Towards Modeling in Large-Scale Genetic and Metabolic Networks.

Marco Polo Castillo-Villalba

Abstract. Mathematical modeling of large-scale genetic and metabolic networks remains an open area of research with profound implications for understanding biochemical signaling systems in both prokaryotic and eukaryotic organisms. In this work, we extend the biochemical modeling framework based on the **Power Law formalism**, originally developed by M.A. Savageau to elucidate design principles in small bacterial networks. This formalism captures gene regulation and metabolic fluxes through two fundamental sets of parameters: *kinetic parameters* and *kinetic orders*. The kinetic parameters determine the magnitudes of fluxes, while the kinetic orders appearing as exponents in the power-law representation encode the regulatory structure of the system. We propose a mathematical technique to characterize the entire family of systems associated with the space of kinetic orders. We show that this exponent space possesses a well-defined structure in **toric geometry** and can therefore be modeled using **algebraic varieties**. To represent large-scale networks, we associate a toric variety with each polynomial equation of the system and then use geometric techniques to coherently combine these varieties into a global structure, or **fan**. This construction enables a modular and geometrically consistent representation of complex biochemical networks.

1. Introduction

2. Theoretical Background: Biochemical Networks as Toric S-systems

One of the main ideas behind the analysis of biochemical systems is the conservation of mass in reactions. In addition, the Michaelis–Menten rate law is one of the fundamental concepts in biochemical systems. However, in recent decades, this law has been refined to provide more detailed biochemical information about the biomolecules participating in the network. For example, it is now possible to consider reactions with arbitrary orders, such as the number of binding sites in gene operator regions or activation parameters from Hill functions.

This is the basis of the Power Law formalism [**11**], which, through parametric extensions, allows for the simultaneous modeling of both metabolism and gene regulation. The formal expression of this model is given below:

2010 *Mathematics Subject Classification.* Primary .

Key words and phrases. Topological Environment, S-systems, toric geometry, quantitative phenotype, kinetic orders, tryptophan gene-metabolic network.

$$\dot{X}_i = \sum_{k=1}^{p_i} \alpha_{ik} \prod_{j=1}^{m} X_j^{g_{ijk}} - \sum_{k=1}^{q_i} \beta_{ik} \prod_{j=1}^{m} X_j^{h_{ijk}}, \quad i = 1, \ldots, n. \tag{2.1}$$

Here, the kinetic parameters α_{ik} and β_{ik} capture the rate constants for production and degradation, respectively, of each species or biomolecule in the network. The kinetic orders g_{ijk} and h_{ijk} represent genotypic information, such as binding sites and the kinetic orders of catalyzed reactions or multimeric regulation sub units in protein structures. Therefore, Equation (2.1) captures all the information required to analyze genetic–metabolic networks.

Definition 2.1. (**S-systems**). We define the number of combinations of dominant terms in Equation (2.1) as:

$$\prod_{j=1}^{m} p_j q_j,$$

which partitions the space of concentrations $(X_1, \ldots, X_n)$. Each partition corresponds to a dominant term, or **S-system**, as defined in [**13**], [**11**], [**7**], [**8**], [**9**]. This is expressed as follows:

$$\dot{X}_i = \alpha_{ip_i} \prod_{j=1}^{m} X_j^{g_{ip_i}} - \beta_{iq_i} \prod_{j=1}^{m} X_j^{h_{iq_i}}, \quad i = 1, \ldots, n. \tag{2.2}$$

In the context of this formalism, an S-system defines a quantitative phenotype and corresponds to a biological reality in molecular biology. These systems are not merely mathematical artifacts; they represent rates of change in molecular concentrations derived from this formal gradient.

2.1. Toric S-systems. In this section, we highlight the importance of studying the exponents g_{ip_i} and h_{iq_i} in an S-system. We assert that the monomial map $g_{ip_i} \longmapsto X_j^{g_{ip_i}}$ represents a **Genotype–Phenotype** relationship with properties of great interest for further study. This central idea forms the foundation of our research. Below, we briefly summarize the definitions introduced in well-known works such as [**7**], [**13**], [**11**], and [**7**].

Definition 2.2. **(Exponent Space or Support)** For an S-system defined by $\dot{X}_i$, we define the *exponent space* of $\dot{X}_i$, also formally known as the *support* of a polynomial, as:

$$\operatorname{supp}(\dot{X}_i) = \{(g_{i,p_1}, \ldots, g_{i,p_n}), (h_{i,q_1}, \ldots, h_{i,q_n})\}. \tag{2.3}$$

Using the support of any S-system $\operatorname{supp}(\dot{X}_i)$, we can construct all linear combinations from the two vectors in the support, $(g_{1,p_1}, \ldots, g_{n,p_n})$ and $(h_{1,q_1}, \ldots, h_{n,q_n})$. The set of all such vector combinations defines the **cone** associated with $\dot{X}_i$, as described in [**4**].

Definition 2.3. **(Cone)** Let $\dot{X}_i$ be an S-system as defined above. We define the cone associated with it as:

$$\sigma = \operatorname{Con}\left(\operatorname{supp}(\dot{X}_i)\right). \tag{2.4}$$

This cone represents the linear space generated by all combinations of the vectors in the support space, including addition, subtraction, and scalar multiplication. If the vectors consist only of integer entries, the cone is called a *lattice cone* [**4**].

Definition 2.4. **(Dual Cone)** Let $\sigma \subseteq \mathbb{Z}^n$ be a lattice cone. The dual cone $\sigma^\vee$ is defined as:

$$\sigma^\vee = \{v \in \mathbb{Z}^n \ : \ \langle v, \sigma \rangle \geq 0\}, \tag{2.5}$$

as described in [**4**].

Definition 2.5. **(Monoid)** By intersecting the dual cone with the n-dimensional integer lattice, we obtain:

$$\sigma^\vee \cap \mathbb{Z}^n,$$

which we call a **monoid**. This monoid, denoted by M, has rank n and is isomorphic to $\mathbb{Z}^n$ [**4**].

Lemma (Gordan's Lemma and the Hilbert Basis). Given a cone σ, the intersection of its dual with the n-dimensional integer space,

$$\sigma^\vee \cap \mathbb{Z}^n,$$

forms a monoid that has a finite set of generators. This generating set is called the **Hilbert basis** [**17**], [**4**].

Definition 2.6. **(Fan or topological environment)** Given a collection of dual cones

$$\{\sigma_1^\vee, \ldots, \sigma_n^\vee\} \subseteq \mathbb{Z}^n,$$

we define the *fan* $P = \bigcup_{i=1}^n \sigma_i^\vee$ as the union of all these cones. We denote this fan by $\Sigma \subseteq \mathbb{Z}^n$ [**4**].

Finally, we construct a coordinate polynomial ring for S-systems. Let $f = \dot{X}_i$ be an S-system defined for any i-th species in a metabolic network. Given the support of this polynomial $\text{supp}(f)$ and its associated cone σ, we define the polynomial ring:

$$R_\sigma = \{f \mid \text{supp}(f) \subseteq \sigma\}.$$

In other words, the exponents of an S-system (i.e., the kinetic orders) are embedded in a geometric space. We will explore new topological properties that arise from this geometric object in the following sections.

Definition 2.7. **(Toric Ideal)**. A toric ideal is the set of prime binomials defined by $I =< X^A - X^B | H * (A - B) = 0 >$, where H is the *Hilbert basis* of the polynomials generated from I, [**17**], [**3**].

Definition 2.8. **(Toric Variety)**. We define a **toric variety** associated with a polynomial f as follows:

$$X_\sigma = \text{Spec}\,(R_\sigma) = \{f_1, \ldots, f_k \in R_\sigma\}. \tag{2.6}$$

Here, the notation $\text{Spec}\,(R_\sigma)$ denotes the set of all prime polynomials (i.e., non-divisible polynomials) $\{f_1, \ldots, f_k\}$ in the ring R_σ for some integer k, see [**4**].

Definition 2.9. **(Environment Cone)**. Let f be an S-system representing the rate law for the change in concentration of any species in a gene–metabolic network. We compute the support of f and the cone associated with it, $\sigma \subseteq \mathbb{Z}^n$. Since $\mathrm{supp}(f) \subseteq \sigma$, we define the **environment cone** associated with f as the dual cone $\sigma^\vee$ of σ. Thus, we have the inclusion relation:

$$\mathrm{supp}(f) \subseteq \sigma \subseteq \sigma^\vee.$$

Relevance in Enzyme Kinetics. Up to this point, we have equipped the set of vectors associated with the kinetic orders of an S-system with new geometric properties. We have defined cones, polynomial rings, and toric varieties. The most important result arises from **Gordan's Lemma** [**4**], which guaranties the existence of a generator basis (the **Hilbert basis**) for the entire kinetic order space defined for any species in the network. Next, we will explore how to exploit the properties induced by this Hilbert basis.

Toric S-systems and Dominant Fluxes.

Given the Hilbert basis $H = \{h_1, \ldots, h_k\}$, we represent it in matrix form considering each h_i as a row vector:

$$h_1 = (u_{11}, u_{12}, \ldots, u_{1n}), \quad h_2 = (u_{21}, u_{22}, \ldots, u_{2n}), \quad \ldots, \quad h_k = (u_{k1}, u_{k2}, \ldots, u_{kn}).$$

$$H = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & u_{k2} & \cdots & u_{kn} \end{pmatrix}$$

We can use this matrix to represent the variables X_i in a new coordinate system using the following monomial transformation:

$$u_{ij} \longmapsto w_i^{u_{ij}},$$

as discussed in [**4**], [**17**], and [**3**]. This transformation reveals dynamical properties encoded within the S-systems through the Hilbert basis [**17**], [**5**].

$$\begin{array}{rcl} X_1 & = & w_1^{u_{11}} w_2^{u_{12}} \cdots w_n^{u_{1n}}, \\ X_2 & = & w_1^{u_{21}} w_2^{u_{22}} \cdots w_n^{u_{2n}}, \\ \vdots & & \vdots \\ X_k & = & w_1^{u_{k1}} w_2^{u_{k2}} \cdots w_n^{u_{kn}}. \end{array} \tag{2.7}$$

This matrix representation allows embedding all S-system equations into a new system of monomial coordinates. Hence, we extend the analysis of molecular networks toward large-scale genetic and metabolic systems.

Theorem 2.10. ***(Toric S-system)***. *If we substitute the above change of coordinates into Equation (2.2), the original S-system equations, we obtain the following expression, valid for all $i = 1, \ldots, n$:*

Proof. *Given an S-system representation of molecular networks as in Equation (2.2), and applying Hironaka's theorem together with the algorithmic construction*

of the Hilbert basis (a finite number of times) to the S-system equations [**1**], [**6**], [**5**], *we find that every S-system is toric and can be described by:*

(2.8)

$$\dot{X}_i(w_1,\ldots,w_k) = \alpha_{ip_i}\prod_{j=1}^{m}\left(w_j^{u_{j1}}w_2^{u_{j2}}\cdots w_n^{u_{jn}}\right)^{g_{ip_i}} - \beta_{iq_i}\prod_{j=1}^{m}\left(w_j^{u_{j1}}w_2^{u_{j2}}\cdots w_n^{u_{jn}}\right)^{h_{iq_i}}.$$

Or, more compactly:

$$\dot{X}_i(w_1,\ldots,w_k) = \alpha_{ip_i}\prod_{j=1}^{m}(W_i^{h_j})^{g_{ip_i}} - \beta_{iq_i}\prod_{j=1}^{m}(W_i^{h_j})^{h_{iq_i}}. \tag{2.9}$$

We thus define a ***Toric S-system*** *according to the expression above, where h_j is an element of the Hilbert basis for some j. We assert that the binomials constructed from the Hilbert basis H,*

$$I_H = \langle W^\alpha - W^\beta\rangle,$$

with $W_i^{h_j g_{ip_i}} = W^\alpha$ and $W_i^{h_j g_{iq_i}} = W^\beta$, are ***toric binomials****.* ***q.e.d.***

The monomial transformation,

$$u_{ij} \longmapsto w_i^{u_{ij}}$$

represents, in this geometric framework, the **Genotype–Phenotype relationship**. These are the coordinate transformations of the geometric basis corresponding to the kinetic orders such as the number of binding sites in regulatory genomic sequences, the order of biochemical reactions in metabolism, or the degree of multimeric regulation in proteins.

The monomials $w_i^{u_{ij}}$ represent the chemical concentrations of the species or biomolecules described by the S-system X_i. The coordinates u_{ij} encode information derived from the genome sequence, which is then expressed through the phenotypic gradient:

$$\dot{X}_j(u_{ij}) = w_i^{u_{ij}}.$$

Thus, the mapping of dynamical behavior occurs through this gradient, linking genotypic parameters to phenotypic observables.

3. Discussion

We have presented new topological properties of biochemical networks by exploiting the structural features of the Power Law and S-system modeling frameworks. In particular, we emphasize the significance of the monomial map: $g_{ij} \longmapsto X_j^{g_{ij}}$, which provides a bridge between biochemical kinetics and algebraic geometry. By focusing our analysis on the exponent space of S-systems, we study the kinetic orders using computational algebraic geometry techniques. This allows us to identify geometric bases that constrain and characterize the possible values of these parameters, thereby facilitating the solution of large systems of differential equations.

The central open questions that can be addressed with this approach are as follows:

- Under what conditions are these geometrical regions valid for assigning values to the kinetic orders?

- Which of these regions represent true biological information and which are merely mathematical artifacts?

Another advantage of the proposed methodology is its potential to model the dynamics of large-scale metabolic and gene-regulatory networks. The monomial map has a formal justification through the mathematical concept of a **toric variety**. Although toric varieties are abstract algebraic objects, they serve as powerful tools for modular analysis: each variety corresponds to a submodule of the biochemical network, and the full network can be reconstructed by *gluing together* these varieties using geometric techniques.

While a single toric variety may provide limited insight into the overall biochemical system, the ensemble of all varieties properly joined captures the global dynamics of large-scale genetic and metabolic networks. This modular, geometrically grounded framework thus offers a new perspective for understanding and simulating the complex interplay between genotype and phenotype.

Finally, we have included in the Supplementary Information the complete set of equations used for the toric analysis of a gene–metabolic network in the allosteric system of the tetrameric protein TrpR (tryptophan repressor) in E. coli K-12. In a future work, we will solve the complete dynamics using toric varieties such as biochemical modules.

4. Acknowledgment

Marco Polo Castillo Villalba is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received the CONAHCYT fellowship 754050. We acknowledge funding from UNAM and from the National Institutes of Health (grant number 5R01GM110597-03). I give a special thanks to Michael A. Savageau for all technical comments during my scholar stay with him, Pedro Miramontes Vidal and Julio Collado-Vides and technical contribution by Laura Gómez Romero.

Declaration of Interest Statement

The author declares that he has no personal, financial, professional, or academic conflicts of interest.

References

[1] M.F. Atiyah. *Resolution of Singularities and Division of Distributions*, Communications On Pure and Applied Mathematics, vol. XXIII, 145-150 (1970).

[2] D. Cox, J. Little and D. O'Shea. *Using Algebraic Geometry.* Springer, 2nd. Edition, August, 2004.

[3] D. Cox, J. Little and H. Schenck. *Toric Varieties* Department Of Mathematics, Amherst College, Amherst MA 01002, 2010.

[4] G. Ewald. *Combinatorial Convexity and Algebraic Geometry.*, Springer-Verlag, New York, Inc., 1996.

[5] Henk M, Weismantel R. On Hilbert bases of polyhedral cones. *Konrad-Zuse-Zentrum für Informationstechnik; 2010.*

[6] T. Fukui. Introduction to Toric Modifications With an Application to Real Singularities. *Department Of Mathematics, Faculty Of Science, Saitama University.* 255 Shimo-Okubo, Urawa 338-8570, Japan, pag. 96-114, 2000.

[7] J. G. Lomnitz, M. A. Savageau. Phenotypic deconstruction of gene circuitry *Chaos.* Vol. 23, Issue 2, May 21, 2013.

[8] J. G. Lomnitz, M. A. Savageau. Strategy Revealing Phenotypic Differences among Synthetic Oscillator Designs. *ACS Synth. Biol.* Volume 3, 686701, July 14, 2014.

[9] J. G. Lomnitz, M. A. Savageau. Elucidating the genotype-phenotype map by automatic enumeration and analysis of the phenotypic repertoire. *npj Systems Biology and Applications.* Vol 1, 15003; doi:10.1038/npjsba.2015.3; published online 28 September 2015.

[10] J. G. Lomnitz, M. A. Savageau. Design Space Toolbox V2: Automated Software Enabling a Novel Phenotype-Centric Modeling Strategy for Natural and Synthetic Biological Systems. *Frontiers in Genetics.* Vol 27, article 117, 12 July, 2015.

[11] Biochemical Systems Analysis: "A Study of Function and Design in Molecular Biology". *Biochemical Systems Analysis.*, volume 1. Addison-Wesley Inc , Reading, Mass, USA, 1976.

[12] Savageau MA. Optimal design of feedback control by inhibition. Dynamic considerations. *J Mol Evol.* 1975;5:199–222.

[13] M. A. Savageau and E. O. Voit. Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear Form. *Mathematical Biosciences.*, Volume 87, Issue 1, pages 83-115, November 1987.

[14] M. A. Savageau, H. O. Voit and D. H. Irvine. Biochemical Systems Theory and Metabolic Control Theory: 1. Fundamental Similarities and Differences. *Mathematical Biosciences.*, Volume 86, Issue(2), pages 127-145, October 1987.

[15] M. A. Savageau. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos.*, Volume 11, Issue 1, December (2001).

[16] M. A. Savageau, P. M. B. M. Coelho, R. A. Fasani and A. Salvador. Phenotypes and tolerances in the design space of biochemical systems. *PNAS.* Vol. 106, no. 16, 6435– 6440, April 21, 2009.

[17] B. Sturmfels. Grobner Bases and Convex Polytopes. *American Mathematical Society.* Providence, RI, 1996.

[18] Voit. E. O. Non Linear Canonical Modeling. S-system Approach To Understanding Complexity. New York. New Nonstrand Reinold, (1991).

Support information

Department of Molecular Biology and Biotechnology, Institute of Biomedical Research UNAM, Circuito Universitario 04510 CDMX, Mexico

Email address: mpolocv@gmail.com

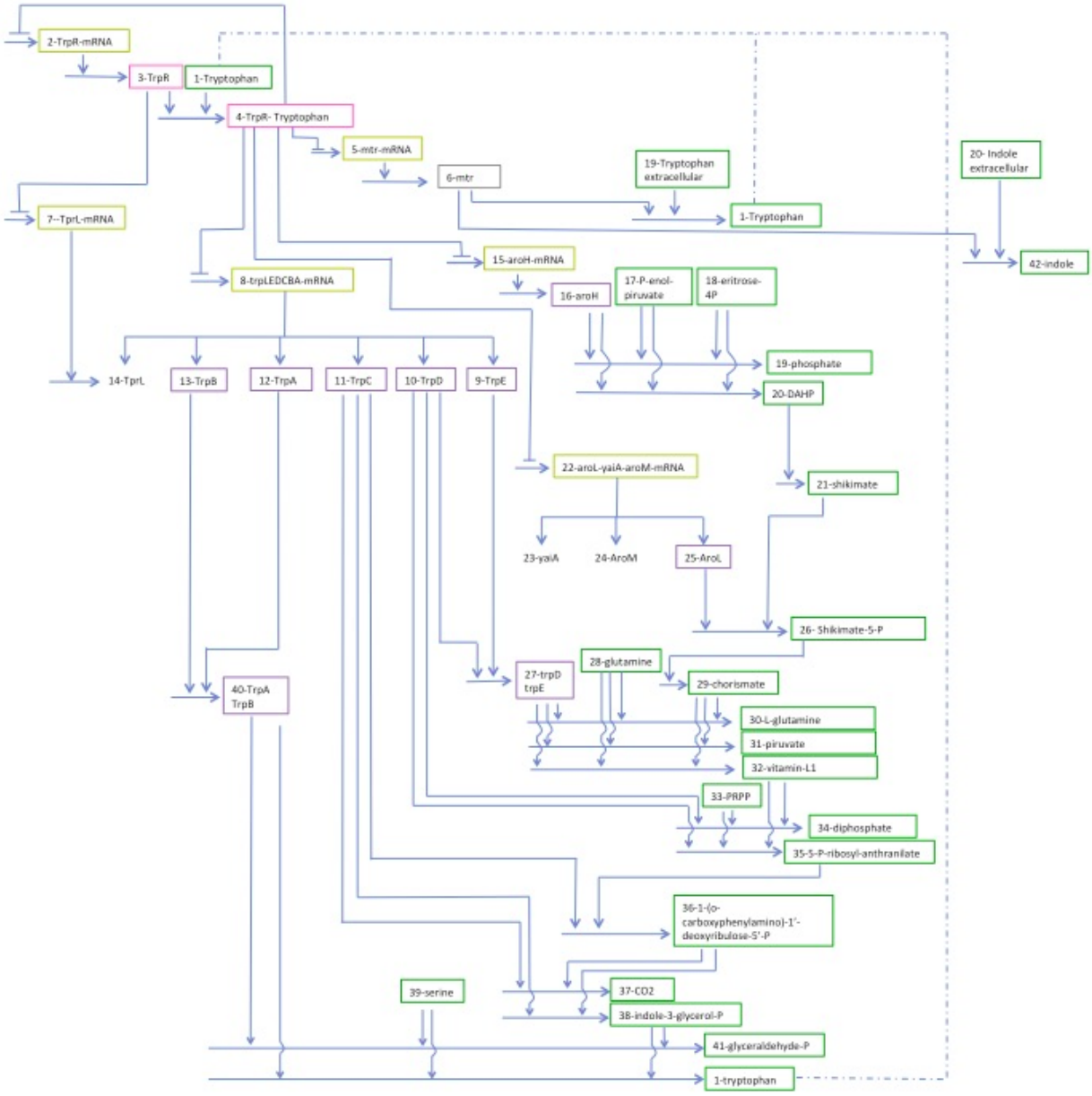

FIGURE 1. Tryptophan genetic and metabolic network

$$\dot{X}_{21} = \alpha_{20} X_{20}^{g_{21,20}} - \beta_{21} X_{21}^{h_{21,21}};$$
$$\dot{X}_{22} = \alpha_{22B} - \beta_{22} X_{22}^{h_{22,22}}, \text{ if } X_4 < X_{4H};$$
$$\dot{X}_{23} = \alpha_{22} X_{22}^{g_{23,22}} - \beta_{23} X_{23}^{h_{23,23}};$$
$$\dot{X}_{24} = \alpha_{22} X_{22}^{g_{24,22}} - \beta_{24} X_{24}^{h_{24,24}};$$
$$\dot{X}_{25} = \alpha_{22} X_{22}^{g_{25,22}} - \beta_{25} X_{25}^{h_{25,25}};$$
$$\dot{X}_{26} = \alpha_{21,25} X_{21}^{g_{26,21}} X_{25}^{g_{26,25}} - \beta_{26} X_{26}^{h_{26,26}};$$
$$\dot{X}_{27} = \alpha_{9,10} X_{9}^{g_{27,9}} X_{10}^{g_{27,10}} - \beta_{27} X_{27}^{h_{27,27}};$$
$$\dot{X}_{28} = \alpha_{28} - \beta_{27,28,29} X_{27}^{h_{28,27}} X_{28}^{h_{28,28}} X_{29}^{h_{28,29}};$$
$$\dot{X}_{29} = \alpha_{26} X_{29}^{g_{29,26}} - \beta_{27,28,29} X_{27}^{h_{29,27}} X_{28}^{h_{29,28}} X_{29}^{h_{29,29}};$$
$$\dot{X}_{30} = \alpha_{27,28,29} X_{27}^{g_{30,27}} X_{28}^{g_{30,28}} X_{29}^{g_{30,29}} - \beta_{30} X_{30}^{h_{30,30}};$$
$$\dot{X}_{31} = \alpha_{27,28,29} X_{27}^{g_{31,27}} X_{28}^{g_{31,28}} X_{29}^{g_{31,29}} - \beta_{31} X_{19}^{h_{31,31}};$$
$$\dot{X}_{32} = \alpha_{27,28,29} X_{27}^{g_{32,27}} X_{28}^{g_{32,28}} X_{29}^{g_{32,29}} -$$
$$\beta_{10,32,33} X_{10}^{h_{32,10}} X_{32}^{h_{32,32}} X_{33}^{h_{32,33}};$$
$$\dot{X}_{33} = \alpha_{33} X_{33} - \beta_{10,32,33} X_{10}^{h_{33,10}} X_{32}^{h_{33,32}} X_{33}^{h_{33,33}};$$
$$\dot{X}_{34} = \alpha_{10,32,33} X_{32}^{g_{34,32}} X_{33}^{g_{34,33}} X_{10}^{g_{34,10}} - \beta_{34} X_{34}^{h_{34,34}};$$

$$\dot{X}_{33} = \alpha_{33} X_{33} - \beta_{10,32,33} X_{10}^{h_{33,10}} X_{32}^{h_{33,32}} X_{33}^{h_{33,33}};$$
$$\dot{X}_{34} = \alpha_{10,32,33} X_{32}^{g_{34,32}} X_{33}^{g_{34,33}} X_{10}^{g_{34,10}} - \beta_{34} X_{34}^{h_{34,34}};$$
$$\dot{X}_{35} = \alpha_{10,32,33} X_{10}^{g_{35,10}} X_{32}^{g_{35,32}} X_{33}^{g_{35,33}} - \beta_{35,11} X_{11}^{h_{35,11}} X_{35}^{h_{35,35}};$$
$$\dot{X}_{36} = \alpha_{35,11} X_{35}^{g_{36,35}} X_{11}^{g_{36,11}} - \beta_{36,11} X_{11}^{h_{36,11}} X_{36}^{h_{36,36}};$$
$$\dot{X}_{37} = \alpha_{36,11} X_{11}^{g_{37,11}} X_{36}^{g_{37,36}} - \beta_{37} X_{37}^{h_{37,37}};$$
$$\dot{X}_{38} = \alpha_{36,11} X_{36}^{g_{38,36}} X_{11}^{g_{38,11}} - \beta_{38,39,40} X_{38}^{h_{38,38}} X_{39}^{h_{38,39}} X_{40}^{h_{38,40}};$$
$$\dot{X}_{39} = \alpha_{39} - \beta_{38,39,40} X_{38}^{h_{39,38}} X_{39}^{h_{39,39}} X_{40}^{h_{39,40}} \text{ if } X_{39} > X_{39L};$$
$$\dot{X}_{40} = \alpha_{12,13} X_{12}^{g_{40,12}} X_{13}^{g_{40,13}} - \beta_{40} X_{40}^{h_{40,40}};$$
$$\dot{X}_{41} = \alpha_{38,39,40} X_{38}^{g_{41,38}} X_{39}^{g_{41,39}} X_{40}^{g_{41,40}} - \beta_{41} X_{41}^{h_{41,41}};$$

FIGURE 2. Toric S-system equations for tetrameric tryptophan protein.