communications biology

ARTICLE

https://doi.org/10.1038/s42003-021-02117-x OPEN

Derivation of stationary distributions of biochemical reaction networks via structure transformation

Hyukpyo Hong [1,2,10], Jinsu Kim [3,4,10], M. Ali Al-Radhawi[5], Eduardo D. Sontag[5,6,7,8,9] & Jae Kyoung Kim [1,2]✉

Long-term behaviors of biochemical reaction networks (BRNs) are described by steady states in deterministic models and stationary distributions in stochastic models. Unlike deterministic steady states, stationary distributions capturing inherent fluctuations of reactions are extremely difficult to derive analytically due to the curse of dimensionality. Here, we develop a method to derive analytic stationary distributions from deterministic steady states by transforming BRNs to have a special dynamic property, called complex balancing. Specifically, we merge nodes and edges of BRNs to match in- and out-flows of each node. This allows us to derive the stationary distributions of a large class of BRNs, including autophosphorylation networks of EGFR, PAK1, and Aurora B kinase and a genetic toggle switch. This reveals the unique properties of their stochastic dynamics such as robustness, sensitivity, and multi-modality. Importantly, we provide a user-friendly computational package, CASTANET, that automatically derives symbolic expressions of the stationary distributions of BRNs to understand their long-term stochasticity.

[1] Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. [2] Biomedical Mathematics Group, Institute for Basic Science, Daejeon, Republic of Korea. [3] Department of Mathematics, University of California, Irvine, CA, USA. [4] NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA, USA. [5] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. [6] Department of Bioengineering, Northeastern University, Boston, MA, USA. [7] Department of Mathematics, Northeastern University, Boston, MA, USA. [8] Department of Chemical Engineering, Northeastern University, Boston, MA, USA. [9] Laboratory of Systems Pharmacology, Program in Therapeutic Science, Harvard Medical School, Boston, MA, USA. [10] These authors contributed equally: Hyukpyo Hong, Jinsu Kim. ✉email: jaekkim@kaist.ac.kr

A standard approach to mathematical modeling of biochemical reaction networks (BRNs) is to use ordinary differential equations (ODEs), whose variables represent concentrations of molecules[1]. However, this deterministic description, while convenient for computation, by its nature cannot capture the inherent randomness of BRNs. In particular, the long-term behavior of ODE systems is characterized by steady states or other attractors, rather than by the stationary distributions statistically observed in real biological systems. As cell biology moves away from bulk averages to single-cell measurements, a focus has shifted to the study of such stationary distributions[2,3]. They can be described by various stochastic approaches[1,4]. In particular, stationary distributions can be described as steady-state solutions of the chemical master equation (CME), which has been widely used to describe the time evolution of the probabilities for the numbers of chemical species in BRNs such as gene regulatory networks and signaling pathways[5].

Since the CME is a differential equation with infinitely many variables, its steady-state solution (i.e., the stationary distribution) can be found analytically only for simple cases, such as linear reaction networks[6] or birth-death processes[7]. Unlike the CME, its deterministic counterpart is a finite dimensional ODE, whose steady-state solutions are relatively easier to calculate. An interesting question, therefore, is whether there is a systematic way of using these deterministic steady states for characterizing the stationary distribution of the stochastic counterpart. There is a positive answer to this question for special networks, called complex balanced networks.

A result from queuing theory[8], reinterpreted in the context of BRNs[9] through the connection between Petri nets and BRNs[10], shows that for complex balanced networks whose kinetics are described by mass action reactions, stationary distributions can be characterized in terms of jointly distributed Poisson random variables with parameters corresponding to deterministic steady states. An independent proof of this result, together with deep applications to CMEs, was developed by Anderson, Craciun, and Kurtz[11]. Complex balancing is difficult to check and depends on rate constant values. However, beautiful work by Horn, Jackson, and Feinberg[12–14] has shown that all networks that have the special structural properties of weak reversibility and zero deficiency are complex balanced, independently of rate constants. Weak reversibility of a network means that the network is a union of closed reaction cycles, and the deficiency of a network is the number of dependent closed reaction cycles, which can be easily checked. Satisfying these two structural properties is a simple condition to derive the stationary distribution of network under mass action reactions with the method in ref.[11].

As various BRNs such as networks of several reversible reactions (e.g., $A + B \leftrightarrow C \leftrightarrow 0$) or cyclic reactions (e.g., $A \to B \to C \to A$) are weakly reversible and deficiency zero, their stationary distributions can be analytically derived[15–20]. These have been used to characterize the stochasticity of various systems, including a genetic oscillator[21] and a competitive inhibition enzyme kinetics model[22]. Unfortunately, the majority of BRNs do not have the special network structure. For instance, only ~0.36% of the Erdös-Rényi random networks of two species with up to bimolecular reactions have a deficiency of zero when the edge probability is 0.5, and the fraction decreases to zero as the number of species increases[23]. Moreover, from a biological standpoint, even simple networks are unlikely to be weakly reversible if they include a bimolecular reaction whose reverse reaction is unimolecular (e.g., autophosphorylation and dephosphorylation).

Here, we develop a framework to derive stationary distributions for a class of networks which do not have the special structure (i.e., weakly reversibility and zero deficiency) by modifying their structures via *network translation*[24,25]. Specifically, by simply merging reactions with a common stoichiometric vector and shifting reactions in the networks, we are able to change their structure to be weakly reversible and deficiency zero while preserving their stochastic dynamics. This allows us to derive the stationary distributions of a large class of BRNs including autophosphorylation networks of EGFR, PAK1, and Aurora B kinase. This derivation reveals key reactions determining the autophosphorylation status, which can seldom be done with a purely numerical approach. Furthermore, we describe how the stochastic dynamics of more complex BRNs can be tracked when our method is applicable for only their subnetworks. Importantly, we provide a user-friendly computational package CASTANET (Computational package for deriving Analytical STAtionary distributions of biochemical reaction networks with NEtwork Translation) that automatically derives the stationary distributions of submitted BRNs via our method. This will provide an effective tool to analyze the stochasticity of BRNs.

Results

Obtaining the desired network structure via network translation. As mentioned in the introduction, the stationary distributions of the stochastic mass action models for BRNs can be derived with any choice of rate constants using the previous method[11] if and only if the networks have two structural properties: *weak reversibility* and *zero deficiency*. However, even very simple networks such as the one shown in Fig. 1a-left fail to satisfy the two properties. Weak reversibility means that if there exists a path from a *complex* (i.e., a node in the reaction graph) to another complex, then there is a reverse path from the second one back to the first one. Because there is no path from $A + A$ to A while there is a path from A to $A + A$, the network in Fig. 1a-left is not weakly reversible. The deficiency of a network is a non-negative integer index calculated by subtracting both the number of linkage classes (i.e., connected components in the reaction graph) and the dimension of the subspace spanned by the stoichiometric vectors from the number of complexes. The deficiency of the network in Fig. 1a-left is one. Therefore, the previous method[11] cannot be used to derive its stationary distribution.

Two different reactions, $0 \to A$ and $A \to A + A$, have the same stoichiometric vector $(1, 0)$ because both reactions produce one molecule of A (Fig. 1a-left). Thus, these two reactions can be merged by unifying the source complexes 0 and A into 0 and summing the propensities of both reactions (Fig. 1a). This procedure is known as *network translation*[24,25], which was proposed to investigate deterministic systems. This procedure is also applicable to stochastic systems as it preserves the stochastic dynamics (see Supplementary Note 1 for details). For instance, the propensities of the production of A are $\alpha_1 + \alpha_4 n_A$ in both the original (Fig. 1a-left) and the translated network (Fig. 1a-right). Although the network translation is simple, it can effectively change the structure of the network to be a weakly reversible deficiency zero network.

Propensity factorization is required. Even though the translated network is weakly reversible and of zero deficiency, the new model no longer follows mass action kinetics since the propensity of the reaction $0 \to A$ is not constant (Fig. 1a-right). In this case, previously, it was known that the method in refs.[11,26] is still applicable if the non-mass action propensity functions can be factorized as a certain form. However, the propensity functions of this translated network do not have the certain form. Thus, we generalize the previous factorization form so that stationary distributions can be derived for a larger class of BRNs. Specifically,

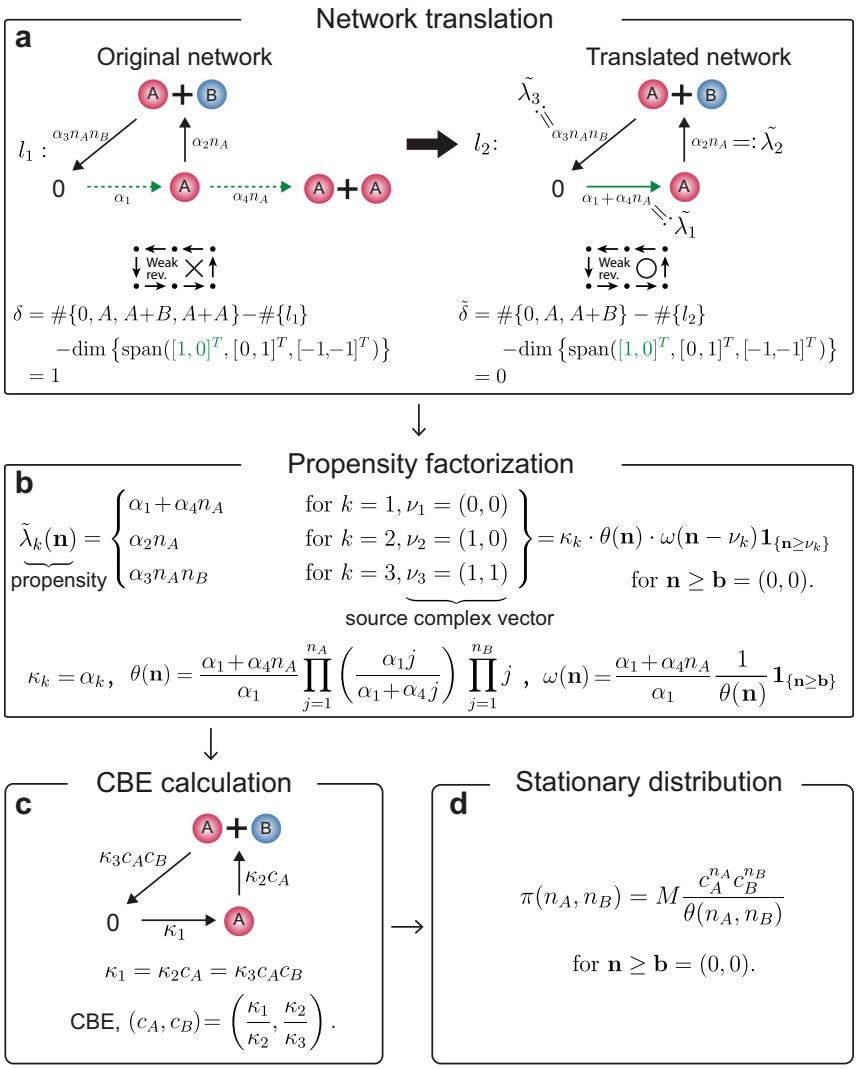


Fig. 1 Derivation of a stationary distribution with network translation. a The non-weakly reversible and deficiency (δ) one network is translated to the weakly reversible deficiency zero network by merging two reactions, which have the same stoichiometric vectors (green dotted lines). $\tilde{\lambda}_k$ denotes the propensities of the translated network. **b** Factorize $\tilde{\lambda}_k$ with constants κ_k and functions $\theta(\mathbf{n})$ and $\omega(\mathbf{n})$ as $\tilde{\lambda}_k(\mathbf{n}) = \kappa_k\theta(\mathbf{n})\omega(\mathbf{n}-\nu_k)\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}}$ on $\Gamma = \{\mathbf{n}|\mathbf{n}\geq\mathbf{b}\}$ at which $\tilde{\lambda}_k(\mathbf{n}) > 0$ if $\mathbf{n}\geq\nu_k+\mathbf{b}$, and $\tilde{\lambda}_k(\mathbf{n}) = 0$ otherwise. ν_k is the source complex vector of the kth reaction. **c** Compute a complex balanced equilibrium (CBE) of the deterministic mass action model for the translated network with rate constants $\{\kappa_k\}$. **d** Using the $\theta(\mathbf{n})$ and the CBE, the stationary distribution can be derived analytically. Here, M is a normalizing constant.

we show that all the propensities of the translated network $\tilde{\lambda}_k(\mathbf{n})$ need to be factorized as

$$\tilde{\lambda}_k(\mathbf{n}) = \kappa_k\theta(\mathbf{n})\omega(\mathbf{n}-\nu_k)\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}} \quad (1)$$

for some constants $\kappa_k > 0$ and functions $\theta(\mathbf{n}) > 0$ and $\omega(\mathbf{n}) \geq 0$ on a set $\Gamma = \{\mathbf{n}|\mathbf{n}\geq\mathbf{b}\}$ where the ν_k is the source complex vector of the kth reaction, the inequality is coordinate-wise, and the $\mathbf{b}$ needs to be chosen so that $\tilde{\lambda}_k(\mathbf{n}) > 0$ if and only if there are sufficient reactants (i.e., $\mathbf{n}\geq\nu_k+\mathbf{b}$) in Γ. For the translated network (Fig. 1a-right), $\tilde{\lambda}_k(\mathbf{n}) > 0$ if and only if $\mathbf{n}\geq\nu_k$ like mass action kinetics, and thus $\mathbf{b} = (0,0)$ and $\Gamma = \mathbb{Z}^2_{\geq 0}$.

Propensity functions satisfying the factorization condition (Eq. (8)) include a generalized mass action kinetics (Eq. (7)). For instance, if a source complex is $0, A, A+A$, or $A+B$, propensity functions following the generalized mass action kinetics are proportional to 1, $f_A(n_A)$, $f_A(n_A)f_A(n_A-1)$, or $f_A(n_A)f_B(n_B)$, respectively. Note that if the f_i's are identity functions then the propensities follow standard mass action kinetics (Eq. (6))[11,27]. The propensity functions following the generalized mass action kinetics can be easily factorized with $\theta(\mathbf{n}) = \omega(\mathbf{n})^{-1} = \prod_{i=1}^{d}\prod_{j=b_i+1}^{n_i} f_i(j)$, where d is the number of the constitutive chemical species (see Eq. (9) for details). However, the translated network (Fig. 1a-right) does not follow the generalized mass action kinetics (Eq. (7)) because the propensity function of the reaction $0\to A$, $\alpha_1+\alpha_4 n_A$, is not proportional to 1 (i.e., it is not constant). Thus, we need to solve recurrence relations as described in Supplementary Note 2 to identify the propensity factorization (Fig. 1b):

$$\begin{aligned}\kappa_k &= \alpha_k,\\ \theta(\mathbf{n}) &= \frac{\alpha_1+\alpha_4 n_A}{\alpha_1}\prod_{j=1}^{n_A}\left(\frac{\alpha_1 j}{\alpha_1+\alpha_4 j}\right)\prod_{j=1}^{n_B} j,\\ \text{and } \omega(\mathbf{n}) &= \frac{\alpha_1+\alpha_4 n_A}{\alpha_1}\frac{1}{\theta(\mathbf{n})}\end{aligned} \quad (2)$$

Derivation of stationary distribution. After identifying κ_k, $\theta(\mathbf{n})$, and $\omega(\mathbf{n})$ via the propensity factorization, we need to find a

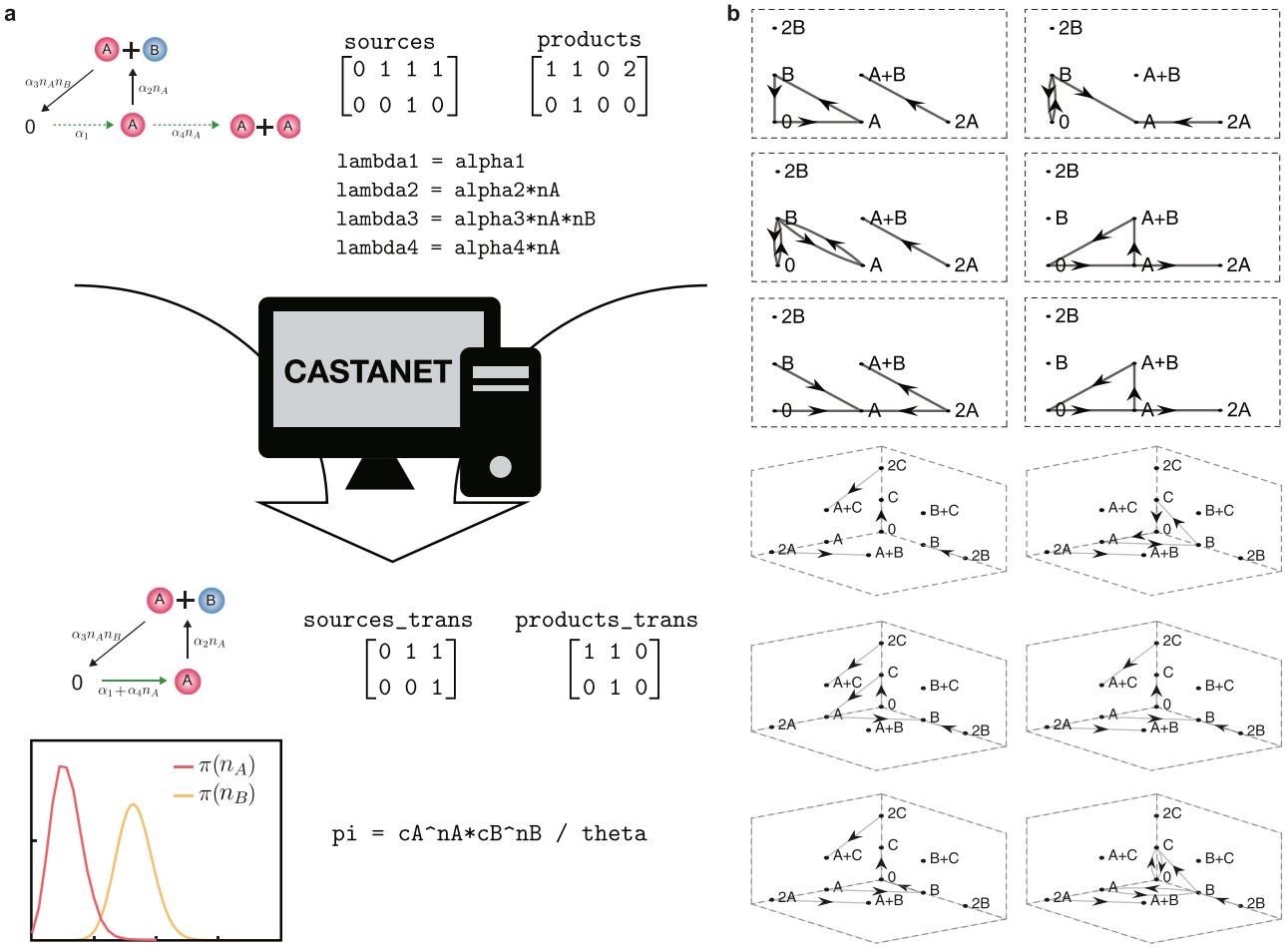


Fig. 2 CASTANET (Computational package for deriving Analytical STAtionary distributions of biochemical reaction networks with NEtwork Translation). a A schematic diagram for the computational package. If users simply enter the source complexes, product complexes, and propensity functions of reactions (`lambda_k`), then the package identifies a weakly reversible deficiency zero translated BRN (`sources_trans` and `products_trans`) and then derives its stationary distribution (`pi`). See Supplementary Note 4 and Supplementary Fig. 2 for a step-by-step manual. **b** BRNs with two species (top) and three species (bottom) whose stationary distributions were calculated by our computational package. The tail and head of each arrow represent the source and product complexes of reactions, respectively. They are assumed to follow the stochastic mass-action kinetics, and the rate constants can take any positive values. See Supplementary Figs. 3 and 4 for more examples. Here, each network is embedded in euclidean space where we present A+A and B+B as 2A and 2B, respectively.

complex balanced equilibrium (CBE) of the deterministic mass action model with rate constants $\{\kappa_k\}$ for the translated network (Fig. 1c). The CBE is a steady state at which for each complex ν, the in-flow rate to ν is equal to the out-flow rate from ν[12]. For instance, based on the deterministic model in Fig. 1c, the complex balance conditions for the complexes 0, A, and $A+B$ are $\kappa_3 c_A c_B = \kappa_1$, $\kappa_1 = \kappa_2 c_A$, and $\kappa_2 c_A = \kappa_3 c_A c_B$, respectively. By solving these equations, we can obtain the CBE, $(c_A, c_B) = (\kappa_1/\kappa_2, \kappa_2/\kappa_3)$. Note that the existence of a CBE is guaranteed because we translate a network to be weakly reversible and deficiency zero[13].

Finally, using the function $\theta(\mathbf{n})$ (Fig. 1b) and the CBE (c_A, c_B) (Fig. 1c), we can derive the stationary distribution of the stochastic model for the translated network, which is the same as that of the original network, as follows:

$$\pi(n_A, n_B) = M \frac{c_A^{n_A} c_B^{n_B}}{\theta(n_A, n_B)} \tag{3}$$

for $n_A \geq 0, n_B \geq 0$ where M is the normalizing constant so that the sum of the stationary distribution is one (see Methods for details). In this example, the distribution $\pi(\mathbf{n})$ is obtained on $\Gamma = \mathbb{Z}^2_{\geq 0}$. This state space is closed as proved in Supplementary information, and it is irreducible (i.e., every state is reachable from every other state; see Supplementary Note 3 for details). On the other hand, if an irreducible state space is a proper subset of Γ, possibly due to a conservation law, then the normalizing constant M is chosen so that the sum of $\pi(\mathbf{n})$ over the subset is one.

Computational package, CASTANET. Applying our theoretical framework (Fig. 1) has two practical difficulties. Translating a given network to a weakly reversible deficiency zero network (Fig. 1a) is not straightforward as prohibitively many candidates of translated networks often exist. Furthermore, it is challenging to check whether the factorization condition holds (Fig. 1b) as it requires to solve associated recurrence relations. Thus, we have developed a user-friendly, open-source, and publicly available computational package, "CASTANET (https://github.com/Mathbiomed/CASTANET)," that automatically performs network translation and propensity factorization and derives stationary distributions (Fig. 2a). With this package, we were able to easily identify hundreds of BRNs and derive analytic forms of their stationary distributions. We have provided some of them in Fig. 2b and Supplementary Figs. 3 and 4. To use this package,

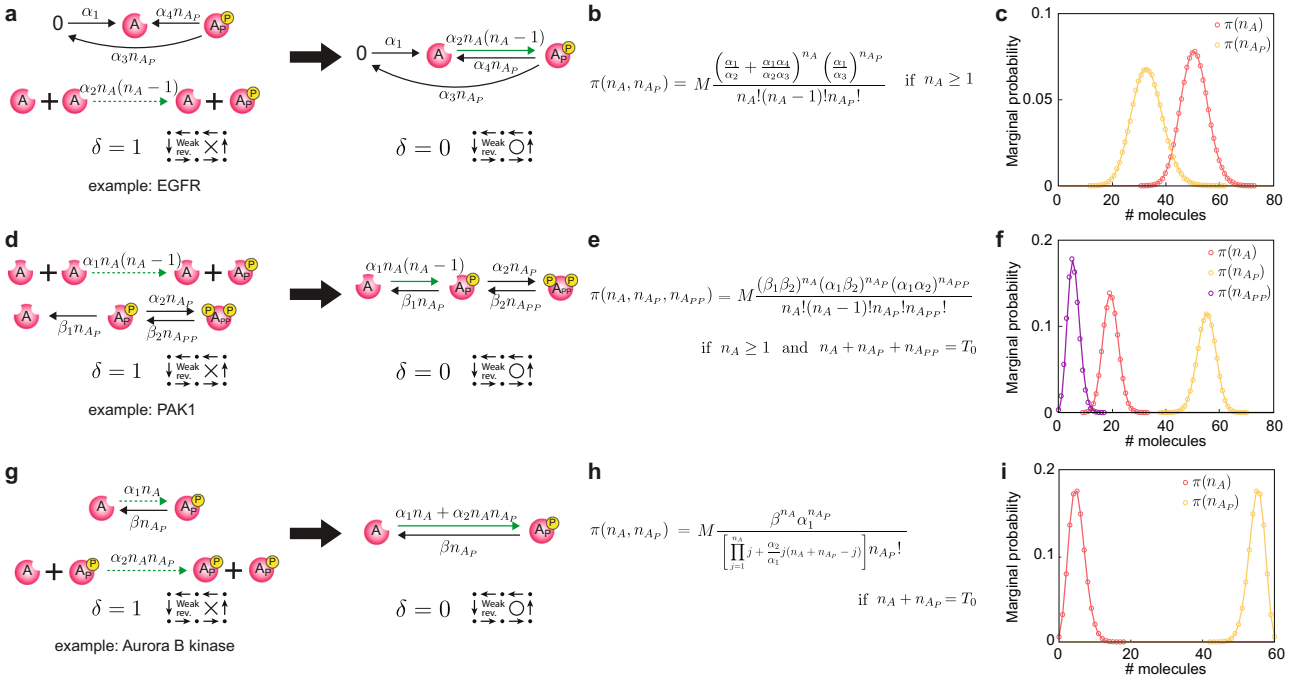


Fig. 3 Stationary distributions of diverse autophosphorylation networks. a, **d**, **g** While the autophosphorylation networks have deficiencies of one and are not weakly reversible, they can be translated to weakly reversible deficiency zero networks. **b**, **e**, **h** Thus, stationary joint distributions can be derived using the method illustrated in Fig. 1. T_0 in **e** and **h** represent the total numbers of proteins, which are conserved. **c**, **f**, **i** The marginal probabilities of the numbers of species derived from the formula (solid lines) and stochastic simulations (dots) are consistent. Here, parameter values are set as follows: **a** $\alpha_1 = 10$, $\alpha_2 = 0.03$, $\alpha_3 = 0.3$, $\alpha_4 = 2$, **d** $\alpha_1 = 0.3$, $\alpha_2 = 0.1$, $\beta_1 = 2$, $\beta_2 = 1$, $T_0 = 80$, **g** $\alpha_1 = 0.001$, $\alpha_2 = 1$, $\beta = 5$, $T_0 = 60$. For each example, 10^5 simulations were performed using the Gillespie algorithm[60].

users only need to enter the source complexes, product complexes, and propensity functions of reactions.

Stationary distributions of autophosphorylation networks. Our theoretical framework and especially CASTANET extend the class of BRNs whose stationary distributions can be analytically derived using CBEs (Figs. 1 and 2). This class includes various autophosphorylations networks (Fig. 3) that are not weakly reversible due to autophosphorylation reactions, which occur in intermolecular (trans), intramolecular (cis), or mixed manners[28,29].

Asymmetric trans-autophosphorylation occurs if two monomers form a homodimer and one of them acts as an 'enzyme' and phosphorylates the other. This type of autophosphorylation occurs in the epidermal growth factor receptor (EGFR), which triggers signal transduction for cell proliferation[30]. The key regulatory reactions for EGFR include its synthesis, trans-autophosphorylation, dephosphorylation, and degradation (Fig. 3a-left). The asymmetric trans-autophosphorylation is a reaction that transforms the complex $A + A$ to $A + A_P$. The dephosphorylation reaction is not the reverse of the previous reaction; instead, it occurs from the complex A_P to the complex A. Thus, the network is not weakly reversible. However, CASTANET automatically identifies a weakly reversible deficiency zero translated network and its propensity factorization (Fig. 3a-right) and then derives the analytic form of stationary distribution (Fig. 3b) that matches what is calculated with stochastic simulations (Fig. 3c).

In addition, having the formula (Fig. 3b) allows us to easily understand the long-term behavior of the system, something that is not possible with a purely computational approach. For instance, $\pi(n_{A_P})$ in Fig. 3c is the Poisson distribution with rate α_1/α_3. This indicates that the synthesis (α_1) and degradation rates (α_3) of A are the determinants of the long-term status of n_{A_P}, which is surprisingly robust to the changes of phosphorylation (α_2) and dephosphorylation rates (α_4). Furthermore, $\pi(n_A)$ is solely determined by $\frac{\alpha_1}{\alpha_2}(1 + \frac{\alpha_4}{\alpha_3})$, and its moments can be calculated with the modified Bessel functions (see Supplementary Note 3 for details). This allows us to identify that the stationary distribution of n_A is sub-Poissonian, and its coefficient of variation attains the maximum at $\frac{\alpha_1}{\alpha_2}(1 + \frac{\alpha_4}{\alpha_3}) \approx 1.8$ (Supplementary Fig. 1).

Trans- and cis-autophosphorylation can occur sequentially. For example, p21-activated kinase 1 (PAK1), which regulates cell motility and morphology, phosphorylates a threonine residue in the kinase domain in a trans manner asymmetrically ($A + A \rightarrow A + A_P$)[28,31] and then phosphorylates a serine residue in the regulatory domain of itself in a cis manner ($A_P \rightarrow A_{PP}$)[32,33] (Fig. 3d-left). While the original network of PAK1 is not weakly reversible (Fig. 3d-left), CASTANET identifies a weakly reversible deficiency zero translated network (Fig. 3d-right) and derives the analytic form of stationary distribution (Fig. 3e) that matches the simulation result (Fig. 3f).

Both trans- and cis-autophosphorylation can occur simultaneously as in Aurora B kinase, which controls mitotic progression[34]. In an Aurora B kinase network, cis-autophosphorylation ($A \rightarrow A_P$) promotes rapid trans-autophosphorylation ($A + A_P \rightarrow A_P + A_P$), which forms a positive feedback in the system[34]. For this network, CASTANET successfully applies our method to derive the analytic form of stationary distribution (Fig. 3g–i).

While mass action kinetics are commonly used to describe autophosphorylations[34,35] as in our examples, the Michaelis–Menten function and Hill function are also often used[36,37]. Moreover, they are also used to describe the effects of phosphatases on dephosphorylation and proteasomes on degradation[38,39], which EGFR, PAK1, and Aurora B kinase undergo[40–45]. Even when the mass action propensities in the

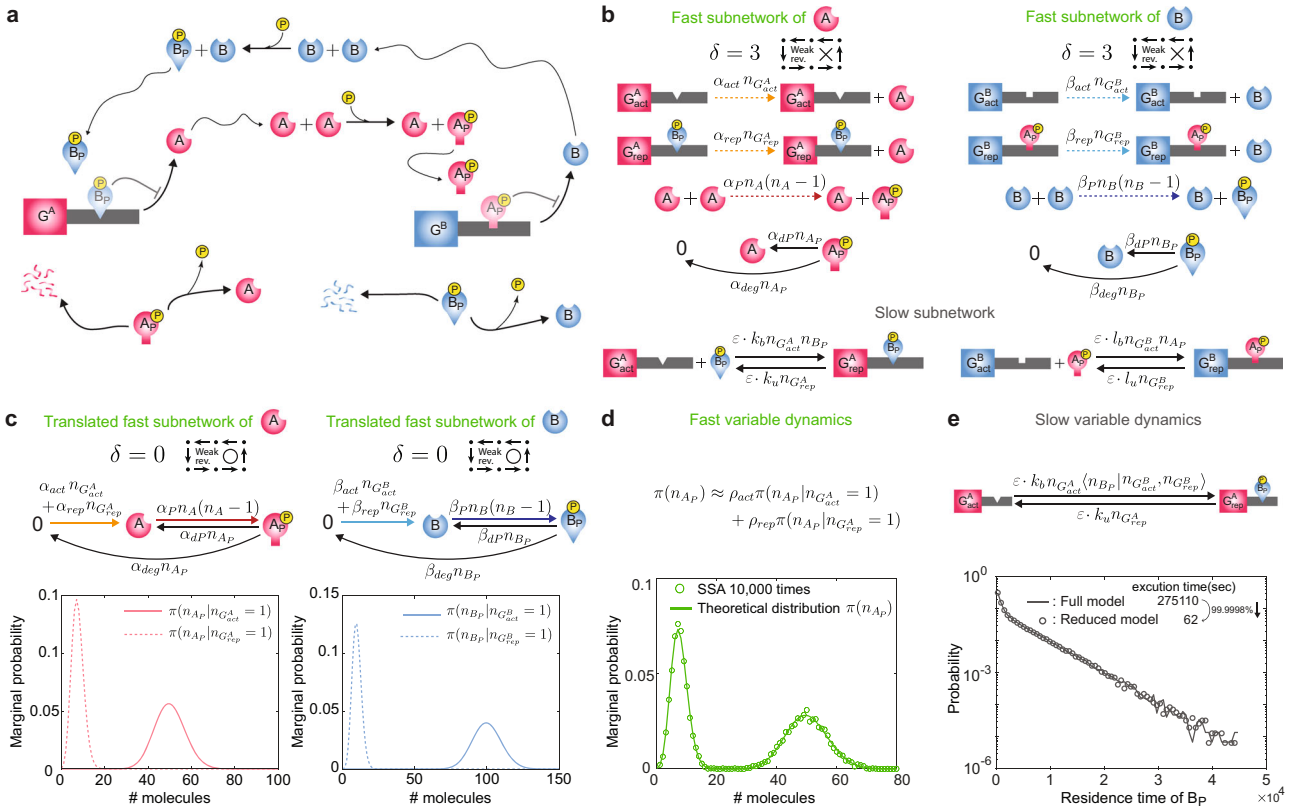


Fig. 4 Fast and slow dynamics of a multi-timescale system are identified via network translation. a Schematic diagram of a toggle switch system. **b** Protein kinetics are fast but promoter kinetics (activation and repression of the genes) are slow. The fast subnetworks for *A* and *B* are not weakly reversible and have deficiencies of three. **c** Translated fast subnetworks, obtained by merging reactions having the same stoichiometric vectors (colored arrows), are weakly reversible and deficiency zero. Stationary distributions of the number of the phosphorylated proteins conditioned on the gene states (activated/repressed) are derived by using the method illustrated in Fig. 1. **d** The unconditional stationary distribution of a fast variable, which is bimodal, can be accurately approximated by the weighted average of the conditional stationary distributions. The weights ρ_{act} and ρ_{rep} are the probabilities that the gene G^A is active and repressed, respectively. **e** By replacing the fast variable n_{B_P} with its conditional stationary moment $\langle n_{B_P}|n_{G^B_{act}}, n_{G^B_{rep}}\rangle$ determined by the current state of the slow variables $n_{G^B_{act}}$ and $n_{G^B_{rep}}$, the reduced model can be derived, which can capture the slow dynamics of the genes. For instance, the reduced model (dots) accurately captures the residence time distributions of the repressor B_P to its target gene G^A of the full model (solid line). The execution times for performing 10^4 simulations with the full and the reduced models are 275110 and 62 seconds, respectively. Parameter values are set as: $\epsilon = 10^{-5}$, $\alpha_{act} = 10$, $\alpha_{rep} = 1.5$, $\alpha_P = 0.2$, $\alpha_{dP} = 1$, $\alpha_{deg} = 0.2$, $\beta_{act} = 10$, $\beta_{rep} = 1$, $\beta_P = 0.3$, $\beta_{dP} = 2$, $\beta_{deg} = 0.1$, $k_b = 1$, $k_u = 30$, $l_b = 1.3$, and $l_u = 20$.

networks (Fig. 3a, d, and g-left) are replaced with the Michaelis-Menten or Hill functions, their stationary distributions can still be derived with the same approach (Supplementary Fig. 5).

When the presented networks are extended by adding reactions, our methods might not be applicable. For instance, if an additional trans-autophosphorylation ($A + A_P \to A_P + A_P$) is added to the example in Fig. 3a, although it can be translated to a weakly reversible deficiency zero network, their propensities cannot be factorized as in Eq. (1). Thus, the stationary distribution of the extended network cannot be derived by our method. However, it can be approximated by the stationary distribution of the original network if the rate constant of the added reaction is small enough (see Supplementary Fig. 6 for details). Such approximation works for the extended networks of the other networks (Fig. 3d, g) as well. This indicates that if the stationary distributions of core subnetworks, which consist of dominant reactions, can be derived by our method, then it could be used to approximate the stationary distributions of their more complex parent networks.

Translation of fast subnetworks reveals both the fast and slow dynamics of a multi-timescale system. As the number of nodes (i.e., complexes) of networks increases, the networks are less likely to be a weakly reversible deficiency zero network even after network translation, and thus our method is less likely to be applicable. However, such large networks commonly consist of reactions occurring at different time scales[46]. In this case, if we can derive the conditional stationary distributions of only fast subnetworks with our method, both the fast and slow dynamics of the full network can be accurately captured.

For gene regulatory networks, if the promoter kinetics (i.e., binding and unbinding of transcription factors to promoters) are fast, the fast subnetwork is a simple reversible binding network (i.e., weakly reversible and of zero deficiency), and thus its stationary distribution can be easily calculated[21]. On the other hand, when the promoter kinetics are slow, the fast subnetwork includes a complex protein reaction network whose stationary distribution is challenging to derive. This can occur for a variety of reasons, e.g., the presence of nucleosomes in eukaryotic cells usually slows down the binding and unbinding of transcription factors[20].

A genetic toggle switch with the slow promoter kinetics, which consists of a pair of genes G^A and G^B, is an example of such multi-timescale system (Fig. 4a). The genes G^A and G^B express proteins *A* and *B*, respectively. Subsequently, these proteins undergo asymmetric trans-autophosphorylation, and they mutually repress gene expression by binding to the promoter

region of each other's gene. The binding and unbinding of the phosphorylated proteins occur at a slower time scale. Therefore, the entire network can be divided into the fast subnetwork consisting of gene expression, phosphorylation, dephosphorylation, and degradation and the slow subnetwork consisting of binding and unbinding (Fig. 4b). While the fast subnetworks for A and B are neither weakly reversible nor deficiency zero, they can be translated to weakly reversible deficiency zero networks (Fig. 4c-top). The propensity functions of these translated networks can be factorized as in Eq. (8), so their stationary distributions conditioned on the slow variables can be derived with CASTANET (Fig. 4c). Depending on the slowly changing gene states, the distributions of the proteins A, A_P, B, and B_P can dramatically change.

When the gene states slowly change, the fast variables rapidly equilibrate to the conditional stationary distributions determined by the current gene states (Fig. 4c). Thus, the weighted average of the conditional stationary distributions with the probabilities of the corresponding gene states accurately approximates the full (i.e., unconditional) stationary distribution under timescale separation (i.e., $\varepsilon \ll 1$)[20]. For instance, the full stationary distribution of A_P can be approximated as

$$\pi(n_{A_P}) \approx \rho_{act}\pi(n_{A_P}|n_{G^A_{act}} = 1) + \rho_{rep}\pi(n_{A_P}|n_{G^A_{rep}} = 1) \quad (4)$$

where ρ_{act} and ρ_{rep} are the probabilities that the gene G^A is active and repressed, respectively. The ρ_{act} becomes larger as the dissociation constant between G^A and its repressor B_P is larger, and the number of repressor B_P is smaller. The ρ_{act} can be calculated by identifying the eigenvector of the matrix consisting of the dissociation constant, and the conditional stationary moments of the repressors obtained from Fig. 4c[20], and $\rho_{rep} = 1 - \rho_{act}$ (see Supplementary Note 5 for details). Thus, using Eq. (4), we can accurately capture the bimodal stationary distribution of the protein A_P (Fig. 4d), leading to phenotypic heterogeneity in isogenic populations. Similarly, the full bimodal stationary distributions of the other fast variables A, B, and B_P can also be accurately captured (Supplementary Fig. 7). Note that these bimodalities cannot be captured by the corresponding deterministic model, which predicted monostability. Such mismatches between the stochastic and deterministic model have been frequently observed in the presence of timescale separation[1,20,47].

The conditional stationary distributions of the fast variables, obtained by using our approach (Fig. 4c), allow us to capture the slow dynamics of the full system as well. On the slow time scale, the slow variables are unlikely to be changed, but the fast variables rapidly equilibrate to their conditional stationary distributions for the given slow variables. Thus, by replacing the fast variables in the propensity functions of the slow reactions with their quasi-steady states (QSSs): conditional stationary moments, we can obtain the reduced model[21,48]. For the toggle switch system, the QSSs of the fast variables A_P and B_P can be computed from their conditional stationary distributions (Fig. 4c). Then by replacing the fast variables A_P and B_P with their QSSs, we can obtain the reduced model with only the slow variables, the active and repressed genes (Fig. 4e). This reduced model accurately captures the slow dynamics of the full model: the binding and unbinding of the repressors to the genes. Both the full and the reduced models yield nearly identical distributions of the residence time of the repressor B_P, which quantifies how long the repressor maintains its binding to the gene G^A (Fig. 4e). Because the reduced model does not simulate the fast reactions, which incur a large computational cost in the full model simulation, computation time decreases by 99.9998%.

Discussion

In this study, we have developed a framework and its computational package that analytically derive stationary distributions of a large class of BRNs. Specifically, we showed that the stationary distribution of a BRN can be derived if two conditions are satisfied: the network can be transformed to a weakly reversible deficiency zero network via network translation (Fig. 1a) and the propensity functions of the translated network satisfies the generalized factorization property of mass action kinetics, identified in this study (Fig. 1b). We found that these conditions are satisfied in numerous BRNs including various autophosphoryaltion networks by using CASTANET (Fig. 2, Supplementary Figs. 3 and 4). Furthermore, even when only a subnetwork of more complex BRNs satisfies the conditions, the stochastic dynamics can often be captured. That is, the stationary distribution of the subnetwork consisting of dominant reactions, derived with our method, can accurately approximate the stationary distribution of its parent network (Supplementary Fig. 6). Furthermore, the derivation of the stationary distribution of a fast subnetwork is enough to capture both the slow and fast stochastic dynamics of its multi-timescale parent network (Fig. 4). With these analytically derived stationary distributions of BRNs, their long-term stochastic behaviors such as their dependence on rate constants can be effectively investigated, and the likelihood function of parameters for Bayesian inference can also be derived[2].

Our work focused on the derivation of steady-state solutions of the CME using the underlying network structure following previous studies[11,26]. However, the CME is not usually used to capture cell division, which should be taken into account to describe single cell behavior in general. Thus, it would be interesting in future work to extend our method to the population balance equation[49,50], which describes stochastic cell population dynamics (e.g., cell division) as well as intracellular dynamics. This extension could be accomplished by averaging stationary distributions from cell populations after a stationary distribution of each cell is derived by our method.

We have translated a network to have the desired structural properties (i.e., weak reversibility and zero deficiency) by merging reactions with a common stoichiometric vector (Figs. 1a and 3g) and shifting a reaction preserving its stoichiometric vector (Fig. 3a, d). While the idea underlying this procedure is simple, it greatly extends the class of networks whose structure can be changed to the desired one. For instance, when the edge probability is $\frac{1}{2}$, the fraction of deficiency zero networks among Erdös-Rényi random networks with two species and at most bimolecular reactions increases more than six times after network translation. The identification of such translation, which is not simple, can be done automatically by the provided computational package, CASTANET. In particular, to efficiently search translated networks, in CASTANET, we use the necessary conditions for network translation toward weakly reversible and deficiency zero networks, derived in this study (see Supplementary Note 4 for details).

Furthermore, CASTNET performs the propensity factorization of translated networks, which is required to derive the stationary distributions of networks with non-mass action kinetics. In this study, by extending the previous factorization condition[11,26] to ours (Eq. (8)), we have been able to derive stationary distributions of various BRNs (Figs. 1, 2, and 3, Supplementary Figs. 3 and 4). Although the factorization condition with non-mass action kinetics have been rarely investigated[11,26] due to its complexity and lack of motivation, our work motivates studies on it as translated networks typically follow non-mass action kinetics. To cover more weakly reversible deficiency zero translated networks, we aim to further generalize our factorization conditions, and

accordingly, we will update our computational package CASTANET.

By changing the network structure while preserving the stochastic dynamics via network translation, we have been able to use the theory, applicable to weakly reversible deficiency zero networks[11], to understand the stochastic dynamics of a larger class of networks with non-zero deficiencies. Similarly, by translating a network to have a deficiency of one, it would be possible to show that the networks have the properties of a network with a deficiency of one, such as absolute concentration robustness: the steady state value of a species is invariant to the overall input of the system[51–53]. Furthermore, network translation of stochastic BRNs can also be used to identify stochastic properties of networks based on their structures, such as positive recurrence[54] and extinction[52,55,56].

Methods

Biochemical reaction network. BRN is a graphical representation of a given biochemical system[12,14,57,58]. It consists of the triple $\{\mathcal{S},\mathcal{C},\mathcal{R}\}$ where $\mathcal{S}=\{S_1,\dots,S_d\}$ is the set of interacting *species*, $\mathcal{C}=\{C_1,\dots,C_m\}$ is the set of *complexes*, and $\mathcal{R}=\{\nu_1\to\nu_1',\dots,\nu_r\to\nu_r'\}$ is the set of *reactions*. A complex is a non-negative linear combination of species (i.e., $C_i=a_{i1}S_1+\cdots+a_{id}S_d$), which is also represented as a d-dimensional non-negative integer-valued vector $(a_{i1},\dots,a_{id})$. A reaction is an ordered pair of complexes. This allows the BRN to be represented as a directed graph $(\mathcal{C},\mathcal{R})$, where complexes are nodes and reactions are directed edges. Hence, a reaction $R_j:=\nu_j\to\nu_j'$, where ν_j and ν_j' are the *source* and *product* complexes of the jth reaction, respectively. The vector $\nu_j'-\nu_j$ is called a *stoichiometric vector* of the jth reaction, which describes the relative change in the number of molecules of reactants and products between the sides of each reaction. A *linkage class* is a connected component of the network when all reactions are regarded as undirected edges. Weak reversibility means that if there is a sequence of reactions from a complex C_i to another complex C_j then there must be a sequence of reactions from C_j to C_i. The deficiency δ is the integer index defined as $|\mathcal{C}|-l-s$, where $|\mathcal{C}|$ is the number of complexes, l is the number of linkage classes, and s is the dimension of the subspace spanned by all stoichiometric vectors (Fig. 1a).

Complex balanced equilibrium. CBE of the deterministic mass action model for a BRN with rate constants $\{\kappa_k\}$ is the steady state $\mathbf{c}\in\mathbb{R}^d_{>0}$ which satisfies the following equality for each complex $z\in\mathcal{C}$ (Fig. 1c):

$$\sum_{k:\,\nu_k'=z}\kappa_k\mathbf{c}^{\nu_k}=\sum_{k:\,\nu_k=z}\kappa_k\mathbf{c}^{\nu_k} \tag{5}$$

where the $\kappa_k\mathbf{c}^{\nu_k}=\kappa_k c_1^{\nu_{k1}}c_2^{\nu_{k2}}\cdots\cdots c_d^{\nu_{kd}}$ is the rate function of the kth reaction following the deterministic mass action kinetics, and ν_{ki} is the ith entry of ν_k[12]. The LHS is the sum of rate functions over reactions whose product complex is z (i.e., $\nu_k'=z$), and the RHS is the sum of rate functions over reactions whose source complex is z (i.e., $\nu_k=z$). In other words, at CBE, the in-and out-flows create a balance for each complex. The deterministic mass action model for a BRN possesses a CBE regardless of rate constants if and only if the BRN is weakly reversible and deficiency zero[13]. Furthermore, even when a BRN has non-zero deficiency and is weakly reversible, the deterministic mass action model for the BRN possesses a CBE with specific choice of rate constants[13].

Stochastic model of biochemical reaction networks. We model a BRN as a continuous-time Markov chain (CTMC) for an isothermal well-stirred system with a constant volume. The state of the CTMC at time t, $\mathbf{n}(t)=(n_1,\dots,n_d)\in\mathbb{Z}^d_{\geq 0}$, represents the copy number of each species. Each reaction is associated with a *propensity function*:

$$\lambda_k:\mathbb{Z}^d_{\geq 0}\to\mathbb{R}_{\geq 0},\ k=1,\dots,r.$$

Specifically, $\lambda_k(\mathbf{n})$ is the probability that the kth reaction occurs in a short interval of length dt if the state at the beginning of the interval was $\mathbf{n}$. Using the propensity functions, we can derive the CME, which describes the time evolution of the probability of the model:

$$\frac{dp_{\mathbf{n}}}{dt}=\sum_{k=1}^{r}\lambda_k(\mathbf{n}-(\nu_k'-\nu_k))p_{\mathbf{n}-(\nu_k'-\nu_k)}-\sum_{k=1}^{r}\lambda_k(\mathbf{n})p_{\mathbf{n}}$$

for $\mathbf{n}\in\mathbb{Z}^d_{\geq 0}$, where $p_{\mathbf{n}}(t)$ denotes the probability that the state of the system equals $\mathbf{n}\in\mathbb{Z}^d_{\geq 0}$ at time t. A *stationary distribution* $\pi(\mathbf{n})$ of a given CTMC is the steady-state solution of the CME that satisfies the following infinite equation:

$$\sum_k\lambda_k(\mathbf{n}-(\nu_k'-\nu_k))\pi(\mathbf{n}-(\nu_k'-\nu_k))=\sum_k\lambda_k(\mathbf{n})\pi(\mathbf{n}).$$

It means that if the CTMC is initialized with its stationary distribution, the vector of probabilities $p(t)$ will stay constant for all time $t>0$.

The stochastic mass action propensity functions are assumed to be proportional to the number of ways in which species can combine to form the source complex. Hence, the kth propensity function with the rate constant α_k can be written as:

$$\lambda_k(n_1,\dots,n_d)=\alpha_k\prod_{i=1}^{d}\frac{n_i!}{(n_i-\nu_{ki})!}\mathbf{1}_{\{n_i\geq\nu_{ki}\}}. \tag{6}$$

Additionally, the propensity functions can have a more generalized form as follows:

$$\lambda_k(n_1,\dots,n_d)=\alpha_k\prod_{i=1}^{d}f_i(n_i)f_i(n_i-1)\cdots f_i(n_i-(\nu_{ki}-1))\mathbf{1}_{\{n_i\geq\nu_{ki}\}} \tag{7}$$

where functions $f_i:\mathbb{Z}_{\geq 0}\to\mathbb{R}_{\geq 0}$. For instance, the translated network in Fig. 3a-right follows this form as $f_A(n_A)=n_A(n_A-1)$ and $f_{A_p}(n_{A_p})=n_{A_p}$. This is called 'generalized' stochastic mass action kinetics since if f_i's are identity functions then it is equivalent to the stochastic mass action kinetics (Eq. (6)).

Network translation. Network translation is a procedure to transform a BRN $\{\mathcal{S},\mathcal{C},\mathcal{R}\}$ to another one $\{\mathcal{S},\tilde{\mathcal{C}},\tilde{\mathcal{R}}\}$ that satisfies the condition: the sum of propensities of a set of reactions sharing the same stoichiometric vector remains identical (Fig. 1a). That is, for each vector $\gamma\in\mathbb{Z}^d$, the propensity functions of the original and the translated networks, $\lambda_k(\mathbf{n})$ and $\tilde{\lambda}_{\tilde{k}}(\mathbf{n})$, satisfy the following:

$$\sum_{k:\,\nu_k'-\nu_k=\gamma}\lambda_k(\mathbf{n})=\sum_{\tilde{k}:\,\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}=\gamma}\tilde{\lambda}_{\tilde{k}}(\mathbf{n})$$

for all $\mathbf{n}\in\mathbb{Z}^d_{\geq 0}$, where $\nu_k'-\nu_k$ and $\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}$ are the stoichiometric vectors of the kth and $\tilde{k}$th reactions of the first and second models, respectively. For example, merging several reactions sharing a common stoichiometric vector into one reaction is network translation. Similarly, shifting reactions preserving their stoichiometric vectors is also an instance of network translation (e.g., $A+B\to 2B$ to $A\to B$). Network translation can change the structural properties of BRNs, such as weak reversibility and the deficiency, but it preserves the associated CME, i.e., stochastic dynamics (see Supplementary Note 1 for details).

Propensity factorization. To derive the stationary distribution with our approach (Fig. 1), all the propensities $\tilde{\lambda}_k(\mathbf{n})$ should be factorized as

$$\tilde{\lambda}_k(\mathbf{n})=\kappa_k\theta(\mathbf{n})\omega(\mathbf{n}-\nu_k)\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}} \tag{8}$$

for some constants $\kappa_k>0$ and functions $\theta(\mathbf{n})>0$ and $\omega(\mathbf{n})\geq 0$ on a set $\Gamma=\{\mathbf{n}\,|\,\mathbf{n}\geq\mathbf{b}\}\subset\mathbb{Z}^d_{\geq 0}$ at which $\tilde{\lambda}_k(\mathbf{n})>0$ if $\mathbf{n}\geq\nu_k+\mathbf{b}$ and $\tilde{\lambda}_k(\mathbf{n})=0$ otherwise. For the stochastic mass action kinetics (Eq. (6)), $\mathbf{b}=\mathbf{0}$ as $\tilde{\lambda}_k(\mathbf{n})>0$ if and only if $\mathbf{n}\geq\nu_k$. For the translated network in Fig. 3a, $\mathbf{b}=(1,0)$ because each propensity is positive if and only if $\mathbf{n}\geq\nu_k+(1,0)$ in $\Gamma=\{\mathbf{n}|(n_A,n_{A_p})\geq(1,0)\}$.

While the propensity factorization can be calculated by solving recurrence relations (see Supplementary Note 2 for details), it can be obtained without solving recurrence relations if all the propensities of a given network follow the generalized mass action kinetics (Eq. (7)). In this case, the factorization can be easily obtained as

$$\kappa_k:=\alpha_k,\ \theta(\mathbf{n}):=\prod_{i=1}^{d}\prod_{j=b_i+1}^{n_i}f_i(j),\ \text{and}\ \omega(\mathbf{n}):=\frac{1}{\theta(\mathbf{n})}\mathbf{1}_{\{\mathbf{n}\geq\mathbf{b}\}}, \tag{9}$$

where $\prod_{j=0}^{-1}a_j=1$ for any $\{a_j\}$ (see Supplementary Note 2 for details).

Derivation of stationary distribution. If a network is weakly reversible and deficiency zero so that it has a CBE $\mathbf{c}\in\mathbb{R}^d_{>0}$[13] and propensity function $\lambda_k(\mathbf{n})$ can be factorized as in Eq. (8) on Γ, a stationary measure of the network can be derived as

$$\pi(\mathbf{n})=\begin{cases}\frac{\mathbf{c}^{\mathbf{n}}}{\theta(\mathbf{n})} & \text{if}\ \ \mathbf{n}\in\Gamma,\\ 0 & \text{if}\ \ \mathbf{n}\in\Gamma^{\mathrm{c}}.\end{cases}$$

Supplementary Note 3 provides the proof and detailed illustration. By scaling this stationary measure with the normalizing constant, which is the reciprocal of the summation of $\pi(\mathbf{n})$ over the irreducible state space, the stationary distribution on the irreducible state space can be obtained. For instance, the normalizing constant for the stationary distribution (Fig. 3e) is calculated by summing $\pi(\mathbf{n})$ over the irreducible state space $\{(n_A,n_{A_p},n_{A_{pp}})\,|\,n_A\geq 1,n_A+n_{A_p}+n_{A_{pp}}=T_0\}$. While computing the normalizing constants is sometimes challenging, a symbolic computation approach using Wilf-Zeilberger theory can be used for the stochastic mass action model[59].

Computational package, CASTANET. We have developed a user-friendly, open-source, and publicly available computational package, CASTANET, that performs the network translation and propensity factorization automatically (Fig. 2a). The

package checks two conditions: whether a given BRN can be made weakly reversible and of zero deficiency after network translation, and whether the propensities of the translated network can be factorized as in Eq. (8). If these two conditions are satisfied, CASTANET then calculates the analytic formula for a stationary distribution.

To efficiently search weakly reversible deficiency zero translated networks, we derived their necessary conditions (see Supplementary Note 4 for details) and incorporated them in the package. Furthermore, CASTANET constructs a candidate for the factorization function $\theta(\mathbf{n})$ in symbolic expression, which allows us to check propensity factorization condition without checking infinite combinations (see Supplementary Note 4).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data used in the current research are available upon request to the corresponding authors. Simulation data underlying plots shown in Figs. 3 and 4 are provided in Supplementary Data 1.

Code availability

The MATLAB (version R2020b) code performing network translation, propensity factorization, and CBE calculation to derive stationary distribution (schematically shown in Figs. 1 and 2) can be found at https://github.com/Mathbiomed/CASTANET. The detailed description and step-by-step manual are provided in Supplementary information.

Received: 23 March 2021; Accepted: 16 April 2021;
Published online: 24 May 2021

References

1. Hahl, S. K. & Kremling, A. A comparison of deterministic and stochastic modeling approaches for biochemical reaction systems: on fixed points, means, and modes. *Front. Genet.* **7**, 157 (2016).
2. Kim, J. K. & Marioni, J. C. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.* **14**, R7 (2013).
3. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
4. Schnoerr, D., Sanguinetti, G. & Grima, R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A: Math. Theor.***50**, 093001 (2017).
5. Yang, J.-M. et al. Integrating chemical and mechanical signals through dynamic coupling between cellular protrusions and pulsed erk activation. *Nat. Commun.* **9**, 4673 (2018).
6. Gadgil, C., Lee, C. H. & Othmer, H. G. A stochastic analysis of first-order reaction networks. *Bull. Math. Biol.* **67**, 901–946 (2005).
7. Allen, L. An Introduction to Stochastic Processes with Applications to Biology (CRC Press, 2010).
8. Kelly, F. *Reversibility and Stochastic Networks*. (Wiley, New York, 1979).
9. Mairesse, J. & Nguyen, H.-T. Deficiency zero Petri nets and product form. In Franceschinis, G. & Wolf, K. (eds.) *Applications and Theory of Petri Nets*, 103-122 (Springer-Verlag, 2009).
10. Angeli, D., de Leenheer, P. & Sontag, E. A Petri net approach to the study of persistence in chemical reaction networks. *Math. Biosci.* **210**, 598–618 (2007).
11. Anderson, D. F., Craciun, G. & Kurtz, T. G. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.* **72**, 1947–1970 (2010).
12. Horn, F. & Jackson, R. General mass action kinetics. *Arch. Rat. Mech. Anal.* **47**, 81–116 (1972).
13. Horn, F. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Arch. Rat. Mech. Anal.* **49**, 172–186 (1972).
14. Feinberg, M. Complex balancing in general kinetic systems. *Arch. Rat. Mech. Anal.* **49**, 187–194 (1972).
15. Wu, S., Fu, J., Li, H. & Petzold, L. Automatic identification of model reductions for discrete stochastic simulation. *J. Chem. Phys.* **137**, 034106 (2012).
16. Ghaemi, R. & Del Vecchio, D. Stochastic analysis of retroactivity in transcriptional networks through singular perturbation. In *2012 American Control Conference (ACC)*, 2731–2736 (2012).
17. Mélykúti, B., Hespanha, J. P. & Khammash, M. Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks. *J. R. Soc. Interface* **11**, 20140054 (2014).
18. Hepp, B., Gupta, A. & Khammash, M. Adaptive hybrid simulations for multiscale stochastic reaction networks. *J. Chem. Phys.* **142**, 034118 (2015).
19. Hwang, H. J. & Velázquez, J. J. L. Bistable stochastic biochemical networks: highly specific systems with few chemicals. *J. Math. Chem* **51**, 1343–1375 (2013).
20. Ali Al-Radhawi, M., Del Vecchio, D. & Sontag, E. D. Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS Comput. Biol.* **15**, 1–27 (2019).
21. Kim, J. K. & Sontag, E. D. Reduction of multiscale stochastic biochemical reaction networks using exact moment derivation. *PLoS Comput. Biol.* **13**, 1–24 (2017).
22. Kan, X., Lee, C. H. & Othmer, H. G. A multi-time-scale analysis of chemical reaction networks: Ii. stochastic systems. *J. Math. Biol.* **73**, 1081–1129 (2016).
23. Anderson, D. F. & Nguyen, T. D. Prevalence of deficiency zero reaction networks in an erdos-renyi framework (2019). 1910.12723.
24. Johnston, M. D. Translated chemical reaction networks. *Bull. Math. Biol.* **76**, 1081–1116 (2014).
25. Johnston, M. D. & Burton, E. Computing weakly reversible deficiency zero network translations using elementary flux modes. *Bull. Math. Biol.* **81**, 1613–1644 (2019).
26. Anderson, D. F. & Cotter, S. L. Product-form stationary distributions for deficiency zero networks with non-mass action kinetics. *Bull. Math. Biol.* **78**, 2390–2407 (2016).
27. Sontag, E. D. Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction. *IEEE Trans. Autom. Control* **46**, 1028–1047 (2001).
28. Beenstock, J., Mooshayef, N. & Engelberg, D. How do protein kinases take a selfie (autophosphorylate)? *Trends Biochem. Sci.* **41**, 938–953 (2016).
29. Dodson, C. A., Yeoh, S., Haq, T. & Bayliss, R. A kinetic test characterizes kinase intramolecular and intermolecular autophosphorylation mechanisms. *Sci. Signal.* **6**, ra54–ra54 (2013).
30. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* **1**, 2005.0010 (2005).
31. Wang, J., Wu, J.-W. & Wang, Z.-X. Structural insights into the autoactivation mechanism of p21-activated protein kinase. *Structure* **19**, 1752–1761 (2011).
32. Dammann, K., Khare, V. & Gasche, C. Tracing paks from gi inflammation to cancer. *Gut* **63**, 1173–1184 (2014).
33. Parrini, M. C., Lei, M., Harrison, S. C. & Mayer, B. J. Pak1 kinase homodimers are autoinhibited in trans and dissociated upon activation by cdc42 and rac1. *Mol. Cell* **9**, 73–83 (2002).
34. Zaytsev, A. V. et al. Bistability of a coupled aurora b kinase-phosphatase system in cell division. *eLife* **5**, e10644 (2016).
35. Doherty, K., Meere, M. & Piiroinen, P. T. Some mathematical models of intermolecular autophosphorylation. *J. Theor. Biol.* **370**, 27–38 (2015).
36. Wang, Z.-X. & Wu, J.-W. Autophosphorylation kinetics of protein kinases. *Biochem. J.* **368**, 947–952 (2002).
37. Mouri, K., Nacher, J. C. & Akutsu, T. A mathematical model for the detection mechanism of dna double-strand breaks depending on autophosphorylation of atm. *PLoS ONE* **4**, 1–14 (2009).
38. Nguyen, L. K., Kolch, W. & Kholodenko, B. N. When ubiquitination meets phosphorylation: a systems biology perspective of egfr/mapk signalling. *Cell Commun. Signal.* **11**, 52 (2013).
39. Luciani, F., Keşmir, C., Mishto, M., Or-Guil, M. & de Boer, R. J. A mathematical model of protein degradation by the proteasome. *Biophys. J.* **88**, 2422–2432 (2005).
40. Tiganis, T. Protein tyrosine phosphatases: dephosphorylating the epidermal growth factor receptor. *IUBMB Life* **53**, 3–14 (2002).
41. King, C. C. et al. p21-activated kinase (pak1) is phosphorylated and activated by 3-phosphoinositide-dependent kinase-1 (pdk1). *J. Biol. Chem.* **275**, 41201–41209 (2000).
42. Sessa, F. & Villa, F. Structure of Aurora B–INCENP in complex with barasertib reveals a potential transinhibitory mechanism. *Acta Crystallogr. Sect. F* **70**, 294–298 (2014).
43. Chen, C.-Y., Yu, Z.-Y., Chuang, Y.-S., Huang, R.-M. & Wang, T.-C. V. Sulforaphane attenuates egfr signaling in nsclc cells. *J. Biomed. Sci.* **22**, 38 (2015).
44. Gully, C. P. et al. Antineoplastic effects of an aurora b kinase inhibitor in breast cancer. *Mol. Cancer* **9**, 42 (2010).
45. Weisz Hubsman, M., Volinsky, N., Manser, E., Yablonski, D. & Aronheim, A. Autophosphorylation-dependent degradation of Pak1, triggered by the Rho-family GTPase, Chp. *Biochem. J.* **404**, 487–497 (2007).
46. Shamir, M., Bar-On, Y., Phillips, R. & Milo, R. Snapshot: timescales in cell biology. *Cell* **164**, 1302–1302.e1 (2016).
47. Bibbona, E., Kim, J. & Wiuf, C. Stationary distributions of systems with discreteness-induced transitions. *J. R. Soc. Interface* **17**, 20200243 (2020).
48. Rao, C. V. & Arkin, A. P. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the gillespie algorithm. *J. Chem. Phys.* **118**, 4999–5010 (2003).

49. Waldherr, S. Estimation methods for heterogeneous cell population models in systems biology. *J. R. Soc. Interface* **15**, 20180530 (2018).
50. Kremling, A. Systems biology: mathematical modeling and model analysis (CRC Press, 2013).
51. Shinar, G. & Feinberg, M. Structural sources of robustness in biochemical reaction networks. *Science* **327**, 1389–1391 (2010).
52. Anderson, D. F., Enciso, G. A. & Johnston, M. D. Stochastic analysis of biochemical reaction networks with absolute concentration robustness. *J. R. Soc. Interface* **11**, 20130943 (2014).
53. Enciso, G. A. Transient absolute robustness in stochastic biochemical networks. *J. R. Soc. Interface* **13**, 20160475 (2016).
54. Anderson, D. F. & Kim, J. Some network conditions for positive recurrence of stochastically modeled reaction networks. *SIAM J. Appl. Math.* **78**, 2692–2713 (2018).
55. Johnston, M. D. A computational approach to extinction events in chemical reaction networks with discrete state spaces. *Math. Biosci.* **294**, 130–142 (2017).
56. Johnston, M. D., Anderson, D. F., Craciun, G. & Brijder, R. Conditions for extinction events in chemical reaction networks with discrete state spaces. *J. Math. Biol.* **76**, 1535–1558 (2018).
57. Feinberg, M. Foundations of Chemical Reaction Network Theory (Springer, 2019).
58. Anderson, D. F. & Kurtz, T. G. Continuous time markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, 3-42 (Springer, 2011).
59. Sontag, E. D. & Zeilberger, D. A symbolic computation approach to a problem involving multivariate poisson distributions. *Adv. Appl. Math.* **44**, 359–377 (2010).
60. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).

Acknowledgements

This work is supported by the Institute for Basic Science (IBS-R029-C3)(J.K.K.), the Samsung Science and Technology Foundation (SSTF-BA1902-01) (J.K.K.), and the National Research Foundation of Korea (Global Ph. D. Fellowship Program 2019H1A2A1075303) (H.H.), as well as US NSF grants 1716623 and 1849588 (E.D.S.). J.K. thanks German Enciso for travel support via NSF grant DMS1616233. H.H. and J.K. thank the 2019 Chemical Reaction Networks summer school and workshop at Politecnico di Torino for providing valuable feedbacks on this work.

Author contributions

H.H., E.D.S., and J.K.K. designed research. H.H., J.K., and J.K.K. developed the theory and algorithm. H.H. performed computation. H.H., J.K., M.A.A., E.D.S., and J.K.K. discussed and analyzed results. H.H., J.K., and J.K.K. drafted the manuscript and designed the figures, and H.H., J.K., M.A.A., E.D.S., and J.K.K. revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02117-x.

Correspondence and requests for materials should be addressed to J.K.K.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information: Derivation of stationary distributions of biochemical reaction networks via structure transformation

Hyukpyo Hong,[1,2,*] Jinsu Kim,[3,4,*] M Ali Al-Radhawi,[5] Eduardo D. Sontag,[5,6,7,8,9] and Jae Kyoung Kim[1,2,†]

[1] *Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea*

[2] *Biomedical Mathematics Group, Institute for Basic Science, Daejeon 34126, Republic of Korea*

[3] *Department of Mathematics, University of California, Irvine, CA 92697, United States*

[4] *NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, CA 92697, United States*

[5] *Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, United States*

[6] *Department of Bioengineering, Northeastern University, Boston, MA 02115, United States*

[7] *Department of Mathematics, Northeastern University, Boston, MA 02115, United States*

[8] *Department of Chemical Engineering, Northeastern University, Boston, MA 02115, United States*

[9] *Laboratory of Systems Pharmacology, Program in Therapeutics Science, Harvard Medical School, Boston, MA 02115, United States*

SUPPLEMENTARY NOTE 1: NETWORK TRANSLATION PRESERVES THE STOCHASTIC DYNAMICS

Let $\{\mathcal{S},\mathcal{C},\mathcal{R}\}$ and $\{\mathcal{S},\tilde{\mathcal{C}},\tilde{\mathcal{R}}\}$ be an original network and a translated network, respectively. Then, from the definition of network translation,

$$\sum_{k:\nu_k'-\nu_k=\gamma}\lambda_k(\mathbf{n})=\sum_{\tilde{k}:\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}=\gamma}\tilde{\lambda}_{\tilde{k}}(\mathbf{n})\quad\text{for each vector }\gamma\in\mathbb{Z}^d, \tag{S1}$$

where λ_k and $\tilde{\lambda}_{\tilde{k}}$ are the propensities of stochastic models (i.e., continuous-time Markov chains) for $\{\mathcal{S},\mathcal{C},\mathcal{R}\}$ and $\{\mathcal{S},\tilde{\mathcal{C}},\tilde{\mathcal{R}}\}$, respectively. Let $p_{\mathbf{n}}(t)$ and $\tilde{p}_{\mathbf{n}}(t)$ be the probabilities of the continuous-time Markov chains associated with $\{\mathcal{S},\mathcal{C},\mathcal{R}\}$ and $\{\mathcal{S},\tilde{\mathcal{C}},\tilde{\mathcal{R}}\}$ being at state $\mathbf{n}$, respectively. Then, $p_{\mathbf{n}}(t)$ is the solution of CME given by:

$$\begin{aligned}\frac{dp_{\mathbf{n}}}{dt}&=\sum_{k=1}^{r}\lambda_k(\mathbf{n}-(\nu_k'-\nu_k))p_{\mathbf{n}-(\nu_k'-\nu_k)}-\sum_{k=1}^{r}\lambda_k(\mathbf{n})p_{\mathbf{n}}\\&=\sum_{\gamma\in\mathbb{Z}^d}\sum_{k:\nu_k'-\nu_k=\gamma}\lambda_k(\mathbf{n}-(\nu_k'-\nu_k))p_{\mathbf{n}-(\nu_k'-\nu_k)}-\sum_{\gamma\in\mathbb{Z}^d}\sum_{k:\nu_k'-\nu_k=\gamma}\lambda_k(\mathbf{n})p_{\mathbf{n}}\\&=\sum_{\gamma\in\mathbb{Z}^d}\sum_{\tilde{k}:\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}=\gamma}\tilde{\lambda}_{\tilde{k}}(\mathbf{n}-(\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}))p_{\mathbf{n}-(\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}})}-\sum_{\gamma\in\mathbb{Z}^d}\sum_{\tilde{k}:\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}=\gamma}\tilde{\lambda}_{\tilde{k}}(\mathbf{n})p_{\mathbf{n}}\\&=\sum_{\tilde{k}=1}^{\tilde{r}}\tilde{\lambda}_{\tilde{k}}(\mathbf{n}-(\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}}))p_{\mathbf{n}-(\tilde{\nu}_{\tilde{k}}'-\tilde{\nu}_{\tilde{k}})}-\sum_{\tilde{k}=1}^{\tilde{r}}\tilde{\lambda}_{\tilde{k}}(\mathbf{n})p_{\mathbf{n}},\end{aligned}$$

where r and $\tilde{r}$ are the total numbers of reactions in the original and the translated reaction networks, respectively, and the third equality follows from Eq. (S1). Thus, $p_{\mathbf{n}}(t)$ and $\tilde{p}_{\mathbf{n}}(t)$ are the solution of the same CME because the

* These authors contributed equally

† jaekkim@kaist.ac.kr

CME associated with the translated network is given by

$$\frac{d\tilde{p}_{\mathbf{n}}}{dt} = \sum_{\tilde{k}=1}^{\tilde{r}} \tilde{\lambda}_{\tilde{k}}(\mathbf{n} - (\tilde{\nu}'_{\tilde{k}} - \tilde{\nu}_{\tilde{k}}))\tilde{p}_{\mathbf{n}-(\tilde{\nu}'_{\tilde{k}}-\tilde{\nu}_{\tilde{k}})} - \sum_{\tilde{k}=1}^{\tilde{r}} \tilde{\lambda}_{\tilde{k}}(\mathbf{n})\tilde{p}_{\mathbf{n}}.$$

Therefore, $p_{\mathbf{n}}(t) \equiv \tilde{p}_{\mathbf{n}}(t)$ for all $\mathbf{n} \in \mathbb{Z}^d_{\geq 0}$ as long as they have the same initial condition.

SUPPLEMENTARY NOTE 2: PROPENSITY FACTORIZATION

As we discussed in the main text, to derive a stationary distribution, all the propensities $\lambda_k(\mathbf{n})$ should be factorized as

$$\lambda_k(\mathbf{n}) = \kappa_k \theta(\mathbf{n})\omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}} \tag{S2}$$

for some constants $\kappa_k > 0$ and functions $\theta(\mathbf{n}) > 0$ and $\omega(\mathbf{n}) \geq 0$ on a set $\Gamma = \{\mathbf{n} \mid \mathbf{n} \geq \mathbf{b}\}$, where $\lambda_k(\mathbf{n}) > 0$ if and only if $\mathbf{n} \geq \nu_k + \mathbf{b}$ in Γ meaning that $\nu_k + \mathbf{b}$ represents the minimal copy numbers for the reaction to occur. This factorization form generalizes the previous result, Eq. (32) in [2], which requires the following factorization:

$$\lambda_k(\mathbf{n}) = \kappa_k \frac{\zeta(\mathbf{n})}{\zeta(\mathbf{n} - \nu_k)}\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}} \tag{S3}$$

for $\zeta : \mathbb{Z}^d_{\geq 0} \to \mathbb{R}_{>0}$ on $\mathbb{Z}^d_{\geq 0}$. Note that if $\omega(\mathbf{n}) = \frac{1}{\theta(\mathbf{n})}$ and $\Gamma = \mathbb{Z}^d_{\geq 0}$ (i.e., $\mathbf{b} = \mathbf{0}$), Eq. (S2) is equivalent to Eq. (S3).

Propensity factorization by solving recurrence relations

To obtain the desired factorization for given propensity functions, we need to solve recurrence relations. Here, we illustrate how to solve the following factorization equations of Fig. 1b.

$$\lambda_1(n_A, n_B) = \alpha_1 + \alpha_4 n_A = \kappa_1 \theta(n_A, n_B)\omega(n_A, n_B), \tag{S4}$$

$$\lambda_2(n_A, n_B) = \alpha_2 n_A = \kappa_2 \theta(n_A, n_B)\omega(n_A - 1, n_B)\mathbf{1}_{\{n_A\geq 1\}}, \text{ and} \tag{S5}$$

$$\lambda_3(n_A, n_B) = \alpha_3 n_A n_B = \kappa_3 \theta(n_A, n_B)\omega(n_A - 1, n_B - 1)\mathbf{1}_{\{n_A\geq 1, n_B\geq 1\}}. \tag{S6}$$

After setting $\kappa_k = \alpha_k$, we divide Eq. (S4) and Eq. (S5) by Eq. (S5) and Eq. (S6), respectively. Then we have $\omega(n_A, n_B) = (\frac{\alpha_4 n_A + \alpha_1}{\alpha_1 n_A})\omega(n_A - 1, n_B)$ and $\omega(n_A - 1, n_B) = \frac{1}{n_B}\omega(n_A - 1, n_B - 1)$. Solving these recurrence relations, we get $\omega(n_A, n_B) = \prod_{j=1}^{n_A} \left(\frac{\alpha_4 j + \alpha_1}{\alpha_1 j}\right) \frac{1}{n_B!}\omega(0, 0)$. Then $\theta(n_A, n_B)$ is determined as $\frac{(\alpha_4 n_A + \alpha_1)}{\alpha_1 \omega(0,0)} \prod_{j=1}^{n_A} \left(\frac{\alpha_1 j}{\alpha_4 j + \alpha_1}\right) n_B!$ from the second equality in Eq. (S4). Note that this factorization cannot be accomplished by the previously identified form (Eq. (S3), $\lambda_k = \kappa_k \theta(\mathbf{n})\theta(\mathbf{n} - \nu_k)^{-1}\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}}$), which requires $\omega(n_A, n_B) = \theta(n_A, n_B)^{-1}$, because Eq. (S4) implies $\omega(n_A, n_B) \neq \theta(n_A, n_B)^{-1}$. Here, $\omega(0, 0)$ can be determined by requiring the sum of the stationary distribution $\sum_{\mathbf{n}} \pi(\mathbf{n})$ to be 1.

Propensities following generalized mass action kinetics can be factorized without solving recurrence relations

If the propensities follow a sort of generalized stochastic mass action kinetics, then their factorization can be easily obtained. In what follows, we use the convention that $\prod_{j=0}^{-1} a_j = 1$ for any $\{a_j\}$.

Theorem 1. *Let a biochemical reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be given, and let $\lambda_k(\mathbf{n})$ be the propensity functions of the associated continuous-time Markov chain. Suppose that there exists a vector $\mathbf{b}$ such that for $\mathbf{n} \in \Gamma = \{\mathbf{n} \mid \mathbf{n} \geq \mathbf{b}\}$, we*

have $\lambda_k(\mathbf{n}) > 0$ *if* $\mathbf{n} \geq \nu_k + \mathbf{b}$ *and* $\lambda_k(\mathbf{n}) = 0$ *otherwise for each* $k = 1, 2, \ldots, r$. *We further assume that there exist constants* $\alpha_k > 0$ *and functions* $f_i : \mathbb{Z}_{\geq 0} \to \mathbb{R}_{\geq 0}, i = 1, \ldots, d$ *such that* $\lambda_k(\mathbf{n}) = \alpha_k \prod_{i=1}^{d} \prod_{j=0}^{\nu_{ki}-1} f_i(n_i - j)$ *for* $k = 1, \ldots, r$ *and all* $\mathbf{n} \in \Gamma$, *where* ν_{ki} *is the* i*th entry of the source complex vector* ν_k. *Then* $\lambda_k(\mathbf{n}) = \kappa_k \theta(\mathbf{n}) \omega(\mathbf{n} - \nu_k) \mathbf{1}_{\{\mathbf{n} \geq \nu_k\}}$ *on the* Γ *with*

$$\kappa_k := \alpha_k, \quad \theta(\mathbf{n}) := \prod_{i=1}^{d} \prod_{j=b_i+1}^{n_i} f_i(j), \quad \textit{and} \quad \omega(\mathbf{n}) := \frac{1}{\theta(\mathbf{n})} \mathbf{1}_{\{\mathbf{n} \geq \mathbf{b}\}}. \tag{S7}$$

Proof. For $\mathbf{n} \in \Gamma$,

$$\begin{aligned}
\kappa_k \theta(\mathbf{n}) \omega(\mathbf{n} - \nu_k) \mathbf{1}_{\{\mathbf{n} \geq \nu_k\}} &= \alpha_k \frac{\prod_{i=1}^{d} \prod_{j=b_i+1}^{n_i} f_i(j)}{\prod_{i=1}^{d} \prod_{j=b_i+1}^{n_i - \nu_{ki}} f_i(j)} \mathbf{1}_{\{\mathbf{n} - \nu_k \geq \mathbf{b}\}} \mathbf{1}_{\{\mathbf{n} \geq \nu_k\}} \\
&= \alpha_k \prod_{i=1}^{d} \prod_{j=n_i - \nu_{ki}+1}^{n_i} f_i(j) \mathbf{1}_{\{n_i - \nu_{ki} \geq b_i\}} \\
&= \alpha_k \prod_{i=1}^{d} \prod_{j=0}^{\nu_{ki}-1} f_i(n_i - j) \mathbf{1}_{\{n_i - \nu_{ki} \geq b_i\}} \\
&= \begin{cases} \alpha_k \prod_{i=1}^{d} \prod_{j=0}^{\nu_{ki}-1} f_i(n_i - j) & \text{if } \mathbf{n} \geq \nu_k + \mathbf{b} \\ 0 & \text{otherwise} \end{cases} \\
&= \lambda_k(\mathbf{n}).
\end{aligned}$$

□

Remark 1. A similar condition was introduced by Eq. (27) in the previous study [2], which is a special case of the generalized mass action kinetics with $\mathbf{b} = 0$. Furthermore, the associated function $f_i(n_i)$ can be thought as the "rate of association" of the ith species as pointed out in [9]. Such rate of association of the species have been also used in order to generalize the Horn-Jackson-Feinberg theory to deterministic non-mass action system [11].

Remark 2. The example in Fig. 3g does not follow the generalized mass action kinetics because the propensity of the reaction $A \to A_P$, $\alpha_1 n_A + \alpha_2 n_A n_{A_P}$, depends on both n_A and n_{A_P}, but the latter is not supported in the source complex. Nevertheless, it can be re-expressed as $\alpha_1 n_A + \alpha_2 n_A (T_0 - n_A)$ using the conservation law: $n_A + n_{A_P} = T_0$ so it now follows the generalized mass action kinetics.

SUPPLEMENTARY NOTE 3: DERIVATION OF STATIONARY DISTRIBUTIONS

If a network is weakly reversible and has zero deficiency, the deterministic mass action model with any rate constants always possesses a CBE $\mathbf{c} \in \mathbb{R}^d_{>0}$ [4, 7, 8]. Then the existence of the propensity factorization ensures that the stationary distribution of the associated continuous-time Markov chain can be derived analytically. In what follows, $\mathbf{x} \ngeq \mathbf{y}$ means there is at least one i such that $x_i < y_i$.

Theorem 2. *For a given biochemical reaction network* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, *suppose that there exists a complex balanced equilibrium* $\mathbf{c} \in \mathbb{R}^d_{>0}$ *for the deterministic mass action model for* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ *with rate constants* $\{\kappa_k\}$. *Let* $\lambda_k(\mathbf{n})$ *be the propensity functions of the continuous-time Markov chain associated with the network* $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$. *Suppose that there exists a vector* $\mathbf{b}$ *such that for* $\mathbf{n} \in \Gamma = \{\mathbf{n} \mid \mathbf{n} \geq \mathbf{b}\}$, *we have* $\lambda_k(\mathbf{n}) > 0$ *if* $\mathbf{n} \geq \nu_k + \mathbf{b}$ *and* $\lambda_k(\mathbf{n}) = 0$ *otherwise for each* $k = 1, 2, \ldots, r$. *Assume further that the propensity functions* $\lambda_k(\mathbf{n})$ *can be factorized as*

$$\lambda_k(\mathbf{n}) = \kappa_k \theta(\mathbf{n}) \omega(\mathbf{n} - \nu_k) \mathbf{1}_{\{\mathbf{n} \geq \nu_k\}} \text{ on } \Gamma = \{\mathbf{n} \mid \mathbf{n} \geq \mathbf{b}\}, \tag{S8}$$

where θ and ω are non-negative functions such that $\theta(\mathbf{n}) > 0$ if $\mathbf{n} \in \Gamma$. Then the Γ is a closed subset, and the stochastic model admits a stationary measure that can be written as

$$\pi(\mathbf{n}) = \begin{cases} \dfrac{\mathbf{c}^{\mathbf{n}}}{\theta(\mathbf{n})} & \text{if } \mathbf{n} \in \Gamma, \\ 0 & \text{if } \mathbf{n} \in \Gamma^c. \end{cases}$$

If π is summable over the state space, then the measure π can be normalized to be a stationary distribution.

Proof. We first show that Γ is closed, i.e., there is no transition from Γ to Γ^c. Assume that there exists a reaction from $\mathbf{n} \in \Gamma$ to $\mathbf{n} + \nu'_k - \nu_k \in \Gamma^c$ with $\lambda_k(\mathbf{n}) > 0$. This implies that $n_i + \nu'_{ki} - \nu_{ki} < b_i$ for some i, and thus $n_i < b_i - \nu'_{ki} + \nu_{ki} \leq b_i + \nu_{ki}$. That is, $\mathbf{n} \not\geq \nu_k + \mathbf{b}$, and $\mathbf{n} \geq \mathbf{b}$ since $\mathbf{n} \in \Gamma$. Therefore, $\lambda_k(\mathbf{n}) = 0$ from the definition of $\mathbf{b}$. This contradicts to $\lambda_k(\mathbf{n}) > 0$, so Γ is closed.

We turn to showing that the given π solves the CME at steady state:

$$\sum_k \lambda_k(\mathbf{n} - (\nu'_k - \nu_k))\pi(\mathbf{n} - (\nu'_k - \nu_k)) = \sum_k \lambda_k(\mathbf{n})\pi(\mathbf{n}) \quad \text{for all } \mathbf{n} \in \mathbb{Z}^d_{\geq 0}. \tag{S9}$$

Case 1: $\mathbf{n} \in \Gamma^c$. The right-hand side in Eq. (S9) is 0 as $\pi(\mathbf{n}) = 0$ for $\mathbf{n} \in \Gamma^c$. For the left-hand side in Eq. (S9), if $\mathbf{n} - (\nu'_k - \nu_k) \in \Gamma$, $\lambda_k(\mathbf{n} - (\nu'_k - \nu_k)) = 0$ because Γ is a closed set, and if $\mathbf{n} - (\nu'_k - \nu_k) \in \Gamma^c$, $\pi(\mathbf{n} - (\nu'_k - \nu_k)) = 0$ by its definition. Therefore, the left-hand side in Eq. (S9) is 0.

Case 2: $\mathbf{n} \in \Gamma$. For each $\mathbf{n} \in \Gamma$, we define $K_\Gamma(\mathbf{n}) = \{k \mid \mathbf{n} - (\nu'_k - \nu_k) \in \Gamma\}$. Then Eq. (S9) can be reduced into

$$\sum_{k \in K_\Gamma(\mathbf{n})} \lambda_k(\mathbf{n} - (\nu'_k - \nu_k))\pi(\mathbf{n} - (\nu'_k - \nu_k)) = \sum_k \lambda_k(\mathbf{n})\pi(\mathbf{n}) \quad \text{for all } \mathbf{n} \in \mathbb{Z}^d_{\geq 0} \tag{S10}$$

because $\pi(n - (\nu'_k - \nu_k)) = 0$ for each $k \in K_\Gamma(\mathbf{n})^c$. After the substitution of $\pi(\mathbf{n}) = \dfrac{\mathbf{c}^n}{\theta(\mathbf{n})}$ and Eq. (S8), the Eq. (S10) becomes

$$\begin{aligned} LHS &= \sum_{k \in K_\Gamma(\mathbf{n})} \kappa_k \frac{\mathbf{c}^{\mathbf{n} - (\nu'_k - \nu_k)}}{\theta(\mathbf{n} - (\nu'_k - \nu_k))} \theta(\mathbf{n} - (\nu'_k - \nu_k))\omega(\mathbf{n} - \nu'_k)\mathbf{1}_{\{\mathbf{n} \geq \nu'_k\}} \\ &= \mathbf{c}^{\mathbf{n}} \sum_{k \in K_\Gamma(\mathbf{n})} \kappa_k \mathbf{c}^{\nu_k - \nu'_k} \omega(\mathbf{n} - \nu'_k)\mathbf{1}_{\{\mathbf{n} \geq \nu'_k\}}, \end{aligned} \tag{S11}$$

$$RHS = \frac{\mathbf{c}^{\mathbf{n}}}{\theta(\mathbf{n})} \sum_k \kappa_k \theta(\mathbf{n})\omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n} \geq \nu_k\}} = \mathbf{c}^{\mathbf{n}} \sum_k \kappa_k \omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n} \geq \nu_k\}}. \tag{S12}$$

We can show that the RHS and LHS are equal by using that $\mathbf{c} \in \mathbb{R}^d_{>0}$ is a CBE and thus for any fixed complex $z \in \mathcal{C}$,

$$\sum_{k : \nu'_k = z} \kappa_k \mathbf{c}^{\nu_k} = \sum_{k : \nu_k = z} \kappa_k \mathbf{c}^{\nu_k}.$$

Multiplying $\mathbf{c}^{-z}\omega(\mathbf{n} - z)\mathbf{1}_{\{\mathbf{n} \geq z\}}$ on the both sides, we have

$$\begin{aligned} \sum_{k : \nu'_k = z} \kappa_k \mathbf{c}^{\nu_k - z} \omega(\mathbf{n} - z)\mathbf{1}_{\{\mathbf{n} \geq z\}} &= \sum_{k : \nu_k = z} \kappa_k \mathbf{c}^{\nu_k - z} \omega(\mathbf{n} - z)\mathbf{1}_{\{\mathbf{n} \geq z\}}, \\ \sum_{k : \nu'_k = z} \kappa_k \mathbf{c}^{\nu_k - \nu'_k} \omega(\mathbf{n} - \nu'_k)\mathbf{1}_{\{\mathbf{n} \geq \nu'_k\}} &= \sum_{k : \nu_k = z} \kappa_k \omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n} \geq \nu_k\}}. \end{aligned}$$

Hence, by summing these up for all $z \in \mathcal{C}$ and multiplying $\mathbf{c}^{\mathbf{n}}$ both sides, we get

$$\mathbf{c}^{\mathbf{n}} \sum_k \kappa_k \mathbf{c}^{\nu_k - \nu'_k} \omega(\mathbf{n} - \nu'_k)\mathbf{1}_{\{\mathbf{n} \geq \nu'_k\}} = \mathbf{c}^{\mathbf{n}} \sum_k \kappa_k \omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n} \geq \nu_k\}} \tag{S13}$$

because every reaction has exactly one complex as its source complex and product complex. Since the right hand side of Eq. (S13) is equal to the RHS (Eq. (S12)), we need to show that the left hand side of Eq. (S13) is equal to the LHS (Eq. (S11)).

If $k \in K_\Gamma(\mathbf{n})^c$, then $\mathbf{n} - (\nu'_k - \nu_k) \in \Gamma^c$ meaning that $n_i - \nu'_{ki} \leq n_i - (\nu'_{ki} - \nu_{ki}) < b_i$ for some i and thus $\mathbf{n} - \nu'_k \in \Gamma^c$. Since the network has the complex balanced equilibrium, the network is weakly reversible [7]. Thus, there exists a reaction $\tilde{k}$ whose source complex $\nu_{\tilde{k}}$ is equal to ν'_k. For such $\tilde{k}$, by the definition of $\mathbf{b}$, $\lambda_{\tilde{k}}(\mathbf{n}) = \kappa_{\tilde{k}}\theta(\mathbf{n})\omega(\mathbf{n} - \nu_{\tilde{k}})\mathbf{1}_{\{\mathbf{n}\geq\nu_{\tilde{k}}\}} = 0$ because $\mathbf{n} \geq \mathbf{b}$ and $\mathbf{n} - \nu_{\tilde{k}} \in \Gamma^c$ (i.e., $\mathbf{n} - \nu_{\tilde{k}} \not\geq \mathbf{b}$). Since $\theta(\mathbf{n}) > 0$, furthermore, $\omega(\mathbf{n} - \nu_{\tilde{k}})\mathbf{1}_{\{\mathbf{n}\geq\nu_{\tilde{k}}\}} = \omega(\mathbf{n} - \nu'_k)\mathbf{1}_{\{\mathbf{n}\geq\nu'_k\}} = 0$. Hence, the left hand side of Eq. (S13) is equal to the Eq. (S11) and therefore, Eq. (S11) and Eq. (S12) are equal. □

Example 1. Here, we illustrate Theorem 2 with the translated EGFR network in Fig. 3a:

$$0 \xrightarrow{\alpha_1} A \underset{\alpha_4 n_{A_P}}{\overset{\alpha_2 n_A(n_A-1)}{\rightleftarrows}} A_P, \qquad A_P \xrightarrow{\alpha_3 n_{A_P}} 0$$

The CBE $\mathbf{c} = (c_A, c_{A_P})$ of the deterministic mass action model satisfies the balancing equations for the complex 0: $\alpha_3 c_{A_P} = \alpha_1$, for the complex A: $\alpha_1 + \alpha_4 c_{A_P} = \alpha_2 c_A$, and for the complex A_P : $\alpha_2 c_A = \alpha_3 c_{A_P} + \alpha_4 c_{A_P}$. Thus, $\mathbf{c} = (c_A, c_{A_P}) = (\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3}, \frac{\alpha_1}{\alpha_3})$. For this example, $\mathbf{b} = (1, 0)$ since $\lambda_k(\mathbf{n}) > 0$ for $\mathbf{n} \geq \nu_k + (1, 0)$, and $\lambda_k(\mathbf{n}) = 0$ for $\mathbf{n}$ such that $\mathbf{n} \geq (1, 0)$ and $\mathbf{n} \not\geq \nu_k + (1, 0)$ for all $k = 1, \ldots, 4$. Letting $f_A(n_A) = n_A(n_A - 1)$ and $f_{A_p} = n_{A_P}$, the condition in Theorem 1 holds. Therefore, the propensity factorization $\lambda_k(\mathbf{n}) = \kappa_k\theta(\mathbf{n})\omega(\mathbf{n} - \nu_k)\mathbf{1}_{\{\mathbf{n}\geq\nu_k\}}$ is given by

$$\kappa_k = \alpha_k, \quad \theta(n_A, n_{A_P}) = n_A!(n_A - 1)!n_{A_P}!$$

and

$$\omega(n_A, n_B) = \frac{1}{\theta(n_A, n_{A_P})}\mathbf{1}_{\{n_A\geq 1\}} = \frac{1}{n_A!(n_A - 1)!n_{A_P}!}\mathbf{1}_{\{n_A\geq 1\}}$$

for $(n_A, n_{A_P}) \in \Gamma = \{(n_A, n_{A_P}) \mid n_A \geq 1, n_{A_P} \geq 0\}$. By Theorem 2, a stationary measure $\pi(n_A, n_{A_P})$ is given by

$$\pi(n_A, n_{A_P}) = \begin{cases} M\dfrac{\left(\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3}\right)^{n_A}\left(\frac{\alpha_1}{\alpha_3}\right)^{n_{A_P}}}{n_A!(n_A - 1)!n_{A_P}!} & \text{if } (n_A, n_{A_P}) \in \Gamma, \\ 0 & \text{if } (n_A, n_{A_P}) \in \Gamma^c. \end{cases}$$

where M is a normalizing constant. $\pi(n_A, n_{A_P})$ is indeed a stationary distribution because it is summable over Γ as follows:

$$\begin{aligned} \sum_{(n_A,n_{A_P})\in\Gamma} \frac{\left(\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3}\right)^{n_A}\left(\frac{\alpha_1}{\alpha_3}\right)^{n_{A_P}}}{n_A!(n_A - 1)!n_{A_P}!} &\leq \sum_{(n_A,n_{A_P})\in\Gamma} \frac{\left(\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3}\right)^{n_A}\left(\frac{\alpha_1}{\alpha_3}\right)^{n_{A_P}}}{n_A!n_{A_P}!} \\ &\leq \sum_{(n_A,n_{A_P})\in\mathbb{Z}^2_{\geq 0}} \frac{\left(\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3}\right)^{n_A}\left(\frac{\alpha_1}{\alpha_3}\right)^{n_{A_P}}}{n_A!n_{A_P}!} \\ &= \exp\left(\frac{\alpha_1}{\alpha_2} + \frac{\alpha_1\alpha_4}{\alpha_2\alpha_3} + \frac{\alpha_1}{\alpha_3}\right) < \infty. \end{aligned}$$

Moreover, since Γ is closed, the states $(n_A, n_{A_P}) \in \Gamma^c$ are transient as each state in Γ^c is reachable to Γ with either $0 \to A$ or $A_p \to A$. This implies that the stochastic process can visit the state $\mathbf{n} \in \Gamma^c$ only finitely many times. Indeed, if $(n_A, n_{A_P}) \in \Gamma^c$, i.e., $n_A = 0$, then only the reactions $0 \to A$, $A_P \to A$ and $A_P \to 0$ can occur in Γ^c. Thus, the process moves along one of the stoichiometric vectors $(1, 0), (1, -1)$, and $(0, -1)$, then it eventually escapes from

Γ^c to Γ and never comes back to Γ^c. Due to the transient states, Theorem 6.6, which is the most general version of the previous work [2], is not applicable to this example.

Let us compute the marginal means and variances of n_A and n_{A_P}. For n_{A_P}, the marginal distribution

$$\pi(n_{A_P}) = M_1 \frac{(\alpha_1/\alpha_3)^{n_{A_P}}}{n_{A_P}!}, \quad n_{A_P} \geq 0,$$

which is a Poisson distribution with rate $\frac{\alpha_1}{\alpha_3}$, so both the mean and the variance are $\frac{\alpha_1}{\alpha_3}$. Thus, the coefficient of variation (CV) is $\sqrt{\frac{\alpha_3}{\alpha_1}}$, and the Fano factor, defined as the variance divided by the mean, is always 1. For n_A, the marginal distribution

$$\pi(n_A) = M_2 \frac{u_0^{n_A}}{(n_A-1)!n_A!}, \quad n_A \geq 1$$

where $u_0 = (\alpha_1/\alpha_2 + \alpha_1\alpha_4/\alpha_2\alpha_3)$, and M_2 is the normalizing constant.

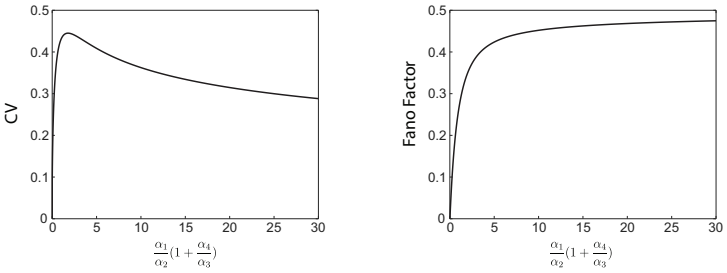
$$M_2 = \left(\sum_{n_A=1}^{\infty} \frac{u_0^{n_A}}{n_A!(n_A-1)!} \right)^{-1} = [\sqrt{u_0} I_1(2\sqrt{u_0})]^{-1}$$

where $I_m(x)$ is the modified Bessel function of the first kind. Then the mean and variance are given by

$$\mathbb{E}[n_A] = \sum_{n_A=1}^{\infty} n_A \pi(n_A) = \sqrt{u_0} \frac{I_0(2\sqrt{u_0})}{I_1(2\sqrt{u_0})},$$
$$\mathrm{Var}(n_A) = \mathbb{E}[n_A^2] - (\mathbb{E}[n_A])^2 = \sqrt{u_0} \frac{I_0(2\sqrt{u_0})}{I_1(2\sqrt{u_0})} + u_0 - u_0 \frac{I_0(2\sqrt{u_0})^2}{I_1(2\sqrt{u_0})^2}.$$

The CV and the Fano factor are also given by

$$\mathrm{CV} = \frac{\sqrt{\mathrm{Var}(n_A)}}{\mathbb{E}[n_A]} = \sqrt{\frac{\mathbb{E}[n_A^2]}{\mathbb{E}[n_A]^2} - 1} = \sqrt{\frac{1}{\sqrt{n_0}} \frac{I_1(2\sqrt{u_0})}{I_0(2\sqrt{u_0})} + \frac{I_0(2\sqrt{u_0})^2}{I_1(2\sqrt{u_0})^2} - 1},$$
$$\text{Fano factor} = \frac{\mathrm{Var}(n_A)}{\mathbb{E}[n_A]} = 1 + \sqrt{u_0} \frac{I_1(2\sqrt{u_0})}{I_0(2\sqrt{u_0})} - \sqrt{u_0} \frac{I_0(2\sqrt{u_0})}{I_1(2\sqrt{u_0})}.$$



Supplementary Figure 1. The coefficient of variation (CV) and Fano factor of the stationary distribution of n_A in Fig. 3A can be determined by the single parameter $u_0 = \frac{\alpha_1}{\alpha_2}(1 + \frac{\alpha_4}{\alpha_3})$. The CV attains the maximum at $u_0 \approx 1.8$, and the Fano factor monotonically increases. Since the Fano factor is always less than 1, this distribution is sub-Poissonian.

Remark 3. By the basic Markov properties [10], Theorem 2 not only provides a stationary measure but also characterizes the status of states. Since $\pi(\mathbf{n}) > 0$ for each $\mathbf{n} \in \Gamma$, every state in Γ is recurrent as long as the process is non-explosive (i.e., well-defined for all time t). Since Γ is a closed set, furthermore, if a state $\mathbf{n} \in \Gamma^c$ is reachable to Γ, then $\mathbf{n}$ is transient. Another special case is that a state $\mathbf{n}$ is isolated state, i.e., $\lambda_k(\mathbf{n}) = 0$ for all k. In this case, the set of the single element $\{\mathbf{n}\}$ is an irreducible subset, and $\pi(\mathbf{n}) = 1$ on $\{\mathbf{n}\}$ is a stationary distribution, obviously.

Remark 4. This theorem does not require nor imply the irreducibility of Γ. If Γ is not irreducible then for an irreducible subset $\Gamma_g \subset \Gamma$, we consider a measure $\pi_{\Gamma_g} = \pi|_{\mathbf{n}\in\Gamma_g}$ restricted on Γ_g. This measure is still a stationary measure and can be a stationary distribution on Γ_g as long as π_{Γ_g} is summable over Γ_g. For instance, in the example in Fig. 3d, Γ is not an irreducible set, but a union of irreducible subsets $\Gamma_{T_0} = \{(x_1, x_2, x_3) \in \mathbb{Z}^3_{\geq 0} \mid x_1 \geq 1 \text{ and } x_1 + x_2 + x_3 = T_0\}$. Since the stationary measure π obtained by Theorem 2 is summable on Γ_{T_0} for each T_0, the restricted stationary measure $\pi_{\Gamma_{T_0}}$ can be normalized to be a stationary distribution on the irreducible subset. In this remark, we prove the irreducibility of $\{(x_1, x_2, x_3) \in \mathbb{Z}^3_{\geq 0} \mid x_1 \geq 1 \text{ and } x_1 + x_2 + x_3 = T_0\}$ for Fig. 3d. We say that for a Markov chain X, state y is *accessible* from state x and write $x \to_{\mathrm{a}} y$ if X can reach y starting from x with positive probability, that is, the probabilities $P(X(t) = y \mid X(0) = x) > 0$ for some $t > 0$. We say that states x and y *communicate* and write $x \leftrightarrow_{\mathrm{c}} y$ if $x \to_{\mathrm{a}} y$ and $y \to_{\mathrm{a}} x$. For the Markov chain $X(t)$ associated with a translated network, we denote by $\lambda_{C_1\to C_2}$ and η the propensity of a reaction $C_1 \to C_2$ and the stoichiometric vector of the reaction, respectively. Note that if $\lambda_{C_1\to C_2}(x) > 0$ then $x \to_{\mathrm{a}} x + \eta$ because for a sufficiently small Δt,

$$P(X(\Delta t) = x + \eta \mid X(0) = x) = \lambda_{C_1\to C_2}(x)\Delta t + o(\Delta t).$$

By using this, for the translated network in Fig. 3d, we have $(n_A, n_{A_P}, n_{A_{PP}}) \leftrightarrow_{\mathrm{c}} (n_A - 1, n_{A_P} + 1, n_{A_{PP}})$ and $(n_A, n_{A_P}, n_{A_{PP}}) \leftrightarrow_{\mathrm{c}} (n_A - 1, n_{A_P}, n_{A_{PP}} + 1)$ for $n_A \geq 2$ and $n_A + n_{A_P} + n_{A_{PP}} = T_0$ because $\lambda_{A\to A_P}(n_A, n_{A_P}, n_{A_{PP}}) > 0, \lambda_{A_P\to A_{PP}}(n_A - 1, n_{A_P} + 1, n_{A_{PP}}) > 0$ and $\lambda_{A_{PP}\to A}(n_A - 1, n_{A_P}, n_{A_{PP}} + 1) > 0$, which imply that $(n_A, n_{A_P}, n_{A_{PP}}) \to_{\mathrm{a}} (n_A - 1, n_{A_P} + 1, n_{A_{PP}}) \to_{\mathrm{a}} (n_A - 1, n_{A_P}, n_{A_{PP}} + 1) \to_{\mathrm{a}} (n_A, n_{A_P}, n_{A_{PP}})$. As $\leftrightarrow_{\mathrm{c}}$ forms an equivalent class, any two states in $\{(x_1, x_2, x_3) \in \mathbb{Z}^3_{\geq 0} \mid x_1 \geq 1 \text{ and } x_1 + x_2 + x_3 = T_0\}$ communicate so it is irreducible when $T_0 \geq 2$. If $T_0 = 1$ then the set consists of a single element $(1, 0, 0)$, and it is irreducible itself.

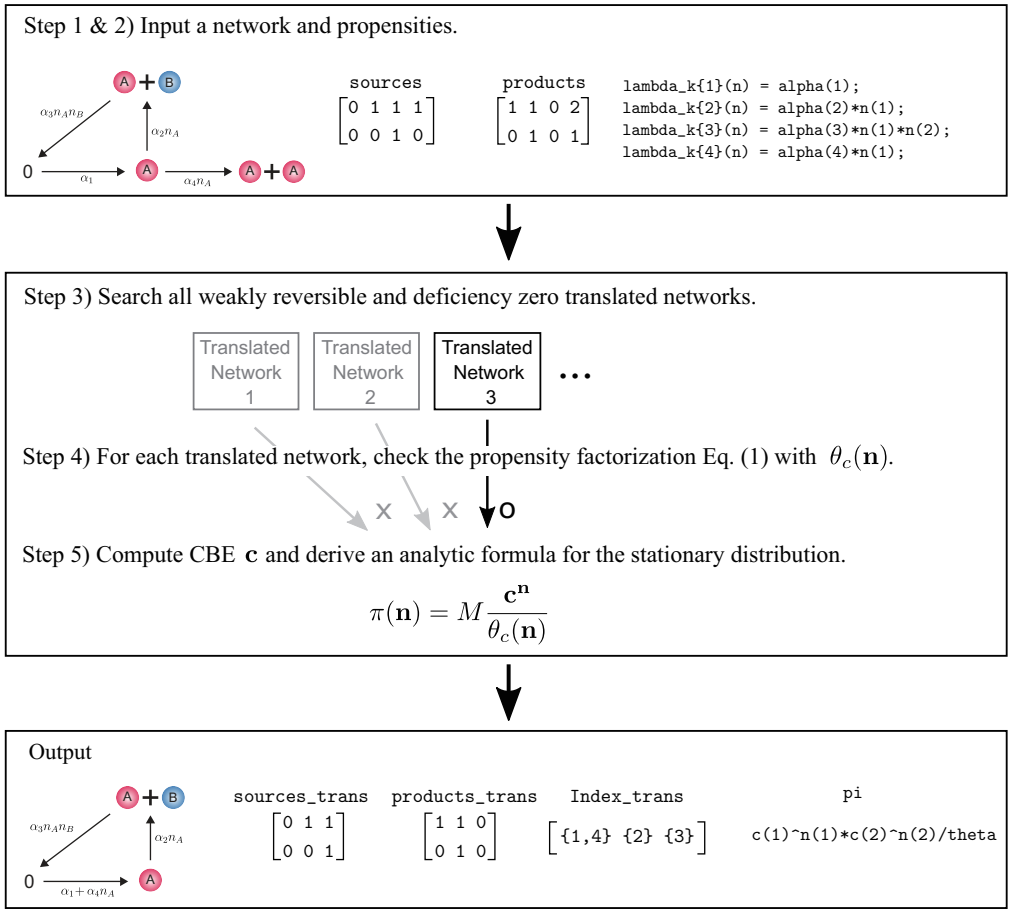
Similarly, for the translated network in Fig. 1 we have that $(n_A, n_B) \leftrightarrow_{\mathrm{c}} (n_A + 1, n_B)$ and $(n_A, n_B) \leftrightarrow_{\mathrm{c}} (n_A, n_B + 1)$ for any $(n_A, n_B) \in \mathbb{Z}^2_{\geq 0}$ because $\lambda_{0\to A}(n_A, n_B) > 0$, $\lambda_{A\to A+B}(n_A + 1, n_B) > 0$ and $\lambda_{A+B\to 0}(n_A + 1, n_B + 1) > 0$, which imply that $(n_A, n_B) \to_{\mathrm{a}} (n_A + 1, n_B) \to_{\mathrm{a}} (n_A + 1, n_B + 1) \to_{\mathrm{a}} (n_A, n_B)$. As $\leftrightarrow_{\mathrm{c}}$ forms an equivalent class, any two states in $\mathbb{Z}^2_{\geq 0}$ communicate and hence $\mathbb{Z}^2_{\geq 0}$ is irreducible. The irreducibilities of the other two examples in Fig. 3a, g can be shown in similar ways.

Remark 5. By introducing a new factor $\omega(\mathbf{n})$, Theorem 2 generalizes Theorem 6.6 in [2], which is the most general version of the previous work [2]. In particular, if the function $\omega(\mathbf{n})$ for the factorization is $\theta(\mathbf{n})^{-1}$ then Theorem 2 becomes equivalent to Theorem 6.6 in [2]. This generalization is demonstrated by three biologically relevant examples in Fig. 3. For instance, $\omega(\mathbf{n}) = \theta(\mathbf{n})^{-1}\mathbf{1}_{\{\mathbf{n}\geq\mathbf{b}\}}$ in Fig. 3a with the nonzero vector $\mathbf{b} = (1, 0)$, so this factorization is not directly covered by the previously identified form, where $\omega(\mathbf{n}) = \theta(\mathbf{n})^{-1}$. However, the additional characteristic function is somehow negligible under a change of variable $\mathbf{n}' = \mathbf{n} - \mathbf{b}$ since the state space can be restricted on $\{\mathbf{n} \geq \mathbf{b}\}$. Nevertheless, the usefulness of our theorem is clearly demonstrated by the examples in Fig. 1 and Supplementary Figs. 3 and 4 .

SUPPLEMENTARY NOTE 4: CASTANET: COMPUTATIONAL PACKAGE FOR DERIVING STATIONARY DISTRIBUTION FORMULAE

Applying our theoretical framework (Fig. 1) has two practical difficulties. Translating a given network to a weakly reversible deficiency zero network (Fig. 1a) is not straightforward as prohibitively many candidates of translated networks often exist. Furthermore, unless propensity functions follow the generalized mass action kinetics, it is challenging to check whether the factorization condition holds (Fig. 1b) as it requires to solve associated recurrence relations. Thus, we have developed an open-source and publicly available MATLAB code (GitHub repository: http://github.com/Mathbiomed/CASTANET) that performs our theoretical analysis to derive stationary distributions. Specifically, the package checks two conditions: whether a given BRN can be made weakly reversible and of

zero deficiency after network translation, and whether the propensities of the translated network can be factorized as in Eq. (S2). If these two conditions are satisfied, the package then calculates the analytic formula for a stationary distribution. One can easily run our code by simply entering the source complexes, product complexes, and propensity functions of reactions.



Supplementary Figure 2. A step-by-step schematic diagram of CASTANET. Here, $\theta_c(\mathbf{n})$ is a candidate for propensity factorization. See the manual for details.

Manual for the code

We explain how to enter the input and run the code `CRN_main.m` with our example in Fig. 1a as follows.

Step 1) Enter the source and the product complexes into `sources` and `products` as column vectors.

```
% Fig.1 example
sources = [0 0; 1 0; 1 1; 1 0]';
products = [1 0; 1 1; 0 0; 2 0]';
```

```
>> disp(sources)
     0 1 1 1
     0 0 1 0

>> disp(products)
```

```
1 1 0 2
0 1 0 0
```

The ith columns of `source` and `product` represent the source complex and the product complex of the ith reaction, respectively. For instance, the fourth columns of `source` ($[1,0]^{\mathrm{T}}$) and `product` ($[2,0]^{\mathrm{T}}$) describe the source and product complexes of the fourth reaction: $A(=1A+0B) \to A+A(=2A+0B)$.

Step 2) Enter the propensity functions of all reactions into the cell variable `lambda_k` and run the section `'Initialization of all input variables and parameters'` to set up all the input variables.

```
lambda_k{1}(n) = alpha(1);
lambda_k{2}(n) = alpha(2)*n(1);
lambda_k{3}(n) = alpha(3)*n(1)*n(2);
lambda_k{4}(n) = alpha(4)*n(1);
```

Note that the propensity functions do not need to follow the mass action kinetics. The symbolic variable `alpha(k)` represents the rate constants of the reactions. Users can also fix the values of `alpha(k)` (e.g., `alpha(1) = 3;`).

Step 3) Run the section `'Performing Network translation'` to obtain translated networks.

```
[Solution,Index] = CRN_translation(sources, products, 2);
```

The third input argument `2` means that the function searches all translated networks whose reaction order is at most two (i.e., bimolecular). It can be changed to any other positive integer. The ith row of `Solution` contains the source and product complexes of the ith translated network. Each element in the ith row of `Index` is a set of indices indicating which reactions in the original network are merged to form a reaction in the ith translated BRN. For instance, if the ij entry of `Index` is $\{1,3\}$, then this indicates that the first and third reactions in the original BRN were merged to form the jth reaction in the ith translated network. As the outcome, the number of identified weakly reversible deficiency zero translated networks is reported:

```
The number of weakly reversible deficiency zero translated networks is 2.
```

Step 4) Run the section `'Performing Propensity factorization'` to identify a translated network whose propensity functions satisfy the factorization condition (Eq. (S2)) among all the translated networks obtained in Step 3. Specifically, for each translated network, the code constructs a candidate ($\theta_{\mathrm{c}}(\mathbf{n})$) for the function $\theta(\mathbf{n})$ and checks the factorization conditions with the candidate $\theta_{\mathrm{c}}(\mathbf{n})$. This candidate is necessarily the desired function $\theta(\mathbf{n})$ if there exists a factorization (see the next subsection for details). The key function `CRN_theta_construction()` provides the candidate, and `CRN_check_factorization_condition()` examines whether the factorization condition (Eq. (S2)) holds. Since the example in Fig. 1 has a translated network satisfying the factorization condition, the code successfully finds the factorization and displays the following line:

```
The factorization condition holds for the translated network 1!
```

If none of the translated networks have the desired propensity factorization, the code displays the following line:

```
No translated network satisfying the factorization condition is
identified.
```

Step 5) If a translated network having the propensity factorization is found in Step 4, run the section `'Compute CBE and derive a stationary distribution pi(n)'` to compute a complex balanced equilibrium and analytically derive a stationary distribution. Then the code will provide the translated network and its stationary distribution formula, which is the same as the stationary distribution of the original network.

Output)

We have five outputs: `source_trans`, `product_trans`, `Index_trans`, `lambda_trans`, and `pi`. The first two outputs represent the stochiometric vectors of the source complexes and the product complexes in the translated network identified in Step 4. `Index_trans` is the row of `Index` corresponding to the translated network. `lambda_trans` contains the propensity functions of the translated network, and `pi` is a symbolic expression for the stationary distribution.

```
The source complexes of the translated network:
      0 1 1
      0 0 1

The product complexes of the translated network:
      1 1 0
      0 1 0

The index of the translated network:
      {1,4} {2} {3}
ni is the number of the ith species.
1st reaction propensity of the translated network:
alpha1 + alpha4*n1

2nd reaction propensity of the translated network:
alpha2*n1

3rd reaction propensity of the translated network:
alpha3*n1*n2

Analytic formula for the stationary distribution:
piecewise(~in((alpha1 - alpha4)/alpha4, 'integer') |
~(alpha1 - alpha4)/alpha4 in Dom::Interval([-n1], [-1]),
(alpha2^(n2 - n1)*alpha4^n1*gamma((alpha1 + alpha4*n1)/alpha4))/
(alpha3^n2*gamma(n1 + 1)*factorial(n2)*gamma(alpha1/alpha4)))
```

Here, `piecewise(condition, value)` means conditionally defined function in MATLAB. If `condition` holds then `value` is the function output. Since MATLAB symbolic expression uses `gamma(a+k+1)/gamma(a+1)` to represent $(a+1)(a+2)\cdots(a+k)$ and the domain of the gamma function does not contain the non-positive integers, the above complicated expression appears. The above expression certainly the same to the theoretical result.

Note that, technically, this is a formula for a stationary *measure*, not a distribution. To get the stationary distribution, the stationary measure needs to be normalized so that the sum of the probabilities over the state space is one, which is possible only when the formula is summable over the state space.

Underlying algorithm of the code

The code generates all possible network translations under a user-defined maximum reaction order (e.g., 1, 2, and 3 for unimolecular, bimolecular, and termolecular reaction networks, respectively). Among these translated networks, the code identifies weakly reversible deficiency zero networks. However, because there are sometimes prohibitively many translated networks (e.g., 864 candidates for the example in Fig. 1 with maximum reaction order 3), checking weak reversibility and deficiency for all translated network is extremely inefficient. In particular, it greatly increases computational cost to check weak reversibility and count the number of linkage classes via a connected components search algorithm (Tarjan's algorithm [12]). Thus, before performing this calculation, we first simply check whether a

translated network can have a desired property by using the following necessary conditions of being a weakly reversible deficiency zero network, which greatly reduce computational cost.

Theorem 3 (Necessary condition of a network to be weakly reversible after network translation)**.** *For a given biochemical reaction network, if there exists a weakly reversible translated network of the given network then for each species index i, one of the following two conditions holds.*

1. *The ith coordinates of the stoichiometric vectors of all reactions, $\nu'_{ik} - \nu_{ik}$, are zero. In other words, $\{\nu'_{i1} - \nu_{i1}, \nu'_{i2} - \nu_{i2}, \ldots, \nu'_{ir} - \nu_{ir}\} = \{0\}$.*
2. *There exist both positive and negative numbers among the ith coordinates of the stoichiometric vectors of all reactions. In other words, the set $\{\nu'_{i1} - \nu_{i1}, \nu'_{i2} - \nu'_{i2}, \ldots, \nu'_{ir} - \nu'_{ir}\}$ has both positive and negative elements.*

Proof. This can be proved by contradiction. Suppose that there exists a species index i such that the set of ith coordinates of all the stoichiometric vectors has only non-negative (non-positive) elements and has at least one positive (negative) element. Note that this is the negation of the above statement. Since the set of stoichiometric vectors is invariant under network translation, the weakly reversible translated network also has the same set of stoichoimetric vectors as the original network does. Note that every reaction in a weakly reversible network belongs to a closed cycle. However, the reaction whose stoichiometric vector has positive ith coordinate cannot be contained in any closed cycle because the ith coordinate of every other stoichiometric vector is non-negative. □

If the necessary condition is not satisfied for a given BRN, we do not need to check weak reversibility of all translated networks, which greatly reduces computational cost.

Theorem 4 (Necessary condition of a network to have zero deficiency)**.** *For a given biochemical reaction network, let n, l, and s be the number of complexes, the number of linkage classes, and the dimension of the subspace spanned by the stoichiometric vectors, respectively. If the deficiency of the network is zero, then $s + 1 \leq n \leq 2s$.*

Proof. The deficiency δ is given by $n - l - s$. Since the deficiency of the given network is zero, $n = l + s$. As long as the given network is not the empty network (i.e. $\mathcal{S} = \mathcal{C} = \mathcal{R} = \emptyset$), there exists at least one linkage class, so $1 \leq l$. Each linkage class consists of at least two complexes so $n(= l + s) \geq 2l$ and thus $l \leq s$. Therefore, $s + 1 \leq n = l + s \leq 2s$. □

This condition might appear useless because n can vary among translated network. Nevertheless, s is invariant under network translation, and furthermore, we no longer need to calculate the number of linkage classes l for the condition in Theorem 4, which requires to perform the graph search algorithm and thus takes high computational cost. Hence, by avoiding this calculation for translated networks that do not satisfy the necessary condition, we can greatly reduce computational cost.

After identifying a weakly reversible and of zero deficiency network, the code checks whether each translated network satisfies the factorization condition (Eq. (S2)) by constructing the explicit formula for a candidate $\theta_{\mathrm{c}}(\mathbf{n})$ for the function $\theta(\mathbf{n})$ in Eq. (S2). The candidate $\theta_{\mathrm{c}}(\mathbf{n})$ can be constructed as follows. We first derive the recurrence relation of the function $\theta(\mathbf{n})$. If there exists a desired propensity factorization, from the factorization conditions for reactions i and j at $\mathbf{n} - \nu_j$ and $\mathbf{n} - \nu_i$,

$$\lambda_i(\mathbf{n} - \nu_j) = \kappa_i \theta(\mathbf{n} - \nu_j)\omega(\mathbf{n} - \nu_j - \nu_i),$$
$$\lambda_j(\mathbf{n} - \nu_i) = \kappa_j \theta(\mathbf{n} - \nu_i)\omega(\mathbf{n} - \nu_i - \nu_j).$$

By dividing the above equations, we get the following recurrence relation:

$$\theta(\mathbf{n} - \nu_j) = \frac{\lambda_i(\mathbf{n} - \nu_j)/\kappa_i}{\lambda_j(\mathbf{n} - \nu_i)/\kappa_j} \times \theta(\mathbf{n} - \nu_i). \tag{S14}$$

Let $\mathbf{n}_0$ be a state satisfying $\lambda_k(\mathbf{n}_0) > 0$ for all k. Let us consider a sequence of pairs of source complexes, $(\nu_{a_1}, \nu_{b_1}), (\nu_{a_2}, \nu_{b_2}), \ldots, (\nu_{a_m}, \nu_{b_m})$, such that

$$\mathbf{n}_0 + \sum_{j=1}^{m} (\nu_{a_j} - \nu_{b_j}) = \mathbf{n}. \tag{S15}$$

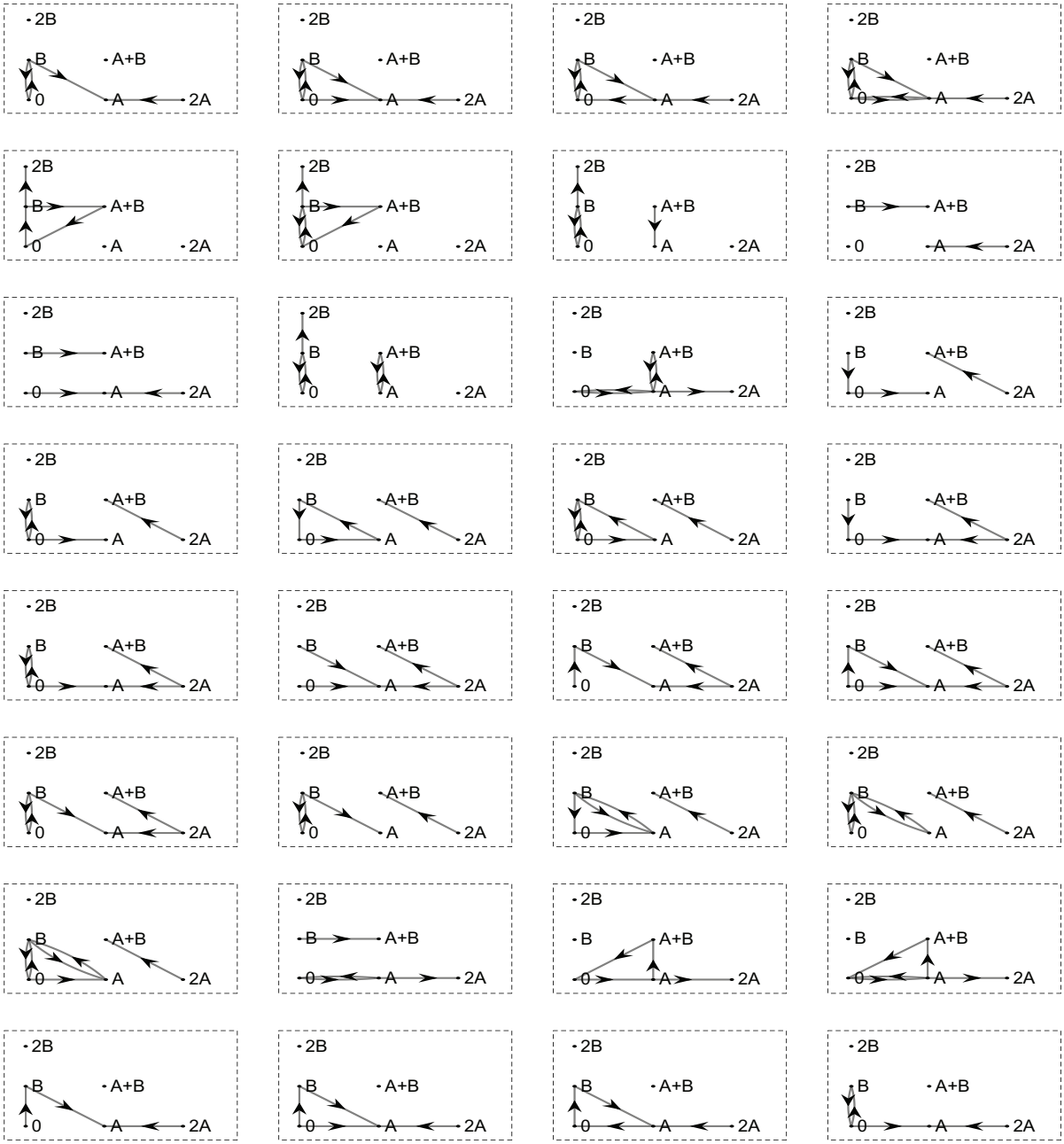
This means that $\mathbf{n}_0 \to \mathbf{n}_0 + (\nu_{a_1} - \nu_{b_1}) \to \cdots \to \mathbf{n}_0 + \sum_{j=1}^{m} (\nu_{a_j} - \nu_{b_j}) = \mathbf{n}$ is a path from $\mathbf{n}_0$ to $\mathbf{n}$. For a pair of $\mathbf{n}$ and $\mathbf{n}_0$, in order to identify this type of paths, we use the row-style Hermite normal form, an analogue of reduced echelon form for a matrix over the integers. Specifically, non-zero rows of the Hermite normal form of the matrix whose jth row is $\nu_{j+1} - \nu_1$ form a basis for the lattice as $\nu_i - \nu_j = (\nu_i - \nu_1) - (\nu_j - \nu_1)$. Thus, by using this basis, one can identify a path from $\mathbf{n}_0$ to $\mathbf{n}$ [3, 5]. Note that such a path is not unique in general but it is enough to consider a single path (see below). We multiply Eq. (S14) along this path, and then we obtain

$$\theta(\mathbf{n}) = \prod_{j=1}^{m} \frac{\lambda_{a_j}(\mathbf{n}_0 + \sum_{i=1}^{j} (\nu_{a_i} - \nu_{b_i}))/\kappa_{a_j}}{\lambda_{b_j}(\mathbf{n}_0 + \sum_{i=1}^{j-1} (\nu_{a_i} - \nu_{b_i}))/\kappa_{b_j}} \times \theta(\mathbf{n}_0). \tag{S16}$$

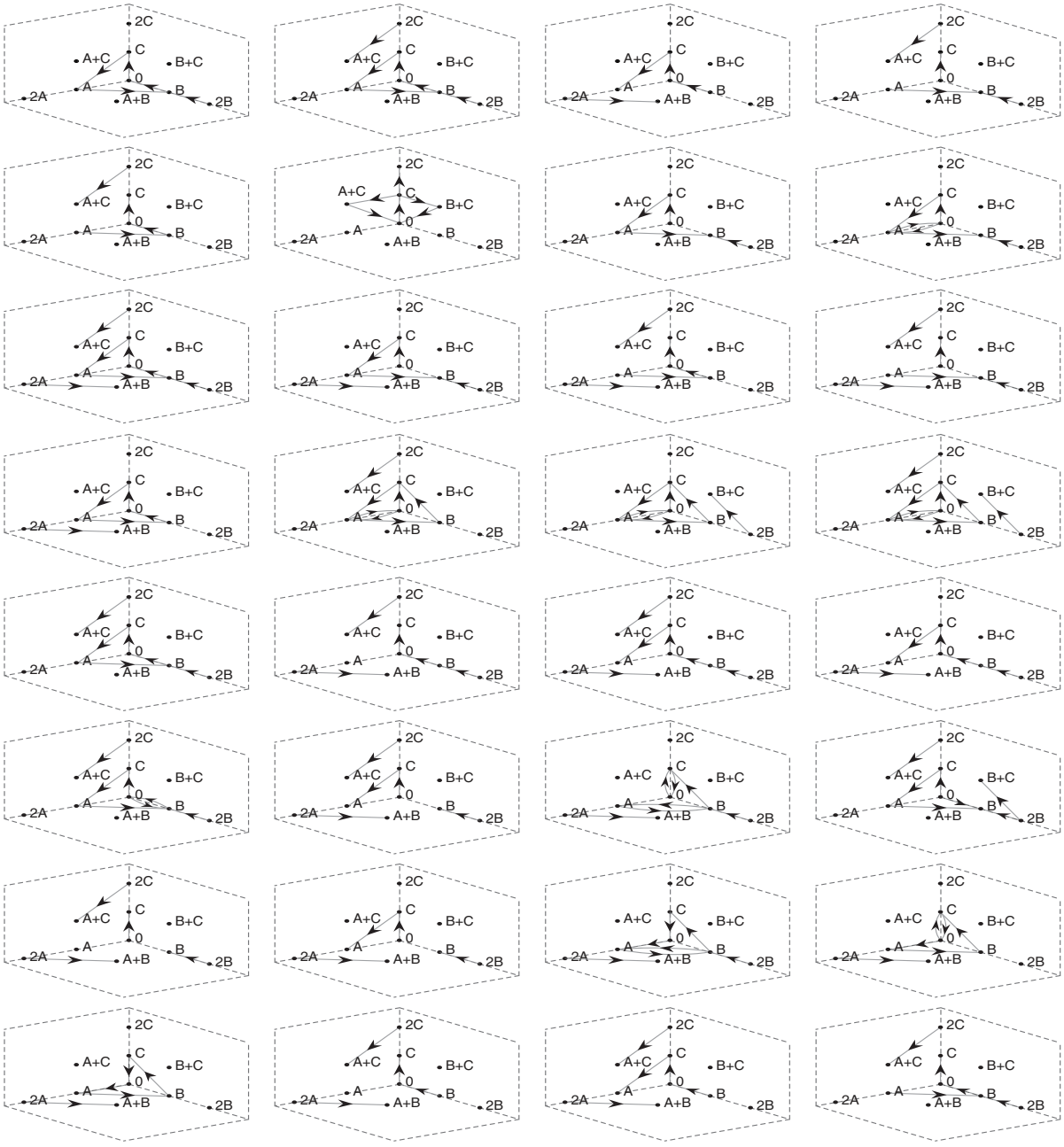
Note that the RHS of Eq. (S16) can be obtained by substituting $\mathbf{n}$ in Eq. (S14) with $\mathbf{n} + \nu_i$ and one can assume $\theta(\mathbf{n}_0) = 1$ without loss of generality as the factorization holds up to constants.

For a special case, CASTANET uses a more efficient way to identify $\theta(\mathbf{n})$. Specifically, if the propensity functions of a translated network follow the generalized stochastic mass action kinetics and all the complexes consist of zero or one species, then CASTANET constructs the function $\theta(\mathbf{n})$ by using Eq. (S7).

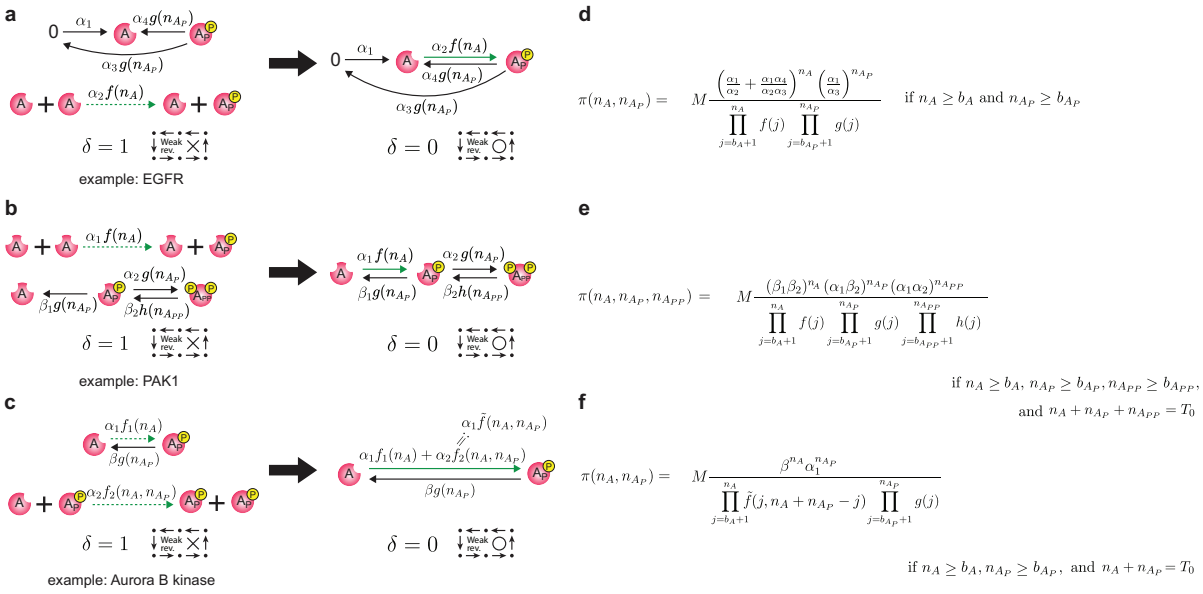
Based on this expression for $\theta(\mathbf{n})$, the code constructs the candidate $\theta_{\mathrm{c}}(\mathbf{n})$ using Eq. (S16) with a path from $\mathbf{n}_0$ to $\mathbf{n}$. If there exists a desired function $\theta(\mathbf{n})$ for propensity factorization then it must be the same as the candidate $\theta_{\mathrm{c}}(\mathbf{n})$ because both should be expressed as Eq. (S16). Thus, the propensity factorization condition (Eq. (S2)) can be checked using $\theta_{\mathrm{c}}(\mathbf{n})$. In the package, `CRN_find_elementary_path()` generates a path from $\mathbf{n}_0$ to $\mathbf{n}$, `CRN_theta_construction()` constructs $\theta_{\mathrm{c}}(\mathbf{n})$, and `CRN_check_factorization_condition()` examines the factorization condition. The functions `CRN_find_elementary_path()`, `CRN_find_elementary_function()` and `CRN_solve_sym_linear()` are auxiliary functions to preprocess the input variables into appropriate forms for the construction code. See `README.md` in the GitHub repository for details. Finally, a complex balanced equilibrium of the deterministic mass action model for the translated network is determined by solving the algebraic equation for the complex balanced equilibrium with the function `CRN_compute_cbe()`.



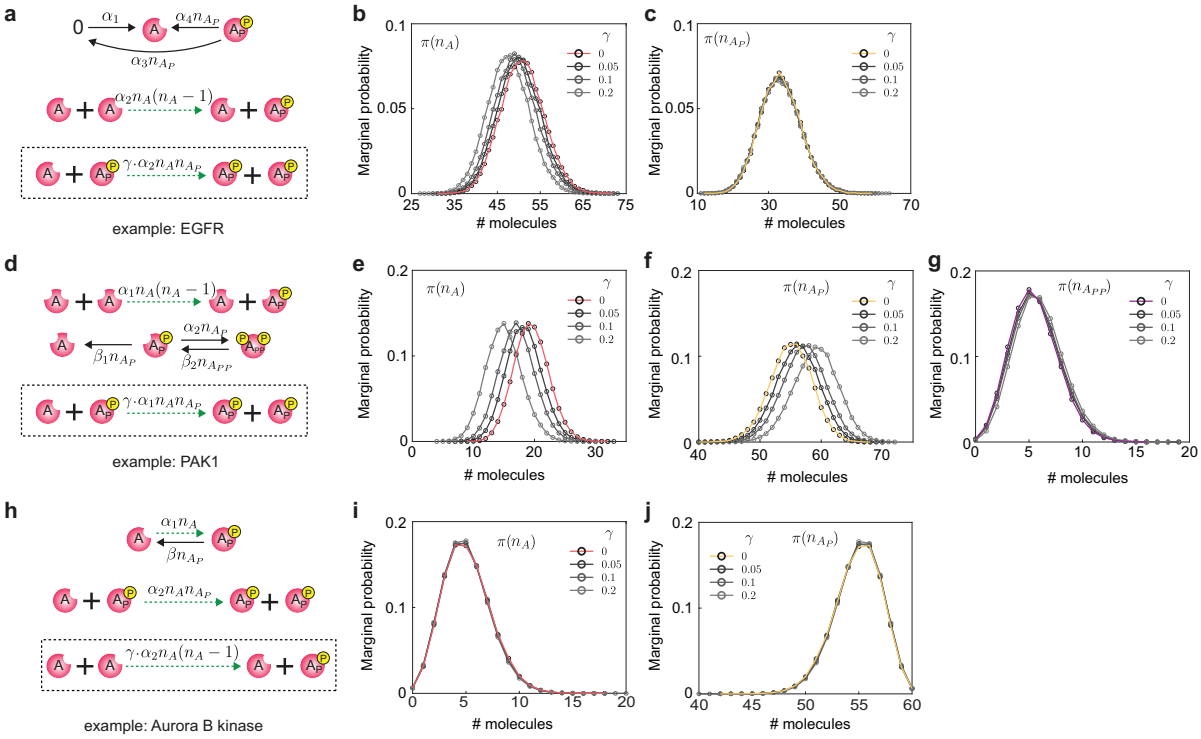
Supplementary Figure 3. Our computational package CASTANET identifies BRNs with two species whose stationary distribution can be analytically derived. Specifically, CASTANET translates these networks to be weakly reversible and of zero deficiency and then finds factorizations for the propensity functions of the translated networks as in Eq. (S8), thus calculating their stationary distributions. The figure shows the first 32 networks identified by our package among randomly selected networks with two species (A and B) and at most bimolecular reactions (e.g., $A + B \to A$). More examples can be found in https://github.com/Mathbiomed/CASTANET. All kinetics are assumed to follow the mass action kinetics. All rate constants are set to be one to reduce complexity while arbitrary rate constants are allowed to derive stationary distributions.



Supplementary Figure 4. Our computational package CASTANET identifies BRNs with three species whose stationary distribution can be derived. Specifically, CASTANET translates these networks to be weakly reversible and of zero deficiency and then finds factorizations for the propensity functions of the translated networks as in Eq. (S8), thus calculating their stationary distributions. The figure shows the first 32 networks identified by our package among randomly selected networks with three species (A, B, and C) and at most bimolecular reactions (e.g., $A + B \to C$). More examples can be found in https://github.com/Mathbiomed/CASTANET. All kinetics are assumed to follow the mass action kinetics. All rate constants are set to be one to reduce complexity while arbitrary rate constants are allowed to derive stationary distributions.



Supplementary Figure 5. Stationary distributions of autophosphorylation networks with non-elementary reactions (e.g., Hill functions). **a, b, c** Autophosphorylation networks with non-elementary propensity functions. The propensity functions f, f_1, f_2, g, and h can be Michaelis-Menten or Hill functions, which are often used to describe autophosphorylation, phosphatase-mediated dephosphorylation, and proteasomal degradation. The propensity functions f, f_1, f_2, g, and h can be Michaelis-Menten or Hill functions, which are often used to describe autophosphorylation, phosphatase-mediated dephosphorylation, and proteasomal degradation. In fact, $f(n_A)$ can be other functions as long as there exists b_A such that $f(n_A) > 0$ if and only if $n_A > b_A$. For the Michaelis-Menten and Hill-type functions, $b_A = 0$ because $f(n_A) > 0$ if and only if $n_A > b_A = 0$. The restriction $f(n_A) > 0$ if and only if $n_A > b_A$ means simply that the propensity is positive when there is a large enough number of A molecules (i.e., $n_A > b_A$), and the value of b_X is, in turn, the minimum number of species A in the recurrent state. This indicates that our method can be applied to any typical propensity function f for these networks. Similarly, $\tilde{f}, g$, and h can be other functions satisfying the natural positivity condition (e.g, $g(n_{A_P}) > 0$ if and only if $n_{A_P} > b_{A_P}$). **d, e, f** Stationary distributions of the autophosphorylation networks with the non-elementary reactions. Note that the stationary distributions of the networks in Fig. 3 are special cases of these formulae.



Supplementary Figure 6. The stationary distributions of modified autophosphorylation networks with additional trans-autophosphorylation reactions (dotted boxes). These can be approximated by the stationary distributions of the original networks as long as the relative magnitude (γ) of the rate constant of a new trans-autophosphorylation reaction (dotted box) is sufficiently small compared to that of the original trans-autophosphorylation reaction. **a** Trans-autophosphorylation reaction $A + A_P \to A_P + A_P$ is added to the original network (Fig. 3a). Due to this modification, although the modified network can be translated to a weakly reversible deficiency zero network, a desired propensity factorization does not exist. **b** Nevertheless, the stationary distributions $\pi(n_A)$ of the modified network ($\gamma = 0.05, 0.1, 0.2$, gray lines) calculated from stochastic simulations can be closely approximated by the stationary distribution of the original network ($\gamma = 0$, colored line), which is derived by our method (Fig. 3b, c). **c** Even the stationary distributions $\pi(n_{A_P})$ of the modified network are nearly identical to those of the original network. **d** Similar to **a**, adding $A + A_P \to A_P + A_P$ makes their propensities unable to be factorized as in Eq. (S2) even after network translation. **e, f, g** The stationary distributions, $\pi(n_A)$, $\pi(n_{A_P})$, and $\pi(n_{A_{PP}})$ of the modified network can be approximated by the stationary distributions of the original network ($\gamma = 0$), which are obtained by our method (Fig. 3e, f). **h** Another trans-autophosphorylation reaction $A + A \to A + A_P$ is added to the original network (Fig. 3g). **i,j** The stationary distributions $\pi(n_A)$ and $\pi(n_{A_P})$ of the modified network are nearly identical to the stationary distributions of the original network ($\gamma = 0$). Note that the stationary distribution of the modified network in **h** can still be analytically derived because there exists a desired propensity factorization for this modified network after network translation. For each example, 10^5 simulations were performed using the Gillespie algorithm. See Fig. 3 for the values of parameters.

SUPPLEMENTARY NOTE 5: STATIONARY DISTRIBUTIONS OF THE GENETIC TOGGLE SWITCH WITH SLOW PROMOTER KINETICS (FIG. 4)

To obtain the probabilities of the gene G^A is active (ρ_{act}) and repressed (ρ_{rep}), we can construct the reduced-order Markov chain infinitesimal generator Λ_r:

$$[\Lambda_r]_{ij} = \left\{ \begin{array}{ll} \varepsilon^{-1}\mathbb{E}\left[\lambda_{ij}|\text{gene state } j\right] & \text{for } i \neq j \\ -\varepsilon^{-1}\sum_{k\neq j}\lambda_{kj} & \text{for } i = j \end{array} \right\},$$

where λ_{ij} is the propensity of the reaction that changes the gene state from j to i, and ε is the ratio between the fast and slow timescale. The gene state $1, 2, 3$, and 4 represent (G^A_{act}, G^B_{act}), (G^A_{act}, G^B_{rep}), (G^A_{rep}, G^B_{act}), and (G^A_{rep}, G^B_{rep}), respectively, where G^A_{act} represents $n_{G^A_{act}} = 1$ and $n_{G^A_{rep}} = 0$, and G^A_{rep} represents $n_{G^A_{act}} = 0$ and $n_{G^A_{rep}} = 1$, and it is defined in the same manner for the gene G^B. Then λ_{21} is the propensity of the reaction $A_P + G^B_{act} \to G^B_{rep}$, which changes the gene state from (G^A_{act}, G^B_{act}) to (G^A_{act}, G^B_{rep}). Thus, $[\Lambda_r]_{21} = \varepsilon^{-1}\mathbb{E}\left[\lambda_{21}|(G^A_{act}, G^B_{act})\right] = l_b\mathbb{E}\left[n_{A_P}|(G^A_{act}, G^B_{act})\right]$ because $\lambda_{21} = \varepsilon \cdot l_b n_{A_P} n_{G^B_{act}}$, and $n_{G^B_{act}} = 1$. Since the conditional moment of n_{A_P} is solely determined by the state of the gene G^A, $[\Lambda_r]_{21} = l_b\mathbb{E}\left[n_{A_P}|G^A_{act}\right]$. Repeating the similar computations, we get the infinitesimal generator

$$\Lambda_r = \begin{bmatrix} -l_b\mathbb{E}\left[n_{A_P}|G^A_{act}\right] - k_b\mathbb{E}\left[n_{B_P}|G^B_{act}\right] & l_u & k_u & 0 \\ l_b\mathbb{E}\left[n_{A_P}|G^A_{act}\right] & -l_u - k_b\mathbb{E}\left[n_{B_P}|G^B_{rep}\right] & 0 & k_u \\ k_b\mathbb{E}\left[n_{B_P}|G^B_{act}\right] & 0 & -k_u - l_b\mathbb{E}\left[n_{A_P}|G^A_{rep}\right] & l_u \\ 0 & k_b\mathbb{E}\left[n_{B_P}|G^B_{rep}\right] & l_b\mathbb{E}\left[n_{A_P}|G^A_{rep}\right] & -k_u - l_u \end{bmatrix}.$$

Note that Λ_r is Metzler i.e., all the entries are non-negative except for those on the diagonal, and $\mathbf{1}^T\Lambda_r = 0$. Hence, the Perron-Frobenius theorem [6] implies that there exists a positive eigenvector $\rho = (\rho_{(act,act)}, \rho_{(act,rep)}, \rho_{(rep,act)}, \rho_{(rep,rep)})$ corresponding to the eigenvalue 0. Furthermore, we can uniquely determine the eigenvector ρ using the normalization: $\rho_{(act,act)} + \rho_{(act,rep)} + \rho_{(rep,act)} + \rho_{(rep,rep)} = 1$. Here, $\rho_{(act(,rep),act(,rep))}$ is the probability that the gene state is $(G^A_{act(,rep)}, G^B_{act(,rep)})$ [1]. From this eigenvector, we can derive the approximation of the full (i.e., unconditional) mariginal stationary distribution of fast variable A_P :

$$\begin{aligned} \pi(n_{A_P}) \approx & \rho_{(act,act)}\pi(n_{A_P}|G^A_{act}, G^B_{act}) + \rho_{(act,rep)}\pi(n_{A_P}|G^A_{act}, G^B_{rep}) \\ & + \rho_{(rep,act)}\pi(n_{A_P}|G^A_{rep}, G^B_{act}) + \rho_{(rep,rep)}\pi(n_{A_P}|G^A_{rep}, G^B_{rep}), \end{aligned} \tag{S17}$$

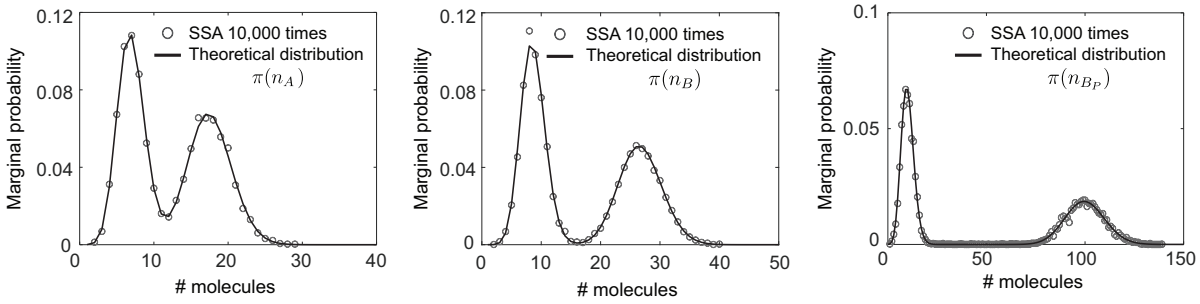
where the approximation becomes more accurate as the timescale separation is larger and becomes exact when $\varepsilon \to 0$. Since the conditional distribution of the A_P solely depends on the state of the gene G^A, it can be reduced as follows:

$$\pi(n_{A_P}) \approx (\rho_{(act,act)} + \rho_{(act,rep)})\pi(n_{A_P}|G^A_{act}) + (\rho_{(rep,act)} + \rho_{(rep,rep)})\pi(n_{A_P}|G^A_{rep}).$$

Finally, the probabilities that the gene G^A is active (ρ_{act}) and repressed (ρ_{rep}) are given by

$$\rho_{act} = \rho_{(act,act)} + \rho_{(act,rep)} \quad \text{and} \quad \rho_{rep} = \rho_{(rep,act)} + \rho_{(rep,rep)}.$$

The full marginal stationary distributions of all other fast variables: A, B, B_P, can also be obtained by computing the linear combination of the conditional stationary distributions of each fast variable, similar to Eq. (S17) (Supplementary Fig. 7). More details on the derivation of the reduced-order Markov chain are discussed in [1].



Supplementary Figure 7. Full marginal stationary distributions formulae (solid lines) and the simulation results obtained with 10^4 times Gillespie algorithms (dots) for A, B, B_P in Fig. 4. See Fig. 4 for the values of parameters.

SUPPLEMENTARY REFERENCES

[1] M. Ali Al-Radhawi, Domitilla Del Vecchio, and Eduardo D. Sontag. Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS Computational Biology*, 15(2):1–27, 2019.

[2] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72(8):1947–1970, 2010.

[3] Henri Cohen. A course in computational algebraic number theory. *Graduate texts in Math.*, 138:88, 1993.

[4] Martin Feinberg. Complex balancing in general kinetic systems. *Archive for Rational Mechanics and Analysis*, 49(3):187–194, 1972.

[5] Ankit Gupta and Mustafa Khammash. Determining the long-term behavior of cell populations: A new procedure for detecting ergodicity in large stochastic reaction networks. *IFAC Proceedings Volumes*, 47(3):1711–1716, 2014. 19th IFAC World Congress.

[6] D. Hinrichsen and N. K. Son. μ-analysis and robust stability of positive linear systems. *Applied Mathematics and Computer Science*, 8:253–268, 1998.

[7] Fritz Horn. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, 49(3):172–186, 1972.

[8] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47(2):81–116, 1972.

[9] F. Kelly. *Reversibility and Stochastic Networks*. Wiley, New York, 1979.

[10] J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[11] Eduardo D Sontag. Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction. *IEEE transactions on automatic control*, 46(7):1028–1047, 2001.

[12] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2):146–160, 1972.

Figure 4e Reduced		Figure 4e Full mode		Figure 4d		Figure 3c			
Resident t	Probabilit	Resident t	Probabilit	Number c	Probability	Number c	Probabilit	Number c	Probabilit
300	0.315059	300	0.310996	1	0.0022	31	4.00E-05	12	2.00E-05
900	0.153582	900	0.150808	2	0.0065	32	7.00E-05	13	1.00E-05
1500	0.089841	1500	0.090143	3	0.0192	33	5.00E-05	14	0.0001
2100	0.062847	2100	0.062829	4	0.0364	34	0.00018	15	0.00021
2700	0.04856	2700	0.052077	5	0.0518	35	0.00044	16	0.00039
3300	0.042816	3300	0.042497	6	0.0697	36	0.00073	17	0.00069
3900	0.036158	3900	0.036508	7	0.0762	37	0.00134	18	0.00141
4500	0.031126	4500	0.031463	8	0.0725	38	0.00275	19	0.00245
5100	0.026961	5100	0.027452	9	0.0566	39	0.00468	20	0.00349
5700	0.022998	5700	0.024226	10	0.0434	40	0.00732	21	0.00664
6300	0.021139	6300	0.020794	11	0.0269	41	0.01163	22	0.00994
6900	0.01827	6900	0.018464	12	0.0175	42	0.01733	23	0.01342
7500	0.016295	7500	0.016252	13	0.0112	43	0.02381	24	0.01878
8100	0.013911	8100	0.014246	14	0.0067	44	0.03209	25	0.02522
8700	0.01204	8700	0.013006	15	0.0028	45	0.04272	26	0.03278
9300	0.011178	9300	0.011055	16	0.0011	46	0.05168	27	0.04026
9900	0.009436	9900	0.009408	17	0.0002	47	0.061	28	0.04779
10 500	0.008322	10 500	0.008257	18	0.0003	48	0.06866	29	0.05526
11 100	0.007228	11 100	0.007478	19	0.0003	49	0.07637	30	0.06152
11 700	0.006496	11 700	0.006334	20	0.0001	50	0.07804	31	0.06663
12 300	0.005388	12 300	0.005817	21	0	51	0.07873	32	0.0691
12 900	0.004818	12 900	0.004797	22	0.0003	52	0.07745	33	0.06862
13 500	0.00476	13 500	0.004342	23	0.0001	53	0.07017	34	0.06767
14 100	0.004054	14 100	0.003935	24	0.0001	54	0.06298	35	0.06357
14 700	0.003186	14 700	0.00335	25	0.0001	55	0.05482	36	0.05973
15 300	0.003044	15 300	0.00315	26	0.0003	56	0.04491	37	0.05398
15 900	0.002604	15 900	0.002578	27	0.0001	57	0.03692	38	0.04762
16 500	0.002312	16 500	0.002426	28	0.0001	58	0.02809	39	0.04094
17 100	0.00206	17 100	0.001971	29	0	59	0.02035	40	0.03347
17 700	0.001658	17 700	0.001578	30	0.0005	60	0.01527	41	0.02765
18 300	0.001658	18 300	0.001378	31	0.0006	61	0.01095	42	0.02061
18 900	0.001386	18 900	0.001254	32	0.0015	62	0.00686	43	0.01621
19 500	0.001069	19 500	0.001206	33	0.0015	63	0.00432	44	0.0131
20 100	0.000984	20 100	0.000944	34	0.0022	64	0.00304	45	0.00957
20 700	0.000797	20 700	0.000848	35	0.0027	65	0.00183	46	0.00648
21 300	0.000764	21 300	0.000765	36	0.0036	66	0.00099	47	0.00497
21 900	0.000706	21 900	0.000669	37	0.0041	67	0.0006	48	0.00327
22 500	0.000427	22 500	0.000558	38	0.0077	68	0.00033	49	0.00239
23 100	0.000583	23 100	0.000503	39	0.0085	69	0.00018	50	0.00142
23 700	0.000324	23 700	0.000524	40	0.0117	70	0.00014	51	0.00101

24 300	0.000447	24 300	0.00042	41	0.0136	71 0.0001	52 0.0006
24 900	0.000369	24 900	0.000372	42	0.0149	72 3.00E-05	53 0.00034
25 500	0.000369	25 500	0.000358	43	0.0194	73 1.00E-05	54 0.00031
26 100	0.00022	26 100	0.000255	44	0.0194		55 6.00E-05
26 700	0.000181	26 700	0.000165	45	0.0224		56 0.00019
27 300	0.00022	27 300	0.000165	46	0.0233		57 7.00E-05
27 900	0.000181	27 900	0.000145	47	0.0266		58 1.00E-05
28 500	0.000149	28 500	0.000124	48	0.0265		59 2.00E-05
29 100	0.000117	29 100	0.000207	49	0.0284		60 1.00E-05
29 700	9.71E-05	29 700	5.51E-05	50	0.031		
30 300	0.00011	30 300	0.000124	51	0.0249		
30 900	5.18E-05	30 900	0.000145	52	0.026		
31 500	7.12E-05	31 500	4.14E-05	53	0.0262		
32 100	8.42E-05	32 100	6.89E-05	54	0.0219		
32 700	7.12E-05	32 700	4.82E-05	55	0.0212		
33 300	5.83E-05	33 300	4.82E-05	56	0.0207		
33 900	4.53E-05	33 900	4.14E-05	57	0.0176		
34 500	4.53E-05	34 500	6.20E-05	58	0.0133		
35 100	3.24E-05	35 100	4.82E-05	59	0.011		
35 700	1.94E-05	35 700	1.38E-05	60	0.0091		
36 300	3.89E-05	36 300	6.89E-06	61	0.0077		
36 900	1.94E-05	36 900	2.07E-05	62	0.007		
37 500	2.59E-05	37 500	3.45E-05	63	0.0044		
38 100	3.24E-05	38 100	2.76E-05	64	0.0044		
38 700	3.24E-05	38 700	1.38E-05	65	0.0035		
39 300	6.48E-06	39 300	1.38E-05	66	0.0013		
39 900	1.30E-05	39 900	6.89E-06	67	0.0024		
40 500	1.30E-05	40 500	1.38E-05	68	0.002		
41 100	0.00E+00	41 100	1.38E-05	69	0.0015		
41 700	1.30E-05	41 700	6.89E-06	70	0.0003		
42 300	6.48E-06	42 300	6.89E-06	71	0.0001		
42 900	0.00E+00	42 900	6.89E-06	72	0.0003		
43 500	6.48E-06	43 500	6.89E-06	73	0.0002		
44 100	6.48E-06	44 100	1.38E-05	74	0		
44 700	0.00E+00	44 700	1.38E-05	75	0		
				76	0		
				77	0.0001		
				78	0		
				79	0		
				80	0.0001		

Figure 3f

y	Number c	Probabilit	Number c	Probabilit	Number c	Probability
	7	1.00E-05	38	2.00E-05	0	0.00323
	8	0	39	0	1	0.01926
	9	8.00E-05	40	4.00E-05	2	0.05706
	10	0.00036	41	9.00E-05	3	0.11194
	11	0.0012	42	0.00021	4	0.15923
	12	0.00416	43	0.00029	5	0.17586
	13	0.01056	44	0.00075	6	0.16314
	14	0.02318	45	0.0021	7	0.12803
	15	0.04299	46	0.00382	8	0.0861
	16	0.06984	47	0.00746	9	0.04957
	17	0.10159	48	0.01475	10	0.02617
	18	0.12419	49	0.024	11	0.01225
	19	0.13987	50	0.03828	12	0.00506
	20	0.13266	51	0.05686	13	0.00216
	21	0.11643	52	0.07597	14	0.00065
	22	0.09123	53	0.09303	15	0.00023
	23	0.06202	54	0.10968	16	5.00E-05
	24	0.03847	55	0.11564	17	1.00E-05
	25	0.02123	56	0.11443		
	26	0.0115	57	0.0992		
	27	0.00485	58	0.08478		
	28	0.00219	59	0.06277		
	29	0.00086	60	0.04321		
	30	0.00034	61	0.02543		
	31	0.00016	62	0.01486		
	32	1.00E-05	63	0.00721		
	33	2.00E-05	64	0.00301		
			65	0.00153		
			66	0.00041		
			67	0.00012		
			68	4.00E-05		
			69	1.00E-05		

Figure 3i

Number c	Probabilit	Number c
0	0.00625	42
1	0.03144	43
2	0.07961	44
3	0.13622	45
4	0.17285	46
5	0.17163	47
6	0.14888	48
7	0.10781	49
8	0.06866	50
9	0.04104	51
10	0.02015	52
11	0.00923	53
12	0.00388	54
13	0.00155	55
14	0.00051	56
15	0.00022	57
16	4.00E-05	58
17	2.00E-05	59
18	1.00E-05	60

Probability
1.00E-05
2.00E-05
4.00E-05
0.00022
0.00051
0.00155
0.00388
0.00923
0.02015
0.04104
0.06866
0.10781
0.14888
0.17163
0.17285
0.13622
0.07961
0.03144
0.00625