HYPERDIGRAPH-THEORETIC ANALYSIS OF THE EGFR SIGNALING NETWORK: INITIAL STEPS LEADING TO GTP:RAS COMPLEX FORMATION

JOSEPH S. OLIVEIRA, JANET B. JONES-OLIVEIRA, DAVID A. DIXON, COLIN G. BAILEY, AND DEAN W. GULL

ABSTRACT. We construct an algebraic-combinatorial model of the SOS compartment of the EGFR biochemical network. A Petri net is used to construct an initial representation of the biochemical decision-making network, which in turn defines a hyperdigraph. We observe that the linear algebraic structure of each hyperdigraph admits a canonical set of algebraic-combinatorial invariants that correspond to the information flow conservation laws governing a molecular kinetic reaction network. The linear algebraic structure of the hyperdigraph and its sets of invariants can be generalized to define a discrete algebraic-geometric structure, which is referred to as an oriented matroid. Oriented matroids define a polyhedral optimization geometry that is used to determine optimal subpaths that span the nullspace of a set of kinetic chemical reaction equations. Sets of constrained submodular path optimizations on the hyperdigraph are objectively obtained as a spanning tree of minimum cycle paths. This complete set of subcircuits is used to identify the network pinch points and invariant flow subpaths. We demonstrate that this family of minimal circuits also characteristically identifies additional significant biochemical reaction pattern features. We use the SOS Compartment A of the EGFR biochemical pathway to develop and demonstrate the application of our algebraic-combinatorial mathematical modeling methodology.

1. INTRODUCTION

Living cells can sense their environment and respond to environmental stimuli. Cell signaling governs how information from the environment is decoded, processed, and transferred so that the cell can adequately respond. Cell signaling mechanisms are defined by sets of biochemical pathways that, in turn, physically define sets of chemical networks made up of biochemical circuits. The biochemical pathways that define cell signaling represent complex chemical kinetic phenomena. These pathways are a fundamentally important part of cellular communications and the basic principles of cell signaling seem to be very similar in essentially all organisms.

One important example is signaling through the epidermal growth factor receptor (EGFR), which regulates various cellular mechanisms, such as cell growth, survival, proliferation, and differentiation. EGF receptors play an important role in mammalian development and have also been implicated in tumor formation. EGFR is a member of the receptor tyrosine kinase (RTK) family of receptors. The signaling pathways of various RTKs are reasonably well established and have common underlying features such as phosphorylation of the receptor and its interaction with signal transducing molecules containing the src homology domain or the phosphotyrosine binding domains. For EGFR, the signal from the receptor is transmitted to the nucleus through a series of protein-protein interactions with

Date: August 20, 2003.

Corresponding author: J. B. Jones-Oliveira (jjo@pnl.gov).

the adaptor signaling molecules, through the MAP kinase cascade, and leads to Erk activation, which controls the gene expression. A number of groups have been developing computational approaches to modeling cell pathways, especially that of EGFR. See Bhalla & Iyengar (1999), Resat *et al.* (2003), Wiley & Cunningham (1981), Gex-Fabry & Lisi (1984), Kholodenko *et al.* (1999), Bajzer *et al.* (1989), Lund *et al.* (1990), Sorkin *et al.* (1991), French & Lauffenburger (1997), and Haugh & Lauffenburger (1998). These approaches have used available biochemical data and have focused on the time-dependence of the various biochemical species that are generated and consumed.

Deviant behavior of a cell can be governed by a cell signaling path or network that is out of control, *e.g.*, leading to uncontrolled cell growth or death. Activation of the EGFR network shown in Figure 1 can induce either cell division or differentiation depending on the strength and duration of the signal. A basic understanding of inter- and intra-cellular signaling processes and the biological response to changes in the signaling process is critical to our understanding of how cells and groups of cells, such as microbial communities or mammalian tissues, respond to environmental insults.

Cell signaling, when viewed as an operational systems process model, is defined by the hierarchical control-feedback mechanisms of complex chemical pathways. Computational models serve as a means for integrating knowledge about a pathway and enabling the prediction of cellular behavior. As part of our efforts to better understand cellular processes and how information is being processed by the cell, we have been developing approaches based on advanced graph theories where we do not need to know all of the kinetic information required to completely model the chemical system. Rather, our approach enables us to study how information is passed through the network based on its connectivity. Although we do not get as much quantitative information from our approach, we do obtain significant amounts of qualitative insights into the information processing of the cell. In addition, our approach could be very useful in analyzing the network and connectivity information being generated in high-throughput experiments. See Ideker *et al.* (2001). Our approach to modeling complex cellular networks is to use Petri nets based on graph theory. Qualitative sets of inferences define the process control logic of a Petri net model of biochemical reaction networks. In every biochemical reaction network, there exist sets of reactions that define the conversion of biochemically reacting species. The replacement and depletion of biochemical species is defined by the set of reactions. The transport of sets of reacting species within a given network is directed by time-ordered sets of operational inferences, the process control logic of the systems' reaction network. We want to identify the extremal (maximal and minimal) sets of sub-circuit paths of a given network in terms of their exchange fluxes, subject to the conservation of system fluxes that act as the balance laws for any chemical reaction pathway. Not every sub-circuit will, in fact, be principal for the regulation of specific staged productions of metabolic species. Some sub-networks will be purely catalytic, and yet others will define maximal and minimal accumulations of metabolites and control pinch points which are analogous to set points in a control system and are referred to as governors.

A critical issue with modeling cell signaling processes is that rarely are all of the rate constants and equilibrium constants known. Thus, we need to develop new optimization methods to search for the biochemical subcircuits that complete a pathway when there is missing kinetic data. We also need to develop new tools that let us understand the control switching logic of biochemical networks. This information will enable us to develop new experiments to better understand such networks. The long term goal of such models is to be able to predict biochemical outcomes based upon perturbed biochemical networks. We need to determine which processes need to be understood in detail so that we can

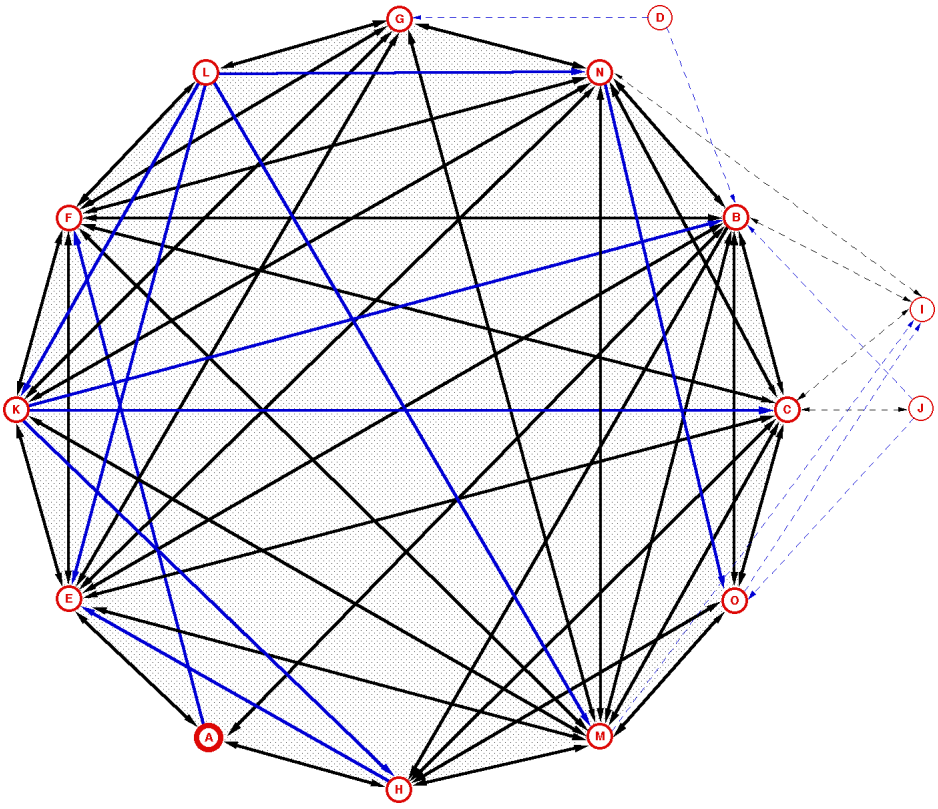


FIGURE 1. The network of signaling pathways between components of the EGF receptor system. Vertices are labeled based on the associated compartment, as specified by Bhalla & Iyengar. The pathways presented in blue denote communication connections which are one-way; all other pathways are bidirectional. The pathways which propagate the initialization of the formation of the GTP:Ras complex, or Compartment A, to the rest of the EGF network are highlighted in bold. The dashed arrows indicate connections to components (D, I, and J) that function as sources by converting a single species, while plain arrows indicate more complex interactions. A complete listing of edges for the Bhalla & Iyengar model is given in Table 1.

experimentally study the most important of these processes. Thus, we need to develop a fundamental understanding of the various reactions governing the biochemical pathways and how they are coupled, from the initial signal-receptor interaction to the initial signal-induced reaction to the various amplification steps to the final biological impact.

Compartment A:
EGFR, SOS

$$A \xrightarrow{GDP:Ras} B$$
$$A \xrightarrow{GTP:Ras} B$$
$$A \xrightarrow{Erk-PP} E$$
$$A \xrightarrow{EGF:EGFR} F$$
$$A \xrightarrow{GTP:Ras} H$$

Compartment B:
Ras

$$B \xrightarrow{GDP:Ras} A$$
$$B \xrightarrow{GTP:Ras} A$$
$$B \xrightarrow{Ca_4CaM} C$$
$$B \xrightarrow{Deph} C$$
$$B \xrightarrow{PKA} C$$
$$B \xrightarrow{PKC} C$$
$$B \xrightarrow{Deph} E$$
$$B \xrightarrow{Deph} F$$
$$B \xrightarrow{GTP:Ras} H$$
$$B \xrightarrow{PKC} H$$
$$B \xrightarrow{Ca_4CaM} I$$
$$B \xrightarrow{Deph} M$$
$$B \xrightarrow{PKC} M$$
$$B \xrightarrow{Ca_4CaM} N$$
$$B \xrightarrow{PKA} O$$

Compartment C:
AC, PDE, cAMP

$$C \xrightarrow{Ca_4CaM} B$$
$$C \xrightarrow{Deph} B$$
$$C \xrightarrow{PKA} B$$
$$C \xrightarrow{PKC} B$$
$$C \xrightarrow{Deph} E$$
$$C \xrightarrow{Deph} F$$
$$C \xrightarrow{PKC} H$$
$$C \xrightarrow{Ca_4CaM} I$$
$$C \xrightarrow{cAMP} J$$
$$C \xrightarrow{Deph} M$$
$$C \xrightarrow{PKC} M$$
$$C \xrightarrow{Ca_4CaM} N$$
$$C \xrightarrow{PKA} O$$

Compartment D:
Glu, mGluR, Gq

$$D \xrightarrow{G_{\beta\gamma}} B$$
$$D \xrightarrow{Gq} G$$

Compartment E:
AA, PLA2

$$E \xrightarrow{Erk-PP} A$$
$$E \xrightarrow{Deph} B$$
$$E \xrightarrow{Deph} C$$
$$E \xrightarrow{Ca} F$$
$$E \xrightarrow{Deph} F$$
$$E \xrightarrow{Ca} G$$
$$E \xrightarrow{AA} K$$
$$E \xrightarrow{Ca} K$$
$$E \xrightarrow{DAG} K$$
$$E \xrightarrow{Deph} M$$
$$E \xrightarrow{Ca} M$$
$$E \xrightarrow{Ca} N$$

Compartment F:
PLC_γ, DAG, IP3

$$F \xrightarrow{Deph} B$$
$$F \xrightarrow{Deph} C$$
$$F \xrightarrow{Ca} E$$
$$F \xrightarrow{DAG} E$$
$$F \xrightarrow{Deph} E$$
$$F \xrightarrow{Ca} G$$
$$F \xrightarrow{DAG} G$$
$$F \xrightarrow{IP3} G$$
$$F \xrightarrow{PIP2} G$$
$$F \xrightarrow{Ca} K$$
$$F \xrightarrow{DAG} K$$
$$F \xrightarrow{IP3} L$$
$$F \xrightarrow{Ca} M$$
$$F \xrightarrow{Deph} M$$
$$F \xrightarrow{Ca} N$$

Compartment G:
PLC, DAG, IP3

$$G \xrightarrow{Ca} E$$
$$G \xrightarrow{DAG} E$$
$$G \xrightarrow{Ca} F$$
$$G \xrightarrow{DAG} F$$
$$G \xrightarrow{IP3} F$$
$$G \xrightarrow{PIP2} F$$
$$G \xrightarrow{Ca} K$$
$$G \xrightarrow{DAG} K$$
$$G \xrightarrow{IP3} L$$
$$G \xrightarrow{Ca} M$$
$$G \xrightarrow{Ca} N$$

Compartment H:	$H \xrightarrow{Erk-PP} A$		$M \xrightarrow{Ca} F$
MAPK Cascade	$H \xrightarrow{PKC} B$		$M \xrightarrow{Deph} F$
	$H \xrightarrow{PKC} C$		$M \xrightarrow{Ca} G$
	$H \xrightarrow{Erk-PP} E$		$M \xrightarrow{PKC} H$
	$H \xrightarrow{PKC} M$		$M \xrightarrow{Ca_4CaM} I$
	$H \xrightarrow{PP2A} O$		$M \xrightarrow{Ca} K$
Compartment I:	$I \xrightarrow{Ca_4CaM} B$		$M \xrightarrow{Ca} N$
CaMKII	$I \xrightarrow{Ca_4CaM} C$		$M \xrightarrow{Ca_2CaM} N$
	$I \xrightarrow{Ca_4CaM} N$		$M \xrightarrow{Ca_3CaM} N$
Compartment J:	$J \xrightarrow{PKA} B$		$M \xrightarrow{Ca_4CaM} N$
PKA	$J \xrightarrow{PKA} C$		$M \xrightarrow{CaM:Ca_2:Ca_4:PP2B} O$
	$J \xrightarrow{PKA} O$		$M \xrightarrow{CaM:Ca_3:Ca_4:PP2B} O$
Compartment K:	$K \xrightarrow{PKC} B$		$M \xrightarrow{CaM:Ca_4:Ca_4:PP2B} O$
PKC	$K \xrightarrow{PKC} C$	**Compartment N:**	$N \xrightarrow{Ca_4CaM} B$
	$K \xrightarrow{Ca} E$	CaN	$N \xrightarrow{Ca_4CaM} C$
	$K \xrightarrow{DAG} E$		$N \xrightarrow{Ca} E$
	$K \xrightarrow{Ca} F$		$N \xrightarrow{Ca} F$
	$K \xrightarrow{Ca} G$		$N \xrightarrow{Ca} G$
	$K \xrightarrow{PKC} H$		$N \xrightarrow{Ca_4CaM} I$
	$K \xrightarrow{Ca} M$		$N \xrightarrow{Ca} K$
	$K \xrightarrow{PKC} M$		$N \xrightarrow{Ca} M$
	$K \xrightarrow{Ca} N$		$N \xrightarrow{CaM:Ca_2:Ca_4:PP2B} M$
Compartment L:	$L \xrightarrow{Ca} E$		$N \xrightarrow{CaM:Ca_3:Ca_4:PP2B} M$
Ca, IP3	$L \xrightarrow{Ca} F$		$N \xrightarrow{CaM:Ca_4:Ca_4:PP2B} M$
	$L \xrightarrow{Ca} G$		$N \xrightarrow{Ca_4PP2B} O$
	$L \xrightarrow{Ca} K$		$N \xrightarrow{CaM:Ca_2:Ca_4:PP2B} O$
	$L \xrightarrow{Ca} M$		$N \xrightarrow{CaM:Ca_3:Ca_4:PP2B} O$
	$L \xrightarrow{Ca} N$		$N \xrightarrow{CaM:Ca_4:Ca_4:PP2B} O$
Compartment M:	$M \xrightarrow{Ca_4CaM} B$	**Compartment O:**	$O \xrightarrow{PKA} B$
CaM	$M \xrightarrow{Deph} B$	PP1	$O \xrightarrow{PKA} C$
	$M \xrightarrow{PKC} B$		$O \xrightarrow{PP2A} H$
	$M \xrightarrow{Ca_4CaM} C$		$O \xrightarrow{PP1} I$
	$M \xrightarrow{Deph} C$		$O \xrightarrow{CaM:Ca_2:Ca_4:PP2B} M$
	$M \xrightarrow{PKC} C$		$O \xrightarrow{CaM:Ca_3:Ca_4:PP2B} M$
	$M \xrightarrow{Ca} E$		$O \xrightarrow{CaM:Ca_4:Ca_4:PP2B} M$
	$M \xrightarrow{Deph} E$		

TABLE 1. Edges between components of the EGF receptor system. A graph representation is shown in Figure 1. Compartment B interacts with twelve other EGF compartments; Compartment M interacts with eleven others; Compartments C, E and N interact with ten others each; Compartments F and K with nine others; G and H with seven; L and O with six, A and I with four, J with three, and Compartment D with only two interactions. The species through which the communications between components act are indicated above the arrows describing the direction of flow of information.

2. MATHEMATICAL APPROACH

In order to demonstrate our approach, Compartment A of the EGFR signaling pathway (as defined by Bhalla & Iyengar) was selected as a rather small but substantive sub-network for illustrative purposes. Our method is quite general and will work for any family of coupled chemical rate equations. An advantage of our approach is that we only need to know how species are connected but do not need to know the details of the chemical processes, either equilibria or kinetics, in order to provide useful insights.

Our approach is to use Petri nets as a tool to understand network decision-making control processes based on connectivity. The concept behind a standard Petri net is that of a set of counters that enumerate tokens as they are gradually moved from one place to another subject to the satisfaction of a set of transition rules. This is simple counting. In this paper, we extend and analyze Petri nets using hyperdigraphs that provably define oriented matroids. Petri nets have associated S-invariants and T-invariants (associated with states and transitions, respectively) coming from a linear algebra viewpoint; the important parts of these invariants are the sign ($\pm$). We can show that the S-invariants and T-invariants define oriented matroids. The S-invariant and T-invariant duality provides a rich combinatorial algebraic-geometry representation theory. This duality can generate constrained-based combinatorial optimization schemes for minimum-maximum information flow laws that are characteristic of every signaling pathway [Oliveira *et al.* (to be published)]. These pathway connections satisfy fixed point objective functions of multiset arrangements of reactive chemical species. All optimizations of Petri net representations of cell signaling pathways are constructed over this combinatorial geometry and use its metric to define the objective functions for a given flow path optimization/regulation.

Oliveira *et al.* (2001) established that Petri nets define a class of hyperdigraphs whose signed supports are identified with an oriented matroid. The oriented matroid associated with the hyperdigraph representation of a Petri net is used to search for minimal cycles. These minimal cycles represent the smallest (irreducible) subcircuits composed of places and transitions that span the network. *Places* represent event locations in the net, and *transitions* (physical rules including rates or decision making rules that represent conditional control laws) represent the rules governing the flow in the system. The topology of a Petri net is completely specified by its *incidence matrix*, an algebraic matrix whose rows are places and columns are transitions of the Petri net. All matrix entries are either 0, 1, or -1; these quantities specify the absence or presence of a connecting edge between two places, as well as its direction. The incidence matrix is equivalent to the Petri net. In addition to its topology or connectivity, a Petri net at a given time has a state or *marking* that is specified by the number of *tokens* in each place. Markings can be thought of as tokens representing information. A place producing the information is referred to as a source; and a place consuming the information is a sink. Flux conservation is achieved when the rate at which tokens are being produced equals the rate at which tokens are being consumed. When the flux, for a given firing sequence, starts and ends at the same point it is called a cycle. These cycles are of interest because they represent the paths through which the network is passing as well as conserving information, *cf.* Oliveira *et al.* (2001, 2003). When a reaction or series of reactions takes place, the corresponding transitions are said to have fired, and the token numbers then change commensurate with the stoichiometry of the reaction.

S-Invariants and T-Invariants. Schuster *et al.* present an algorithm to find (in their terminology) the *elementary flux modes* of a reaction network, *cf.* Schuster *et al.* (2000, 2002a and 2002b). These are exactly the minimal positive T-supports of the network. Their algorithm is fast and effective. The algorithm is based (loosely) on the underlying

oriented matroid as it is relies upon the underlying partial order of T-supports and uses careful bookkeeping on the linear algebra side to keep track of the supports required. The major difference between their algorithm and ours is that we are retaining the entire matroid, determining all of the S-supports and T-supports, including but not limited to just the minimal positive T-supports.

The Petri net approach introduced here provides us with an extensive set of combinatorial tools for deducing the qualitative control logic of biochemical networks. This approach defines states in the system to be marked places and the tokens are colored with markings that symbolically represent concentrations of biomolecular species such as metabolites, enzymes, and cofactors, *etc.*. In other words, the generation, storage and depletion of biomolecular species that define the discrete compartmentalized components in a biochemical reaction network are the corresponding colored balls or tokens that are arranged into marked boxes or places of our Petri net. The pathways that define the "trajectories" of reacting species through a biochemical reaction network correspond to graph cycles or sub-network circuits of sets of reacting metabolites. The systematic nature of this modeling approach studies the circuit arrangements or partitions of a biochemical network as functions of marked balls (biochemical species) being arranged into marked places, subject to a set of process control rules defined by the transition conditionals of the Petri net. The tokens are symbolic representations of biomolecular concentrations.

Every linear system of chemical rate equations is subject to the conservation laws of system flux. Every linear system of chemically balanced equations can be written as an $m \times n$ conservation matrix N^t, where n corresponds to the number of reacting species in the chemical system and m corresponds to the number of reactions that are taking place within the chemical system.

For every system of biochemical reactions, the set of reactions that defines the system is referred to as a set of fluxes. In deriving a systems analysis model, it is necessary to define a boundary around the set of reactions. This boundary imposes a set of bounding conditions for physically discretizing the system of chemical equations and, as such, defines two classes of system fluxes: the internal set of system fluxes defined by the set of internal sources and sinks that correspond to our sets of marked places in the Petri net model; and the external fluxes which indicate the transport of the biochemical species *e.g.*, metabolites, that can exist outside of the bounded system as potential input sources to a new sub-system of biochemical reaction equations. Note that each of the external fluxes has a positive value.

The boundedness property of the Petri net representation identifies which combinations of paths have intermediate maximum and minimum accumulations. This is the case since we are enumerating multisets of molecular species into marked place holders subject to reaction and conservation constraints. The arrangements define a partitioning of kinetic reaction space that corresponds to the set of sub-circuits that span the network.

If we construct a Petri net model for the system of reactions, the transpose of the incidence matrix N^t of the net is exactly this conservation matrix. We observe that the conservation matrix N^t is a linear transformation on the n-dimensional rational or $\mathbb{Q}^n$-vector space. For every conservation matrix N^t, there exists a set of stoichiometrically defined basis vectors v, which is an n-vector of rational numbers known as the S-supports. A set of S-supports is an S-invariant. If N^t is an $m \times n$ matrix, the nullspace of N^t, denoted by $\mathcal{N}(N^t)$, consists of all vectors v in rational n-dimensional space, $\mathbb{Q}^n$, such that

$$N^t v = 0. \tag{1}$$

It is important to note that the incidence matrix, N, is not invertible because of the Eulerian nature of the network. We observe that the linear transformation defined by equation (1) is a subspace of $\mathbb{Q}^n$. The dimension of the nullspace of N^t depends on the rank of N^t, where the rank r of N^t is defined as the number of free variables that exist in the system of equations that defines the matrix N^t.

$$r + \mathcal{N}(N^t) = m. \tag{2}$$

An S-invariant has as its dual a T-invariant, m-vector of rational numbers w, such that

$$Nw = 0. \tag{3}$$

If equation (1) defines a linear algebraic balance equation that represents the stoichiometric conservation of system flux in the linear system of biochemical reaction equations, then the matrix equation (1) is solved when we obtain the linear combination of linearly independent basis vectors $p_1, \ldots, p_n$ that spans the equational nullspace $\mathcal{N}(N^t)$ of the conservation matrix N^t.

When the dimension of the nullspace is small, it is a simple exercise to compute the solution to equation (1). If the dimensionality of the system of linear conservation equations and hence the nullspace of equation (1) is large, then we have to use computational methods to calculate a basis for the nullspace. The computational hardness associated with computing all of the linearly independent sets of admissible solutions that span $\mathcal{N}(N^t)$ grows as the dimension of N^t. It is notable that Petri net representations of signaling networks are scale-free. The Petri net model is therefore a combinatorial, scale-independent representation of the multiscale physics defined by a system of chemical rate equations which in turn defines a biomolecular signaling network.

We can construct a Turing complexity measure, in both time and space, that determines how computationally hard it is in all generality to identify the principal sub-circuits within a multicoupled biochemical signaling network. The best computational decision procedure for finding principal sub-circuits in these types of networks is equivalent to the Hamiltonian circuit identification problem that is well-known to be NP hard. See Garey & Johnson (1979). If we are missing any parametric data in the network then the circuit identification problem computationally explodes superexponentially to become PSPACE hard. In systems analysis models, this case is often referred to as the circuit reachability problem. See Oliveira *et al.* (2001) and Jones *et al.* (1977).

The nullspaces of equation (1) contain all of the solutions that satisfy the balance equations defined by (1). The principal problem is to be able to optimally generate the entire nullspace and then search for all of the biochemically meaningful basis sets that span the nullspace as linear combinations of sub-cycle paths. As indicated above, this nullspace corresponds to the space of cycles of some graph, and so we take biochemically meaningful to mean a positive cycle or path.

Ideally, we want all minimal positive cycles. The method used in this paper is based on a combinatorial geometric analysis of cycles in graphs. It first finds a suitable graph, generates the cycle matroid, and translates back to find positive cycles in the Petri net. The computational construction of the nullspace of interest requires that a decision procedure exist that generates a set of signed vectors, those corresponding to the signed sets of tuples that define an incidence matrix N^t that characterizes the state representation of the system of equations, that in turn defines the biochemical reaction system. The algorithm must efficiently generate a minimum number of linearly independent signed vectors that defines a candidate basis set as a signed sub-cycle. The optimal search of an n-set of points in $\mathbb{Q}^n$ space, that satisfies the dimensional requirements that the number of basis vectors

necessary to define the basis set is equal to the dimensions of $\mathcal{N}(N^t) \subseteq \mathbb{Q}^n$, is provably polynomial. This is not true for finding an optimal set of basis vectors over $\mathbb{R}^n$. The S-invariants can also be used to identify the existence of all non-reachable sub-circuits in the reaction network. We note that the reachability of one marked place to another marked place in a Petri net is subject to the condition that all place markings are reachable when the conservation of fluxes is defined by the network S-invariants. The T-invariants of the network determine the set of conditional transitions that have to be evaluated to identify the set of sub-circuits that span the entire reaction network as circuits. The T-invariants are defined by an equation that is identified with equation (1), given by equation (3).

3. Analysis of Compartment A of the EGFR Network

Model of EGFR Compartment A. We apply these methods to the SOS subreaction network (Compartment A) of the EGFR pathway [Bhalla & Iyengar (1999). This complex biochemical process can be considered to be composed of two fundamental types of biochemical building blocks, which are both noted to be reversible: molecular reactions and complex formation. See Oliveira *et al* (2001, 2003). For example, an enzyme reaction is composed of one molecular reaction and two complex formations. We have chosen the Bhalla & Iyengar model because it is a consistent representation for all fifteen compartments of the EGFR network. It is our intent to build the hyperdigraph mathematical representations of all fifteen compartments and subsequently attempt to link them all together.

A graph representation of the communication between compartments of the EGFR signaling network is presented in Figure 1. The edges represent shared species within the network. Table 1 contains a listing of these edges. The edges have a directionality based on a given compartment producing a species which is then distributed, as input, to other compartments. In the case of neither compartment producing a given species, yet both making use of it, there is an assumed bi-directional edge connecting the compartments to illustrate sharing. In this way, we are able to model the control of the network in terms of timing.

Thirty places and 46 transitions are required to specify Compartment A, and we note that we are not tracking the ATP, ADP, water molecules of the reaction nor the CO_2 released. Our primary concern is with the connectivity of the network made up from biochemical stencils. See Oliveira *et al.* (2001, 2003). Two place nodes are connected subject to a transition node simply if it is possible for the second place to be reached from the first place through some reversible physical/chemical mechanism. Of course, the actual amount of chemical system reversibility may be very small and is dependent on the equilibrium constant and/or kinetic rate constants. Since all of the reactions under consideration are potentially reversible, then these paired sets of transitions are identified explicitly. The implications of the complexity in the transition rules is deferred to a subsequent paper, and, in this discussion, only the existence of the species moving through the places of the pathways are indicated. In other words, we do not assign weights or probabilities to the paths of the reactions; we only indicate the existence of the reaction paths and their directions. We are not considering the quantities of molecular species present, the equilibrium constants, their reaction rates, nor the timed sequence of events. Forward paths are designated by directed solid lines and backward paths are designated by directed dashed lines. Forward and backward paths are specified without prejudice, acknowledging that one or the other of the paths may be significantly dominant. We consider that two places, or even subreactions, are connected regardless of whether that communication of

Place	Species	Type
p_1	Epidermal growth factor [EGF]	
p_2	Epidermal growth factor receptor [EGFR]	
p_3	EGF : EGFR	Enzyme
p_4	EGF : EGFR$_{internalized}$	
p_5	Src homolog and collagen protein [SHC]	Protein
p_6	Phosphorylated SHC [SHC-P]	
p_7	Phosphotase(1) [P'ase(1)]	Enzyme
p_8	SHC : EGF : EGFR	
p_9	EGF : EGFR : SHC-P	
p_{10}	SHC : P'ase (1)	
p_{11}	P'ase(1) : SHC-P	
p_{12}	SHC-P : SOS : GRB2	Enzyme
p_{13}	Guanosine diphosphate : Ras proteins [GDP : Ras]	Protein
p_{14}	Guanosine triphosphate : Ras [GTP : Ras]	
p_{15}	Ras GAP	Enzyme
p_{16}	GDP : Ras : SHC-P : SOS : GRB2	
p_{17}	SHC-P : SOS : GRB2 : GTP : Ras	
p_{18}	GDP : Ras : Ras GAP	
p_{19}	Ras GAP : GTP : Ras	
p_{20}	SOS : GRB2	
p_{21}	Growth factor receptor binding protein 2 [GRB2]	Protein 2
p_{22}	SOS* : GRB2	
p_{23}	Son of sevenless homolog protein [SOS]	Protein
p_{24}	Phosphorylated SOS [SOS*]	
p_{25}	Extracellular signal-regulated kinase-peroxisome proliferator [Erk-PP]	Enzyme
p_{26}	Phosphotase(2) [P'ase(2)]	Enzyme
p_{27}	SOS : Erk-PP	
p_{28}	Erk-PP : SOS*	
p_{29}	SOS : P'ase(2)	
p_{30}	P'ase(2) : SOS*	

TABLE 2. Petri net place definitions for Compartment A of the EGFR network. Commonly used abbreviations are indicated in brackets. Note the three groupings of two enzymes and a protein.

information is via a forward or a backward path or combinations thereof, indicating only whether a physical connection is possible. Our assumption of reversibility of the system requires the adding of extra edges to ensure that every connection between two places is both forward and backward, *i.e.* a pair of edges.

Specifications of the Places. The list of 30 places is presented in Table 2. The places have been color-coded according to the subreactions comprising the cycle. The places represent the various key chemical species produced or consumed in the cycle. Note that there are six enzymes in this compartment.

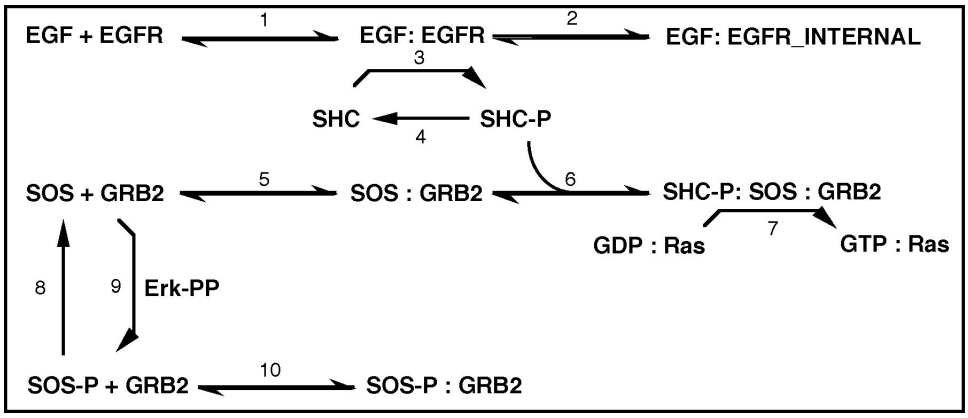


FIGURE 2. Original Diagram of the MAPK Compartment A of RAS-RAF.

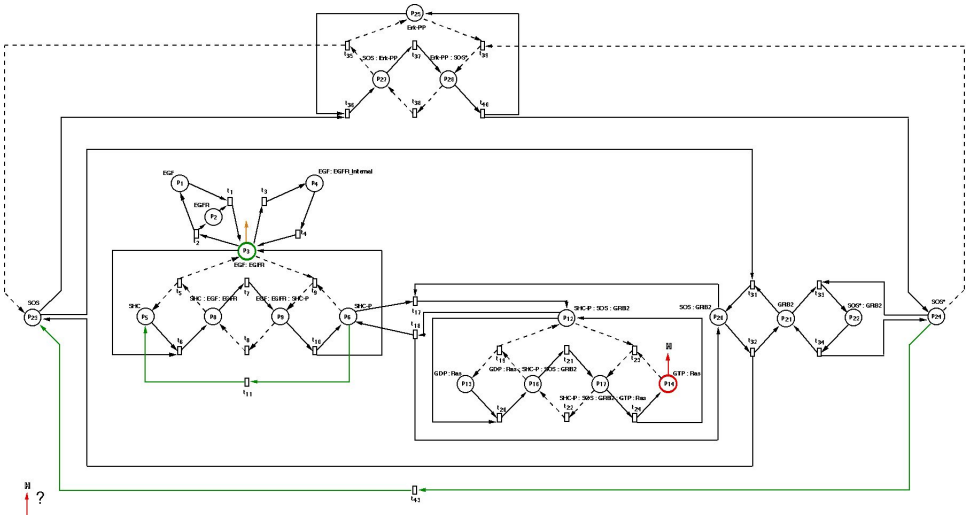


FIGURE 3. Petri Net Representation of original diagram of Compartment A, EGFR pathway, with places and transition conditions labelled. Note that the input signal from Compartment H is missing.

Petri Net Representation. A pictorial biological description is given in Figure 2, and the Petri Net representation of the biology is given in Figure 3. The corresponding incidence matrix is given in Table 3.

In the process of building the three representations – the biological, the Petri Net, and the incidence matrix – several irregularities were discovered. In the biological representation, there was no feedback of information from GAP:RAS and there were two uni-directional pathways. In the Petri net representation, this led to the appearance of incomplete sub-circuits leading to an imbalanced incidence matrix. The conservation of inputs and outputs for the entire network is obtained by demonstrating that either all column vectors, or alternatively row vectors, of the connectivity matrix of sub-matrices must sum to zero.

	p_1	p_2	p_3	p_4	p_5	p_6	p_8	p_9	p_{12}	p_{13}	p_{14}	p_{15}	p_{17}	p_{20}	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{27}	p_{28}
t_1	−1	−1	1																		
t_2	1	1	−1																		
t_3			−1	1																	
t_4			1	−1																	
t_5			1		1		−1														
t_6			−1		−1		1														
t_7							−1	1													
t_8							1	−1													
t_9			−1			−1		1													
t_{10}			1			1		−1													
t_{11}					−1	−1															
t_{17}						−1			1					−1							
t_{18}						1			−1					1							
t_{19}									1	1		−1									
t_{20}									−1	−1		1									
t_{21}												−1	1								
t_{22}												1	−1								
t_{23}									−1		−1		1								
t_{24}									1		1		−1								
t_{31}														1	−1		−1				
t_{32}														−1	1		1				
t_{33}															−1	1		−1			
t_{34}															1	−1		1			
t_{35}																	1		1	−1	
t_{36}																	−1		−1	1	
t_{37}																				−1	1
t_{38}																				1	−1
t_{39}																		−1	−1		1
t_{40}																		1	1		−1
t_{41}																	1	−1			

TABLE 3. Original transpose of the Incidence Matrix for Compartment A, EGFR pathway. The matrix transpose is presented, rather than the incidence matrix itself, for convenience.

Satisfying this condition implies that the incidence matrix of the network is balanced and hence all of the corresponding sub-circuits are complete. If this condition is not true for the incidence matrix representation of the network then the incidence matrix is said to be imbalanced which leads to the existence of incomplete sub-circuits. In view of these facts, the latter case was true. The sub-matrix blocks of the incidence matrix of the Compartment A network were irregular (specifically, some sub-blocks were 7×5 while others were 6×5); and Compartment H would have had no input location. We note that each column summing to zero is equivalent to having an S-invariant consisting of a vector of all ones. Upon closer inspection, it was determined that the biological representation was incomplete; and the Petri Net and incidence matrix pointed out precisely where the missing information had to be included. Precipitated by the mathematical formulation, this resulted in revisions to the network as depicted in Figures 4 and 5.

The biological network now includes the RasGAP enzymatic reaction and the Phosphatase (1) and (2) enzymatic reactions. The biological description is given in Figure 4, and Figure 5 shows the associated Petri net. The incidence matrix is given in Table 4. The now complete biological description of Compartment A is balanced biologically and mathematically.

As shown in the Petri net representation in Figure 5, the dominant features are the three double or paired enzymatic reactions. For example, consider the (ATP bound) EGF:EGFR and the (ATP bound) P'ase(1) enzymatic reactions which act upon SHC and SHC-P. This subnetwork is referred to as the SHC subreaction and is enlarged in Figure 14, presented in the Appendix. The EGF:EGFR provides output to Compartment F in the overall EGFR

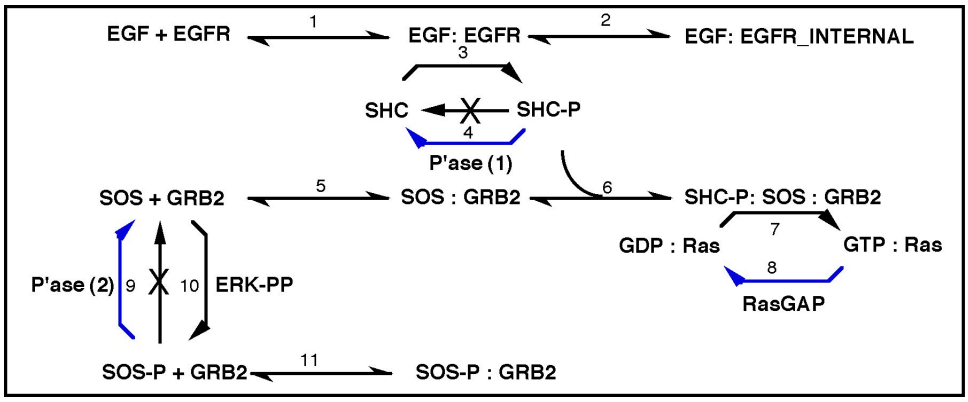


FIGURE 4. Corrected Compartment A, EGFR pathway.

network. The second paired enzymatic reaction, referred to as the GDP:Ras subreaction, is enlarged in Figure 15, also found in the Appendix. The species GTP:Ras provides input to Compartment H, as well as communication with Compartment B. Finally, the third paired reaction, the SOS subreaction, is enlarged in Figure 16 in the Appendix and it receives input from Compartment H at species P'ase(2). As a convenience, care was given in drawing the Petri net representation to minimize the number of pathway crossings. The crossings within a pathway speak to the dimensionality of the network as a whole. There is no reason to expect a naturally occurring biochemical network to be two-dimensional or planar at any level. The associated incidence matrix sub-block structures in Table 4 are each now of like dimension, namely 12×8, and the structure of the incidence matrix itself is block-banded, which is naturally advantageous for computational efficiency. We observe that the biochemical systems analyzed thus far have exhibited similar inherent symmetries and structures. When irregularities have been encountered, further inspection has invariably led to the subsequent inclusion of missing species and reactions which have resulted in this highly symmetric and block-banded structure.

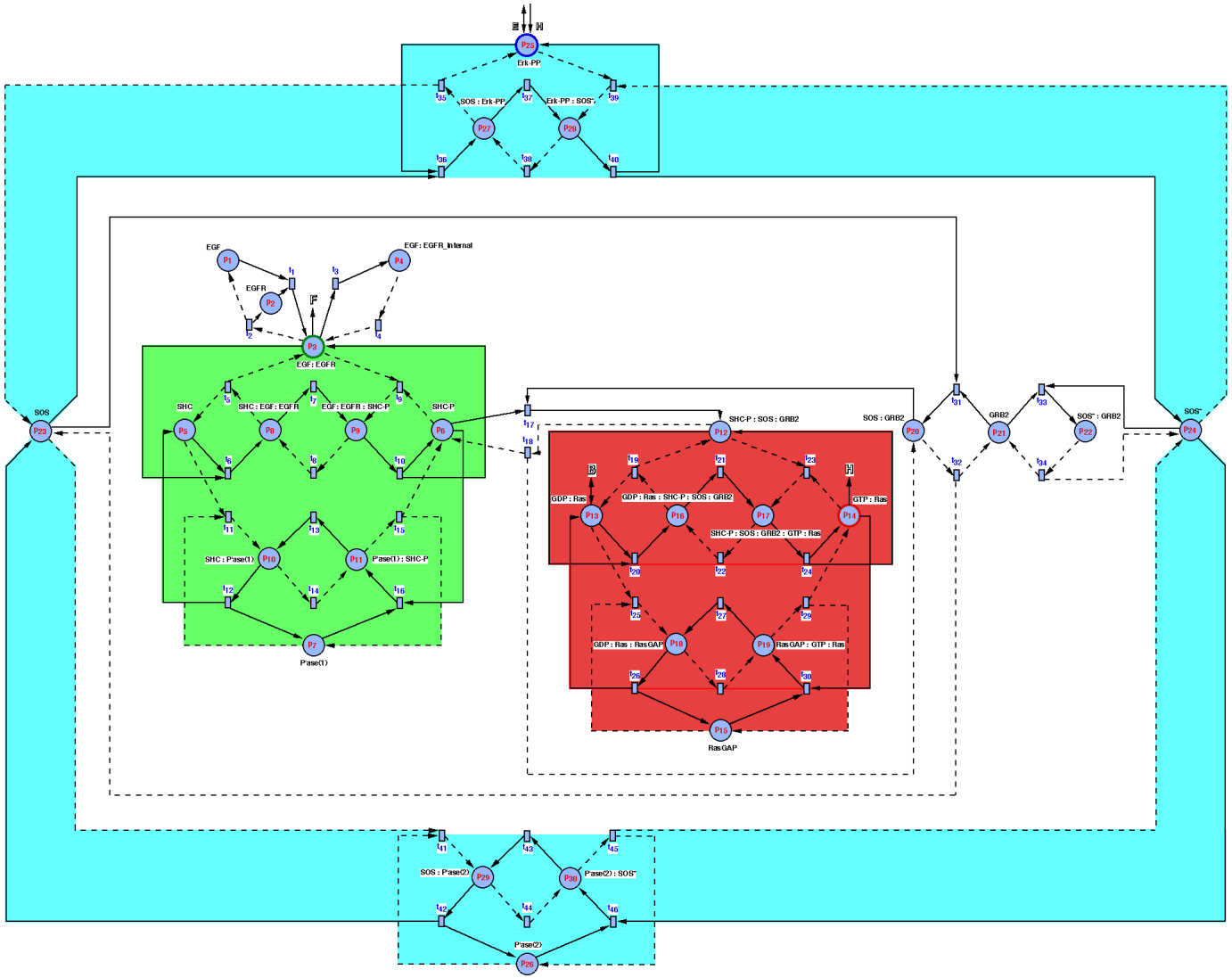


FIGURE 5. Corrected Petri net of Compartment A for the biological representation presented in Figure 4.

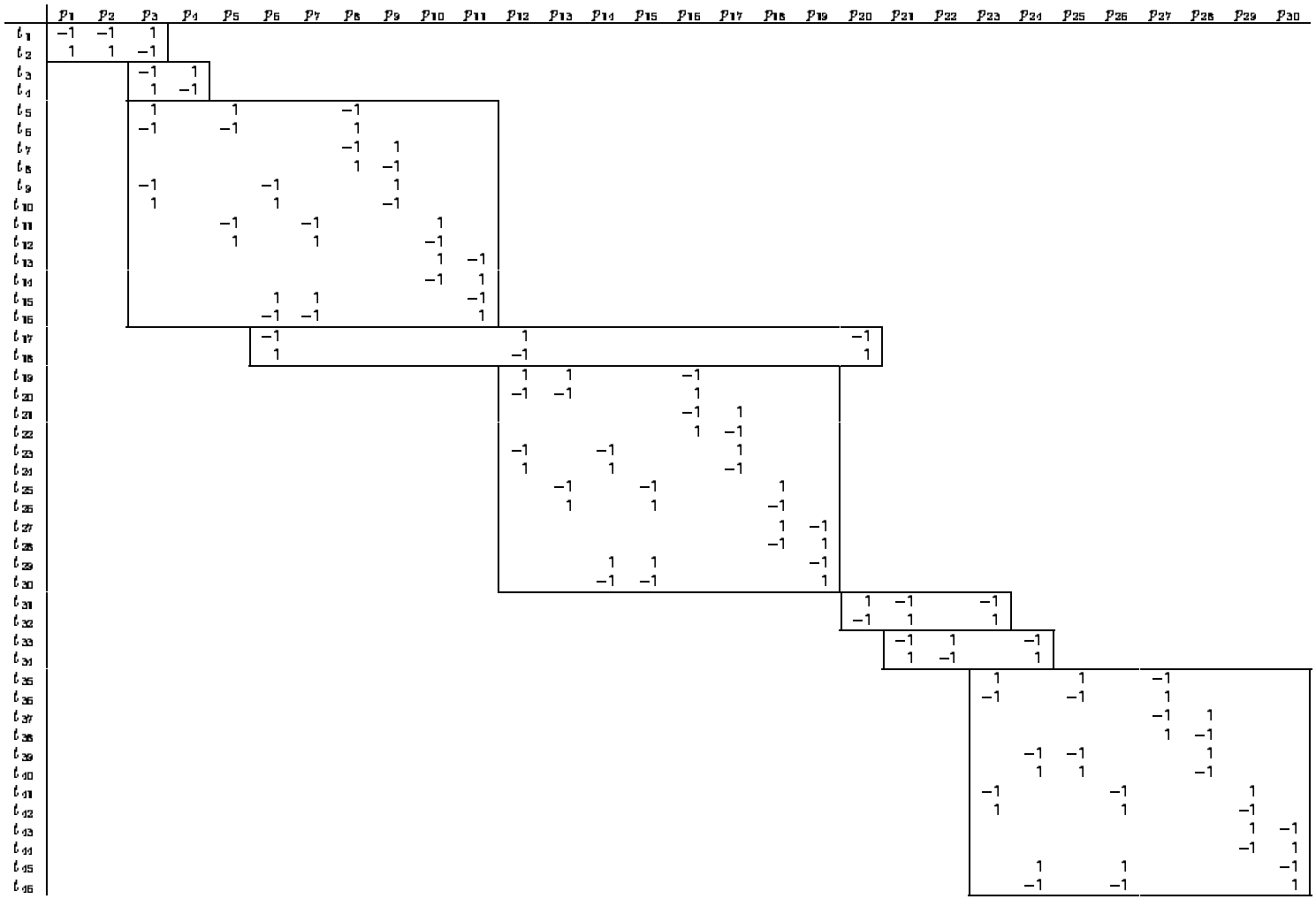


TABLE 4. Corrected transpose of the Incidence Matrix, N^t, for Compartment A of the EGFR pathway. Note that each 12×8 submatrix is composed of two opposite-signed 6×5 matrices; these are coupled enzymatic reactions, each of which is composed of two complex formations and a molecular reaction. There are several 2×3 submatrices representing complex formations and a 2×2 matrix representing a molecular reaction.

Computational Identification of the Minimal Cycles. The unique collection of distinct minimal cycles is the complete listing of closed walks that collectively move information through the network. The minimal pathways are the shortest, non-repeating, non-looping, non-revisited paths. All other cycles in the hyperdigraph can be generated from linear combinations of the set of minimal cycles. Once the decomposition into this set of minimal cycles has been identified, search strategies can be specified as linear objective functions over the database; *i.e.,* the flow network can then be analyzed to identify the set of paths which are of particular interest via optimization analysis using path algebras, linear programming and/or oriented matroid programming.

The process for determining the unique and complete set of minimal pathways begins with specification of the directed connections between places through transitions, as depicted for Compartment A in Figure 5. This specification may be represented as a listing of the directed arrows from places to transitions, as the incidence matrix, or as the Petri net diagram. These directed connections are input to our computer code, MICAH for Matroid Identification Code for the Analysis of Hyperdigraphs. To reduce the size of the otherwise infinite nullspace to a manageably finite number, we have devised a search algorithm, which we call "successive simplification". The algorithm is heavily based on the mathematical ideas presented in Oliveira *et al.* (2001). Beginning with the transpose of the incidence matrix N^t, the $\mathbb{Z}_2$ nullspace $\mathcal{N}(N^t)$ has dimension $m - n + 1$, where m is the number of transitions and n the number of places. Hence, the space has cardinality 2^{m-n+1}. For this example, $m = 46$ and $n = 30$, so the vector space has 2^{17} elements. We construct a faithful transformed digraph representation of the hyperdigraph, which has exactly the same cycles as the Petri net. We next use the fact that the cycles of an undirected graph form a $\mathbb{Z}_2$-vector space, which has the same nullspace as the original N^t, given by a $\mathbb{Z}_2$-matrix. The $\mathbb{Z}_2$-reduced nullspace has dimension less than $m - n + 1$. The basis set is determined for the undirected graph using row reduction, and its minimal cycles are determined using a spanning tree. A list of all cycles in the digraph may then be generated with a simple backtracking search algorithm. A simple bookkeeping procedure is used to keep track of the labels, *i.e.,* the assignments of specific numerical entries in the undirected graph are related to the appropriate places and transitions in the original hyperdigraph. Although this can be a tedious process, a linear complexity order is still provably guaranteed. Our reversibility assumption simplifies this process considerably as now every undirected cycle gives rise to exactly two positive cycles, one in each direction. [Of course, our list will contain non-minimal cycles; for example, figure eights may arise. We avoid listing non-minimal cycles, not by checking for minimality, but by directly listing a minimal subcycle of every cycle we find. This generates repetitions, which are easily removed later, retaining a linear complexity order in both polynomial time and space. See Sipser (1997).] The successive simplification algorithm is further developed below for this example.

A hyperdigraph representation of the Petri net is first constructed from the incidence matrix that defines the Petri net. This hyperdigraph is then transformed into a multigraph by replacing the transition nodes of the hyperdigraph with regular directed edges. Each directed flow through a transition is replaced by a single directed labeled edge. Diagrammatically, this transformation is demonstrated in Figure 6. We next transform this directed multigraph into an undirected simple graph. This is accomplished by suppressing multiple edges and removing directionality while retaining the labeled information. The process of transforming the hyperdigraph to the multigraph results in the same number of vertices as places, $n = 30$. However, in order to differentiate multiple paths through a given transition, 32 edges are added to the original 46, for a total of 78 edges. We next determine that the basis size of the reduced nullspace of the transformed graph is ten, resulting in ten

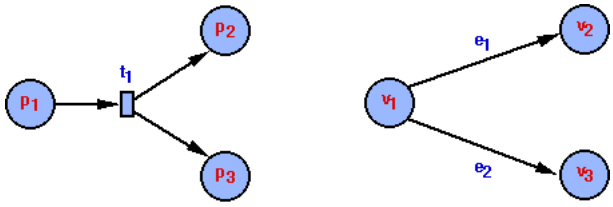


FIGURE 6. Transformation from a Petri net multigraph representation of information flowing through a single transition to multiple places to a simple Eulerian graph using labelled edges. Places got to vertices while shared transitions go to individual edges.

78×1 vectors. This implies that there are $2^{10} - 1$ possible minimal cycles through the undirected transformed graph or $2^{10+1} - 2$ possible minimal cycles through the directed graph. There exists a bijective correspondence between the set of edges in this simple Eulerian graph and the set of obtainable two-cycles. Given this mapping, it then becomes a simple exercise to enumerate all of the two-cycles. Once repeated edges have been eliminated from the undirected multigraph, a spanning tree is constructed from the graph. A spanning tree of the undirected graph for Compartment A is given in Figure 7. Note that there are 30 vertices and 29 edges. The identification of larger cycles, *i.e.*, longer than the two-cycles already in the listing, next requires the use of this spanning tree. By definition, a fundamental cycle is a cycle having exactly one edge not contained in the spanning tree. There exists a theorem which demonstrates the case that the addition of any edge to the tree, which has no cycles, creates exactly one cycle, and so the set of fundamental cycles is in one-to-one correspondence with the edges adjoined to the tree, *cf.* Gibbons (1985). So, every edge which is not listed in the spanning tree of the simple Eulerian graph is now added to generate a longer cycle, *cf.* Bachem & Kern (1991). Every edge is added to the spanning tree, exhaustively. Each time a cycle is added to the listing, MICAH eliminates from the listing any cycles which can be composed from combining any of the other smaller cycles, thereby retaining only the shortest cycles. Each minimal cycle corresponds to a nullvector of the incidence matrix of the transformed graph. There is an empty cycle (*i.e.*, with no edges) in the undirected graph and two empty cycles in the directed graph, which the algorithm does not count. Finally, the minimal cycles of the hyperdigraph are restored by reassigning direction using the retained labeled information.

The cycle listing generated in this way defines the set of irreducible minimal cycles. The complete set of cycles of the hyperdigraph is obtained by taking linear combinations of the set of minimal cycles. Observe that each linear combination generates two paths, one forward and one backward. It follows that each of these minimal cycles spans the nullspace of the incidence matrix of the original hyperdigraph. The final list of minimal cycles is the smallest set of cycles required to compose any physically meaningful circuit that the network system is capable of performing. This listing is guaranteed to be unique and complete. In theory, all of the closed circuits achievable in this network may be composed from the set of minimal circuits. In fact, the size of the directed characterizing data set is considerably smaller than the theoretical value of 2^{11}.

Unique Set of Distinct Minimal Cycles. MICAH first identifies just the 39 two-cycles because the algorithm suppresses multiple edges and would not identify them otherwise. These same two-cycles can be recognized as all of the molecular reactions in the Petri net

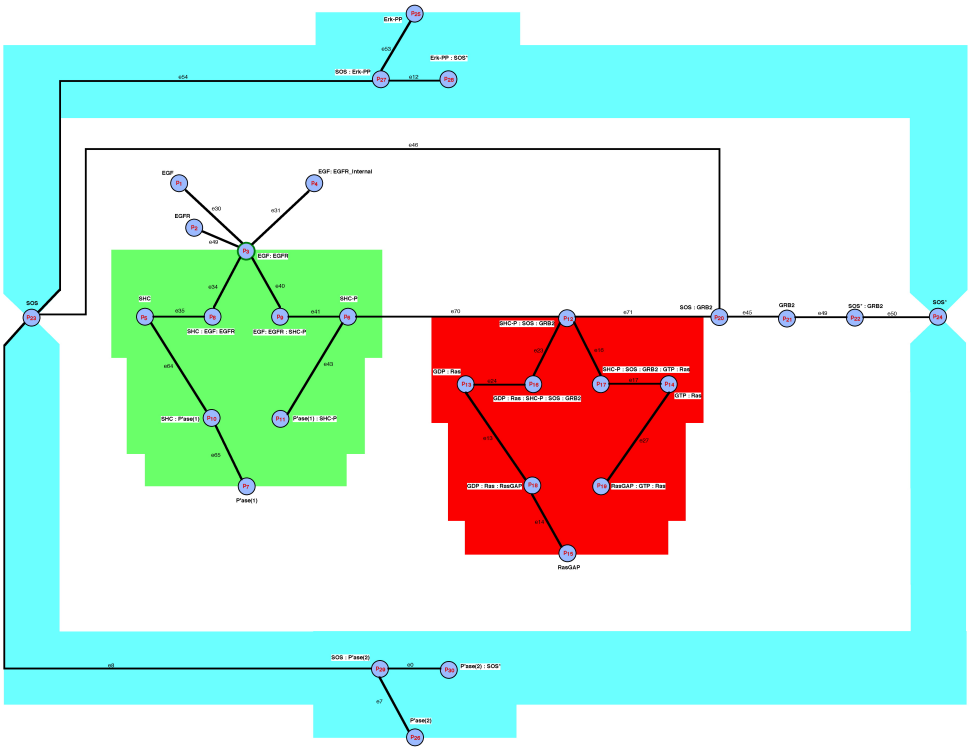


FIGURE 7. Spanning tree for EGFR Compartment A.

diagram, noting further that each complex formation contains two two-cycles and each enzyme reaction contains five two-cycles. We identify the remaining set of minimal cycles by finding minimal subcycles inside every cycle, ignoring duplicates which are removed from the listing via a script.

The resulting 39 two-cycles are given in Table 5 and the 44 other cycles are given in Table 6, grouped by length. There are 10 pathways of length 3, 6 pathways of length 6, 16 pathways of length 7, and 10 pathways of length 8. It is interesting to note that there are no pathways of length 4 or 5. The total number of distinct minimal cycles is nearly a factor of twenty-five less than the theoretical number possible due to the removal of non-minimal cycles. Setting aside the 0-cycle, all 83 of these minimal cycles, with their label information, are required in combination to completely characterize Compartment A of the EGFR network.

Cycle No.	Cycle Specification	Cycle No.	Cycle Specification
1a	$p_{29} \overset{t_{44}}{\dashrightarrow} p_{30} \overset{t_{43}}{\rightarrow} p_{29}$	21a	$p_3 \overset{t_3}{\rightarrow} p_4 \overset{t_4}{\dashrightarrow} p_3$
2a	$p_{30} \overset{t_{45}}{\dashrightarrow} p_{24} \overset{t_{46}}{\rightarrow} p_{30}$	22a	$p_8 \overset{t_5}{\dashrightarrow} p_3 \overset{t_6}{\rightarrow} p_8$
3a	$p_{30} \overset{t_{45}}{\dashrightarrow} p_{26} \overset{t_{46}}{\rightarrow} p_{30}$	23a	$p_8 \overset{t_5}{\dashrightarrow} p_5 \overset{t_6}{\rightarrow} p_8$
4a	$p_{28} \overset{t_{40}}{\rightarrow} p_{24} \overset{t_{39}}{\dashrightarrow} p_{28}$	24a	$p_8 \overset{t_7}{\rightarrow} p_9 \overset{t_8}{\dashrightarrow} p_8$
5a	$p_{28} \overset{t_{40}}{\rightarrow} p_{25} \overset{t_{39}}{\dashrightarrow} p_{28}$	25a	$p_3 \overset{t_9}{\dashrightarrow} p_9 \overset{t_{10}}{\rightarrow} p_3$
6a	$p_{26} \overset{t_{41}}{\dashrightarrow} p_{29} \overset{t_{42}}{\rightarrow} p_{26}$	26a	$p_6 \overset{t_9}{\dashrightarrow} p_9 \overset{t_{10}}{\rightarrow} p_6$
7a	$p_{23} \overset{t_{41}}{\dashrightarrow} p_{29} \overset{t_{42}}{\rightarrow} p_{23}$	27a	$p_6 \overset{t_{16}}{\rightarrow} p_{11} \overset{t_{15}}{\dashrightarrow} p_6$
8a	$p_{28} \overset{t_{38}}{\dashrightarrow} p_{27} \overset{t_{37}}{\rightarrow} p_{28}$	28a	$p_7 \overset{t_{16}}{\rightarrow} p_{11} \overset{t_{15}}{\dashrightarrow} p_7$
9a	$p_{18} \overset{t_{26}}{\rightarrow} p_{13} \overset{t_{25}}{\dashrightarrow} p_{18}$	29a	$p_{21} \overset{t_{31}}{\rightarrow} p_{20} \overset{t_{32}}{\dashrightarrow} p_{21}$
10a	$p_{18} \overset{t_{26}}{\rightarrow} p_{15} \overset{t_{25}}{\dashrightarrow} p_{18}$	30a	$p_{23} \overset{t_{31}}{\rightarrow} p_{20} \overset{t_{32}}{\dashrightarrow} p_{23}$
11a	$p_{19} \overset{t_{27}}{\rightarrow} p_{18} \overset{t_{28}}{\dashrightarrow} p_{19}$	31a	$p_{21} \overset{t_{33}}{\rightarrow} p_{22} \overset{t_{34}}{\dashrightarrow} p_{21}$
12a	$p_{17} \overset{t_{24}}{\rightarrow} p_{12} \overset{t_{23}}{\dashrightarrow} p_{17}$	32a	$p_{24} \overset{t_{33}}{\rightarrow} p_{22} \overset{t_{34}}{\dashrightarrow} p_{24}$
13a	$p_{17} \overset{t_{24}}{\rightarrow} p_{14} \overset{t_{23}}{\dashrightarrow} p_{17}$	33a	$p_{27} \overset{t_{35}}{\dashrightarrow} p_{25} \overset{t_{36}}{\rightarrow} p_{27}$
14a	$p_{17} \overset{t_{22}}{\dashrightarrow} p_{16} \overset{t_{21}}{\rightarrow} p_{17}$	34a	$p_{27} \overset{t_{35}}{\dashrightarrow} p_{23} \overset{t_{36}}{\rightarrow} p_{27}$
15a	$p_{12} \overset{t_{20}}{\rightarrow} p_{16} \overset{t_{19}}{\dashrightarrow} p_{12}$	35a	$p_{10} \overset{t_{12}}{\rightarrow} p_5 \overset{t_{11}}{\dashrightarrow} p_{10}$
16a	$p_{13} \overset{t_{20}}{\rightarrow} p_{16} \overset{t_{19}}{\dashrightarrow} p_{13}$	36a	$p_{10} \overset{t_{12}}{\rightarrow} p_7 \overset{t_{11}}{\dashrightarrow} p_{10}$
17a	$p_{19} \overset{t_{29}}{\dashrightarrow} p_{14} \overset{t_{30}}{\rightarrow} p_{19}$	37a	$p_6 \overset{t_{17}}{\rightarrow} p_{12} \overset{t_{18}}{\dashrightarrow} p_6$
18a	$p_{19} \overset{t_{29}}{\dashrightarrow} p_{15} \overset{t_{30}}{\rightarrow} p_{19}$	38a	$p_{20} \overset{t_{17}}{\rightarrow} p_{12} \overset{t_{18}}{\dashrightarrow} p_{20}$
19a	$p_1 \overset{t_1}{\rightarrow} p_3 \overset{t_2}{\dashrightarrow} p_1$	39a	$p_{11} \overset{t_{13}}{\rightarrow} p_{10} \overset{t_{14}}{\dashrightarrow} p_{11}$
20a	$p_2 \overset{t_1}{\rightarrow} p_3 \overset{t_2}{\dashrightarrow} p_2$		

TABLE 5. The labeled pathway listing of the 39 minimal two-cycles cycles of EGFR Compartment A. The forward paths are indicated by solid arrows and the backward paths are indicated by dashed arrows. Each two-cycle is a loop from one place to another and then back to the first via a forward and then a backward path, or *vice versa*.

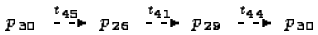
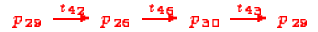
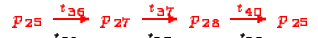
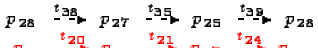
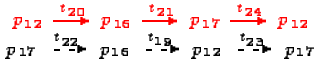
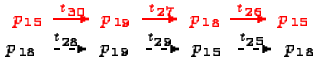
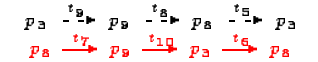
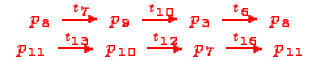
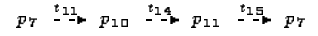
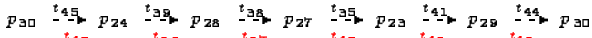
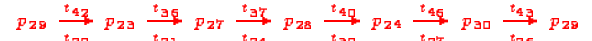
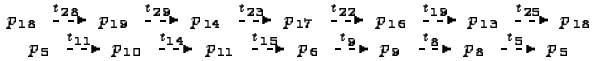
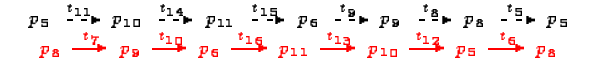
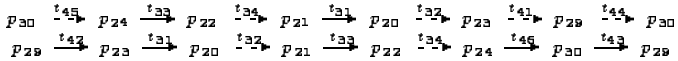
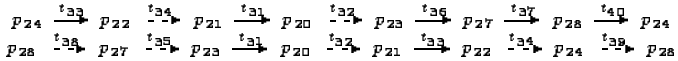
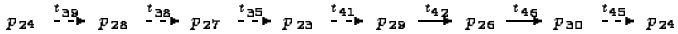
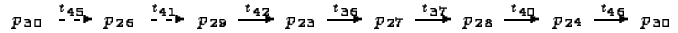
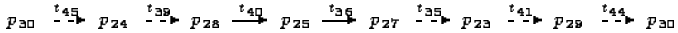
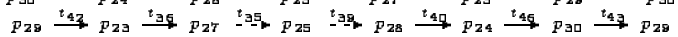
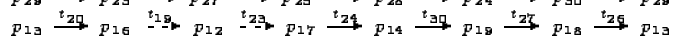
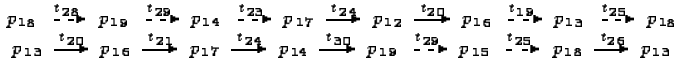
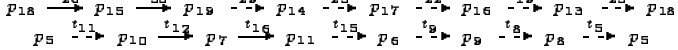
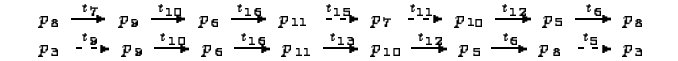
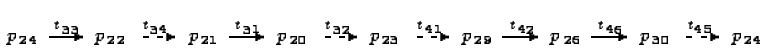
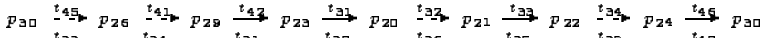
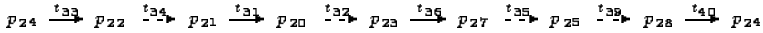
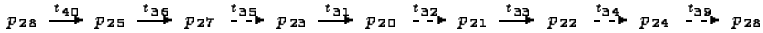
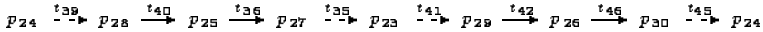
Number	Length	Cycle Specification
1b	3	
2b	3	
3b	3	
4b	3	
5b	3	
6b	3	
7b	3	
8b	3	
9b	3	
10b	3	
11b	3	
12b	3	
13b	6	
14b	6	
15b	6	
16b	6	
17b	6	
18b	6	
19b	7	
20b	7	
21b	7	
22b	7	
23b	7	
24b	7	
25b	7	
26b	7	
27b	7	
28b	7	
29b	7	
30b	7	
31b	7	
32b	7	
33b	7	
34b	7	
35b	8	
36b	8	
37b	8	
38b	8	
39b	8	
40b	8	
41b	8	
42b	8	
43b	8	
44b	8	

TABLE 6. Labeled pathway listing of the remaining 44 minimal cycles. Pathways listed in red are the nine positive forward-only cycles.

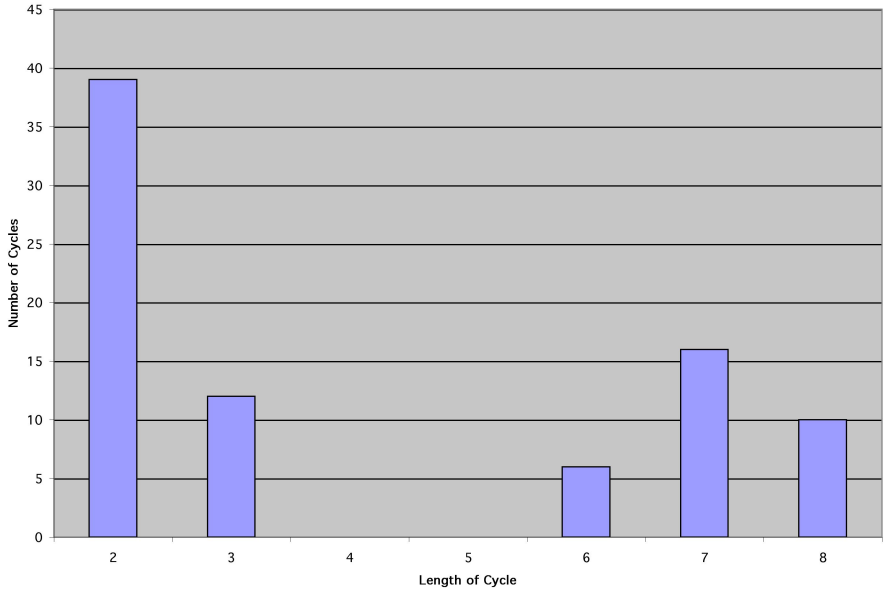


FIGURE 8. Discrete distribution of the number of cycles of a given cycle length. The two-cycles are the molecular reactions and the three-cycles are the forward-only and backward-only paths through the enzymatic reactions. Note that there is the indication of an emerging binomial distribution for the longer length cycles.

We observe that half of the cycles of length 3 presented in Table 6 are the forward-only and half are the backward-only pathways through the enzymatic reactions. Similarly, half of the cycles of length 6 are forward-only and half are backward only, through the paired enzymatic reactions. The cycles of lengths 7 and 8 are mixed with some forward and some backward paths, connecting the three paired enzymatic reactions.

Our study of additional biochemical networks has revealed a pattern where the number of cycles of a given length is high for two-cycles and three-cycles and then drops off. The remaining number of cycles of a given length is represented by the discrete analog of a Gaussian distribution with many cycles of considerable length, as shown in Figure 8. This pattern will be further and more dramatically substantiated in subsequent publications of the analysis of other components of the larger EGFR network.

Frequency distributions are next presented indicating the number of occurrences of each transition (Table 7) or place (Table 8) in the list of minimal cycles. Extreme values in the network analysis are of interest, be they valleys (*i.e.* infrequent occurences) or peaks (*i.e.* frequent occurrence).

Transition	No. Occur.	Percent Occur.	Transition	No. Occur.	Percent Occur.
t_{35}, t_{36}	12	14.4	t_{25}, t_{26}	9	10.8
t_{39}, t_{40}	12	14.4	t_{29}, t_{30}	9	10.8
t_{41}, t_{42}	12	14.4	t_{37}, t_{38}	5	6.0
t_{45}, t_{46}	12	14.4	t_{43}, t_{44}	5	6.0
t_{31}, t_{32}	10	12.0	t_{7}, t_{8}	4	4.8
t_{33}, t_{34}	10	12.0	t_{13}, t_{14}	4	4.8
t_{5}, t_{6}	9	10.8	t_{21}, t_{22}	4	4.8
t_{9}, t_{10}	9	10.8	t_{27}, t_{28}	4	4.8
t_{11}, t_{12}	9	10.8	t_{1}, t_{2}	2	2.4
t_{15}, t_{16}	9	10.8	t_{17}, t_{18}	2	2.4
t_{19}, t_{20}	9	10.8	t_{3}, t_{4}	1	1.2
t_{23}, t_{24}	9	10.8			

TABLE 7. EGFR Compartment A cycle transition counts. The transitions are color-coded to match the subreaction identification. Note that the paired transitions essentially split into two distinct categories: those greater than 10% and and those less than 6%.

Transitions. There are 46 transitions required to specify the EGFR Compartment A network. The transitions, the count of their occurrences in the 70 minimal cycles, and the percentage of their occurrences in the 70 minimal cycles are presented in Table 7. They are listed in order of decreasing occurrence. A bar plot for transitions 1 through 46 versus percentage of occurrence in the 70 minimal cycles is given in Figure 9 in the order of transition number. Because we are assuming that all reactions are reversible with $K = 1$, each pair of transitions occurs with the same frequency, as shown in Table 7 and Figure 9. We have listed the transition occurrences in pairs; however, in subsequent analyses for which the equilibrium constant equal to one assumption will not be made, the values will obviously be different. At this point, we are merely looking at the possible pathways along which information can flow.

The key features in the transition bar plot enable us to assign and compare the low number of transitions with the high number of transitions. A low percentage of transitions means that information does not pass as readily in these parts of the cycle and a high percentage means that these transitions are readily passing information. This assumes that the equilibrium constants are unity as well as the rate constants.

The first set of transition pairs that occur most frequently are (t_{35}, t_{36}); (t_{39}, t_{40}); (t_{41}, t_{42}); and (t_{45}, t_{46}). These transitions are the activation of SOS either through P'ase(2) or ERK-PP via phosphorylation reactions. We note that the ERK-PP is coupled to both the E and H Compartments. The next most likely transitions involve the actuation of GRB2 and formation of SOS*:GRB2. The middle set of transitions occurring at approximately 11% frequency are those in the SHC subnetwork and in the GDP:Ras subnetworks. The first set of low valued transitions is associated with the internalization of EGFR. The internalization leads to subsequent biochemistry and our results clearly show that the rate at which this occurs will be important as this serves as the control point for subsequent reactions. The next set represents the formation of the EGF:EGFR complex as expected as this is the initiation step in the cycle. As observed previously, initiation steps tend to occur with low frequency as they are not coupled to many cycles. The final set corresponds to flux flowing

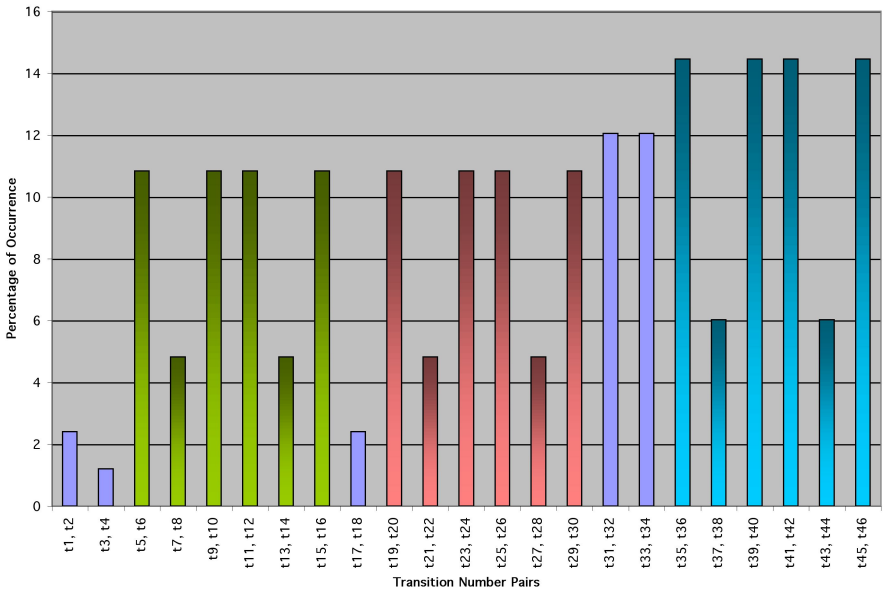


FIGURE 9. Transition occurrence bar plot. The distribution demonstrates extremal bounds. In other words, the data falls into two bins: above 6% and below 6%. This indicates the existence of a Bernoulli process.

from the SHC subreaction to the GDP:Ras subreaction effectively defining the branching ratio leading to SHC formation. It is interesting to note that there is not a high percentage occurrence with any of the transitions and that 15 out of the 23 pairs of transitions are in the range of 11 - 14% at the high-end. Only one transition pair (t_3, t_4) occurs only once and it is involved with the internalization of the EGF:EGFR complex. Basically, except for a few low spots as described above, once the cycle in initiated it sustains itself as shown in Figure 9. The lower values are usually associated with formation of larger complexes (molecular machines).

Places. Thirty places specify the EGFR Compartment A network. The places, the count of their occurences in the 83 minimal cycles, and the percentage of their occurrences in the set of 83 minimal cycles are presented in Table 8. The table is presented in descending order of occurrence. A bar plot for places 1 through 30 versus percentage of occurrence in the 83 minimal cycles is presented in Figure 10. The plot is presented in order of ascending place number.

There are a number of places that have a high percentage of occurrences, although no places appear in greater than 25 percent of the minimal cycles. As expected, from the transitions, the highest percentages are those associated with p_{23} and p_{24}, which are SOS and SOS*, followed by p_{27} through p_{30}, which are the complexes SOS:Erk-PP, Erk-PP:SOS*, SOS:P'ase(2) and P'ase(2):SOS*, respectively. The next highest percentages involve the SHC and Ras complexes. As expected from the transition data, places p_1, p_2 and p_4, corresponding to EGF, EGFR and EGF:EGFR$_{internalized}$, appear least often. This is consistent with the fact that the signal process is initialized by binding of EGF to

Place	No. Occur.	Percent Occur.	Place	No. Occur.	Percent Occur.
p_{23}	19	22.9	p_6	11	13.3
p_{24}	19	22.9	p_{20}	11	13.3
p_{27}	17	20.5	p_5	10	12.0
p_{28}	17	20.5	p_{12}	10	12.0
p_{29}	17	20.5	p_{13}	10	12.0
p_{30}	17	20.5	p_{14}	10	12.0
p_8	13	15.7	p_{21}	10	12.0
p_9	13	15.7	p_{22}	10	12.0
p_{10}	13	15.7	p_{25}	10	12.0
p_{11}	13	15.7	p_{26}	10	12.0
p_{16}	13	15.7	p_7	8	9.6
p_{17}	13	15.7	p_{15}	8	9.6
p_{18}	13	15.7	p_1	1	1.2
p_{19}	13	15.7	p_2	1	1.2
p_3	11	13.3	p_4	1	1.2

TABLE 8. EGFR Compartment A cycle place counts. The transitions are color-coded to match the subreaction identification.

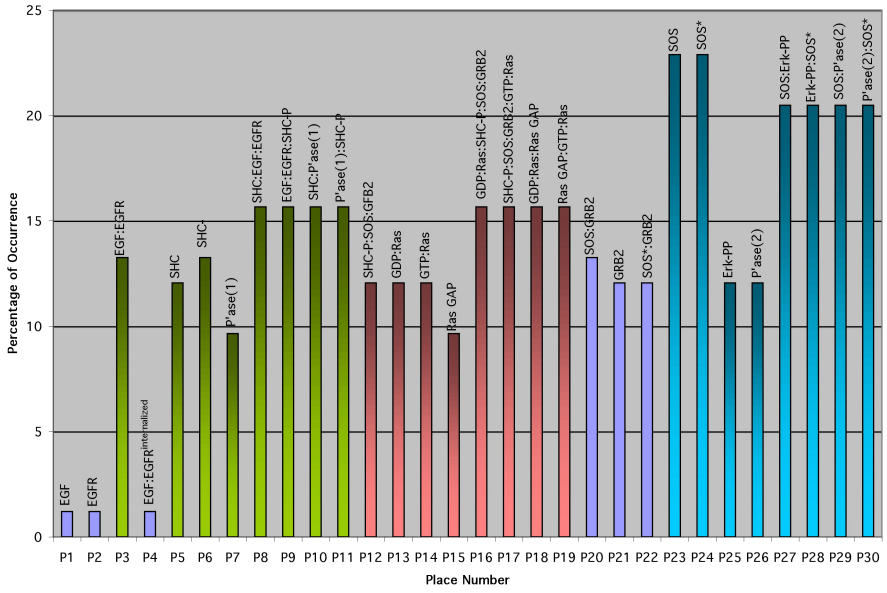


FIGURE 10. Place occurrence bar plot.

EGFR and that internalization of the complex is a control point for a different biochemical pathway.

The regularity for the places is consistent with the results for the transitions. The initialization (EGF and EGFR) are low as they are the start of the sequence. The formation of

EGF:EGFR$_{internalized}$ is also low as there is only one path connecting it to the network. Obviously, a key control point is here as once EGF:EGFR$_{internalized}$ is formed, it can only go back through the EGF:EGFR place. Thus, internalization is a way to control the initiation of the entire pathway as well as a separate pathway. The only two places with low occurrence (just below 10%) are P'ase(1) involved in the phosphorylation of SHC and Ras GAP itself.

Forward Positive Pathways. There are nine positive forward-only minimal cycle, which are highlighted in red in Table 6. They are repeated in Figure 9 to facilitate further discussion. There are two forward three-cycles associated with each of the three subreactions – SHC, GTP:Ras, and SOS. In addition, there are three six-cycles associated with each of the three subreactions. They may be easily identified by tracing them through their respective enlarged Figures 14, 15, and 16, which are included in the Appendix.

Number	Cycle Specification	Subreaction Description
10b	$p_8 \xrightarrow{t_7} p_9 \xrightarrow{t_{10}} p_3 \xrightarrow{t_6} p_8$	SHC:EGF:EGFR
11b	$p_{11} \xrightarrow{t_{13}} p_{10} \xrightarrow{t_{12}} p_7 \xrightarrow{t_{16}} p_{11}$	SHC:P'ase(1)
5b	$p_{12} \xrightarrow{t_{20}} p_{16} \xrightarrow{t_{21}} p_{17} \xrightarrow{t_{24}} p_{12}$	GDP:Ras:SHC-P:SOS:GRB2
7b	$p_{15} \xrightarrow{t_{30}} p_{19} \xrightarrow{t_{27}} p_{18} \xrightarrow{t_{26}} p_{15}$	GDP:Ras:RasGAP
3b	$p_{25} \xrightarrow{t_{36}} p_{27} \xrightarrow{t_{37}} p_{28} \xrightarrow{t_{40}} p_{25}$	SOS:Erk-PP
2b	$p_{29} \xrightarrow{t_{42}} p_{26} \xrightarrow{t_{46}} p_{30} \xrightarrow{t_{43}} p_{29}$	SOS:P'ase(2)
18b	$p_8 \xrightarrow{t_7} p_9 \xrightarrow{t_{10}} p_6 \xrightarrow{t_{16}} p_{11} \xrightarrow{t_{13}} p_{10} \xrightarrow{t_{12}} p_5 \xrightarrow{t_6} p_8$	SHC
15b	$p_{13} \xrightarrow{t_{20}} p_{16} \xrightarrow{t_{21}} p_{17} \xrightarrow{t_{24}} p_{14} \xrightarrow{t_{30}} p_{19} \xrightarrow{t_{27}} p_{18} \xrightarrow{t_{26}} p_{13}$	GDP:Ras
14b	$p_{29} \xrightarrow{t_{42}} p_{23} \xrightarrow{t_{36}} p_{27} \xrightarrow{t_{37}} p_{28} \xrightarrow{t_{40}} p_{24} \xrightarrow{t_{46}} p_{30} \xrightarrow{t_{43}} p_{29}$	SOS

TABLE 9. The labeled pathway listing of the nine positive forward-only minimal cycles. These pathways were listed previously in Table 6 and retain that enumeration.

It is interesting to note that six of the thirty places are not contained in the nine positive forward-only minimal cycles and they are p_1 [EGF], p_2 [EGFR], p_4 [EGF:EGFR], p_{20} [SOS:GRB2], p_{21} [GRB2] and p_{22} [SOS*:GRB2]. These biochemical species are critical to the functioning of this compartment in that they act as either sources or sinks to the network. Note that p_1, p_2 and p_4 are places with low probabilities of occurrence as they are involved with initiation of the network and loss of the EGF:EGFR complex through internalization. They appear in pathways only in combinations of both forward and backward pathways, and they specifically occur in two-cycles. This observation will be relevant to the discussion on S-supports.

Analyses of the forward-only pathways shows flux (information content) transferring from Compartment A to Compartment F. Therefore, the concentration of species within this subreaction pool is a limiting factor for the productivity of the network as a whole. There is not a positive forward-only cycle throughout Compartment A of the EGFR network. Compartment A acts as the input to the entire EGFR network with initiation of EGF binding to EGFR. However, Compartment A is a subnetwork of the EGFR network into which Compartments H feeds back biochemical information, and Compartment A feeds biochemical information into Compartments F and H. Compartment A also shares information with Compartment B (GTP:Ras and GDP:Ras) and Compartment E (Erk-PP).

S- and T-Invariant Duals. Incidence matrix representations define a framework for determining the combinatorial invariants that characterize both the flow and conditional regulation of information in an operational system modeled by a Petri net. The net invariants can be used to identify the existence of bi-stable switches in the molecular network.

S-invariants and S-supports. Beginning with the connectivity incidence matrix transpose N^t, a complete list of S-supports of the corresponding oriented matroid and the invariant vector associated with each support are determined. First, the basis over the rationals $\mathbb{Q}$ is found using Gaussian elimination, resulting in a basis which is in row-reduced echelon form. This form looks like the first few positions are rather random, but the rest of the matrix is a square identity. The signed supports of this basis are determined by the procedure described in Oliveira *et al.* (2001). The oriented matroid is built by closing off the listing of signed supports under the three oriented matroid operations listed in Oliveira *et al.* (2001) Definition 23. The algorithm works as follows: Take the signed supports of the basis and apply the three operations. If a new signed support is generated, add it to the listing. This process is performed until no new signed supports are generated. Thereafter, a search may be performed to determine the positive-only minimal supports, *etc*.

The S-supports represent those vectors satisfying equation (1). They represent a listing of the places and directions through which information must pass in conservation form. Each S-invariant vector v_n is of dimension 30×1. A basis for the set of S-invariants, $\mathcal{N}(N^t)$, is given by Table 11.

There were over 400,000 distinct S-supports identified. The distribution of S-support cycle lengths versus number of occurrences of that length has an emerging binomial-like distribution which is shifted toward the longer length cycles (see Figure 11). This observation of the higher-end group is consistent with the higher-end distribution of minimal cycles (Figure 8). Further, this same distribution can be seen within the clusters of S-supports of a given length, when broken down by type, as in Figure 12. The types are specified by a three-tuple: <number of non-zero entries in the support vector, number with positive value, number with negative value>. Figure 13 provides an enlarged view of the binomial-like subdistribution within the binomial-like overall distribution, which reinforces the argument that these networks contain a tremendous amount of symmetry and pattern despite their inherent interacting complexities. Figure 13 depicts the detailed information for the S-supports of length 21, which is the peak cluster shown in Figure 11.

The existence of an underlying binomial distribution is anticipated since we are enumerating the subsets of the $n-$sets of the network's place and transition vertices. From this we obtain binomial posets of size 2^n. It follows that a binomial generating function is the canonical generator of the distributions of place and transition occurrences in the network. From this observation, we immediately obtain that there are $\binom{n}{k}$ arrangements, since we are counting the occurrences of subsets of $n-$sets each with given cardinality k. We also observe that the distribution favors the existence of the larger subsets, since we have placed an additional condition (restriction), which requires that a candidate subset must generate a cycle in the network if it is to be included in the enumeration. This constraint eliminates many of the smaller subsets, as shown in Figures 11) and 12, resulting in the shift of the distribution to the longer length vectors. In contrast, the distribution for a given cycle length is symmetric, as shown in Figure 13. Further, we observe that as the network model goes from discrete to continuous, as n goes to infinity, the binomial distributions converge in the limit to Gaussian distributions. The existence of the underlying binomial distributions implies that the discrete sample space, that the network employs to self-organize, is generated by a Bernoulli process. In the stochastic limit for the continuum kinetic model

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$	$\mathbf{v}_7$	$\mathbf{v}_8$	$\mathbf{v}_9$	$\mathbf{v}_{10}$
p_1	0	1	0	−1	−1	1	1	0	0	0
p_2	−1	1	0	−1	−1	1	0	0	0	0
p_3	1	0	0	0	0	0	0	0	0	0
p_4	0	1	0	−1	−1	1	1	0	0	0
p_5	0	0	0	1	1	−1	−1	0	0	0
p_6	0	1	0	0	0	0	0	0	0	0
p_7	0	1	0	0	0	0	0	0	0	0
p_8	0	0	0	1	1	−1	−1	0	0	0
p_9	0	0	1	0	0	0	0	0	0	0
p_{10}	0	0	1	−1	−1	1	1	0	0	0
p_{11}	0	0	1	0	0	0	0	0	0	0
p_{12}	0	0	0	1	1	−1	0	0	0	0
p_{13}	0	0	0	0	0	0	1	0	0	0
p_{14}	0	0	0	0	−1	1	0	0	0	0
p_{15}	0	0	0	1	0	0	0	0	0	0
p_{16}	0	0	0	1	0	0	0	0	0	0
p_{17}	0	0	0	0	−1	1	0	0	0	0
p_{18}	0	0	0	0	0	1	0	0	0	0
p_{19}	0	0	0	0	1	0	0	0	0	0
p_{20}	0	0	0	0	0	1	0	0	0	0
p_{21}	0	0	0	0	0	0	1	0	1	−1
p_{22}	0	0	0	0	0	0	0	0	−1	1
p_{23}	0	0	0	0	0	0	1	0	0	0
p_{24}	0	0	0	0	0	0	0	0	−1	1
p_{25}	0	0	0	0	0	0	0	1	1	−1
p_{26}	0	0	0	0	0	0	0	1	0	0
p_{27}	0	0	0	0	0	0	0	1	0	0
p_{28}	0	0	0	0	0	0	0	0	0	1
p_{29}	0	0	0	0	0	0	0	0	1	0
p_{30}	0	0	0	0	0	0	0	0	0	1

TABLE 11. A basis of dimension ten for the space of S-invariants.

of the network, the discrete Bernoulli process passes to a Poisson process. Finally, we observe that this passage indicates that the network lives in a class of fundamental queueing networks, where these networks are characterized by Markov processes.

Of the 400411 distinct S-supports, only 295 of them are positive-only. Further, of the 295 positive-only S-supports, only ten are minimal. Here, the minimal listing means the smallest number of vectors such that each place is visited at least once. Recall in the discussion of the nine minimal forward-only pathways that it was noted that some of the places were not visited in the listing. Specifically, the sources and the sinks of the network were not contained in forward only pathways. Therefore, it is no surprise that the listing of nine minimal forward-only pathways is not identical to the ten minimal positive-only S-supports. However, they are related. The minimal positive S-supports enumerate the hypercycles, which are provably obtained as linear combinations of positive forward-only minimal cycles.

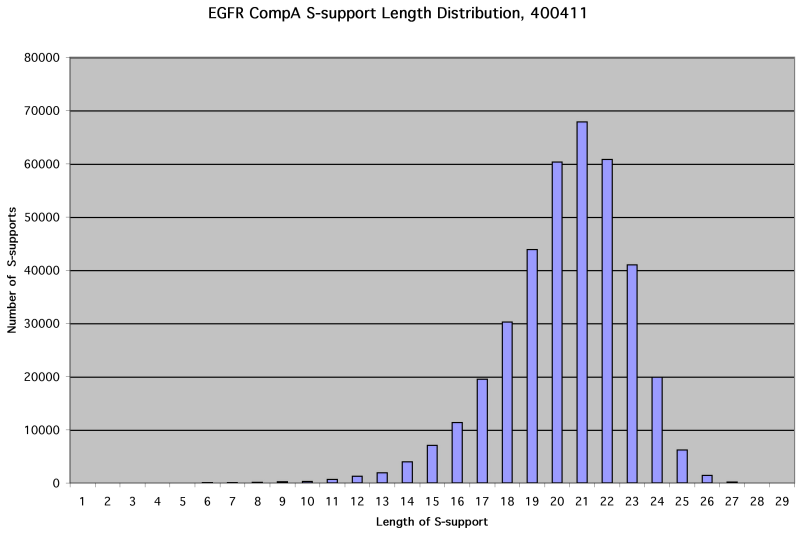


FIGURE 11. Discrete distribution of the number of S-support cycles of a given length. For example, there are 67,860 S-supports of length 21.

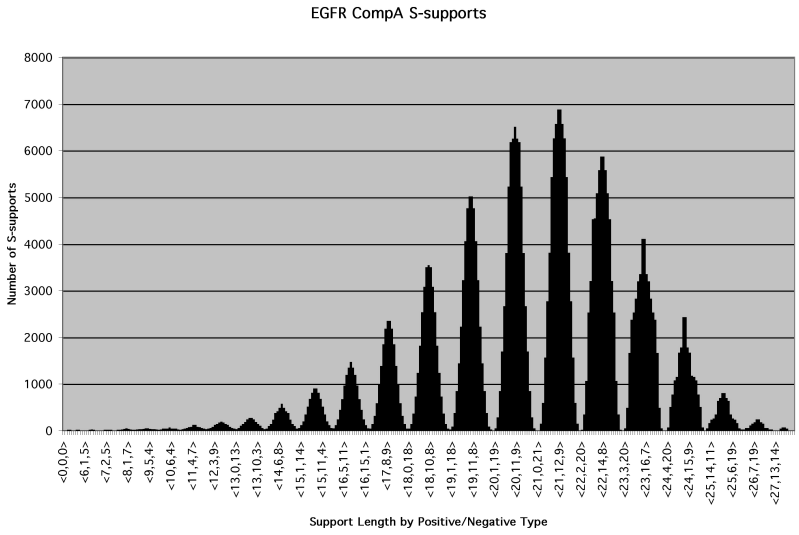


FIGURE 12. Discrete distribution of the number of S-support cycles of a given length, broken out into the individual types denoted <number of non-zero entries, number of positive values, number of negative values>.

The ten minimal positive-only S-supports may be represented in terms of vectors or as sets of places containing unit values only. Note that the order of the places is irrelevant. These ten S-supports are presented in Table 12.

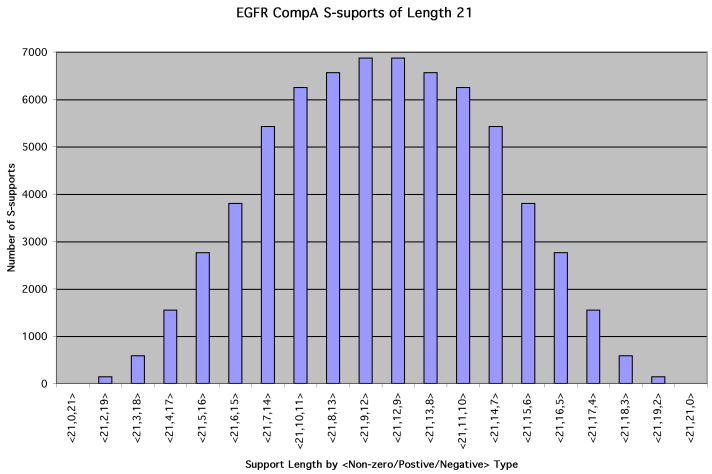


FIGURE 13. Discrete distribution of the number of S-support cycles of length 21.

S-support No.	S-support Specification	S-support No.	S-support Specification
1c	p_7, p_{10}, p_{11}	6c	$p_5, p_6, p_8, p_9, p_{10}, p_{11}, p_{12}, p_{16}, p_{17}$
2c	p_{25}, p_{27}, p_{28}	7c	p_1, p_3, p_4, p_8, p_9
3c	p_{15}, p_{18}, p_{19}	8c	p_2, p_3, p_4, p_8, p_9
4c	p_{26}, p_{29}, p_{30}	9c	$p_{12}, p_{16}, p_{17}, p_{20}, p_{21}, p_{22}$
5c	$p_{13}, p_{14}, p_{16}, p_{17}, p_{18}, p_{19}$	10c	$p_{12}, p_{16}, p_{17}, p_{20}, p_{22}, p_{23}, p_{24}, p_{27}, p_{28}, p_{29}, p_{30}$

TABLE 12. The ten minimal positive-only S-supports as represented as a list of places whose entries are unitary.

S-support $1c$ contains the same places as pathway $11b$; S-support $2c$ contains the same places as pathway $3b$; S-support $3c$ contains the same places as pathway $7b$; S-support $4c$ contains the same places as pathway $2b$; and S-support $5c$ contains the same places as pathway $15b$. However, in order to include all of the places while listing vectors which satisfy equation (1) with unary values only, the remaining four S-supports must be augmented by two-cycles. Therefore, S-support $7c$ contains the same places as pathways $10b+19a+21a$; S-support $7c$ contains the same places as pathways $10b + 20a + 21a$; S-support $9c$ contains the same places as pathways $5b + 38a + 29a + 31a$; and finally, S-support $10c$ contains the same places as pathways $14b + 5b + 38a + 32a$. All nine of the positive forward-only minimal cycles presented in Table 6 and repeated in Table 9 are contained in the ten positive-only minimal S-supports, but in order for the vectors to satisfy equation (1) they must include information from either the sources or the sinks of the network.

It is observed that the number of positive-only minimal S-supports is the same as the basis size for the S-invariants; however, the mathematical significance, if any, of this has not yet been resolved. The S-supports of the network are obtained from the S-invariants of the network; where the S-invariants define the canonical conservation law for molecular species reactants and their products. We observe that this conservation law characterizes the flow of the network at equilibrium. The S-supports are a minimal basis subset of the

basis set that spans the S-invariant nullspace. Therefore, each minimal positive (forward) basis corresponds to a minimal cycle that, in turn, canonically corresponds to a local equilibrium flow in the network. This condition is the combinatorial analog of Nash equilibria for a conservative non-cooperative n-person game [Oliveira *et al.* (to be published) and Szép & Forgó (1985)].

T-invariants and T-supports. Beginning with the connectivity incidence matrix, a complete list of T-supports of the oriented matroid and the invariant vector associated with each support can be determined. The T-supports represent those vectors satisfying equation (3). They represent a listing of the transitions and directions through which information must pass in conservation form. Each invariant vector w_m is of dimension 46×1. A basis for the space of T-invariants is given by Table 13.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}	w_{13}	w_{14}	w_{15}	w_{16}	w_{17}	w_{18}	w_{19}	w_{20}	w_{21}	w_{22}	w_{23}	w_{24}	w_{25}	w_{26}
t_1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−1	0	0	0	0	0	0	0	0
t_2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−1	0	0	0	0	0	0	0	0
t_3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	1	1	0	0	0	0	0	0	0
t_5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−1	0	0	0	0	0	0	0	0
t_6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_7	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	−1	0	0	0	0	0	0	0	0
t_9	0	0	0	1	0	0	0	−1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{10}	0	0	0	0	0	0	1	−1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{11}	0	0	0	0	1	0	0	−1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{12}	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{13}	0	0	0	0	0	1	0	1	0	0	0	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{14}	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{15}	0	0	0	0	0	0	0	−1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
t_{16}	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{18}	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{19}	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{20}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{21}	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{22}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	−1	0	1	0	0
t_{23}	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{24}	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	−1	0	0
t_{25}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{26}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−1	0	1	0	0
t_{27}	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{28}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	−1	0	0
t_{29}	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
t_{30}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
t_{31}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
t_{32}	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
t_{33}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
t_{34}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
t_{35}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
t_{36}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
t_{37}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
t_{38}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	−1	0	0
t_{39}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
t_{40}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
t_{41}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
t_{42}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
t_{43}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
t_{44}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
t_{45}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
t_{46}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

TABLE 13. A basis of dimension 26 for the space of T-invariants.

4. Conclusion

We have performed a Petri net analysis of the initial subnetwork in the EGFR signaling path. The Petri net shows how information is passed through the network based only on connectivity assuming $K_{eq} = 1$ with no regard for the value of the rate constants. Analysis of the network shows that complex formation and internalization of the EGF:EGFR complex are key regulators for the network, *i.e.*, the initiation of the network is not found in many cycles. As compared to other networks, the frequency of transition pairs is not very high, reaching a maximum of $< 15\%$, which indicates a balanced load sharing of the work performed by the network. The methodology identifies the network pinch points, which, in turn, identify the points of control in the network. Those events which occurred with the greatest frequency in the network are phosphorylation reactions leading to activation of SOS, an effector protein. The production of SOS leads to the signal eventually being transmitted to the kinase cascade, so it is not surprising that it is a key element showing up in a number of cycles and that activation of SOS is so important in the network.

5. Acknowledgements

This research was supported by the Laboratory Directed Research and Development program at the Pacific Northwest National Laboratory (PNNL) and performed in part in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL. The EMSL is funded by the Office of Biological and Environmental Research in the U. S. Department of Energy under Contract DE-AC06-76RLO 1830. The authors would like to thank the ERULF program of the DoE for funding the undergraduate research of Dean W. Gull, Southern Utah University, Cedar City, UT. The authors would like to thank Doug Nordwall for his computer support, and Chad Scherrer and the external reviewers for their invaluable review comments.

Appendix A. Enlarged views of the EGFR Compartment A Subreactions

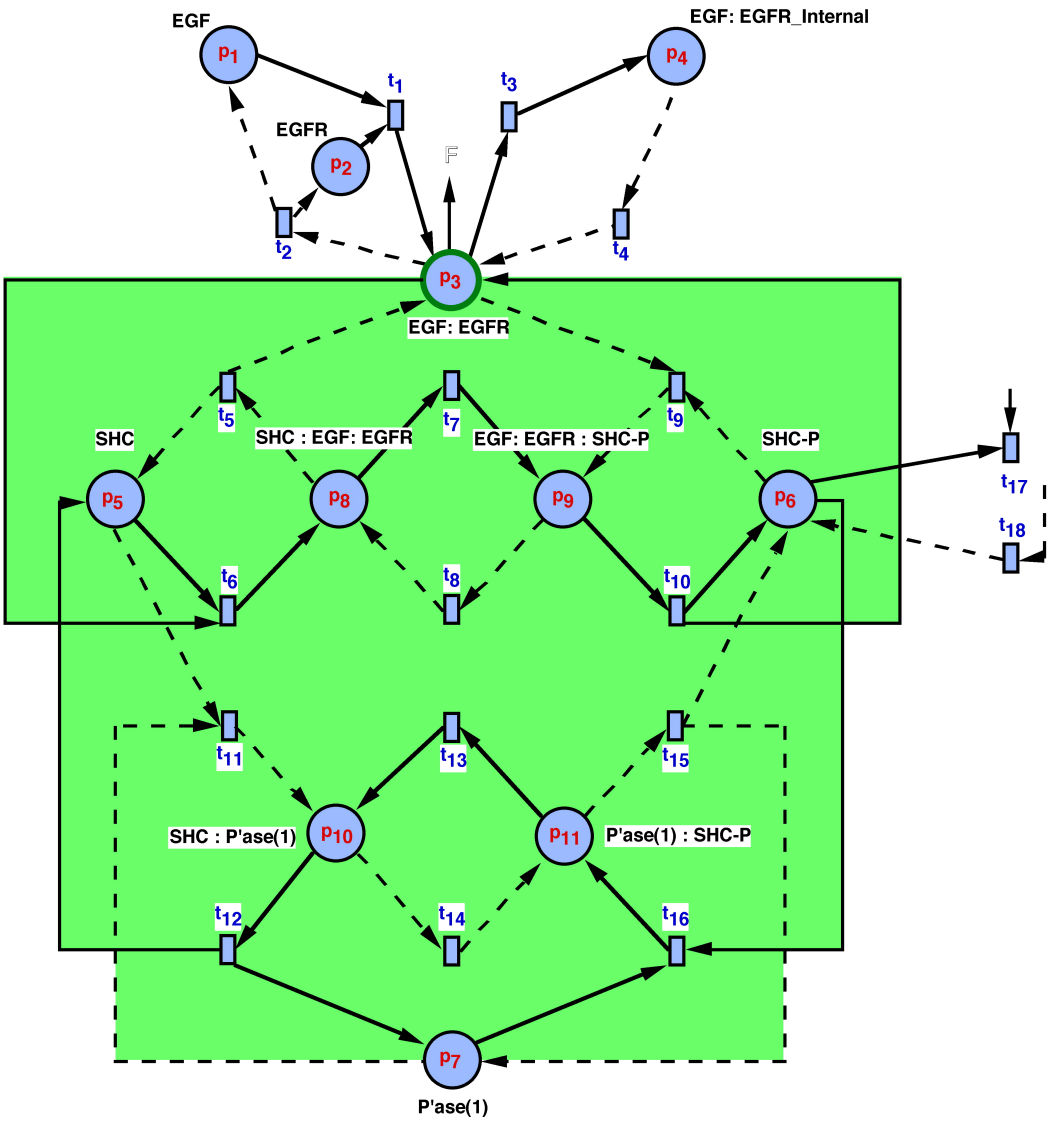


FIGURE 14. SHC Subreaction Petri net of the EGFR Compartment A.

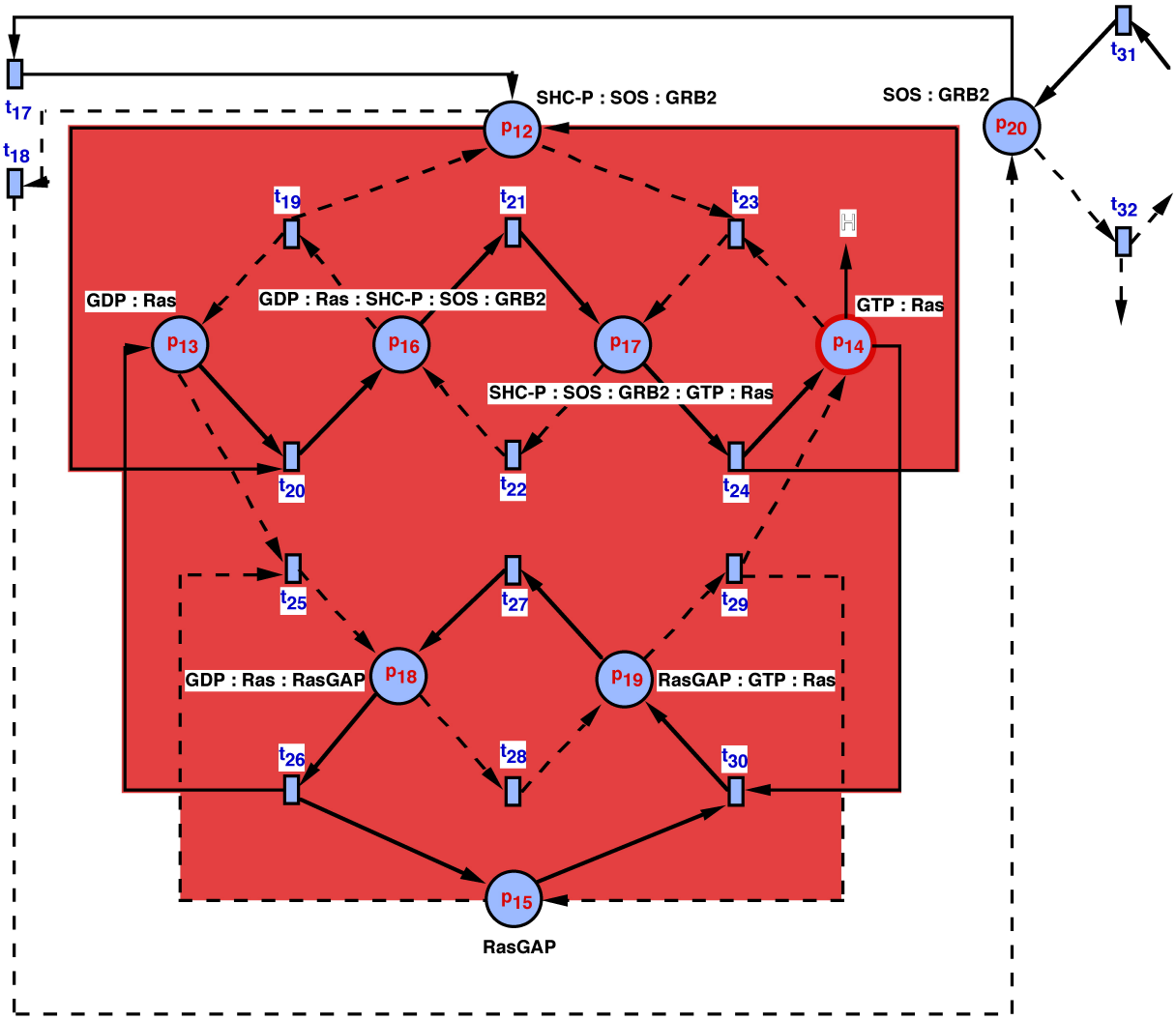


FIGURE 15. GDP:Ras Subreaction of the EGFR Compartment A.

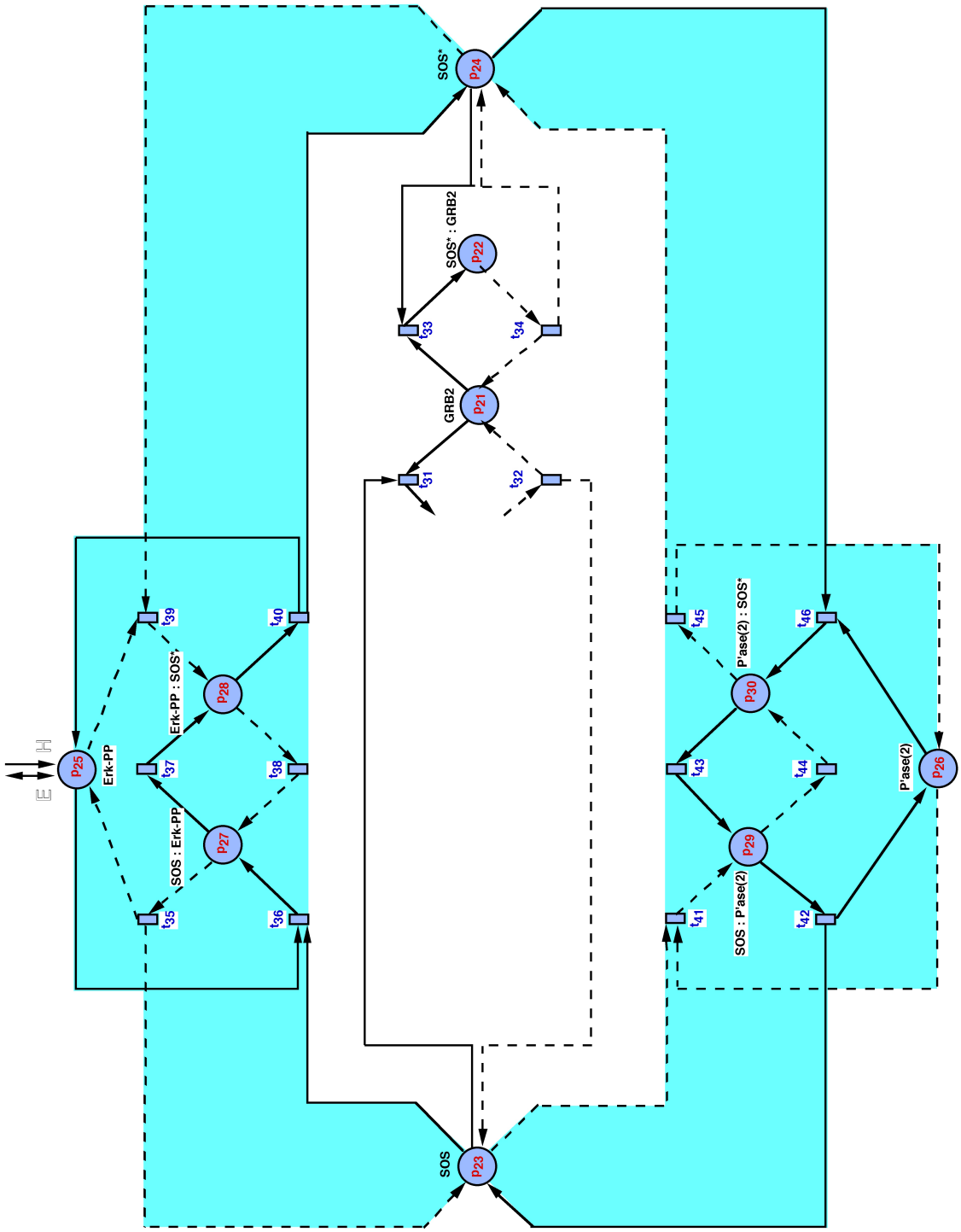


FIGURE 16. SOS Subreaction with GRB2 of the EGFR Compartment A.

REFERENCES

[1] Bachem, A. and W. Kern (1991). *Linear Programming Duality: An Introduction to Oriented Matroids*, Springer-Verlag, New York, NY.

[2] Bajzer, Z., A.C. Myers, and S. Vuk-Pavlovic (1989). *Binding, internalization, and intracellular processing of proteins interacting with recycling receptors a kinetic analysis*, J. Biol. Chem., 264, 13623-13631.

[3] Bhalla, U. S. and R. Iyengar (1999). *Emergent Properties of Networks of Biological Signaling Pathways*, Science, Vol. 283, 15 Jan. 1999, pp. 381-387.

[4] Berge, C. (1973). *Graphs and Hypergraphs*, American Elsevier Publishing Co., New York, NY.

[5] Diestel, R. (2000). *Graph Theory*, 2nd Ed., Springer, New York, NY.

[6] French, A. R. and D. A. Lauffenburger (1997). *Controlling receptor/ligand trafficking: Effects of cellular and molecular properties on endosomal sorting*. Ann. Biomedical Eng. 25, 690-707.

[7] Garey, M. and D. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co., San Francisco, CA.

[8] Gex-Fabry, M. and C. De Lisi (1984). *Receptor-mediated endocytosis: A model and its implications for experimental analysis*, Am. J. Physiol., 247:R768-R779.

[9] Gibbons, A. (1985). *Algorithmic Graph Theory*, Cambridge Univ. Press, New York, NY.

[10] Haugh, J. M., and D. A. Lauffenburger (1998). *Analysis of receptor internalization as a mechanism for modulating signal transduction*, J. Theor. Biol. 195:187-218.

[11] Ideker, T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood (2001). *Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network*,Science, 292, 929-934.

[12] Jones, N. D., L. H. Landweber, and Y. E. Lien (1977). *Complexity of Some Problems in Petri-nets*, Theor. Comp. Sc., North Holland Publishing Co., 4, 277-299.

[13] Kholodenko, B.N., O.V. Demin, G. Moehren, and J.B. Hoek (1999). *Quantification of short term signaling by the epidermal growth factor receptor*, J. Biol. Chem., 274:30169-30181.

[14] Lund, K. A., L. K. Opresko, C. Starbuck, B. J. Walsh, and H. S. Wiley (1990). *Quantitative analysis of the endocytic system involved in hormone-induced receptor internalization*, J. Biol. Chem., 265:15713-15723.

[15] Narayanan, H. (1997). *Submodular Functions and Electrical Networks*, Annals of Discrete Mathematics, North-Holland, Elsevier, New York, NY.

[16] Oliveira, J. S., C. G. Bailey, J. B. Jones-Oliveira, and D. A. Dixon (2001). *An Algebraic-Combinatorial Model for the Identification and Mapping of Biochemical Pathways*, Bull. of Math. Bio., 63, 1163-1196.

[17] Oliveira, J. S., C. G. Bailey, J. B. Jones-Oliveira, D. A. Dixon, D. W. Gull, and M. L. Chandler (2003). *A Computational Model for the Identification of Biochemical Pathways in the Krebs Cycle*, J. of Comp. Bio., Vol. 10, No. 1, 57-82.

[18] Reddy, V. N. , M. N. Liebman and M. L. Mavrovouniotis (1996). *Qualitative Analysis of Biochemical Reaction Systems*, Comput. Biol. Med., 26, 9-24.

[19] Reisig, W. (1985). *Petri Nets, An Introduction*, Brauer, W., G. Rozenberg and A. Salomaa, Editors, Springer-Verlag, New York, NY.

[20] Resat, B. H., J. A. Ewald, D. A. Dixon, and H. S. Wiley (2003). *An Integrated Model of EGF Receptor Trafficking and Signal Transduction*, Biophys. J., accepted 2003.

[21] Schuster, S., D. A. Fell and T. Dandekar (2000). *A General Definition of Metabolic Pathways Useful for Systematic Organization and Analysis of Complex Metabolic Networks*, Nature Biotechnology, Vol. 18, March 2000, pp. 326-332.

[22] Schuster, S., C. Hilgetag, J. H. Woods and D. A. Fell (2002a). *Reaction Routes in Biochemical Reaction Systems: Algebraic Properties, Validated Calculation Procedure and Example from Nucleotide Metabolism*, J. Math. Biol., Vol. 45, 153-181 (2002).

[23] Schuster, S., T. Pfeiffer, F. Moldenhauer, I. Koch and T. Dandekar (2002b). *Exploring the Pathway Structure of Metabolism: Decomposition into Subnetworks and Application to Mycoplasma Pneumoniae*, BioInformatics, Vol. 18, No. 2, 351-361 (2002).

[24] Szép, J. and F. Forgó (1985). *Introduction to the Theory of Games*, D. Reidel Publishing Co., Kluwer Academic Publishers Group, Boston, MA.

[25] Sipser, M. (1997). *Introduction to the Theory of Computation*, PWS Publishing Co., An International Thomson Publishing Co., Boston, MA.

[26] Sorkin, A., C. Waters, K.A. Overholser, and G. Carpenter (1991). *Multiple autophosphorylation site mutations of the epidermal growth-factor receptor analysis of kinase-activity and endocytosis*, J. Biol. Chem. 266:8355-8362.

[27] Wiley, H. S. and D. D. Cunningham (1981). *A steady-state model for analyzing the cellular-binding internalization and degradation of polypeptide ligands*, Cell, 25:433-440.

RADIOLOGICAL AND CHEMICAL SCIENCES GROUP NATIONAL SECURITY SCIENCE AND TECHNOLOGY DIVISION, NATIONAL SECURITY DIRECTORATE, PACIFIC NORTHWEST NATIONAL LABORATORIES, RICHLAND, WASHINGTON, U.S.A.
E-mail address: jso@pnl.gov

INTERNATIONAL TECHNOLOGIES ASSESSMENTS GROUP, NATIONAL SECURITY SCIENCE AND TECHNOLOGY DIVISION, NATIONAL SECURITY DIRECTORATE, PACIFIC NORTHWEST NATIONAL LABORATORIES, RICHLAND, WASHINGTON, U.S.A.
E-mail address: jjo@pnl.gov

BATTELLE FELLOW, FUNDAMENTAL SCIENCES DIRECTORATE, PACIFIC NORTHWEST NATIONAL LABORATORIES, RICHLAND, WASHINGTON, U.S.A.
E-mail address: David.Dixon@pnl.gov

SCHOOL OF MATHEMATICAL AND COMPUTING SCIENCES, VICTORIA UNIVERSITY OF WELLINGTON, WELLINGTON, NEW ZEALAND
E-mail address: Colin.Bailey@vuw.ac.nz

APPLIED MATHEMATICS GROUP, COMPUTATIONAL SCIENCES AND MATHEMATICS DIVISION, FUNDAMENTAL SCIENCES DIRECTORATE, PACIFIC NORTHWEST NATIONAL LABORATORIES, RICHLAND, WASHINGTON, U.S.A.
E-mail address: dean_gull@yahoo.com