the chemical potentials of all bulk species.

Consider now a scenario with more reservoirs. A minimal way to create a NESS is to add one emergent cycle. For example, if we add carbon monoxide then we create one emergent cycle

$$CO_2 + H_2 \rightleftharpoons CO + H_2O. \tag{47}$$

Then when coupled to these reservoirs, the CRN can either grow indefinitely or evolve to a NESS, performing some of the effective reactions. A NESS can be sought by looking for steady states of the slow dynamics. In terms of the net fluxes into the system from the reservoirs $J_i = r_i(z_i^{\mathscr{C}} - n_i^0/\Omega)$ these are

$$0 = J_{CO_2} + J_{CO}, \tag{48}$$

$$0 = 2J_{H_2O} + 2J_{H_2} + 3J_{NH_3}, \tag{49}$$

$$0 = 2J_{CO_2} + J_{CO} + J_{H_2O}, \tag{50}$$

$$0 = J_{NH_3} \tag{51}$$

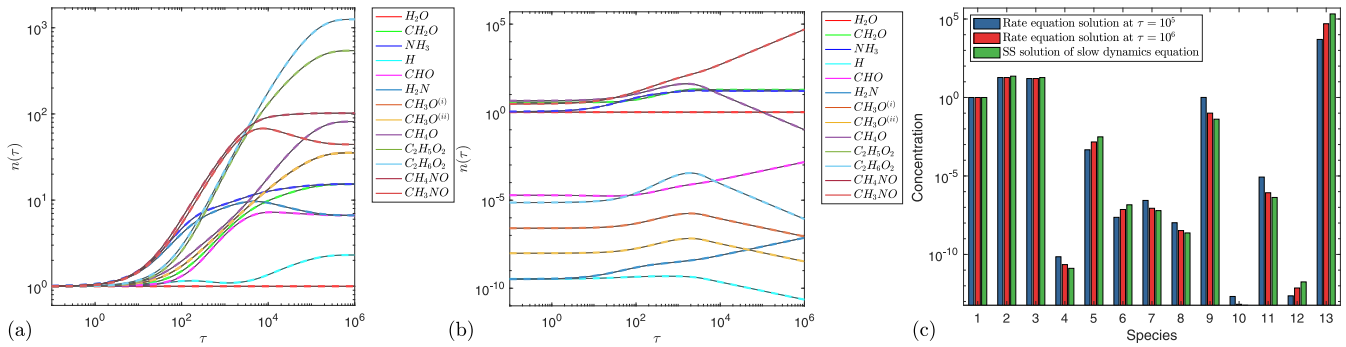

FIG. 4. The slow-dynamics equation captures extremely slow relaxations to steady states. For the early Earth CRN with 13 species and 40 reactions, reservoirs of CH_2O, NH_3, H, CHO, H_2N lead to very slow dynamics. (a) Particle number versus τ at $T = \infty$, where all bulk rates $k_\alpha = 1$ in appropriate units. Here $\epsilon = 10^{-2}$. Solutions of Eq. (25) (thin black solid line) track solutions to the rate equations (colored dashed lines), even over the six decades in τ needed to reach a steady state. (b) The same CRN is simulated at $T = 1000$ K, where the relaxation is slower and the numerical equations are very stiff. Over the simulated time, solutions to Eq. (25) (thin black solid line) track solutions to the rate equations (colored dashed lines). As shown in (c), this solution appears to be converging towards a nonequilibrium steady state, found from Eq. (25) by looking for steady states. Species are labeled in the same order as in the legends of (a) and (b). For each species, three bars are shown (from left to right): the solution to the rate equations after $\tau = 10^5$, the solution to the rate equations after $\tau = 10^6$, and the steady-state solution of Eq. (25).

for C, H, O, and N, respectively. Equation (51) means that the system must equilibrate to be at the reservoir concentration of methane. Explicitly,

$$0 = r_{\mathrm{NH_3}}\left(z^{\mathscr{C}}_{\mathrm{NH_3}} - n^{\mathrm{eq}}_{\mathrm{NH_3}}\, e^{\eta_{\mathrm{N}}} e^{3\eta_{\mathrm{H}}}/\Omega\right)$$

implies $e^{\eta_{\mathrm{N}}} = e^{-3\eta_{\mathrm{H}}} z^{\mathscr{C}}_{\mathrm{NH_3}}\Omega/n^{\mathrm{eq}}_{\mathrm{NH_3}}$. Note that $e^{\eta_{\mathrm{N}}}$ depends on $e^{\eta_{\mathrm{H}}}$, so it is still nontrivial, but since $J_{\mathrm{NH_3}} = 0$, it drops out of the remaining equations for the NESS.

To solve the remaining equations it is convenient to absorb equilibrium concentrations into the rates and reservoir concentrations, viz.,

$$\begin{aligned} J_i &= r_i\left[z^{\mathscr{C}}_i - n^{\mathrm{eq}}_i\, \exp\left(\sum_e \zeta_{ie}\eta_e\right)\Big/\Omega\right] \\ &\equiv \tilde{r}_i\left[\tilde{z}_i - \exp\left(\sum_e \zeta_{ie}\eta_e\right)\right]. \end{aligned} \tag{52}$$

Then labeling the reservoirs in the order H_2O, H_2, CO_2, NH_3, and CO, the remaining equations reduce to

$$0 = \tilde{r}_3(\tilde{z}_3 - e^{\eta_{\mathrm{C}}+2\eta_{\mathrm{O}}}) + \tilde{r}_5(\tilde{z}_5 - e^{\eta_{\mathrm{C}}+\eta_{\mathrm{O}}}), \tag{53}$$

$$0 = \tilde{r}_1(\tilde{z}_1 - e^{2\eta_{\mathrm{H}}+\eta_{\mathrm{O}}}) + \tilde{r}_2(\tilde{z}_2 - e^{2\eta_{\mathrm{H}}}), \tag{54}$$

$$0 = \tilde{r}_3(\tilde{z}_3 - e^{\eta_{\mathrm{C}}+2\eta_{\mathrm{O}}}) + \tilde{r}_1(\tilde{z}_1 - e^{2\eta_{\mathrm{H}}+\eta_{\mathrm{O}}}). \tag{55}$$

Solving Eqs. (53) and (54) as

$$e^{\eta_{\mathrm{C}}} = e^{-\eta_{\mathrm{O}}}\frac{\tilde{r}_3\tilde{z}_3 + \tilde{r}_5\tilde{z}_5}{\tilde{r}_3 e^{\eta_{\mathrm{O}}} + \tilde{r}_5}, \tag{56}$$

$$e^{2\eta_{\mathrm{H}}} = \frac{\tilde{r}_1\tilde{z}_1 + \tilde{r}_2\tilde{z}_2}{\tilde{r}_1 e^{\eta_{\mathrm{O}}} + \tilde{r}_2}, \tag{57}$$

we reduce the problem to

$$0 = \tilde{r}_3\tilde{z}_3 + \tilde{r}_1\tilde{z}_1 - \tilde{r}_3\frac{\tilde{r}_3\tilde{z}_3 + \tilde{r}_5\tilde{z}_5}{\tilde{r}_3 e^{\eta_{\mathrm{O}}} + \tilde{r}_5}e^{\eta_{\mathrm{O}}} - \tilde{r}_1\frac{\tilde{r}_1\tilde{z}_1 + \tilde{r}_2\tilde{z}_2}{\tilde{r}_1 e^{\eta_{\mathrm{O}}} + \tilde{r}_2}e^{\eta_{\mathrm{O}}}, \tag{58}$$

which becomes a quadratic equation for $e^{\eta_{\mathrm{O}}}$. Thus the NESS can easily be found. Now, for given reservoir rates and concentrations, we must check that each e^{η_e} is positive; otherwise the solution is not physical.

Physical solutions can further be divided into two classes. If all the η_e are negative, then the equilibrium concentrations of molecules at infinite temperature decrease with increasing atomic number; larger molecules are less abundant. If instead some $\eta_e > 0$, then that element leads to an increasing concentration with increasing atomic number; large molecules can become exponentially more abundant in the NESS. This opens the possibility for a phase transition separating such regimes, which can be probed with the slow-dynamics equation, with further subtleties at finite T depending on the behavior of chemical potentials with atomic composition. We leave the detailed study of this effect to the future. Here we simply solve (58) for a variety of reservoir rates and concentrations, find solutions with all $\eta_e < 0$, and then compare the result with that of the full rate equations. They agree, showing that the general structure of a potential NESS can be probed without knowledge of the CRN bulk, even for very large CRNs, in the slow-driving limit.

For this CRN, we also look at examples with numerous reservoirs. In such cases, without fine-tuning of parameters, we find that the system grows over a range of τ; this behavior is then confirmed by solution of the full rate equations. Examples are shown in the Supplemental Material [24].

VI. EXTENSION TO FAR-FROM-EQUILIBRIUM STATES

Although above we considered initial states that were near equilibrium before coupling to external reservoirs, this is not essential to obtain a reduction to conserved quantities. Suppose instead that there is a leading-order coupling to reservoirs, which we assume is stationary. We write $\vec{r} \to \frac{1}{\epsilon}\vec{r}^{\,0} + \vec{r}$ so that the rate equation, in the two-time ansatz, is

$$\partial_t\vec{n} + \epsilon\partial_\tau\vec{n} = \underbrace{(\mathbb{S}^0)\cdot\vec{J}^0(\vec{n}) + [\vec{r}^{\,0}(\vec{n}^{\mathscr{C}} - \vec{n})]}_{\tilde{\mathbb{S}}\cdot\tilde{\vec{J}}^0(\vec{n})} + \epsilon[\vec{r}(\vec{n}^{\mathscr{C}} - \vec{n})].$$

Expanding $\vec{n} = \vec{n}^0 + \epsilon\vec{n}^1 + \cdots$, at leading order $\partial_t\vec{n}^0 = \tilde{\mathbb{S}}\cdot\tilde{\vec{J}}^0(\vec{n}^0)$. With time-independent coupling to reservoirs, this will either describe relaxation to a NESS or blow up. Assume that we reach a NESS. Then at the next order we have an equation of the form (22), with $\mathbb{S}$ replaced by $\tilde{\mathbb{S}}$. This will generally have long-time divergences unless $\vec{n}^0$ depends on the slow time τ. Let ζ be a basis of the conserved quantities of $\tilde{\mathbb{S}}$, i.e., $\tilde{\mathbb{S}}^T\cdot\zeta = 0$. Then the slow-dynamics equation is again (23). The differences from the previous analysis are that now (i) the conserved quantities do not necessarily correspond to elements, since the leading-order reservoirs will break some conservation laws, and (ii) $\vec{n}^0$ is not a known function of the slow DOF $\vec{\eta}$. This latter fact means that although true, Eq. (23) cannot be solved without constitutive information on the $\vec{n}^0(\vec{\eta})$ relationship. Moreover, this unknown relationship $\vec{n}^0(\vec{\eta})$ will in general involve the bulk reaction rates, unlike the detailed-balanced case. The main result here is that we know the dimension of the reduced dynamics, equal to the number of conserved quantities of the NESS.

A special case that can be fully analyzed is that of an open system with no emergent cycles. In this case the system will settle to a detailed-balanced equilibrium, in which all reservoirs are equilibrated. For example, consider reservoirs of H_2O, H_2, CO_2, and NH_3. The solution (13) will hold if the reservoir fluxes all vanish: $0 = J_k$ for the four reservoir species. Labeling these species as 1, 2, 3, and 4 and absorbing equilibrium concentrations into the reservoir concentrations as in Sec. V E, these become

$$0 = \tilde{z}_1 - e^{2\eta_{\mathrm{H}}}e^{\eta_{\mathrm{O}}}, \tag{59}$$

$$0 = \tilde{z}_2 - e^{\eta_{\mathrm{H}}}, \tag{60}$$

$$0 = \tilde{z}_3 - e^{\eta_{\mathrm{C}}}e^{2\eta_{\mathrm{O}}}, \tag{61}$$

$$0 = \tilde{z}_4 - e^{\eta_{\mathrm{N}}}e^{3\eta_{\mathrm{H}}}, \tag{62}$$

which are trivially solved for the η. This case then reduces to that of closed detailed-balanced systems.

VII. DISCUSSION AND CONCLUSIONS

We have shown that slowly driven CRNs are governed by the conserved quantities of the corresponding closed system. The latter are generally the element concentrations, giving a huge reduction in DOF in large CRNs. The natural dynamical variables of the slow dynamics are chemical potentials $\vec{\eta}$, which evolve according to (25). From the solution of this equation, which does not involve the bulk reaction rates, one can obtain the full dynamics, at leading order in driving. Moreover, in this limit one can easily probe the structure of the NESS for large realistic CRNs.

This framework may be useful to understand free-energy transduction in open CRNs [15]. Indeed, open CRNs can be considered as chemical machines that interconvert species between the reservoirs. For example, the early Earth CRN considered above, when coupled to reservoirs of CH_2O, H, CH_4O, and $C_2H_6O_2$, has effective reactions (emergent cycles)

$$CH_2O + 2H \rightleftharpoons CH_4O, \tag{63}$$

$$CH_2O + CH_4O \rightleftharpoons C_2H_6O_2. \tag{64}$$

The free-energy change in an emergent cycle is obtained straightforwardly from the chemical potentials of the species, but to understand the efficiency of the chemical machine, one requires the flux through the cycle, except in special cases [15]. Generally, this necessitates the entire suite of reaction rates and a numerical solution of the rate equations. A limiting factor in the analysis is then that many rates are not known for biochemical networks of interest.

Our analysis provides an alternative. For any given forcing, one can solve the slow-dynamics equation, without knowledge of any bulk reaction rates. With this solution in hand, one can evaluate the reaction fluxes and then the efficiency. This solution is guaranteed to work in the limit of slow driving and can provide a benchmark value at finite-rate driving.

For similar reasons, our theory may be useful in conjunction with a circuit theory for CRNs [16]. In the latter, a CRN is coarse grained by treating subsets of CRNs as chemical modules, connected to each other by particular species. For each module, the theory requires the relationship between the flux through emergent cycles and the concentrations of chemostatted species. If a module is treated as fast compared to its external connections, then our theory can be used to find the current-concentration relations, as shown in [24] for an example from [16]. This is particularly useful for large, poorly characterized modules where our method does not require the bulk reaction rates.

How does one know when a CRN is slowly driven? First, it follows immediately from (13) that ratios of concentrations that are stoichiometrically equivalent, for example, products and reactants of any bulk reaction, will be constant, to leading order in ϵ. This is useful if many concentrations can be tracked. More generally, one can attempt to estimate the underlying dimension of the CRN [39]. If small, it is a strong indication that the dynamics is occurring on a slow manifold due to timescale separation.

Independence of the slow dynamics with respect to bulk activation energies is a strong generic form of robustness applicable to all chemical machines whose core is detailed balanced. Moreover, dissipation is minimized, since the system is always close to some equilibrium state. The price of this robustness is that the system responds slowly to outside forcing. Whether this slow dynamics is relevant for real-world chemical machines then depends on system-specific tradeoffs between robustness and speed.

We note that an alternative weak-driving theory has been obtained in [40], also using the Hamilton-Jacobi equation. This theory is based upon the log-probability correction ρ_1, but without the multiple-timescale analysis. This is sufficient for nonequilibrium steady states as considered in [40], but in dynamical problems it will generally suffer from long-time divergences.

Our reduction of complex dynamics to that of the conserved quantities is reminiscent of hydrodynamics. There are however some differences. First, in hydrodynamics one assumes that the system is coupled to other systems that differ only weakly from it. Here instead we do not assume that the reservoirs are near the system; their concentrations can be arbitrarily far from the corresponding concentration in the system. We only assume that they react slowly with the system. Second, in hydrodynamics one considers systems that interact spatially, whereas our system is well mixed and interacts with external reservoirs without any explicit spatial coupling. The extension of our results to include spatial effects is left for future work.

VIII. METHODS

For constructing the methane combustion model, the toolbox Stoichiometry Tools in MATLAB was used [41]. The input $\{CH_4, CO_2, H_2O, CO, O_2, H_2\}$ led to the reactions (41).

For larger models, we used Reaction Mechanism Generator (RMG) [42,43]. Given an input pool of species, RMG iteratively finds possible reactions between the species and new species that can be produced. Rates, enthalpies, and entropies are either looked up in a database or estimated using additivity methods. For the early Earth model, we used the input set $\{CH_2O, NH_3, H_2O\}$, with H_2O as a solvent, into RMG. It results in the 40 reactions and 10 new species shown in [24].

We note that RMG is not guaranteed to find all possible reactions among species. In testing against known CRNs relevant to the origin of life, we found that RMG sometimes failed to find reactions known to be possible. Thus we used it as a method to benchmark our framework against CRNs with valid stoichiometry and a broad range of realistic rates.

For the autotrophic core CRN, we worked only at $T \to \infty$ so that no bulk reaction rates or chemical potentials were needed.

All ordinary differential equations were integrated in MATLAB using the `ode15s` solver.

In the Supplemental Material [24] the autotrophic core reservoirs are listed along with their KEGG IDs [44].

ACKNOWLEDGMENTS

We are grateful to Mark Persic for his preliminary work on this project. This work was funded by NSERC Discovery Grant No. RGPIN-2020-04762 (E.D.G.).

[1] M. S. Okino and M. L. Mavrovouniotis, Simplification of mathematical models of chemical reaction systems, Chem. Rev. **98**, 391 (1998).

[2] O. Radulescu, A. N. Gorban, A. Zinovyev, and V. Noel, Reduction of dynamical biochemical reactions networks in computational biology, Front. Genet. **3**, 131 (2012).

[3] T. J. Snowden, P. H. van der Graaf, and M. J. Tindall, Methods of model reduction for large-scale biological systems: A survey of current methods and trends, Bull. Math. Biol. **79**, 1449 (2017).

[4] K. R. Schneider and T. Wilhelm, Model reduction by extended quasi-steady-state approximation, J. Math. Biol. **40**, 443 (2000).

[5] P. C. Bressloff, Stochastic switching in biology: From genotype to phenotype, J. Phys. A: Math. Theor. **50**, 133001 (2017).

[6] P. Liangrong and L. Hong, Thermodynamics for reduced models of chemical reactions by PEA and QSSA, Phys. Rev. Res. **6**, 013296 (2024).

[7] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, 2000).

[8] Z. Ren, S. B. Pope, A. Vladimirsky, and J. M. Guckenheimer, The invariant constrained equilibrium edge preimage curve method for the dimension reduction of chemical kinetics, J. Chem. Phys. **124**, 114111 (2006).

[9] C. H. Lee and H. G. Othmer, A multi-time-scale analysis of chemical reaction networks: I. deterministic systems, J. Math. Biol. **60**, 387 (2010).

[10] A. Desoeuvres, A. Iosif, C. Lüders, O. Radulescu, H. Rahkooy, M. Seiss, and T. Sturm, Reduction of chemical reaction networks with approximate conservation laws, SIAM J. Appl. Dynamical Syst. **23**, 256 (2024).

[11] T. Schmiedl and U. Seifert, Stochastic thermodynamics of chemical reaction networks, J. Chem. Phys. **126**, 044101 (2007).

[12] U. Seifert, Stochastic thermodynamics, fluctuation theorems and molecular machines, Rep. Prog. Phys. **75**, 126001 (2012).

[13] C. van den Broeck, *Stochastic Thermodynamics: A Brief Introduction* (IOS, Amsterdam, 2013), pp. 155–193.

[14] A. Wachtel, R. Rao, and M. Esposito, Thermodynamically consistent coarse graining of biocatalysts beyond Michaelis–Menten, New J. Phys. **20**, 042002 (2018).

[15] A. Wachtel, R. Rao, and M. Esposito, Free-energy transduction in chemical reaction networks: From enzymes to metabolism, J. Chem. Phys. **157**, 024109 (2022).

[16] F. Avanzini, N. Freitas, and M. Esposito, Circuit theory for chemical reaction networks, Phys. Rev. X **13**, 021041 (2023).

[17] M. Polettini and M. Esposito, Irreversible thermodynamics of open chemical networks. I. Emergent cycles and broken conservation laws, J. Chem. Phys. **141**, 024117 (2014).

[18] R. Rao and M. Esposito, Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics, Phys. Rev. X **6**, 041064 (2016).

[19] R. D. Astumian, Stochastic pumping of non-equilibrium steady-states: How molecules adapt to a fluctuating environment, Chem. Commun. **54**, 427 (2018).

[20] R. G. Gilbert and S. C. Smith, *Theory of Unimolecular and Recombination Reactions* (Blackwell Scientific, Oxford, 1990).

[21] A. De Martino, C. Martelli, and F. A. Massucci, On the role of conserved moieties in shaping the robustness and production capabilities of reaction networks, Europhys. Lett. **85**, 38007 (2009).

[22] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (Springer, Berlin, 1999), Vol. I.

[23] E. J. Hinch, *Matched Asymptotic Expansions*, 1st ed. (Cambridge University Press, Cambridge, 1991).

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.109.044105 for additional analytical derivations and details of example reaction networks.

[25] F. Schlögl, Chemical reaction models for non-equilibrium phase transitions, Z. Phys. **253**, 147 (1972).

[26] H. Ge and H. Qian, Thermodynamic limit of a nonequilibrium steady state: Maxwell-type construction for a bistable biochemical system, Phys. Rev. Lett. **103**, 148103 (2009).

[27] M. Vellela and H. Qian, Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: The Schlögl model revisited, J. R. Soc. Interface **6**, 925 (2009).

[28] M. Doi, Second quantization representation for classical many-particle system, J. Phys. A: Math. Gen. **9**, 1465 (1976).

[29] L. Peliti, Path integral approach to birth-death processes on a lattice, J. Phys. (Paris) **46**, 1469 (1985).

[30] E. De Giuli and C. Scalliet, Dynamical mean-field theory: From ecosystems to reaction networks, J. Phys. A: Math. Theor. **55**, 474002 (2022).

[31] R. Kubo, K. Matsuo, and K. Kitahara, Fluctuation and relaxation of macrovariables, J. Stat. Phys. **9**, 51 (1973).

[32] K. Kitahara, *Advances in Chemical Physics*, edited by G. Nicolis and R. Lefever (Wiley, New York, 1975), Vol. 29, p. 85.

[33] M. I. Dykman, E. Mori, J. Ross, and P. Hunt, Large fluctuations and optimal paths in chemical kinetics, J. Chem. Phys. **100**, 5735 (1994).

[34] E. Smith, Intrinsic and extrinsic thermodynamics for stochastic population processes with multi-level large-deviation structure, Entropy **22**, 1137 (2020).

[35] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. **81**, 2340 (1977).

[36] W. Martin, J. Baross, D. Kelley, and M. J. Russell, Hydrothermal vents and the origin of life, Nat. Rev. Microbiol. **6**, 805 (2008).

[37] E. Smith and H. J. Morowitz, *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere* (Cambridge University Press, Cambridge, 2016).

[38] J. L. Wimmer, A. d. N. Vieira, J. C. Xavier, K. Kleinermanns, W. F. Martin, and M. Preiner, The autotrophic core: An ancient network of 404 reactions converts H_2, CO_2, and NH_3 into amino acids, bases, and cofactors, Microorganisms **9**, 458 (2021).

[39] A. Blokhuis, M. van Kuppeveld, D. van de Weem, and R. Pollice, On data and dimension in chemistry I—irreversibility, concealment and emergent conservation laws, arXiv:2306.09553.

[40] N. Freitas, G. Falasco, and M. Esposito, Linear response in large deviations theory: A method to compute non-equilibrium distributions, New J. Phys. **23**, 093003 (2021).

[41] J. Kantor, Stoichiometry Tools, MATLAB Central File Exchange (2024), https://www.mathworks.com/matlabcentral/fileexchange/29774-stoichiometry-tools.

[42] C. W. Gao, J. W. Allen, W. H. Green, and R. H. West, Reaction mechanism generator: Automatic construction of chemical kinetic mechanisms, Comput. Phys. Commun. **203**, 212 (2016).

[43] M. Liu, A. Grinberg Dana, M. S. Johnson, M. J. Goldman, A. Jocher, A. M. Payne, C. A. Grambow, K. Han, N. W. Yee, and E. J. Mazeau, Reaction mechanism generator v3. 0: Advances in automatic mechanism generation, J. Chem. Inf. Model. **61**, 2686 (2021).

[44] M. Kanehisa and S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. **28**, 27 (2000).