A Query-Based Technique for Interpreting Reachable Sets for Hybrid Automaton Models of Protein Feedback Signaling

Ronojoy Ghosh and Claire Tomlin

Abstract— Intercellular protein signaling networks involving feedback have been modeled successfully using hybrid automata. Algorithms have also been developed to compute symbolic reachable sets for the biologically observed steady states of the models. However, the complexity and size of the reachable sets have made it difficult to interpret them. This paper presents a systematic and efficient algorithm for interpreting large reachable sets by treating them as a database, and using logical queries that encode biologically meaningful properties, to elucidate specific behavior of the modeled system. The results of several query computations for a four cell Delta-Notch protein signaling network model are also given, to illustrate the technique and its utility.

I. INTRODUCTION

In the recent past, there have been many research projects geared toward modeling and analyzing feedback-based biological systems using methods from control theory [1], [2]. The authors' research focuses on a specific biological mechanism known as intercellular protein signaling. Found in all multicellular organisms from an early embryo stage, intercellular signaling is a feedback network which interrelates the fate of a single cell and its neighbors. In particular, the authors have modeled and analyzed two signaling pathways active during development: (1) The Delta-Notch signaling mechanism, responsible for pattern formation in many different biological systems, such as the emergence of ciliated cells in *Xenopus* embryonic skin [3], and (2) the planar cell polarity (PCP) signaling mechanism that controls the position of *trichome*, or hair, growth in *Drosophila melanogaster* pupal wing cell arrays [4].

Both systems have been modeled by the authors [5], [6] using the mathematical framework of hybrid automaton theory. The physical properties of the processes involved in cell signaling make it well-suited for hybrid systems modeling, using a two-level hierarchical framework. At a low level, the spatio-temporal evolution of protein concentration dynamics are governed by linear differential equations; the second tier comprises the logic network that controls the continuous dynamics through discrete switches. The variables of the logic network are the protein concentrations from the lower level. Thus, the two levels are strongly interconnected, yet the model as a whole is more amenable to analysis.

In addition, the authors have developed an algorithm [5] to analytically derive constraints on the system kinetic parameters necessary for biologically feasible steady states to exist, in multi-stable systems. The parameters can be symbolic constants, i.e. numerical values need not be assigned to them. The constraints are expressed in terms of ratios of these parameters and switching thresholds. An abstraction algorithm has also been developed [7] to compute backward reachable sets of states, which are expressed in terms of the system's symbolic parameters. These backward reachable sets, when computed for the steady states of the system, represent regions of state space that are guaranteed to converge to one particular steady state or the other. Biologically, this implies that one can identify sets of initial protein concentrations from which a biologically interesting steady state can be achieved. For example, the dosage of a particular medicine and its effectiveness in achieving a cure (steady state) could be determined using this analysis. Since these reachable sets are large and complex, it is a challenge to interpret them in a biologically meaningful way.

This paper presents a systematic and efficient method of interpreting the reachable sets computed using the authors' abstraction algorithm. The key is to use the reachable set as a *database*, and *query* it for biologically meaningful properties. For example, a valid query would be "if concentration of protein A is greater than protein B in neighboring cells, which steady state pattern will the Delta-Notch lateral inhibition network display?" The importance of the method lies in the fact that *specific* questions can be asked about the behavior of the system, and the results will indicate automatically whether additional constraints are necessary for a biologically interesting steady state event to happen. Computationally, it involves efficiently checking each reachable set to see whether it satisfies a particular query, constructed using logical expressions involving protein concentrations. In the following sections, the key mathematical concepts used are defined first, followed by a brief description of the Delta-Notch hybrid automaton model, and the reachability computation for a four cell Delta-Notch signaling network. This is followed by a formal algorithm and discussion of query based interpretation, and the results of several example queries for the four cell network reachable sets.

II. REACHABILITY COMPUTATION

A. Definitions

Definition 1: A piecewise affine hybrid automaton, $H = (Q, X, \Sigma, Init, f, Inv, R)$, is defined such that

1) $Q = \{q_1, q_2, \ldots, q_m\}$ is the set of discrete states;

2) $X \subset \Re^n$ is the set of continuous state variables;
3) $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ is the set of discrete inputs;
4) $Init = Q_0 \times X_0$ is the set of initial conditions;
5) $f(q, x) = A_q x + b_q$ is the continuous vector field associated with each discrete state, where $A_q \in \Re^{n \times n}$ is a diagonal matrix, and $b_q \in \Re^n$;
6) $Inv(q) = (\bigwedge_i (p_i < 0)) \wedge (\bigwedge_j (p_j = 0)) \wedge (\bigwedge_k (p_k > 0)) \wedge (\bigwedge_l (p_l \leq 0)) \wedge (\bigwedge_m (p_m \geq 0))$, where $p_i \in P_{lt}(q), p_j \in P_{eq}(q), p_k \in P_{gt}(q), p_l \in P_{le}(q), p_m \in P_{ge}(q)$, is the invariant defining each discrete state, where $p_{()} : X \times \Sigma \to \Re$ is a polynomial;
7) $R : Q \times X \times \Sigma \to 2^{Q \times X}$ is the transition map.

This is the class of hybrid automata used by the authors to model protein signaling networks such as the Delta-Notch network described in Section II-C [5]. The state transition matrix A_q is restricted to be diagonal with real eigenvalues. However, the elements of A_q and b_q are free to be symbolic. Constraints may be imposed on these symbolic constants to restrict the behavior of the model. The polynomials defining the invariant of each hybrid state, can be separated into five classes: $P_{lt}(q)$, $P_{eq}(q)$, $P_{gt}(q)$, $P_{le}(q)$, and $P_{ge}(q)$, according to their signs in the state. For example, all polynomials $p_i \in P_{lt}(q)$ are negative in state q, and similar definitions hold for the other classes $P_{eq}(q)$, etc. $P_{lt}(q), P_{eq}(q), \ldots, P_{ge}(q)$ are mutually disjoint, and $\forall q, P_{lt}(q) \cup P_{eq}(q) \cup P_{gt}(q) \cup P_{le}(q) \cup P_{ge}(q)$ is invariant. This implies that the polynomials defining each state are identical, their sign alone varies from state to state.

Definition 2: A discrete transition system, $T = (Q, \Sigma, \to, Q_0, Q_F)$, is defined such that
1) $Q = \{q_1, q_2, \ldots, q_n\}$ is a set of states;
2) $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ is a set of events;
3) $\to \subseteq Q \times \Sigma \times Q$ is a transition relation;
4) $Q_0 \subseteq Q$ is the set of initial states;
5) $Q_F \subseteq Q$ is the set of final states.

A dual representation of a finite transition system is an adjacency matrix $A \in \{0,1\}^{n \times n}$, where $i \in 1, 2, \ldots, n$ represents a discrete state. In the adjacency matrix, $a_{i,j} \in A : a = 1$ means that a transition $q_i \to q_j$ exists, and $a_{i,j} = 0$ means no transition exists from q_i to q_j. The final states, $q \in Q_F$, of the transition system are states that have no transitions out of them. The adjacency matrix can be used to efficiently compute reachable states from a final state of the system.

Definition 3: A state $\hat{q} \in Q$, of a transition system (that may be a hybrid automaton H or a discrete transition system T), is said to be *reachable* from another state q if there exists a finite sequence of transitions $q \xrightarrow{\sigma_1} q_1 \xrightarrow{\sigma_2} \cdots \xrightarrow{\sigma_N} \hat{q}$, from initial state q to state $\hat{q}$. The state q is said to be *backward reachable* from the state $\hat{q}$. A region $P \subseteq Q$ of the hybrid automaton H or discrete transition system T, is *backward reachable* from a state $q \in Q$, if there exist sequences of transitions leading from P to q.

Since computing reachability on hybrid automaton models directly is difficult, the authors have proposed and implemented an abstraction process [7], which converts the hybrid automaton into a discrete transition system that preserves the transition structure while abstracting away the continuous dynamics. The reachability computation is then performed on the discrete abstraction. When the hybrid automaton H is abstracted to the discrete transition system T, its equilibria are abstracted to the states in the set of final states Q_F of T. For a final state $q \in Q_F$, the region P is *uniquely* backward reachable from state q if all sequences of transitions from P terminate only in the state q. The regions of attraction of the steady states of the hybrid automaton correspond to the unique backwards reachable sets for the final states of the abstracted discrete transition system.

B. Reachable Sets: Structure

The backward reachable sets are computed from the adjacency matrix A of the abstracted discrete transition system. This is done by tracing all possible sequences of transitions backward from the final states. The traces are continued until the reachable set becomes invariant. The reachable set thus computed is expressed as the *union* of all the states backward reachable from a particular final state. Recall that from the definition of the hybrid automaton model H, each state is defined by the conjunction of several inequalities involving symbolic polynomials. Therefore, the logical structure of the statement expressing the reachable sets is the disjunctive normal form.

Definition 4: A statement, or formula, is in *disjunctive normal form* if it is a disjunction (sequence of ORs) consisting of one or more disjuncts, each of which is a conjunction (AND) of one or more literals (i.e., logical statements and negations of logical statements [8]).

The reachable set thus takes the form: $Reach = Inv(q_1) \vee Inv(q_2) \vee \ldots \vee Inv(q_k)$, where each $Inv(q_i)$ is a conjunction of polynomial inequalities. The number of states in each reachable set can be large, and it increases with the dimensionality of the transition system, making direct interpretation of the set correspondingly difficult.

C. Delta-Notch Lateral Inhibition Model

The Delta-Notch signaling pathway has been modeled in detail by the authors [5]. To model the lateral inhibition Delta-Notch feedback network, experimentally observed rules governing the biological phenomenon have to be implemented. First, cells have to be in direct contact for
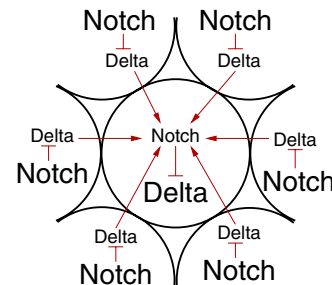

Fig. 1. Influence diagram for Delta-Notch protein signaling network.

Delta-Notch signaling to occur. This implies that a cell is directly affected by, and directly affects in turn, only immediate neighbors. Second, Notch production is turned on by high Delta levels in the immediate neighborhood of the cell and Delta production is switched on by low Notch concentrations in the same cell. Third, at steady state, a cell with high Delta levels must have low Notch level and vice versa. In the model, the cells are assumed to be hexagonal close packed, i.e. each cell has six neighbors in contact with it. An influence diagram showing the signaling network is given in Fig. 1.

Each biological cell is modeled as a four state piecewise affine hybrid automaton. The four states capture the property that Notch and Delta protein production can be individually switched on or off. The formal definition of the one cell hybrid automaton is given by:

$$
\begin{aligned}
H_{one\_cell} &= (Q, X, \Sigma, Init, f, Inv, R) \\
Q &= \{q_1, q_2, q_3, q_4\} \\
X &= (x_1, x_2)^T \in \Re^2 \\
\Sigma &= \left\{ u_N = \sum_{i=1}^{6} x_{Delta,i} \right\} \\
Init &= Q \times \left\{ X \subset \Re^2 : x_1, x_2 > 0 \right\} \\
f(q, x) &= \begin{cases}
[-\lambda_D x_1; -\lambda_N x_2]^T & \text{if } q = q_1 \\
[R_D - \lambda_D x_1; -\lambda_N x_2]^T & \text{if } q = q_2 \\
[-\lambda_D x_1; R_N - \lambda_N x_2]^T & \text{if } q = q_3 \\
[R_D - \lambda_D x_1; R_N - \lambda_N x_2]^T & \text{if } q = q_4
\end{cases} \\
Inv &= \{q_1, \{-x_2 < h_D, u_N < h_N\}\} \\
&\cup \{q_2, \{-x_2 \geq h_D, u_N < h_N\}\} \\
&\cup \{q_3, \{-x_2 < h_D, u_N \geq h_N\}\} \\
&\cup \{q_4, \{-x_2 \geq h_D, u_N \geq h_N\}\}
\end{aligned}
$$

where, x_1 and x_2: Delta and Notch protein concentrations, respectively, in a cell; $x_{Delta,i}$: Delta protein concentration in i^{th} neighboring cell; λ_D and λ_N: Delta and Notch protein decay constants respectively; R_D and R_N: constant Delta and Notch protein production rates, respectively; h_D and h_N: switching thresholds for Delta and Notch protein production, respectively. The switching thresholds h_D and h_N are unknown and possible ranges for them have been derived by the authors [5] using equilibrium analysis. Modeling multiple cell networks involves composing several single cell hybrid automata, with coupling through the input u_N that is the sum of Delta protein concentrations of all neighboring cells.

D. Four Cell Delta-Notch Automaton

The four cell Delta-Notch hybrid model has 8 continuous state variables $x_1, x_2, \ldots, x_8$, and 256 discrete states. The variables x_1, x_3, x_5 and x_7 denote the Delta protein levels and x_2, x_4, x_6 and x_8 represent the Notch protein levels in cells 1, 2, 3 and 4, respectively (Fig. 2). The system has periodic boundary conditions, which implies that the last row and column of cells are wrapped around to provide boundary conditions (protein concentrations) to the cells in
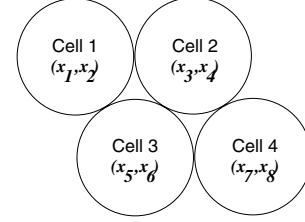

Fig. 2. Layout of four cell Delta-Notch network showing the variables associated with each cell.

the first row and column of the network, and similarly the first row and column are also extended to give boundary protein concentrations to cells in the last row and column.

The network has four possible steady states that are biologically feasible, as shown listed in Table I. Fig. 3 displays the four equilibria; a shaded cell in each case represents a cell with high Delta protein and low Notch protein concentration, and an unshaded cell has low Delta protein and high Notch protein concentration at steady state.

TABLE I
STEADY STATE PROTEIN CONCENTRATIONS IN FOUR CELL
DELTA-NOTCH NETWORK.

Equilibrium 1 $\left(x_1 = 0, x_2 = \frac{R_N}{\lambda_N}, x_3 = 0, x_4 = \frac{R_N}{\lambda_N}, \right.$
$\left. x_5 = 0, x_6 = \frac{R_N}{\lambda_N}, x_7 = \frac{R_D}{\lambda_D}, x_8 = 0 \right)$

Equilibrium 2 $\left(x_1 = 0, x_2 = \frac{R_N}{\lambda_N}, x_3 = 0, x_4 = \frac{R_N}{\lambda_N}, \right.$
$\left. x_5 = \frac{R_D}{\lambda_D}, x_6 = 0, x_7 = 0, x_8 = \frac{R_N}{\lambda_N} \right)$

Equilibrium 3 $\left(x_1 = 0, x_2 = \frac{R_N}{\lambda_N}, x_3 = \frac{R_D}{\lambda_D}, x_4 = 0, \right.$
$\left. x_5 = 0, x_6 = \frac{R_N}{\lambda_N}, x_7 = 0, x_8 = \frac{R_N}{\lambda_N} \right)$

Equilibrium 4 $\left(x_1 = \frac{R_D}{\lambda_D}, x_2 = 0, x_3 = 0, x_4 = \frac{R_N}{\lambda_N}, \right.$
$\left. x_5 = 0, x_6 = \frac{R_N}{\lambda_N}, x_7 = 0, x_8 = \frac{R_N}{\lambda_N} \right)$

The discrete abstraction has 6561 states, and the unique backwards reachable sets for equilibrium 1, 2, 3, and 4 contain 780, 1565, 1566 and 920 states, respectively. The reachability computation for the remaining states is undecidable, because they cannot be subpartitioned using the refinement procedure and lead to more than one steady state. Each state is defined by 8 inequality invariants, hence the size of the reachable sets are quite large and difficult to interpret by inspection.

III. REACHABILITY: QUERY BASED INTERPRETATION

In this section, the authors propose a feasible alternative to direct interpretation of reachability analysis results. The key idea is to *query* the computed reachable sets with logical expressions that encode biologically interesting conditions involving protein concentrations. By verifying whether those expressions are true in each unique backwards reachable set, the analysis can (a) find the steady states that can be attained from those conditions, and (b) deduce any additional constraints on the protein concentrations that are necessary to reach those steady states.
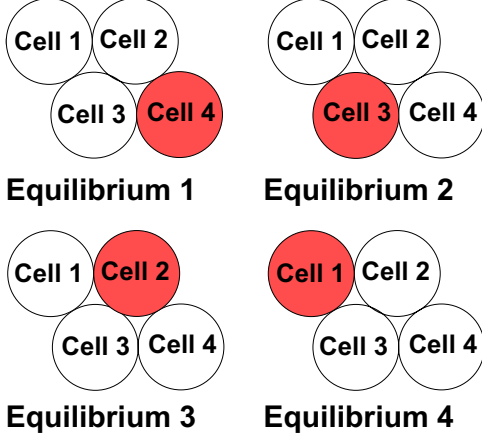
Fig. 3. Biologically consistent steady states of the four cell network with periodic boundary conditions. A shaded cell represents a high steady state concentration of Delta protein and low level of Notch protein, and an unshaded cell has low Delta protein and high Notch protein at steady state.

Algorithm 1: Reach Set Query Algorithm

Input: A set of states, $Reach = Inv(q_1) \vee Inv(q_2) \vee \ldots \vee Inv(q_j)$, backward reachable only from state q_k, and a query statement $Query$

Output: A logical statement $Output$

foreach disjunct $Inv(q_i) \subset Reach$, $i = 1, 2, \ldots j$

   compute $f_i = Query \wedge Inv(q_i)$

return $Output = \bigvee_{i=1}^{j} f_i$

The computation procedure is straightforward, and it takes advantage of the fact that the reachable set is in disjunctive normal form:

1) The query expression is a logical statement that involves protein concentrations. An example of a valid query is: $Query = Conc_a - Conc_b > Conc_c \wedge Conc_d \leq Conc_b$, where $Conc_i$ is concentration of protein species i.

2) The computation attempts to find a simplified truth value for the expression $Query \wedge Reach$. This is an efficient computation because $Reach$ is in disjunctive normal form, $Reach = Inv(q_1) \vee Inv(q_2) \vee \ldots \vee Inv(q_k)$, and hence, through distributivity, $Query \wedge Reach \equiv (Query \wedge Inv(q_1)) \vee (Query \wedge Inv(q_2)) \vee \ldots \vee (Query \wedge Inv(q_k))$. This implies that each of the disjuncts can be simplified separately in an efficient manner, and the results combined.

3) If the resultant expression is *false*, then the query is not valid for that reachable set. If it is *true*, then it always leads to the particular steady state whose backward reachable set was queried. If, the answer returns additional constraints on the protein concentrations, then those constraints will have to be satisfied, in addition to the query expression, for the system to reach the steady state that was queried.

The computational complexity of the query analysis

grows linearly with the number of states in each reachable set. Since each disjunct can be evaluated separately, the computation algorithm is highly parallelizable. The formula simplification can be done by a symbolic decision procedure involving quantifier elimination, the authors implemented it using QEPCAD [9], which is a quantifier elimination program using partial cylindrical algebraic decomposition. In Section IV, the results of several query computations are presented, to demonstrate their use when biologically interesting properties involving protein concentrations can be expressed as logical statements.

IV. RESULTS

All the computations in this section were performed on the unique backwards reachable sets for the four biologically feasible equilibria of the four cell Delta-Notch system, given in Table I. Each query takes around two hours to verify using a 2GHz Pentium 4 workstation with 1GB memory. The following description states and explains each query, and the results for each computation.

Query 1: $x_1 > x_3 \wedge x_1 > x_5 \wedge x_1 > x_7$

This expression encodes the property "Delta protein concentration in cell 1 is higher than Delta concentrations in the other three cells." The query tests the classic lateral inhibition property, i.e. amplification of initial protein concentration differences. The higher Delta protein concentration in cell 1 is expected to suppress Delta protein growth in all its neighbors. Therefore, all initial conditions satisfying query 1 are expected to converge to the biological steady state represented by equilibrium 4.

Query 1 result

Equilibrium 1: Unreachable

Equilibrium 2: Unreachable

Equilibrium 3: Unreachable

Equilibrium 4: $(x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_6 \geq 0 \wedge h_D + x_8 \geq 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_2 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge (h_N - 2x_5 - 2x_3 - 2x_1 \geq 0 \vee h_N - 2x_7 - 2x_3 - 2x_1 \geq 0 \vee h_N - 2x_7 - 2x_5 - 2x_1 \geq 0 \vee (h_D + x_4 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (h_D + x_6 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (h_D + x_2 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (h_D + x_6 > 0 \wedge h_D + x_8 > 0 \wedge h_D + x_4 > 0 \wedge h_D + x_2 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0) \vee (h_D + x_8 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0)))$

The query results confirm that the fundamental lateral inhibitory property of the network is modeled faithfully by the hybrid automaton model. Only the backward reachable set of equilibrium 4 satisfies query 1, the reachable sets for the other three equilibria do not. Furthermore, the result of the query returns additional constraints that are sufficient to guarantee that the system converges to equilibrium 4. For example, the total Delta protein concentration in cells 2, 3 and 4 (Fig. 2) is required to be less than half the switching threshold value for turning on Notch production. Similar conditions hold for the total Delta protein concentrations of cells 1, 3 and 4, and 1, 2 and 4. In biological terms,

this means that there has to be an initial asymmetry in protein concentrations for the lateral inhibition mechanism to amplify. Also, convergence to a particular equilibrium depends on how strong or how weak, relative to the switching threshold, the initial protein concentrations are.

Query 2: $x_1 > x_3 \wedge x_1 > x_5 \wedge x_7 > x_3 \wedge x_7 > x_5$

This query asks a broader question that may result in more than one steady state. Here, the Delta protein concentrations in cells 1 and 4, x_1 and x_7 respectively, are greater than the Delta concentrations in the other two cells, x_1 and x_5. The point of interest here is: which of the steady states are achievable and under what conditions, i.e. what differentiates the reachable sets?

Query 2 results

Equilibrium 1: $(x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_1 \geq 0 \wedge x_8 - x_6 \leq 0 \wedge x_8 - x_2 \leq 0 \wedge x_8 - x_4 \leq 0 \wedge h_D + x_2 \geq 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_6 \geq 0 \wedge h_D + x_8 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0 \wedge ((h_D + x_2 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (h_D + x_4 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (x_7 - x_1 > 0 \wedge h_D + x_8 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (x_7 - x_1 > 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0) \vee (h_D + x_2 > 0 \wedge h_D + x_4 > 0 \wedge h_D + x_6 > 0 \wedge h_D + x_8 < 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0) \vee (h_D + x_6 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0)))$

Equilibrium 2: Unreachable

Equilibrium 3: Unreachable

Equilibrium 4: $(x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_3 > 0 \wedge x_7 - x_5 > 0 \wedge x_7 - x_1 \leq 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_6 \geq 0 \wedge h_D + x_8 \geq 0 \wedge h_D + x_2 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge ((x_7 - x_1 < 0 \wedge h_D + x_4 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_7 - x_1 < 0 \wedge h_D + x_6 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_7 - x_1 < 0 \wedge h_D + x_2 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_7 - x_1 < 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0) \vee (h_D + x_4 > 0 \wedge h_D + x_6 > 0 \wedge h_D + x_8 > 0 \wedge h_D + x_2 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0) \vee (x_7 - x_1 < 0 \wedge h_D + x_8 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0)))$

The results are consistent with the lateral inhibition property. The only equilibria that are guaranteed reachable from the given query condition are 1 and 4. Recall from Table I, that equilibrium 1 represents high Delta concentration only in cell 4, and equilibrium 4 has high Delta protein concentration in cell 1. However, the query results also *automatically* provide the additional conditions that are necessary for the system to reach one or the other. In addition to constraints on the initial Notch protein concentrations, the query results clearly indicate that for equilibrium 1 to be reachable, the Delta protein concentration in cell 4 cannot be less than the Delta concentration in cell 1, i.e. the constraint $x_7 - x_1 \geq 0$ has to be satisfied. The converse condition, $x_7 - x_1 \leq 0$, has to hold for the system to converge to equilibrium 4.

Query 3: $x_3 > x_5 \wedge x_2 < x_8$

Query 3 is very interesting because it proposes differences in protein concentration for both Delta and Notch. The conditions are fairly relaxed; they only stipulate that the Delta concentration in cell 2 is greater than that in cell

3, and Notch levels in cell 1 are less than that in cell 4. The results are somewhat non-intuitive, as discussed below:

Query 3 results

Equilibrium 1: Unreachable

Equilibrium 2: Unreachable

Equilibrium 3: $(x_3 - x_1 \geq 0 \wedge x_4 - x_2 \leq 0 \wedge x_5 - x_3 < 0 \wedge x_6 - x_4 \geq 0 \wedge x_7 - x_3 \leq 0 \wedge x_8 - x_2 > 0 \wedge h_D + x_2 \geq 0 \wedge h_D + x_4 \leq 0 \wedge h_D + x_6 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge ((x_5 - x_1 \geq 0 \wedge x_7 - x_3 < 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \geq 0) \vee (x_6 - x_2 > 0 \wedge h_D + x_2 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0) \vee (x_3 - x_1 > 0 \wedge x_7 - x_3 < 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0) \vee (x_3 - x_1 > 0 \wedge x_7 - x_3 < 0 \wedge h_D + x_4 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0) \vee (x_3 - x_1 > 0 \wedge x_7 - x_3 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0) \vee (x_6 - x_4 > 0 \wedge h_D + x_2 > 0 \wedge h_D + x_4 < 0 \wedge h_D + x_6 > 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0) \vee (x_3 - x_1 > 0 \wedge x_7 - x_3 < 0 \wedge h_D + x_6 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0)))$

Equilibrium 4: $(x_3 - x_1 \leq 0 \wedge x_5 - x_3 < 0 \wedge x_7 - x_1 \leq 0 \wedge x_8 - x_2 > 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_6 \geq 0 \wedge h_D + x_8 \geq 0 \wedge h_D + x_2 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge ((x_3 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_4 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_7 - x_1 < 0 \wedge x_7 - x_3 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0) \vee (x_3 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_6 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_3 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_2 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_3 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \geq 0) \vee (h_D + x_4 > 0 \wedge h_D + x_6 > 0 \wedge h_D + x_8 > 0 \wedge h_D + x_2 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0) \vee (x_3 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_8 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0)))$

From the query results, it can be seen that two equilibria are guaranteed to be reachable from the initial conditions specified by the query. The first one, equilibrium 3, corresponds to a biological steady state where Delta protein concentration in cell 2 alone is high. The other reachable steady state, equilibrium 4, is where the Delta level in cell 1 alone is high. The query results predict that the former or the latter are reachable depending on whether Delta concentration in cell 1 is greater or less than that in cell 2, i.e. $x_1 > x_3$ or $x_1 < x_3$. This agrees with classical lateral inhibition, and so does the result that equilibrium 2 is not guaranteed to be reachable, because of the condition $x_3 > x_5$.

However, the non-intuitive result is that equilibrium 1 is not guaranteed to be reachable. Since there are no restrictions on the Delta protein in cell 4, x_7, it could be expected that the system converges to equilibrium 1 with the additional constraint that $x_7 > x_3$. This does not happen, because of the effect of the constraint on Notch protein levels in query 3. Since cell 4 has a higher Notch concentration than cell 1, this acts to repress Delta production in cell 4 preferentially with respect to cell 1. As a result, there is not sufficient Delta protein in cell 4, under the given query constraints, to compete with the surrounding cells and achieve a high equilibrium level. This underlines

the importance of Notch receptors in lateral inhibition, even though they do not directly affect protein concentrations in neighboring cells (Fig. 1).

Query 4: $x_6 > x_2 \wedge x_6 > x_4 \wedge x_6 > x_8$

Query 4 encodes conditions only on Notch protein concentrations. Even though Notch acts through an intermediary (Delta ligand) to influence protein concentrations in neighbors, it has an observable effect on the reachability of different steady states, as seen in the previous query. Query 4 constrains Notch concentration in cell 3 to be greater than in all its neighbors.

Query 4 results

Equilibrium 1: $(x_6 - x_2 > 0 \wedge x_6 - x_4 > 0 \wedge x_7 - x_5 \geq 0 \wedge x_7 - x_3 \geq 0 \wedge x_7 - x_1 \geq 0 \wedge x_8 - x_2 \leq 0 \wedge x_8 - x_4 \leq 0 \wedge h_D + x_8 \leq 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_2 \geq 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0 \wedge ((h_D + x_8 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (h_D + x_2 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (h_D + x_8 < 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (x_7 - x_5 > 0 \wedge x_7 - x_3 > 0 \wedge x_7 - x_1 > 0 \wedge h_D + x_8 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (h_D + x_8 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0) \vee (h_D + x_8 < 0 \wedge h_D + x_4 > 0 \wedge h_D + x_2 > 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 = 0) \vee (h_D + x_4 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 > 0)))$

Equilibrium 2: Unreachable

Equilibrium 3: $(x_3 - x_1 \geq 0 \wedge x_4 - x_2 \leq 0 \wedge x_5 - x_3 \leq 0 \wedge x_6 - x_2 > 0 \wedge x_7 - x_3 \leq 0 \wedge x_8 - x_6 < 0 \wedge x_8 - x_4 \geq 0 \wedge h_D + x_2 \geq 0 \wedge h_D + x_4 \leq 0 \wedge h_D + x_8 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \geq 0 \wedge ((x_5 - x_3 < 0 \wedge x_5 - x_1 \geq 0 \wedge x_7 - x_3 < 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \geq 0) \vee (x_7 - x_3 < 0 \wedge x_7 - x_1 \geq 0 \wedge x_7 - x_5 \geq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0) \vee (x_8 - x_2 > 0 \wedge h_D + x_2 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0) \vee (x_3 - x_1 > 0 \wedge x_5 - x_3 < 0 \wedge x_7 - x_3 < 0 \wedge h_D + x_4 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0) \vee (x_3 - x_1 > 0 \wedge x_5 - x_3 < 0 \wedge x_7 - x_3 < 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0) \vee (x_8 - x_4 > 0 \wedge h_D + x_2 > 0 \wedge h_D + x_4 < 0 \wedge h_D + x_8 > 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0) \vee (x_3 - x_1 > 0 \wedge x_5 - x_3 < 0 \wedge x_7 - x_3 < 0 \wedge h_D + x_8 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 > 0)))$

Equilibrium 4: $(x_3 - x_1 \leq 0 \wedge x_5 - x_1 \leq 0 \wedge x_6 - x_4 > 0 \wedge x_6 - x_2 > 0 \wedge x_7 - x_1 \leq 0 \wedge x_8 - x_6 < 0 \wedge h_D + x_2 \leq 0 \wedge h_D + x_4 \geq 0 \wedge h_D + x_8 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \leq 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \leq 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \leq 0 \wedge ((h_D + x_2 < 0 \wedge h_D + x_4 > 0 \wedge h_D + x_8 > 0 \wedge h_N - 2x_7 - 2x_3 - 2x_1 \geq 0) \vee (h_D + x_2 < 0 \wedge h_D + x_4 > 0 \wedge h_D + x_8 > 0 \wedge h_N - 2x_5 - 2x_3 - 2x_1 \geq 0) \vee (x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_4 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_2 \geq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0) \vee (h_D + x_2 < 0 \wedge h_D + x_4 > 0 \wedge h_D + x_8 > 0 \wedge h_N - 2x_7 - 2x_5 - 2x_1 \geq 0) \vee (h_D + x_2 < 0 \wedge h_D + x_4 > 0 \wedge h_D + x_8 > 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 \leq 0) \vee (x_3 - x_1 < 0 \wedge x_5 - x_1 < 0 \wedge x_7 - x_1 < 0 \wedge h_D + x_8 \leq 0 \wedge h_N - 2x_7 - 2x_5 - 2x_3 > 0)))$

The query results demonstrate the effect on a cell of having a higher Notch concentration than its neighboring cells. Given the approximate reachable sets, they imply that cell 3 will remain undifferentiated, i.e. the network will not converge to equilibrium 2, a steady state where cell 3 has a high Delta protein concentration and all its neighbors have low Delta levels. All the other equilibria may be reached, depending on the initial asymmetry of Delta protein concentrations in neighboring cells. However, since the reachable sets are under-approximations, there may exist initial conditions from which equilibrium 2 is reachable, but which cannot be identified because they fall in the undecidable state space of the abstracted discrete transition system.

V. Conclusion

This paper demonstrates a useful and efficient method to biologically interpret analytical data from hybrid models of protein signaling. However, even though the examples presented here are protein feedback networks, the algorithms for reachable set computation and query are general in nature, and can also be applied to other biological models, for example, gene regulatory networks. The queries have to be constructed in a meaningful way, and require domain knowledge for the particular biological system under study. Current research is focusing on extending the query based technique to higher dimensional reachability sets of larger networks, including the PCP signaling network in *Drosophila*, where the authors have access to biological data for different PCP signaling protein localization patterns, and some of the predictions from the query tests may even be experimentally verifiable.

References

[1] H. El-Samad, H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash, "Surviving heat shock: control strategies for robustness and performance," *PNAS*, vol. 102, no. 8, pp. 2736–2741, 2004.

[2] H. De Jong, J. L. Gouze, C. Hernandez, M. Page, T. Sari, and J. Geiselmann, "Qualitative simulation of genetic regulatory networks using piecewise-linear models," *Bulletin of Mathematical Biology*, vol. 66, no. 2, pp. 301–340, March 2004.

[3] G. Marnellos, G. A. Deblandre, E. Mjolsness, and C. Kintner, "Delta-notch lateral inhibitory patterning in the emergence of ciliated cells in *Xenopus*: experimental observations and a gene network model," in *Pacific Symposium on Biocomputing*, 2000, pp. 326–337.

[4] D. R. Tree, J. M. Shulman, R. Rousset, M. P. Scott, D. Gubb, and J. D. Axelrod, "Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling," *Cell*, vol. 109, no. 3, pp. 371–381, May 2002.

[5] R. Ghosh and C. J. Tomlin, "Lateral inhibition through delta-notch signaling: a piecewise affine hybrid model," in *Proceedings of the 4th International Workshop on Hybrid Systems: Computation and Control*, ser. LNCS 2034, M. D. D. Benedetto and A. Sangiovanni-Vincentelli, Eds. Rome: Springer Verlag, March 2001, pp. 232–246.

[6] R. Ghosh, K. Amonlirdviman, and C. J. Tomlin, "A hybrid system model of planar cell polarity signaling in drosophila melanogaster wing epithelium," in *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, December 2002, pp. 1588–1594.

[7] R. Ghosh and C. Tomlin, "Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: delta-notch protein signalling," *Systems Biology*, vol. 1, no. 1, pp. 170–183, June 2004.

[8] E. Mendelson, *Introduction to Mathematical Logic*. London: Chapman & Hall, 1997.

[9] H. Hong, "An improvement of the projection operator in cylindrical algebraic decomposition," in *Proc. ISAAC 90*, 1990, pp. 261–264.