



formalism [7]. While amenable for interesting analysis, the methods mentioned above are based on assumptions which disregard important biochemical phenomena. Most of them only capture protein dynamics but cannot accommodate chemical reactions [4]. When the regulatory systems are not spatially homogeneous, partial differential equations and other spatially distributed models such as reaction-diffusion equations [8] and the gene circuit method [9] can be used. The modeling approach based on differential equations implicitly assumes that the concentrations of the species in the network vary continuously and deterministically. The continuity assumption is compromised in cases when the number of molecules of a certain species is small or due to fluctuations in the timing of cellular events. The use of discrete stochastic models is proposed in [10–12]. However, these methods are computationally expensive and cannot handle high dimensional systems.

Our modeling approach is deterministic and based on *hybrid systems* [13–16], *i.e.*, systems in which discrete events are combined with continuous differential equations to capture the switching behavior that is observed in phenomena such as transcription [17], protein-protein interactions, and cell division and growth. We also propose the use of hybrid system as the natural framework giving a global description of a biological system described locally around operating points by simpler dynamics, which are easier to approach for analysis. We successfully used hybrid systems to model, simulate and perform preliminary analysis on small dimensional genetic networks for the quorum sensing system in the marine bacterium *Vibrio fischeri* [18–22]. More specifically, as in [1, 3], we assume a simplified model of regulation, but, in addition to step functions, we allow for piece-wise linear activation functions, which seem more realistic. Moreover, we dramatically enlarge the class of regulatory systems by allowing for coupling and product type nonlinearities in the vector field. Consequently, chemical reactions among the species involved in the regulation can be included. The vector fields are *multi - affine*, *i.e.*, have only product - type nonlinearities to accommodate chemical reactions, and are defined in rectangular partitions of the state space, called *invariants*. The facets of the rectangles correspond to changes in the behavior of a gene or enzyme. For such systems, we show that reachability and safety verification problems can be formulated and solved (conservatively) in an elegant and computationally inexpensive way, based on some interesting properties of multi-affine functions on rectangular regions of the Euclidean space of arbitrary dimension: a multi-affine function is uniquely determined everywhere in a rectangle by its value at the vertices.

We apply this method to the stringent response system, namely, to the question of the existence of an upper limit to the concentration of (p)ppGpp and to the number of stalled translational complexes following a nutritional downshift.

The paper is organized as follows. In the following section we describe our model. We give a brief biological background. We then describe the basic reaction groups and substances that we included in our model, and we formulate the model as a hybrid system. In Section 3 we present the reachability algorithm we use in this paper. Section 4 gives our analytical and numerical results. We

identify the steady states of the model, analytically and numerically. We present results of simulations in a setting similar to what triggers the stringent response in nature. We use reachability analysis to investigate the existence of an upper limit for one of the biologically interesting variables of the model. Finally we summarize our results and outline directions for future work.

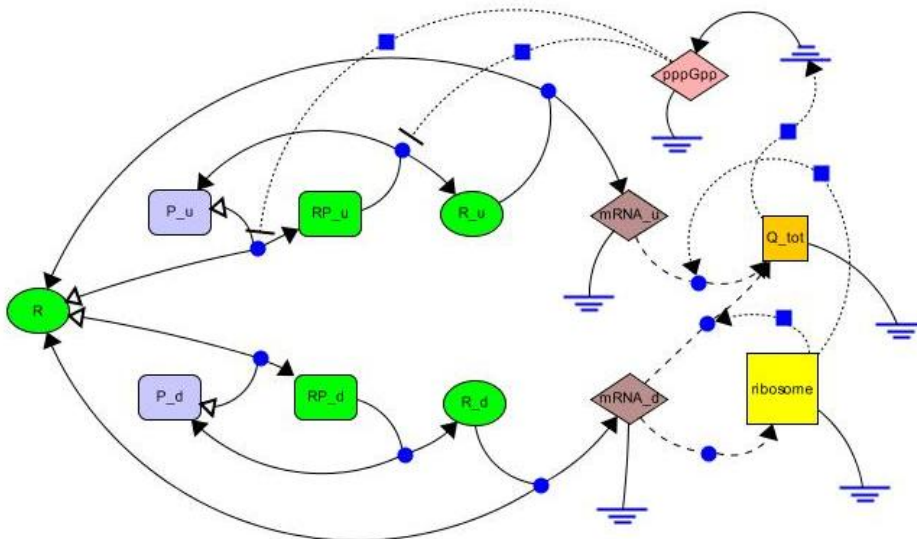
## 2 Biological background and model definition

The stringent response has become an important area of study in *M. tuberculosis* (*M.tb*), where it is has been found to be crucial for latent survival in the infected host [23]. The stringent response occurs in the setting of nutrient deprivation. It is characterized by reduced amino acid consumption and protein production, suppression of anabolic processes, and rerouting of metabolic pathways towards those most essential for survival [24]. Bacteria which lack the stringent response are unable to survive in nutrient deprived media and die from either unrestrained growth or accumulation of misfolded proteins.

In *M.tb*, as in many other bacteria, the stringent response is mediated via the level of (p)ppGpp. In the setting of amino acid deprivation, the levels of (p)ppGpp increase, resulting in a set of genetic regulatory events which trigger cell-wide metabolic changes. The level of (p)ppGpp in the cell is set by the equilibrium of two reactions, synthesis and hydrolysis, which synthesize and destroy (p)ppGpp, respectively. In *M.tb*, both of these functions are performed by the enzyme RelMtb [25]. The rate of the synthesis reaction increases dramatically in the presence of the Rel Activating Complex (RAC) [26]. This complex is formed when the translation machinery of the cell, the ribosome, encounters an uncharged tRNA during protein synthesis. The level of uncharged tRNA is thought to be the mechanism through which *M.tb* senses the abundance of amino acids in its environment. The increase in (p)ppGpp levels has several effects on transcription, which results in a dramatic downregulation of stable RNA and a less dramatic downregulation of mRNA [27]. Downregulation of stable RNA results in the reduction of ribosome concentration, and thus protein synthesis. Though global mRNA levels decrease in the stringent response, a small subset of genes are upregulated [28]. These include genes coding for proteins involved in amino acid biosynthesis and antigen production.

### 2.1 Model equations

A map of the reactions accounted for in our model is shown in Figure 1. Our *dynamical variables* are: the concentrations of two species of promoters  $P_u$  and  $P_d$ , free RNA polymerase  $R$ , four intermediate complexes  $RP_u, RP_d, R_u^*, R_d^*$ , the ribosomes  $Rib$ , and the number of translational complexes  $Q_{tot}$ . Several other substances (including two species of mRNA and the signaling substance (p)ppGpp, denoted by diamonds in Figure 1) are calculated at steady state so they are algebraic functions of the former. We follow two *observables*, one proportional to the concentration of (p)ppGpp,  $g$ , and one which gives a measure



**Fig. 1.** Screenshot from Charon's model design interface with the schematic of the model.

of the transcriptional activity in the cell,  $q$ . We distinguish an *external parameter*  $r$  which is defined as the ratio of uncharged versus charged tRNA, a measure of the availability of resources. The lower  $r$ , the more nutrients are available. There is a significant number of fixed parameters that control the reaction kinetics.

We distinguish two generic species of mRNA, 'upregulated' (u) and 'downregulated' (d), each with its own species of promoter  $P_x$  ( $x = u, d$ ) and intermediate complexes. Free RNA polymerase binds to the promoters to form the  $RP_x$  complexes. The eventual liberation of the promoter site results in the respective transcriptional complexes,  $R_x^*$ . The RNAP is regained at the end of the ensuing elongation process, which results in the transcript  $mRNA_x$ . The total amount of mRNA and the concentration of ribosomes (which is in equilibrium with the 'd' transcript concentration) determine the rate of translation initiation. The number of translational complexes  $Q_{tot}$  is controlled by the initiation rate and the average elongation time. The latter varies with the outside parameter  $r$ , proportionally to  $1 + r$  (elongation slows down when nutrients are scarce). The number of stalled complexes is a fraction of  $Q_{tot}$ , in a ratio of  $r/(1+r)$ . Finally, the concentration of stalled complexes determines the equilibrium concentration of (p)ppGpp, which increases linearly with it, up to a saturation value. The feedback loop is closed by the influence of (p)ppGpp on the kinetics of transcription. The transcription of the 'd' species is inhibited when the (p)ppGpp concentration is high, leading to a decrease in ribosomes which in turn reduces the translation initiation rate.

We map the concentrations of  $\{RP_u, RP_d, R_u^*, R_d^*, Rib, Q_{tot}, R, P_u, P_d\}$  on a nine-dimensional state space  $\{x_k\}_{k=1,9}$ . The result is a hybrid system with two

modes, depending on the value of one of the variables,  $x_6$ .

$$\begin{aligned}
\frac{dx_1}{dt} &= k_{1u} x_7 x_8 - (k_{3u} + k_{2u}) x_1 \\
\frac{dx_2}{dt} &= \bar{k}_{1d}(g(x_6)) x_7 x_9 - (\bar{k}_{3d}(g(x_6)) + \bar{k}_{2d}(g(x_6))) x_2 \\
\frac{dx_3}{dt} &= k_{3u} x_1 - \frac{1}{\tau} x_3 \\
\frac{dx_4}{dt} &= \bar{k}_{3d}(g(x_6)) x_2 - \frac{1}{\tau} x_4 \\
\frac{dx_5}{dt} &= \frac{f}{\tau} x_4 - \frac{1}{C_{rib}} x_5 \\
\frac{dx_6}{dt} &= k_4 \frac{1}{\tau \lambda_{mRNA}} x_5 (x_3 + x_4) - \frac{1}{\theta(1+r)} x_6 \\
\frac{dx_7}{dt} &= \frac{1}{\tau} (x_3 + x_4) - x_7 (k_{1u} x_8 + k_{1d} x_9) \\
\frac{dx_8}{dt} &= (k_{3u} + k_{2u}) x_1 - k_{1u} x_7 x_8 \\
\frac{dx_9}{dt} &= (\bar{k}_{3d}(g(x_6)) + \bar{k}_{2d}(g(x_6))) x_2 - \bar{k}_{1d}(g(x_6)) x_7 x_9
\end{aligned} \tag{1}$$

All quantities in (1), other than the  $x_k$ , are constant model parameters<sup>6</sup>, except for the 'down-regulated' kinetic constants,  $\{\bar{k}_{xd}, x = \overline{1,3}\}$ , which depend on the observable  $g(x_6)$ . The values of the variables  $x_k$  are always positive (since they are concentrations of chemical substances). There are three conservation laws which are automatically satisfied, giving the conservation of the total RNAP,  $R_{tot} = x_1 + x_2 + x_3 + x_4 + x_7$ , and the promoters,  $P_u^{tot} = x_1 + x_8$ , and  $P_d^{tot} = x_2 + x_9$ .

The observable  $g(x_6)$  is a dimensionless quantity defined as the concentration of the signaling substance (p)ppGpp times a kinetic constant. The latter is a function of  $x_6$ , and it increases linearly with it from a basal value  $g_{low}$  (at  $x_6 = 0$ ) to a saturation value  $g_{high}$ , corresponding to a reference value of  $x_6^{ref} = (1+r)Q_{ref}/r$ . The kinetic constants  $\bar{k}_{xd}$  that depend on  $g$  have a similar dependence. Their value changes linearly from the one at  $g = 0$ ,  $k_x$ , to that at  $g = 1$ ,  $k'_x$ , and stays there for  $g > 1$ . The upper value  $g_{high}$  for the parameter set used in this calculation is larger than 1, so the true threshold for the  $x_6$ -dependence is the value that gives  $g(x_6^0) = 1$ . Below the threshold, the constants  $\bar{k}_{xd}$  ( $x = \overline{1,3}$ ) depend linearly on  $x_6$ , and above they are simply equal to  $k'_x$ :

$$\bar{k}_{xd} = \begin{cases} k_x^{min} + \Delta_x x_6 & ; x_6 < x_6^0 \\ k'_x & ; x_6 \geq x_6^0 \end{cases} \tag{2}$$

where  $k_x^{min}$  and  $\Delta_x$  are appropriate functions of  $k_x, k'_x, x_6$  and the limiting values of  $g$ . Our other observable,  $q$ , is a measure of transcriptional activity, or rather, the rate at which translations are initiated. We define it as the first term in the r.h.s. of the equation for  $x_6$  in (1), normalized so that it is comparable to  $x_6/x_6^{ref}$ ,

$$q = \frac{k_4 \theta}{\tau \lambda_{mRNA}} \frac{1}{Q_{ref}} x_5 (x_3 + x_4) ; y \equiv \frac{x_6}{r x_6^{ref}} = \frac{1}{1+r} \frac{x_6}{Q_{ref}} \tag{3}$$

This rather cumbersome definition has the advantage that it does not depend explicitly on  $r$ . At equilibrium we have  $q = y$ .

In summary, the model has two modes defined by the values of  $x_6$ , with the mode transitions occur at  $x_6 = x_6^0$ , corresponding to  $g = 1$ . The equations

<sup>6</sup> For a definition of all the constants refer to a detailed description of the model [29]

in both modes are obtained from (1) with the 'up-regulated' kinetic constants  $k_{xu}, x = \overline{1, 2, 3}$  replaced by  $k_x$ , and the 'down-regulated' constants given by (2). The transition map is trivial, i.e., all of the dynamical variables continue on with the same values they had just before the transition.

### 3 Algorithm for analysis of hybrid multi-affine systems with rectangular invariants

The model described above is defined in rectangular regions of the state space and it is affine in each of the state variables  $x_i$ , i.e., it is multi-affine. This section formally introduces the definition of a multi-affine rectangular hybrid system and extends some results presented in [19].

#### 3.1 Hybrid rectangular multi-affine systems

**Definition 1 (Multi-affine function).** *A multi-affine function  $f : \mathbb{R}^N \rightarrow \mathbb{R}^m$  is a polynomial in the indeterminates  $x_1, \dots, x_N$  with the property that the degree of  $f$  in any of the indeterminates  $x_1, \dots, x_N$  is less than or equal to 1. Stated differently,  $f$  has the form*

$$f(x_1, \dots, x_N) = \sum_{i_1, \dots, i_N \in \{0, 1\}} c_{i_1, \dots, i_N} x_1^{i_1} \cdots x_N^{i_N}, \quad (4)$$

with  $c_{i_1, \dots, i_N} \in \mathbb{R}^m$  for all  $i_1, \dots, i_N \in \{0, 1\}$  and using the convention that if  $i_k = 0$ , then  $x_k^{i_k} = 1$ .

An  $N$ -dimensional rectangle in  $\mathbb{R}^N$  is characterized by two vectors  $a = (a_1, \dots, a_N) \in \mathbb{R}^N$  and  $b = (b_1, \dots, b_N) \in \mathbb{R}^N$ , with the property that  $a_i < b_i$  for all  $i \in \{1, \dots, N\}$ :

$$R_N(a, b) = \{x = (x_1, \dots, x_N) \in \mathbb{R}^N \mid \forall i \in \{1, \dots, N\} : a_i \leq x_i \leq b_i\}. \quad (5)$$

The set of vertices of  $R_N(a, b)$  is denoted by  $V_N(a, b)$ , and may be characterized as

$$V_N(a, b) = \prod_{i=1}^N \{a_i, b_i\} \quad (6)$$

Let  $\xi : \{a_1, \dots, a_N, b_1, \dots, b_N\} \rightarrow \{0, 1\}$  be defined by

$$\xi(a_k) = 0, \quad \xi(b_k) = 1, \quad k = 1, \dots, N \quad (7)$$

Then  $R_N(a, b)$  has  $2N$  facets described by

$$F_N^{j\xi(w_j)}(a, b) = R_N(a, b) \cap \{x \in \mathbb{R}^N \mid x_j = w_j\}, \quad w_j \in \{a_j, b_j\}, \quad j = \overline{1, N} \quad (8)$$

The outer normal of facet  $F_N^{j\xi(w_j)}(a, b)$  is given by

$$n_N^{j\xi(w_j)} = (-1)^{\xi(w_j)+1} e_j, \quad w_j \in \{a_j, b_j\}, \quad j = \overline{1, N} \quad (9)$$

where  $e_j, j = 1, \dots, N$  denote the Euclidean basis of  $\mathbb{R}^N$ .

A rectangular partition of the state space  $(x_1, \dots, x_N)$  is defined as follows. Each axis  $Ox_i$ ,  $i = 1, \dots, N$  is divided into  $n_i \geq 1$  intervals by the thresholds  $0 = \theta_0^i < \theta_1^i < \dots < \theta_{n_i}^i$ . The  $j^{\text{th}}$  interval on the  $Ox_i$ ,  $i = \overline{1, N}$  axis is therefore defined as  $\theta_{j-1}^i \leq x_i < \theta_j^i$ ,  $j = \overline{1, n_i}$ . By convention,  $\theta_0^i = 0$  and  $\theta_{n_i}^i$  is an upper bound giving a physical limit of  $x_i$ . The division of the axes determines a partition of the state space into  $\prod_{i=1}^N n_i$  rectangles. If we let

$$a_{k_1 \dots k_N} = (\theta_{k_1-1}^1, \dots, \theta_{k_N-1}^N) \in \mathbb{R}^N, \quad b_{k_1 \dots k_N} = (\theta_{k_1}^1, \dots, \theta_{k_N}^N) \in \mathbb{R}^N \quad (10)$$

then following the notation in (5), (6), and (8), an arbitrary rectangle from the partition is given by  $R_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N})$ , the corresponding set of vertices by  $V_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N})$ , and the facets by  $F_N^{j \xi_j(w_j)}(a_{k_1 \dots k_N}, b_{k_1 \dots k_N})$ . Explicitly,

$$R_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N}) = \{(x_1, \dots, x_N) \in \mathbb{R}^N \mid \theta_{k_i-1}^i \leq x_i \leq \theta_{k_i}^i, \quad i = \overline{1, N}\} \quad (11)$$

In each rectangle  $R_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N})$ , the system evolves along specific dynamics described by

$$\dot{x} = f_{k_1 \dots k_N}(x), \quad x \in R_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N}) \quad (12)$$

*Remark 1.* A convenient way of representing a hybrid multi-affine system with rectangular invariants (12), (11) is as a *simple graph* with  $\prod_{i=1}^N n_i$  nodes. Node  $(k_1 \dots k_N)$  corresponds to rectangle  $R_N(a_{k_1 \dots k_N}, b_{k_1 \dots k_N})$  and has associated dynamics (12). An edge in the graph connects nodes corresponding to adjacent rectangles, *i.e.*, there is an edge between any pair of nodes that differ by a Hamming distance of 1.

### 3.2 Multi-affine functions on rectangles

Let  $R_N(a, b)$  denote an arbitrary rectangle as defined by (5) and  $f : R_N(a, b) \rightarrow \mathbb{R}^N$  a multi-affine function as in (4).

**Proposition 1.** *A multi-affine function  $f : R_N(a, b) \rightarrow \mathbb{R}^N$  is a convex combination of its values at the vertices of  $R_N(a, b)$ .*

$$f(x_1, \dots, x_N) = \sum_{(v_1, \dots, v_N) \in V_N(a, b)} \prod_{k=1}^N \left( \frac{x_k - a_k}{b_k - a_k} \right)^{\xi(v_k)} \left( \frac{b_k - x_k}{b_k - a_k} \right)^{1 - \xi(v_k)} f(v_1, \dots, v_N), \quad (13)$$

and

$$1 = \sum_{(v_1, \dots, v_N) \in V_N(a, b)} \prod_{k=1}^N \left( \frac{x_k - a_k}{b_k - a_k} \right)^{\xi(v_k)} \left( \frac{b_k - x_k}{b_k - a_k} \right)^{1 - \xi(v_k)}. \quad (14)$$

The proof of this proposition can be found in [19].

**Corollary 1.** *The projection of a multi-affine vector field defined on a rectangle along a given direction is positive (negative) everywhere in the rectangle if and only if its projection along that direction is positive (negative) at the vertices.*

### 3.3 Analysis

We assume that the piecewise defined vector field (12) (possibly non-differentiable), is continuous everywhere, *i.e.*, the vector fields in adjacent rectangles coincide on the common facet. A simple consequence of Corollary 1 can be used to qualitatively analyze the system.

*Problem 1 (Safety verification).* Consider a rectangular partition of the state space (11) with specific dynamics (12) in each rectangle. For a given target rectangle, determine a set of rectangles with the property that if the system starts from arbitrary initial states in any of these rectangles, it will never reach the target rectangle.

The target rectangle in the above Problem can be thought of as a collection of “bad” states. The goal of Problem 1 is to determine collections of initial states which are “safe” to start from.

Corollary 1 is applied to the facets of the  $N$  - rectangles (11), which are, of course,  $N - 1$  - rectangles. A simple and elegant solution to Problem 1 can be described by defining an orientation for the simple graph of the network defined in Remark 1. We allow for both unidirectional and bidirectional edges in the oriented graph. The semantics of the orientation are defined as follows. Let  $r$  and  $p$  be two adjacent nodes in the graph and  $R$  and  $P$  the corresponding adjacent rectangles. A unidirectional edge from  $r$  to  $p$  means that there exists at least one trajectory originating in  $R$  that penetrates into  $P$  through the separating facet, and there is no trajectory starting in  $P$  going to  $R$  through that facet. A bidirectional edge insures the existence of both trajectories originating in  $R$  penetrating in  $P$  and originating in  $P$  penetrating into  $R$ .

---

#### Algorithm 1 Define an oriented graph

---

```

for each node  $(k_1 \dots k_N)$ ,  $k_i = \overline{1, n_i}$ ,  $i = \overline{1, N}$  do
  for each incident edge do
    for each vertex of the corresponding facet do
      calculate the projection of  $f_{k_1 \dots k_N}$  along the outer normal of the facet
    end for
    if the projections are positive at all vertices then
      the edge is unidirectional oriented out of  $(k_1 \dots k_N)$ 
    end if
    if the projections are negative at all vertices then
      the edge is unidirectional oriented towards  $(k_1 \dots k_N)$ 
    end if
    if there is a sign change among the projections at the vertices then
      the edge is bidirectional
    end if
  end for
end for

```

---

Note that, in the oversimplified description above, Algorithm 1 seems inefficient. Indeed, if we apply it to all the rectangles in the partition, most of the vertices are

visited more than ones, and this is not necessary because the vector fields in adjacent rectangles match on the separating facet. A more efficient description would require more complicated notation which is beyond the scope of this paper.

Using the oriented graph, we can now provide a solution to Problem 1. Let  $T$  denote the target rectangle, or, equivalently, the target node in the graph. The following algorithm constructs a set  $\mathcal{R}$  of nodes with the property that if the system starts in any of the corresponding rectangles, then it is possible to reach  $T$ . The complement of  $\mathcal{R}$  with respect to the set of all nodes is a safe set  $\mathcal{S}$ .

---

**Algorithm 2** Construct a safe set  $\mathcal{S}$

---

```

initialize  $\mathcal{R}$  with T
repeat
  for each element  $R$  of  $\mathcal{R}$  do
    for all incident nodes  $P$  connected with an edge (uni or bi-directional) oriented
    towards  $R$  do
      if  $P$  is not already in  $\mathcal{R}$  then
        add  $P$  to  $\mathcal{R}$ 
      end if
    end for
  end for
until cardinality of  $\mathcal{R}$  increases
 $\mathcal{S}$  is the complement of  $\mathcal{R}$  with respect to the set of all nodes

```

---

*Remark 2.* Algorithm 2 for the construction of a safe set  $\mathcal{S}$  might be too conservative, *i.e.*, the set  $\mathcal{S}$  might be unnecessarily small. Indeed, the theory developed in this paper guarantees the existence of a trajectory from a reactangle  $R$  to an adjacent rectangles  $P$  if the (unidirectional or bidirectional) edge between the corresponding nodes in the oriented graph has an arrow from  $R$  to  $P$ . But if there is an edge from  $R$  to  $P$  and also an edge from  $P$  to  $T$ , we cannot guarantee that there is a trajectory of the system from  $R$  to  $T$ . In our analysis, we simply say that there might be a trajectory from  $R$  to  $T$ , and so  $R$  is classified as “unsafe” for  $T$ .

## 4 Results

### 4.1 Steady states and dynamical simulations

The steady-state equations of (1) together with the three conservation laws imply<sup>7</sup>

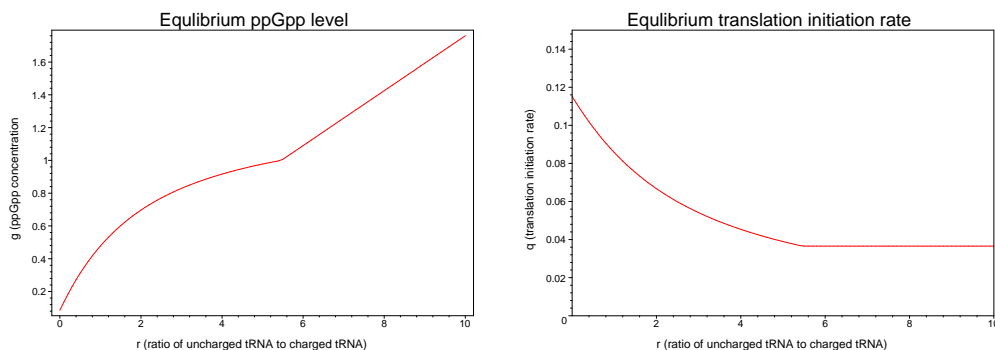
$$R^{tot} = x_7 \left( 1 + P_u^{tot} \frac{(1 + \tau \bar{k}_{3u}(g)) \alpha_u}{1 + \alpha_u x_7} + [P_d^{tot}] \frac{(1 + \tau \bar{k}_{3d}(g)) \alpha_d(g)}{1 + \alpha_d(g) x_7} \right) , \quad (15)$$

where we solved for  $x_1, x_2, x_3, x_4$  and used the conservation laws for  $x_8, x_9$ . The r.h.s. of this equality is a monotonic function of  $x_7$ . It is zero for  $x_7 = 0$  and exceeds  $R^{tot}$  for  $x_7 = R^{tot}$ . Therefore, *independently of  $g$* , this equation has exactly one solution for  $x_7$

<sup>7</sup>  $\alpha_u, \alpha_d$  are notations for  $(k_{2u} + k_{3u})/k_{1u}$  and  $(\bar{k}_{2d}(g) + \bar{k}_{3d}(g))/\bar{k}_{1d}(g)$ , respectively.

in the interval  $[0, R_{tot}]$ . With this result we can solve uniquely for  $x_5$  and finally, for  $x_6$ .

We can hence determine the quantity  $q$  (as defined in (3)), as a function of only  $g$  (or  $x_6$ ). We illustrate this in Figure 4, where we plot  $q(g)$  that results from the above derivation (thick line). On the other hand, from the steady-state condition on  $x_6$ ,  $q$  should also be equal to  $y = x_6/Q_{ref}(1+r)$ . The patterned lines in Fig. 4 represent the values of  $q^{equilibrium} = y(g)$  as determined by the dependence of  $g$  on  $x_6$ , for two different  $r$  values. The steady states (for the respective  $r$  values) are at the intersection of the two types of lines.



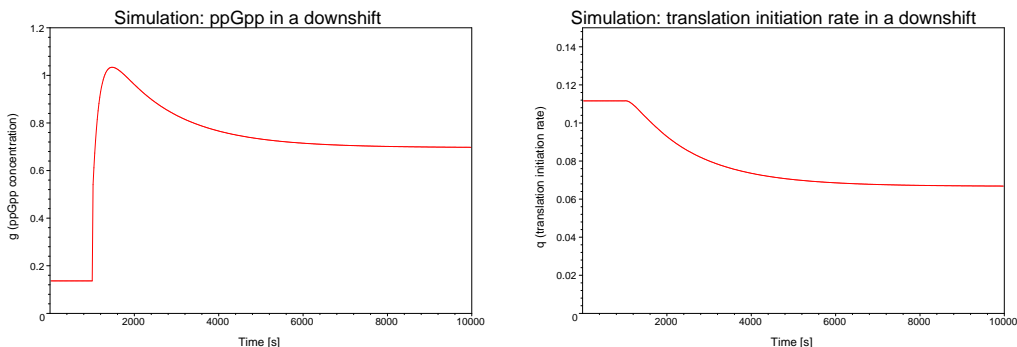
**Fig. 2.** Steady state values of the normalized ppGpp concentration (left) and of the normalized translation initiation rate (right), as a function of the outside quantity  $r$ .

We have verified the steady states described above numerically. The calculations have been performed in part by the software package Charon. Finding steady states in a hybrid system can be accomplished by finding steady states for each set of dynamics and discarding those that fall outside of the region that their dynamics hold over. Charon does this using the NLEQ package, which uses a damped Newton’s method to search for roots of systems of equations. For each set of dynamics, Charon selects several initial guesses in the appropriate region, looks for equilibrium points, and compares any results to the region in question.

There is only one valid steady state for a given value of the parameter  $r$  and conserved quantities  $R^{tot}, P_d^{tot}, P_u^{tot}$ . It follows the thick line in Figure 4. For  $r < r_{trans}$ <sup>8</sup> the steady state lies in the ‘low’ mode and then it crosses over to the ‘high’ mode. In fact both modes have their own steady state, but both lie on the same side of the mode dividing line  $g = 1$ . The dependence of our two observables (the (p)ppGpp concentration  $g$  and the transcription rate  $q$ ) on the outside parameter  $r$  is given in Figure 2. As  $r$  increases, so does  $g$  and the transcription rate  $q$  decreases accordingly. The higher  $r$ , the fewer nutrients are available to the bacterium. This illustrates the main function of the switching mechanism. (p)ppGpp is a signalling substance, its concentration increases as

<sup>8</sup>  $r_{trans}$  is the value of  $r$  for which  $q^{equilibrium} = 1$ .

an indicator of an adverse condition. The result of the adjustment is a decrease in transcriptional activity, to match the reduced availability of nutrients.



**Fig. 3.** Simulation output from the model. See text for details.

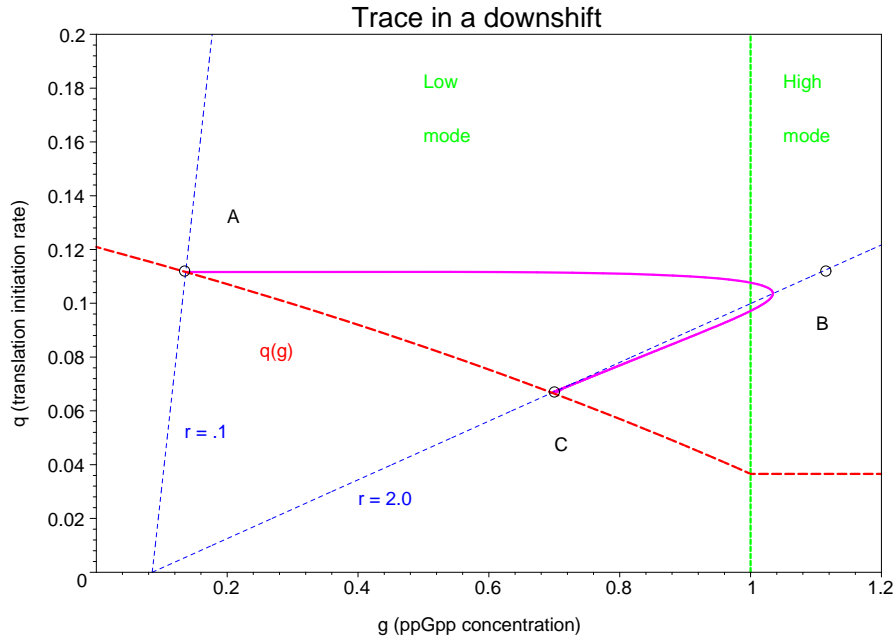
Outputs of a time-dependent simulation are presented in Figures 3 and 4. We follow the observables  $g$  and  $q$  discussed earlier. We emulate the situation which triggers the stringent response in nature. We start with an environment rich in nutrients, represented by a low value of the parameter  $r$ . After the system equilibrates (i.e., the steady state is reached), we change the parameter  $r$  to a high value, corresponding to a sudden downshift in the available nutrients<sup>9</sup>

In Figure 3 we plot our two observables as a function of time. At the downshift  $g$  increases very fast, then tapers off to stabilize at the equilibrium level. By contrast,  $q$  slowly decreases to reach the equilibrium value on the same timescale as  $g$ . This illustrates how the signal is immediately triggered by the sudden change in the outside conditions, while the state of the system (the transcription rate) lags behind, changing to the new value over a longer timescale. The pairs of  $(g, q)$  values for the same simulation are traced out in Figure 4, where we also indicated the locations of the equilibria. The initial and final states correspond to the points A and C respectively. In the course of the simulation the system crosses the mode boundary, when  $g$  briefly exceeds the threshold value of 1. There are no qualitative consequences of this transition. This should not be surprising since the model has been designed to seamlessly match the two modes.

## 4.2 Reachability analysis

The role of the stringent response mechanism is to adjust the cell's transcriptional output to the level of available nutrients. This dependence is illustrated in Figure 2 for our model. The real cell has a similar characteristic. Nature constrains the bacterium very tightly to the optimal output: if the output is too high, stalled translational

<sup>9</sup> From here on the system evolves entirely in the dynamics determined by the higher  $r$  value, with the steady-state corresponding to the low  $r$  as initial condition. Hence all the following discussion refers to the dynamics with  $r$  fixed at the higher value.



**Fig. 4.** A trace of the states of the simulation shown in Figures 2 and 3. The initial state is A and the final state is C. See text for details.

complexes will pile up and will eventually poison it; if it is too low, competitor species will gain an edge by using the nutrients more efficiently. Furthermore, the speed with which the adjustment takes place may also be a matter of life and death.

In *E.coli* the onset of the stringent response is accompanied by a surge in the intracellular level of (p)ppGpp. This high level persists for a relatively short time and then decreases to the equilibrium level for the new conditions. The final (p)ppGpp concentration is higher than the original one, but significantly lower than the peak value immediately following the onset. The surge of (p)ppGpp may have a role in achieving a fast downshift in transcription.

We are interested in how the surge in (p)ppGpp is determined in our model, which lacks an explicit mechanism that completely shuts down the transcription machinery (as can be seen from the horizontal section of the equilibrium line in Figure 4). Is our model capable of preventing the uncontrolled accumulation of stalled translational complexes? The number of stalled complexes is proportional to our variable  $x_6$ , the same as the one  $g$  is linked to. Therefore the height of the surge observed  $x_6$  has a dual significance. It links to a directly measurable quantity (p)ppGpp which can serve as an experimental check, and it is also a measure of the potentially lethal (if too high) amount of translational complexes.

In applying the reachability algorithm to our model, we posed the following question: *given a value of  $r$  and an initial state of the system, is there a value of the variable  $x_6$  which can not be reached in its subsequent evolution?* In the simulation described in Figure 3 the system is initially in the steady state for

$r = 0.1$ , and in the second part of the simulation the value of  $r$  is set to 2. So we are interested in the dynamics of the system at  $r = 2$ , and we would like to determine the upper limit for  $x_6$  for various initial states, in a range that includes the steady state for  $r = 0.1$ .

We implemented various rectangular partitions of the nine-dimensional state space. Typically we used a small number, 2 or 3, of partitions for the variables other than  $x_6$ , and a large number of partitions (up to 16) for  $x_6$ . In a larger run, we used a set of  $3^5 \times 6^2 \times 2^3 \times 17 = 1189728$  rectangular boxes, but only half of them were accessible because of a diagonal conservation law constraint (see below). With the boundaries chosen restrictively, only 30239 of these were reachable, or 'unsafe', yielding an upper limit of approximately  $g^{max} = 1.4$ . This should be compared to the actual maximum hit by the simulation, close to  $g = 1.05$ , and the built-in maximum close to  $g^{hi} = 4.6$ . Unconstrained calculations would lead to either full reachability, or exceed the limit of 100000 'unsafe' hyper-rectangles imposed by memory constraints. We conclude that the model does provide a significant limitation to the height of the surge in  $x_6$  which is in reasonable agreement with the results of traditional simulations.

We found that limits imposed on the other variables have a strong influence on the outcome of the highest-value reachability analysis. In fact, a lack of such limits (i.e., lower limit zero and upper limits beyond those imposed by the conservation laws) leads to practically all points being reachable, hence no upper limit on  $x_6$ . This should not be surprising since a limitation on  $x_6$ , if it exists, can probably be traced back to a limitation on the other variables. The calculations that yielded reasonable upper limits had the allowed ranges for every variable set based on the initial equilibrium state (i.e., 0.2, 0.6, 1.2 of the initial value). Furthermore, constraints on the variables that feed into  $x_6$ , that is  $x_3, x_4, x_5$ , had dramatic effects. In imposing the limits derived from a conservation law,  $x_3 + x_4 < R_{tot}$ , we had to limit the set of allowed rectangles to a set bounded by the 'discrete diagonal', i.e., the sum of the labels in the 3 and 4 directions were constrained to be less than a fixed integer.

## 5 Conclusions

We have built a model for the stringent response in bacteria such as *M.tb* and *E.coli*. Analytical and numerical study shows that it indeed reproduces the main features of the stringent response. One striking feature is an upsurge in the concentration of (p)ppGpp following a nutritional downshift. This raises the question of an upper limit for the number of translational complexes, which is one of the dynamical variables in our model. In our framework, this question can be answered by performing a (conservative) reachability analysis of the switched rectangular multi-affine system. We plan to expand the capabilities of our software tools to perform series of reachability analyses targeted at refining a limiting extremal value. Also, we will examine further the issue of conservation laws which are characteristic to networks of chemical reactions.

*Acknowledgement* This work was supported by the DARPA Biocomp grant, AF F30602-01-2-0563, and NSF grant EIA01-30797.

## References

1. Glass L. Classification of biological networks by their qualitative dynamics. *Journal of Theoretical Biology*, 54:85–107, 1975.
2. Mestl T., E. Plathe, and S. W. Omholt. Periodic solutions in systems of piecewise-linear differential equations. *Dynamics and stability of systems*, 10(2):179–193, 1995.
3. de Jong H., J. L. Gouze, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: a qualitative approach. In *Hybrid Systems: Computation and Control*, 2003.
4. Kauffmann S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.
5. Thomas R. Regulatory networks seen as asynchronous automata: a logical description. *Journal of Theoretical Biology*, 153:1–23, 1991.
6. Kuipers B. Qualitative simulation. *Artificial intelligence*, 29:289–388, 1981.
7. Brutlag D. L., A. R. Galper, and D. H. Millis. Knowledge-based simulation of *dna* metabolism: prediction of enzyme action. *Computer Applications in the Biosciences*, 7(1):9–19, 1991.
8. Kauffmann S. A. *The origins of order: self-organization and selection in evolution*. Oxford University Press, New York, 1993.
9. Mjolsness E., D. H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152:429–453, 1991.
10. Gillespie D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.
11. McAdams H. M. and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Science of the USA*, 94:814–819, 1997.
12. Gillespie D. T. Fluctuation and Dissipation in Brownian Motion. *Am. J. Phys*, 61():1077–1083, 1993.
13. R. Alur, T.A. Henzinger, and E.D. Sontag, editors. *Hybrid Systems III: Verification and Control*. LNCS 1066. Springer-Verlag, 1996.
14. T. Henzinger and S. Sastry, editors. *Hybrid Systems: Computation and Control*. LNCS 1386. Springer, 1998.
15. F. Vaandrager and J. van Schuppen, editors. *Hybrid Systems: Computation and Control*. LNCS 1569. Springer, 1999.
16. N. Lynch and B.H. Krogh, editors. *Hybrid Systems: Computation and Control*. LNCS 1790. Springer, 2000.
17. Tomlin C. and R. Ghosh. Lateral inhibition through delta-notch signaling: A piecewise affine hybrid model. *LNCS vol. 2034, Springer Verlag*, 2001.
18. R. Alur, C. Belta, F. Ivancic, V. Kumar, H. Rubin, J. Schug, O. Sokolsky, and J. Webb. Visual programming for modeling and simulation of bioregulatory networks. In *International Conference on High Performance Computing*, Bangalore, India, 2002.
19. C. Belta, L. Habets, and V. Kumar. Control of multi-affine systems on rectangles with applications to hybrid biomolecular networks. In *41st IEEE Conference on Decision and Control*, Los Angeles, NV, 2002.

20. C. Belta, J. Schug, T. Dang, V. Kumar, G. J. Pappas, H. Rubin, and P. V. Dunlap. Stability and reachability analysis of a hybrid model of luminescence in the marine bacterium *vibrio fischeri*. In *40th IEEE Conference on Decision and Control*, Orlando, FL, 2001.
21. R. Alur, C. Belta, V. Kumar, M. Mintz, G. J. Pappas, H. Rubin, and J. Schug. Modeling and analyzing biomolecular networks. *Computing in Science and Engineering*, pages 20–30, Jan/Feb 2002.
22. R. Alur, C. Belta, F. Ivancic, V. Kumar, M. Mintz, G. Pappas, and J. Schug. Hybrid modelling and simulation of biomolecular networks. In *Lecture Notes in Computer Science*, volume 2034, pages 19–32. 2001.
23. Primm T. P., S. J. Andersen, V. Mizrahi, D. Avarbock, H. Rubin, and C. E. Barry 3rd. The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. *J. Bacteriol. Sep;182(17):4889-98*, 2000.
24. Gourse R. L., Gaal T., Bartlett M. S., Appleman J.A., and Ross W. rna transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu. Rev. Microbiol.*, 50:645-77, 1996.
25. Avarbock D., J. Salem, L. Li, Z. Wang, and H. Rubin. Cloning and characterization of a bifunctional *RelA/SpoT* homologue from *Mycobacterium tuberculosis*. *Gene* 233, 261-269., 1999.
26. Avarbock D., A. Avarbock, and H. Rubin. Differential regulation of opposing activities by the amino-acylation state of a *tRNA-Ribosome-mRNA-Rel<sub>Mtb</sub>* complex. *Biochemistry* 39, 11640-11648., 2000.
27. Barker M. M., T. Gaal, C. A. Josaitis, and R. L. Gourse. Mechanism of regulation of transcription initiation by ppGpp. i. effects of (p)ppGpp on transcription initiation in vivo and in vitro. *J. Mol. Biol. Jan 26;305(4):673-88*, 2001.
28. Betts J. C., P. T. Lukey, L. C. Robb, R. A. McAdam, and K. Duncan. Evaluation of a nutrient starvation model of mycobacterium tuberculosis persistence by gene and protein expression profiling. *Mol Microbiol Feb;43(3):717-31*, 2002.
29. C. Belta, P. Finin, Á. Halász, M. Imieliński, V. Kumar, and H. Rubin. Stringent Response in *Mycobacterium tuberculosis*, in preparation. A technical document is available upon request.