Final Report

– Superhelicity –

AMÉLIE VÉRON, YANNICK WURM

June 25, 2003

Abstract

Variations in DNA supercoiling have been shown to control gene expression, probably because transcriptional machinery has better access to transcription sites on a linear strand than on one which is heavily coiled. *In silico*, we have reproduced supercoiling-linked behavior which has previously been determined experimentally in *Escherichia coli*: Using biological data found in scientific literature, we established a model of the genetic regulation sub-network involving the *fis*, *topA*, *gyrA*, and *gyrB* genes. To mathematically represent the interactions between these genes, we then constructed a set of piece-wise linear equations (PLEs). These were then injected into Genetic Network Analyzer (GNA), a tool developed by Helix-INRIA.
Our *in silico* regulation network reproduced homeostatic control of supercoiling as well as the known pattern of gene expression levels when starved *Escherichia coli* cells are submitted to nutritional upshift.

Contents

List of Figures

1 Introduction

1.1 Context

Modeling control of superhelicity in *Escherichia coli*, is a project by two fourth year students at Insa de Lyon, under close guidance of people at **HELIX INRIA Rhône-Alpes** and **Université Joseph Fourier** in Grenoble. Given the increasing quantity of available gene expression data as well as information concerning regulatory networks, it is worthwhile to using this data. Regulatory networks are large and complex; they cannot be correctly understood and analyzed with only a pen and paper. Through a collaboration between computer scientist Hidde de Jong at Inria Rhône Alpes and biologists Johannes Geiselmann and Dominique Schneider at Université Joseph Fourier in Grenoble, **Genetic Network Analyzer** was born. This free java tool helps model, simulate and analyze behavior of genetic networks. It's first application was successful predicting the sequence of events leading to sporulation versus stationary growth in *Bacillus subtilis* [3]. The regulatory network concerning sporulation in *B. subtilis* had been studied with detail in the past); GNA's adequacy for modeling genetic networks was thus proven. Another project was thus started between the same collaborators: modeling the regulatory network of *Escherichia coli*. This project goes beyond validating GNA's adequacy: it will help better understand the biology in *coli*'s regulation network. If fuzzy points appear, GNA will help test alternative solutions, and will guide wet-lab research.

Where do this paper's authors fit in? Fourth year students in Bio-informatics and Modeling at Insa de Lyon have the opportunity of teaming up in small groups to work on a two-semester project. Our six-person group's major concern was working on a useful, biologically-oriented project. Contact between Hidde de Jong and ourselves was made possible through one of our professors, Hubert Charles. Hidde suggested our trying to model *coli*'s genetic regulation network, so we gave it a shot.
In this paper, we report biological evidence, modeling and simulation results concerning only a subset of the genes identified as interesting concerning genetic regulation in *Escherichia coli*. What follows is a list of the sub-networks studied (ours is first of the three):

- **Superhelicity**: Superhelicity, *gyrA/gyrB*, *topA*, *fis*
- **RpoS**: RpoS, RssB, CRP-cAMP, ppGpp, H-NS, starvation
- **Crp-Fis**: Crp, Fis, Cya, starvation

For results concerning the second network see [10]. third see [6].

Our initial goal was combining these three sub-networks before the end of the second semester. Unfortunately, due to time-constraints, this was not possible.

1.2 Approach

The approach we adopted can be summed up in the following points:

A network of interest is chosen : In our case, global regulation in *E. coli*. A starting-point can be a set of known genes of interest.

Bibliographic Research : Finding, reading and understanding publications concerning the chosen network. A model of interactions can thus be constructed.

Graph of Interactions is established using information obtained through reading. Critical thinking must be used since different papers can report contradictive behavior, and experimental conditions are not always representative of what goes on inside the cell.

Mathematical Model : For several reasons, Piecewise linear equations (PLEs) seem to be the mathematical tool best adapted for modeling networks of genetic regulation. Their major advantage is that they do not require numeric values but work with symbolic thresholds.

GNA Model uses a very specific formalism (text-file)

Simulation of model in GNA given a set of initial conditions (described in a different text-file).

This partitioning is theoretical. In practice it is necessary to repeatedly go back to previous steps. This helps perfect the model, and better understand the network, through testing of different hypotheses.
For further detail on usage of GNA and the mathematical tools (PLE) used to formalize the models used, we strongly recommend reading our tutors following publication: [3] . It is required for understanding the description of our models in the second part of this paper.

1.3 Overview of report

What follows is a description of the network of genetic regulation we have studied. Using an established formalism [9], we constructed a graphic representation of interactions. These are based on biological evidence described later. Modeling the complete network right away was not possible, so we started out by modeling a simplified network where neither Fis concentration nor ATP/ADP ratio could vary. We controlled ATP/ADP ratio, and could thus direct the simulation between starvation and nutritional upshift conditions. We were able reproducing homeostatic control of supercoiling involving Gyrase and Topoisomerase I as observed *in vivo*. Our next step was permitting Fis concentration to vary, but we met some hurdles in our simulation. This article ends with a discussion of the limits met concerning biological information as well as with the mathematical and software tools we used.

2 Regulatory Network

2.1 Process

It is generally accepted that DNA is structured in a double helix, but in reality the degree of DNA's coiling can vary substantially. This phenomenon is known as supercoiling or superhelicity. During exponential growth, bacterial DNA was found to have a higher degree of negative supercoiling than when stationary: the two strands are more linear than would be expected. This is likely to lead to higher accessibility by machinery required for transcription and replication.

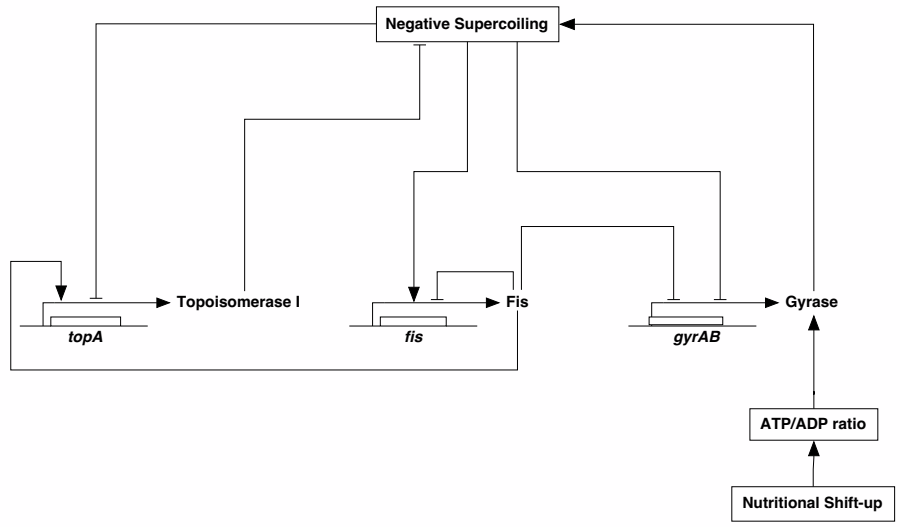
Previous reviewers [4] have identified the major actors of DNA supercoiling regulation in *Escherichia coli*: Gyrase, Topoisomerase I, and Fis. Gyrase actively introduces negative supercoils into double-stranded closed-circular DNA, while Topoisomerase I plays the opposite role by relaxing DNA back towards its standard helical configuration. Both DNA Gyrase and Toposiomerase I belong to the family of four topoisomerases identified in *Escherichia coli*. Their activity is especially crucial during DNA transcription and replication, when the DNA helix must be unwound to allow proper function of large enzymatic machinery, and topoisomerases have indeed been shown to maintain both transcription and replication.
Topoisomerase I changes the degree of DNA supercoiling by causing single-strand breaks and re-ligation, whereas type II topoisomerases (such as Gyrase) cause double-strand breaks. Gyrase reaction cycle normally requires the use of ATP. It leads to the introduction of two negative supercoils. Gyrase subunit A is thought to be principally involved in the DNA breakage and reunion aspects of the supercoiling reaction, while subunit B is responsible for the ATP hydrolysis reaction [1] (further reading: [1, 16, 17]).

Expression of genes coding for these two proteins is regulated in to a certain extent by Fis, a pleiotropic regulator. Fis (Factor for Inversion Stimulation) is a 98 amino acid, homo-dimeric DNA-bending protein known to be a regulator of the metabolic reorganization of the cell during the early exponential growth phase. [19, 18]. Because Fis binding to DNA generates microloops and microdomains, it has direct influence on local supercoiling [23].
Fis usual role is to recruit RNA polymerase into an initial complex. [23] propose a model : "mechanistically the role of Fis in facilitating the initiation process could be explained by assuming that Fis stabilizes a left-handed writhe. In this model, the writhe micro-loop captures the polymerase in the initial complex." However, depending on where Fis' binding site is, Fis can also act as a transcription inhibitor (as for the Gyrase genes *gyrA* and *gyrB*).

Synthesis of the three proteins, Gyrase, Topoisomerase I and Fis, is influenced by chromosomal superhelicity. What results is a complex system of homeostatic regulation. Our first aim here is to review the current experimental knowledge concerning the interactions between these proteins and their genes, and the influence of a sudden variation in ATP/ADP ratio. As a second step, the acquired knowledge will be used to contruct a mathematical model using piece-wise linear equations. Finally, we will use

Genetic Network Analyser (hereafter called GNA) to run simulations of gene expression in *E. coli*.

2.2 Interaction Graph



Arrows represent transcriptional activation or protein production (depending on whether the other end is a gene or a protein), and flat-headed arrows represent transcriptional inhibition.

Figure 1: Graph of Interactions

2.3 Biological Evidence for Interactions

The activations and inhibitions shown in **figure 1**, are based on evidence determined *in vivo* or *in vitro*. What follows is a short description of some of the experiments which brought these interactions to light:

1. Increase in ATP/ADP ratio leads to increased Gyrase synthesis and activity
2. Gyrase activity leads to increased negative supercoiling
3. Low negative supercoiling leads to increase in *gyrA* and *gyrB* expression and Gyrase activity
4. Low negative supercoiling reduces *topA* expression
5. Topoisomerase I reduces the level of negative supercoiling
6. High negative supercoiling leads to increase in *fis* expression
7. Fis is sensitive to nutritional upshift
8. *fis* is negatively autoregulated
9. Fis represses *gyrA* and *gyrB* expression
10. Fis activates *topA* expression

2.3.1 Nutritional shift-up leads to sudden increase in ATP/ADP ratio

Although we have not found direct experimental proof of the link between nutritional upshift and ATP/ADP ratio (i.e. eliminating all other possible factors), it is logical and widely accepted that a sudden increase in ATP/ADP ratio occurs shortly after nutritional shift-up.

2.3.2 Increase in ATP/ADP ratio leads to increased Gyrase synthesis and activity

To show that negative supercoiling is determined by the intracellular ATP/ADP ratio, stationary *E. coli*'s cellular free-energy state was varied by submitting them to stress in three different manners: incubation with potassium cyanide, incubation with DNP (dinitrophenol) and shift to anaerobic conditions. In all three cases, negative supercoiling was determined by the intracellular ATP/ADP ratio. This was not due to transcription or translation, as behavior was unchanged when inhibitors choramphenicol or rifampicin were also applied [25].The changes in supercoiling were very probably due to Gyrase' activity; it has been confirmed that Gyrase subunit B's ATPase activity is a function of ATP/ADP ratio and not directly of ATP concentration [1], because ATP and ADP are competitors.

2.3.3 Gyrase activity leads to increased negative supercoiling

Gyrase appears to be the major source of negative supercoiling *in vivo*, since inhibitors of Gyrase block the introduction of negative supercoils into bacteriophage lambda DNA during infection of a lysogen [4, 7]. By inhibiting Gyrase with known inhibitors Novobiocin and Coumermycin, it has been shown that the link between ATP/ADP ratio and degree of supercoiling is provided by Gyrase: presence led to DNA relaxation [25, 13]. Others modulated *in vivo* gyrase expression by IPTG induction: they were able to direct DNA supercoiling of a probe plamid to linking numbers between -15 and -6 [8].

Despite this, some authors maintain that Gyrase influence on supercoiling is negligeably sensitive towards ATP/ADP ratio variation at *in vivo* levels. Homeostatic regulation of physiological DNA structure appears to dominate [8]. It has been shown that homeostatic control of DNA supercoiling is 0.87, which means that a 10% change of either Topoisomerase I or Gyrase activity results in only a 1.3% change of DNA supercoiling [21]. This is the type of behavior we will want to reproduce in our model.

2.3.4 Low negative supercoiling leads to increase in *gyrAB* expression and Gyrase activity

The rate at which DNA Gyrase supercoils DNA probably depends on superhelical density: Gyrase is more active on relaxed than on supercoiled substrates [27, 22].

A decrease in negative supercoiling through Gyrase inhibition using a temperature sensitive strain as well as through chemical inhibitors led to an up to 10-fold increase in the rate of synthesis of both Gyrase subunits *in vivo* [13]. Different chemical inhibitors were used. Gyrase synthesis was higher with more efficient inhibitors. For coumermycin in particular, delay between addition of inhibitor and optimal synthesis was very close to the time required for relaxation of the *E. coli* chromosome following the addition of Coumermycin (given by [5]). To rule out other Gyrase-synthesis activation mechanisms, a cell-free protein synthesis system was used on plasmids bearing either *gyrA* or *gyrB*. Synthesis of Gyrase subunits increased when DNA was relaxed by Gyrase inhibitors and by addition of Topoisomerase I [13]. Although several other protein synthesis pattern *in vivo* were similar to Gyrase' when submitted to supercoiling variation, this was not a general tendency (some varied in the other direction).

2.3.5 Low negative supercoiling reduces *topA* expression

Negative supercoiling can be lowered through the use of Gyrase inhibitors or mutants. In such cases, galactokinase activity is reduced *in vivo* when a plasmidic $topApromoter-galactokinase$ fusion is carried [24]. Likewise, Topoisomerase I production went up with increased negative supercoiling [21].

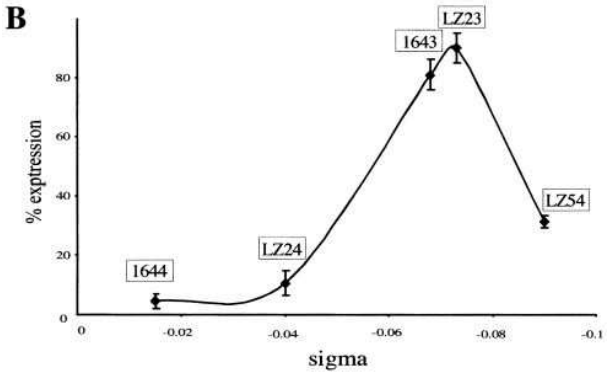
In separate experiments, four different 5' control-regions of the $topA-promoter$ were isolated, recombined with *bal31-nuclease* and studied, both *in vivo* and *in vitro*. In each case, negative supercoiling was required for efficient transcription. Greatest dependency was obtained when combining all four promoters.

2.3.6 Topoisomerase I reduces the level of negative supercoiling

The level of negative supercoiling in bacteria which do not have Topoisomerase is very high [15]. Bacteria possessing an inactive form of Topoisomerase I are not viable (growth is very slow), unless a compensatory mutation in Gyrase appears [12, 11].
Action of Topoisomerase I on supercoiling was confirmed by using a modified promoter: IPTG-induced Topoisomerase production led to a decrease of negative supercoiling level. Influence was minimal though (a thousand-fold range of expression of Topoisomerase I led to an increase of only 6 in linking number); this is certainly due to homeostatic control [21].

2.3.7 High negative supercoiling leads to increase in *fis* expression

When treating cells with a Gyrase inhibitor known to lead to a strong relaxation of DNA, *fis* expression was reduced, as was *fis* promoter activity [20]. This was examined in greater detail: using a set of isogenic strains carrying different combinations of alleles of the genes coding for Gyrase, Topoisomerase IV and Topoisomerase I, a collection of *Escherichia coli* were prepared, covering a wide range of different superhelical DNA densities. Expression of *fis* varied according to **figure 2.3.7**. This behavior was confirmed *in vitro*, and in each case a second gene's expression was also measured; this second gene's behavior differed greatly. Separately, using a *fis-promoter-lacZ*-construct, optimal superhelical density for transcription from plasmids was almost identical to that found for the chromosomal *fis* promoter.



In [20], *in vivo fis* expression was strongly dependent on the negative superhelical density of DNA. Expression is low at superhelical densities below $\sigma = -0.04$. It increases up to 10-fold over a range of -0.04 to -0.07 (maximum at -0.075) and decreases to 30% of maximum at $\sigma = -0.09$.

Figure 2: Expression of *fis* varies greatly in regard to superhelical density *in vivo*

2.3.8 *fis* is negatively autoregulated

Fis binding to an upstream site was proved by DNAse I footprinting; two Fis binding sites were found [14]. The same authors used two $fis - promoter$ - $lacZ$ fusions and compared β-galactosidase activity in bacteria expressing Fis as well as in a Δfis mutant: The promoters differed by their length; the differences in transcription strength were linked to presence and absence of Fis and Fis-binding sites in the Fis promoter: expression was higher when either Fis was absent or the promoter was short (i.e. missing Fis binding sites).
Other authors showed that purified Fis can prevent stable complex formation by RNA polymerase at the *fis*-promoter *in vitro* [2].

2.3.9 Fis represses *gyrA* and *gyrB* expression

In normal cells, *gyrA* and *gyrB* mRNA production peaks between 15 and 30 minutes after nutritional shift-up before stabilizing at near to pre-shift-up levels. In CSH50Δfis cells, which lack Fis, expression was higher and stayed high 50 minutes after shift-up [18]. Using plasmid-borne *gyrA*- and *gyrB-promoter-lacZ* fusions, it was also found that β-galactosidase activity was significantly higher in cells lacking Fis than in wild-type constructs [18].
The same authors then proceeded to DNase I and potassium permanganate foot-printing to find Fis binding sites on linear *gyrA* and *gyrB* promoters : one Fis-binding site was found in the *gyrA*, two that of *gyrB*.

2.3.10 Fis activates *topA* expression

Fis needs to bind DNA to activate a promoter: [19] used the DNA binding-deficient mutant Fis protein R85C in CSH50*fis:kan*. In such cells, Fis usual effect was not observed.

A computer search for Fis binding sites in the promoter region of *topA* P1 indicated several possible sites. To test for Fis binding to the promoter region of *topA*, DNase I foot-printing experiments were

carried out using purified Fis. Fis was found to protect three sites centered at approximately 62, 93 and 129 upstream of the transcription start site of P1.[26]. This was confirmed experimentally: the response to hydrogen peroxide is activation of P1 promoter. Activation only occurs when Fis is present in the cell [26].

3 A first Model : Homeostasis

The most important property of our network is the homeostatic control of negative supercoiling. Our first aim is to establish a model able to reproduce this phenomenon. The only proteins acting on superhelicity in our network are Topoisomerase A (TopI) and Gyrase (Gyr) associated with ATP. They will be the only two regulated proteins in our first model. By doing such a simplification, we can check that for a given concentration of Fis and a given ATP/ADP ratio, a stable state QS can be found, and that when this state is perturbed by adding or removing Gyr or TopI, the system gets back to QS. This is homeostasis.

3.1 Modeling

In this section, we will explain our modeling choices and the resulting equations.

3.1.1 Fis and ATP/ADP ratio

To observe negative supercoiling, we decided to first work with a fixed Fis concentration and ATP/ADP ratio. In GNA syntax, this choice is represented by setting Fis and ATP/ADP ratio as input variables : their "concentration" is defined once for the simulation and will not change. The concentration of an input variable is $u_{input\,variable}$.
From now on, we will only speak of ATP/ADP ratio as "ATP concentration" (u_{ATP}), which is consistent with GNA syntax even though it is not biologically correct. This way, high concentrations of ATP will symbolize high values of ATP/ADP ratio, corresponding to a nutritional upshift, whereas low ATP level will simulate starvation conditions. There will be only two possible states for ATP : low and high. When ATP level is high, Gyrase can be active, otherwise Gyrase is inactive.
On the contrary, Fis has two different thresholds : *gyrAB* and *topI* don't have the same sensibility to Fis concentrations. Indeed, activation of *topI* requires a smaller amount of Fis than repression of *gyrAB* does. Although this might not have been proved in our bibliographic report, this is a logical relation : we want Gyrase to have a chance to be produced before Topoisomerase A level rises and counteracts its action. If *gyrAB* was repressed before *topI* was activated, our system would be static : nothing would happen.

3.1.2 Modeling negative superhelicity

As superhelicity is neither a gene nor a protein, and because it doesn't act as one of those, we can not represent it in GNA as we do for for other variables. Indeed, a gene needs a certain time to produce the protein, and then the action can take place. On the contrary, as soon as Gyrase or Topoisomerase A act, negative superhelicity level changes and target genes immediately react to superhelicity modification. See section 2 for more biological evidence.
A possible way to model negative supercoiling is then to take it out of the model, and just observe the levels of Gyrase, ATP/ATP and Topoisomerase A. For example, we will say that high concentrations of Gyr and ATP and low concentration of TopI correspond to high negative supercoiling whereas the same protein concentrations with low ATP level represent low negative supercoiling.
Although we have suppressed supercoiling from our model, we still have to represent its action on *gyrAB* and *topI* and the actions of Gyrase, ATP and Topoisomerase A on supercoiling. Gyrase concentration will therefore have a direct effect of *gyrAB* and *topI*, exactly as Topoisomerase.
More details on how negative superhelicity is represented are given in sections **3.1.4** and **3.1.5**.

3.1.3 Thresholds

We've already seen two thresholds for Fis concentration. Now, we need to define some for Gyrase and Topoisomerase. Each has a different answer to the level of supercoiling : Gyr will be produced when negative supercoiling is low, whereas TopI will be produced when negative supercoiling is high. We need to define high and low levels of negative supercoiling in terms of ATP, Gyrase and TopI concentrations. *gyrAB* will be sensible to high levels of Gyrase (and become inactive) but sensible to low levels of Topoisomerase (and become active). This hypothesis corresponds to the definition of two thresholds for Gyrase. Identically, we need two thresholds for Topoisomerase. These four thresholds are the base of negative supercoiling homeostasis (see **fig 4**).

3.1.4 Topoisomerase A synthesis

As we can see in the interaction graph (**fig 1**), *topA* is activated by Fis and by high levels of negative supercoiling. We have decided to separate the synthesis in two parts, each representing one promoter's action.

Fis activated synthesis We represent the synthesis of a protein by a step function. In the case of TopI by the Fis sensible promoter, it is quite simple : when Fis concentration goes above the threshold concentration θ_{Fis}^{topA}, TopoisomeraseI is produced, otherwise nothing happends. We also introduce a degradation term in our model, which is proportional to TopI concentration. Our model is not quantitative, but we still need equilibrium inequalities to define the system.

If we call κ_{TopI}^{byFis} the synthesis parameter of Fis activated synthesis and γ_{TopI} the degradation term, we find :

- partial equation :

$$\dot{x}_{TopI} = \kappa_{TopI}^{byFis} s^+(u_{Fis}, \theta_{Fis}^{TopI}) + (supercoiling\, activated TopI\, synthesis) - \gamma_{TopI} x_{TopI}$$

- equilibrium inequalities :

$$\theta_{TopI}^{sg} \leq \frac{\kappa_{TopI}^{byFis}}{\gamma TopI} \leq \theta_{TopI}^{st}$$

Where *sg* means "for negative supercoiling for *gyrAB*" and *st* is "for negative supercoiling for *topoisomeraseI*".

Negative supercoiling activated synthesis *topoisomeraseI* is activated by high levels of negative supercoiling. In our model, negative supercoiling is be evaluated by the relative concentrations of TopI, Gyr and ATP. We consider the level of negative supercoiling to be high when both Gyr and ATP are above a certain threshold, and TopI is under another threshold. Now, we can write :

- complete equation :

$$\dot{x}_{TopI} = \kappa_{TopI}^{byFis} s^+(u_{Fis}, \theta_{Fis}^{TopI}) + \kappa_{TopI}^{bySup} s^+(x_{Gyr}, \theta_{Gyr}^{st}) s^+(u_{ATP}, \theta_{ATP}) s^-(x_{TopI}, \theta_{TopI}^{st}) - \gamma_{TopI} x_{TopI}$$

- equilibrium inequalities :

$$\theta_{TopI}^{sg} \leq \frac{\kappa_{TopI}^{bySup}}{\gamma_{TopI}} \leq \theta_{TopI}^{st}$$

$$\theta_{TopI}^{st} \leq \frac{\kappa_{TopI}^{bySup}}{\gamma_{TopI}} \leq max_{TopI}$$

3.1.5 Gyrase synthesis

In the case of Gyrase, the problem is quite different since both Fis and negative supercoiling high levels repress *gyrAB*. Expression of *gyrAB* is then represented by one high synthesis rate effective when both negative supercoiling and Fis are at low levels. The degradation term is proportional to Gyrase concentration. Here is the simple equation we obtain :

$$\dot{x}_{Gyr} = \kappa_{Gyr} s^-(sup, \theta_{sup}^g) s^-(x_{Fis}, \theta_{Fis}^g) - \gamma_{Gyr} x_{Gyr}$$

However, as said before, we've taken negative supercoiling out of the model, so we have to write an equation with ATP, Fis and TopI instead of negative supercoiling. In TopI synthesis, we've seen that high negative supercoiling level was modeled by

$$s^+(x_{Gyr}, \theta_{Gyr}^{sg}) s^+(u_{ATP}, \theta_{ATP}^s) s^-(x_{TopI}, \theta_{TopI}^{sg})$$

To simplify our demonstration, let's write it this way : $(a \wedge \overline{b})$, where a represents active Gyrase and b *topoisomeraseI*. Now, we want to have low negative supercoiling, which is $\neg(a \wedge \overline{b})$. To understand what this negation means, let's look at the truth tables (**fig 3**)

$(a \wedge \overline{b})$

	$\overline{b}$	b
$\overline{a}$	0	0
a	1	0

high neg. supercoiling

$\neg(a \wedge \overline{b})$

	$\overline{b}$	b
$\overline{a}$	1	1
a	0	1

low neg. supercoiling

Figure 3: Truth tables used for Gyrase synthesis equation

The truth table applying to Gyrase synthesis corresponds to the following equation :

$$\dot{x}_{Gyr} = \kappa_{Gyr}[1 - s^+(x_{Gyr}, \theta^{sg}_{Gyr})s^+(u_{ATP}, \theta^s_{ATP})s^-(x_{TopI}, \theta^{sg}_{TopI})] \times s^-(u_{Fis}, \theta^g_{Fis}) - \gamma_{Gyr}x_{Gyr}$$

In the square brackets, we have the regulation by negative supercoiling : Gyrase will be synthesised all the time, except when all three conditions are verified : high concentrations of both Gyrase and ATP, and low concentrations of TopI. Moreover, the Fis regulated part states that Gyrase can only be produced when Fis concentration is under the threshold θ^g_{Fis}

3.2 Simulation results

3.2.1 General presentation

We use GNA to calculate all possible states corresponding to our model : this is a "complete analysis" of the system. The first step is to filter all non-instantaneous states. We won't keep instantaneous states because they don't have any biological signification. GNA produces a graph displaying all states and their relationships (**fig 5**).

As we can see on fig 5, there are six different groups of states, each corresponding to a unique final stable state. If we display the graph of all initial states, we can notice that each group corresponds to a combination of the possible concentrations of the input variables Fis and ATP (see **fig 6**).

This structure of the states is particularly interesting : it means that the final state only depends on the input variables Fis and ATP. Whatever the initial level of Gyrase or Topoisomerase A, we will obtain the same equilibrium state for a given concentration of Fis and ATP. Practically, if we take a stable state and modify Gyrase or Topoisomerase concentration (introduce perturbation in the system), it will go back to the same stable state. This proves that our system reproduces homeostasis : it is able to regulate the level of negative supercoiling.
Moreover, we've checked that the concentrations of the final states were constistent with ATP and Fis concentrations, and they are (see **fig 7**).

3.2.2 Example : Fis and ATP low

Let's look at **fig 11**. In this system, both Fis and ATP are low. It thus represents starving condition. In this state, Gyrase is not active (it requires high ATP concentrations) and *topoisomeraseI* is not activated by Fis. The level of negative supercoiling is low. However, as can be seen in the beginning of graph in **fig 11**, we introduce a high concentration of Topoisomerase A and very low concentration of Gyrase. With time, we obtain the final state where we have high concentrations of Gyrase and low concentrations of TopI.
In terms of negative supercoiling, the final state is quite perfect : Gyrase is inactive, but the concentration of TopI is low, so the level of negative supercoiling is simply low. However, we have an accumulation of Gyrase, which will be immediately available in case of nutrient upshift.

3.2.3 Reconstitution of the cycle

We are interested in the response of negative supercoiling to starvation and nutient upshift. A natural cycle would be : starvation → nutrient upshift → response → starvation. All these conditions correspond to one of the different combinations of Fis and ATP levels. However, high Fis concentrations when ATP

u_{ATP} :

Threshold : θ_{ATP}

u_{Fis} :

Thresholds : $\theta^t_{Fis}, \theta^g_{Fis}$

Thresholds inequalities : $\theta^t_{Fis} < \theta^g_{Fis}$

x_{TopI} :

Thresholds : $\theta^{sg}_{TopI}, \theta^{st}_{TopI}, \theta^f_{TopI}$

Thresholds inequalities : $\theta^{sg}_{TopI} < \theta^{st}_{TopI} < \theta^f_{TopI}$

State equation :

$$\dot{x}_{TopI} = \kappa^{byFis}_{TopI} s^+(u_{Fis}, \theta^{TopI}_{Fis}) + \kappa^{bySup}_{TopI} s^+(x_{Gyr}, \theta^{st}_{Gyr}) s^+(u_{ATP}, \theta_{ATP}) s^-(x_{TopI}, \theta^{st}_{TopI}) - \gamma_{TopI} x_{TopI}$$

Equilibrium inequalities :

$$\theta^{sg}_{TopI} \leq \frac{\kappa^{byFis}_{TopI}}{\gamma_{TopI}} \leq \theta^{st}_{TopI}$$

$$\theta^{sg}_{TopI} \leq \frac{\kappa^{bySup}_{TopI}}{\gamma_{TopI}} \leq \theta^{st}_{TopI}$$

$$\theta^{st}_{TopI} \leq \frac{\kappa^{bySup}_{TopI} + \kappa^{byFis}_{TopI}}{\gamma_{TopI}} \leq max_{TopI}$$

x_{Gyr} :

Thresholds : $\theta^{sg}_{Gyr}, \theta^{st}_{Gyr}, \theta^f_{Gyr}$

Thresholds inequalities : $\theta^f_{Gyr} < \theta^{st}_{Gyr} < \theta^{sg}_{Gyr}$

State equation :

$$\dot{x}_{Gyr} = \kappa_{Gyr}[1 - s^+(x_{Gyr}, \theta^{sg}_{Gyr}) s^+(u_{ATP}, \theta^s_{ATP}) s^-(x_{TopI}, \theta^{sg}_{TopI})] \times s^-(u_{Fis}, \theta^g_{Fis}) - \gamma_{Gyr} x_{Gyr}$$

Equilibrium inequalities :

$$\theta^{sg}_{Gyr} \leq \frac{\kappa_{Gyr}}{\gamma_{Gyr}} \leq \max_{Gyr}$$

Figure 4: First Model summary

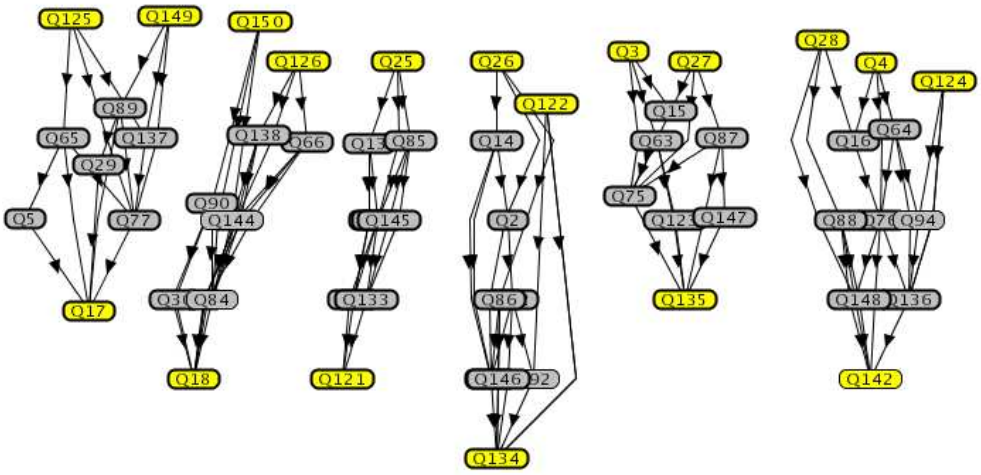


Figure 5: Display of all non-instantaneous states

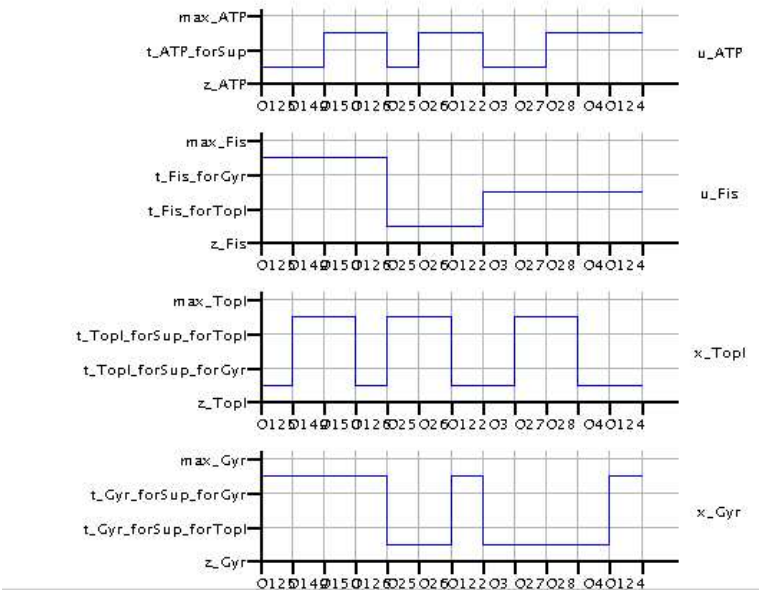


Figure 6: Graph of all initial states : order is not significatif

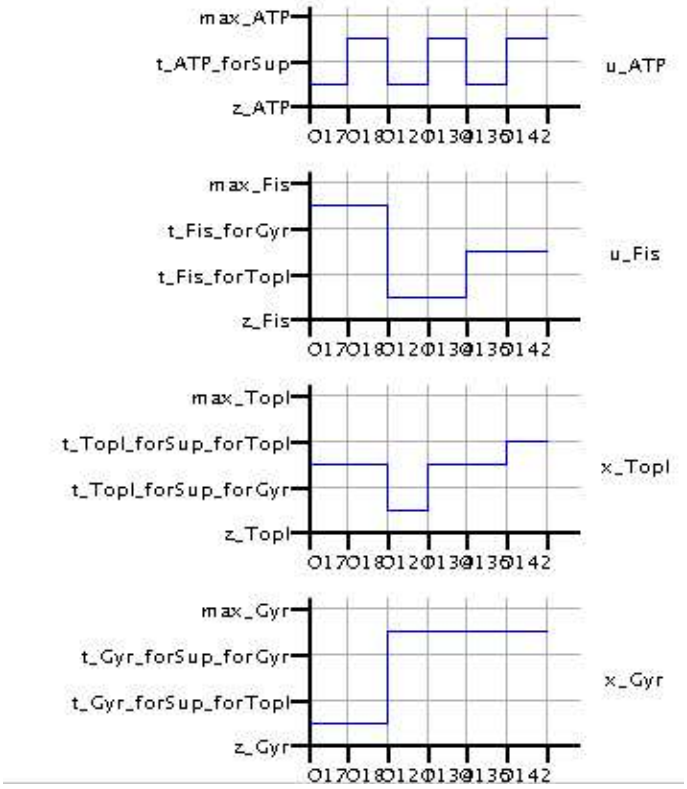


Figure 7: Graph of all final states. It allows us to compare the level of Fis and ATP and the level of Gyr and TopI to see if they are consistent

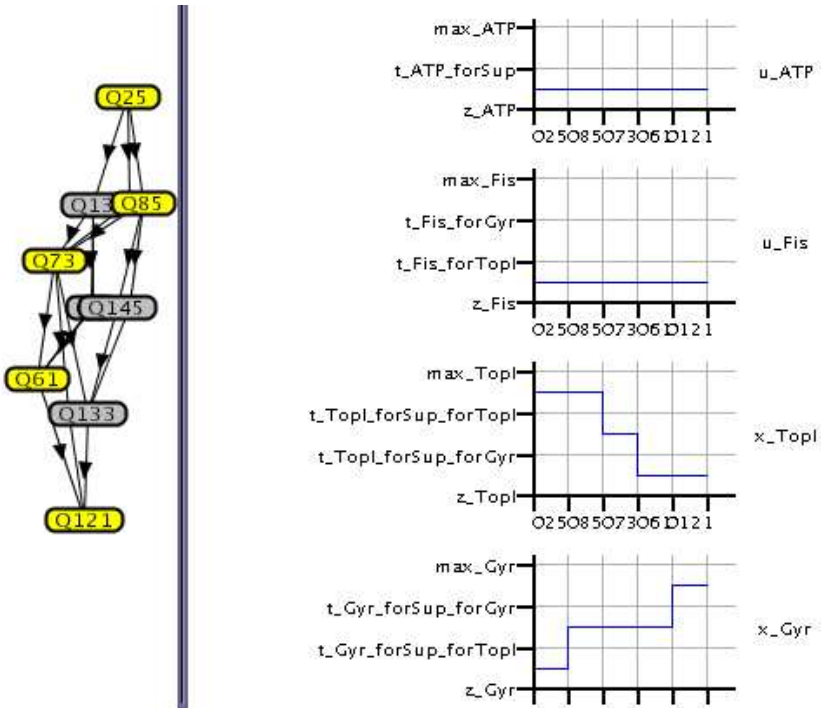


Figure 8: Example

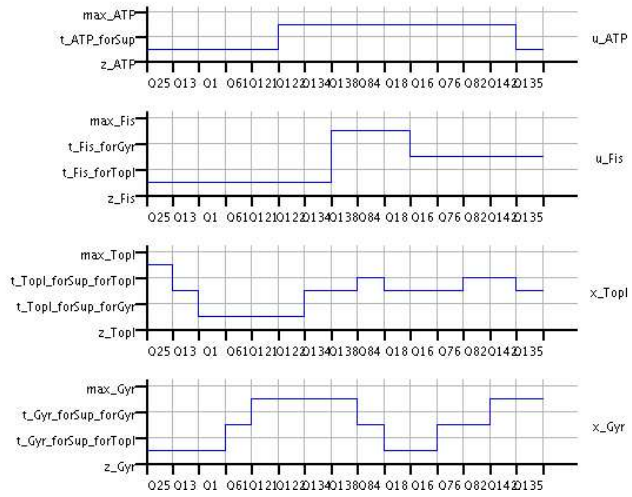


Figure 9: Reconstitution of the cycle : response of Gyr and TopI to ATP and Fis variations

is low don't have any biological signification, so we won't discuss them. The cycle is shown in **fig 9**. Schneider et al. [20] give cycle example : decrease of supercoiling level would decrease *fis* expression and hence relieve the repression of the *gyrA* and *gyrB* promoters by Fis [18]. The increased expression of gyrA and gyrB would then increase the Gyrase level to introduce negative supercoils (see 2.3.3). Because **the maximal activity of the fis promoter is achieved at higher than physiological superhelical densities** (σ= 0.075), the increase of supercoiling beyond the physiological level will, at least initially, activate the *fis* promoter and reinforce the repression of *gyrA* and *gyrB* by Fis. Eventually the physiological level of supercoiling will be reset and stabilized by steady-state levels of Fis and active Gyrase in the cell.

Starvation : ATP and Fis low This case was presented in the previous example (section 3.2.2). Its final state contain high concentrations of inactive Gyrase and low concentrations of TopI.

Upshift : ATP high and Fis low Immediately after the nutrient upshift, ATP level rises and so does negative supercoiling. Indeed, in the graph of initial states (unpublished results GRAPHE INITIAL STATES **fig 6**), we can see that state Q122 is the same as Q121, the final state of starvation, except that the level of ATP is now high. Consequently, with high active Gyrase concentrations, negative supercoiling level increases. Gyrase stays high but TopI starts being produced.

Fis level rises : ATP high and Fis high Biologically, we know that there is a spike of Fis approximately 15 minutes after the nutritional upshift. In our simulation, this corresponds to a moderate concentration of Topoisomerase and a really low concentration of Gyrase, since *gyrAB* is repressed by Fis. Supercoiling relaxes.

End of Fis spike : ATP high and Fis moderate In this condition, both TopI and Gyrase are high, and negative supercoiling level stays high.

Starvation Back to the beginning, with ATP and Fis low.

4 Complete model of our network

The first aim of our network was to reproduce negative supercoiling homeostatic control by Gyr and TopI, which was done in previous section. The next step was to introduce Fis as a state variable, instead of an input variable. Indeed, we want *fis* gene expression to be regulated by negative supercoiling, adapting Fis level to the situation. In this model, we want to see Fis concentration rise as soon as ATP is imported in the system, as negative supercoiling level becomes high. We should find afterwards that Fis level drops back to normal followed by a stabilisation of negative supercoiling.

4.1 Modeling

We globally kept the same model as in the previous section, except that we now consider *fis* regulation. Fis is sensitive to nutritional upshift: when submitted to nutritional upshift, Fis levels increase from less than 100 to over 50,000 copies per cell prior to the first cell division [2]. However, no direct control can be established, and we consider here that this reaction is due to variations in supercoiling. Fis is more sensitive to negative supercoiling than *topoisomerase I*, which means that it will start being translated at a lower negative supercoiling threshold. This has to be expressed in terms of TopI, Gyrase and ATP. As negative supercoiling is low without ATP, Fis can not be produced under the usual ATP threshold. However, it requires a smaller quantity of Gyrase to be produced, and *fis* can accept greater TopI concentrations before its transcription stops. Consequently, we had to add two new thresholds : θ^{sf}_{TopI} and θ^{sf}_{Gyr}, with sf "for negative supercoiling for *fis*".

4.2 Simulation results

This model is much more important than the last one, spawning so many states that it is not possible to do the complete analysis. To reduce the number of states calculated by GNA, we have to define initial conditions. In **fig 11**, we saw that final state Q121 corresponds to the bacteria starving condition stable state. To simulate nutritional upshift, we decided to use this stable state and simply switch the ATP level to high. It still gives us an impressive number of non instantaneous states that all belong to a cycle (see *fig 10*).

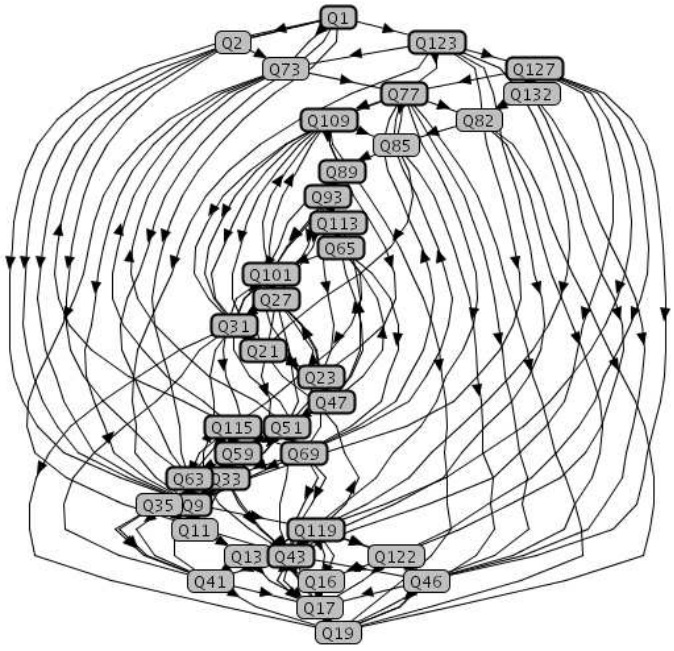


Figure 10: Non instantaneous states obtained with initial conditions corresponding to nutritional upshift after starvation equilibrium was atteined.

We can choose a path starting in Q1, and that comes back to Q31 after a while (see *fig* **??**). This is particularly interesting, because we can actually see what we were looking for.
We can explain the presence of the cycle, which has no biological reality, by the fact that our model is not complete. In particular, it doesn't contain Crp, which also regulates Fis production. It will be a next step in our modelling to put together all 3 small networks in order to have results closer to reality.

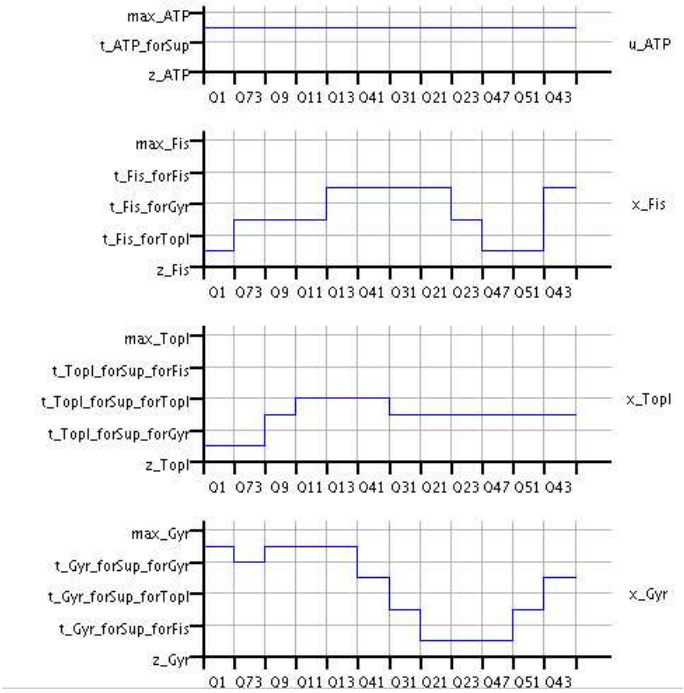


Figure 11: One of the possible paths possible just after nutritional upshift. Initial conditions in Q1 correspond to Q121 of the example. At the end, the path goes back to Q31, making a cycle.

5 Discussion

5.1 Our results

We had to work around GNA's limitations, notably the fact that we were not able to directly access or visualize the our model's most important element, the level of supercoiling. Despite this, we were able to accurately modelize our system and the results obtained were those we expected.

Supercoiling is homeostatically controlled through Topoisomerase I and Gyrase. When adding control over Fis to our system, we ended up stuck in a never-ending cycle of events. *E. coli* doesn't get stuck; this proves that our model is not sufficient as such to explain *E. coli*'s reaction to nutritional upshift. This did not come as a surprise, since Fis' expression is very closely linked to Crp, another pleiotropic regulator.

As stated in the introduction to this paper, our original plan was to assemble our network with two other sub-networks known to be involved in global regulation in *Escherichia coli*. This would certainly have been very interesting; the global network is so complex, it is hard to predict it's behavior. It is for larger networks that a tool such as GNA becomes essential. An interesting test of the complete model would have been "deleting" genes *in silico* and seeing whether real-life mutant's behavior is reproduced. Another possibility would be constitutive expression of a given gene such as *gyrAB* or *topA*.

5.2 Personal Experience

This project fit into our curriculum in Bio-informatics and modeling at Insa de Lyon. Work on this project was intermittent throughout the school-year. It taught us many things that will be useful to us in our future careers, including the following:

- Concerning modeling: we realized how important it is to know as much as possible about the biological system before trying to create a model describing it. One shouldn't wait until one knows everything because: one will never know everything, one learns by trial and error, and most importantly, deadlines are getting nearer!
- Concerning literature: we got the hang of reading scientific articles (after having read 20+ each!). Reading and synthesizing so large an amount of information is not easy!

- Scientists collaborate - they don't work by themselves. In particular, biologists have a growing need for computer scientists able to understand biological issues.
- Modeling is fun when it works!

6 Conclusion

Although not being able to assemble our network with the two modeled by our classmates is very frustrating, all is not lost. Indeed, Eva Laget, one of our classmates, will continue this project as part of her DEA next year (DEA is another French Masters-equivalent diploma).

7 Appendices

7.1 GNA model with Fis and ATP/ADP ratio as input variables

```
input-variable: u_ATP
  zero-parameter: z_ATP
  box-parameter: max_ATP
  threshold-parameters: t_ATP_forSup

input-variable: u_Fis
  zero-parameter: z_Fis
  box-parameter: max_Fis
  threshold-parameters: t_Fis_forTopI,  t_Fis_forGyr
  threshold-inequalities:
t_Fis_forTopI < t_Fis_forGyr;

state-variable: x_TopI
  zero-parameter: z_TopI
  box-parameter: max_TopI
  production-parameters: k_TopI_byFis, k_TopI_bySup
  degradation-parameters: g_TopI
  threshold-parameters: t_TopI_forSup_forGyr, t_TopI_forSup_forTopI
  state-equation:
d/dt x_TopI = k_TopI_byFis * s+(u_Fis, t_Fis_forTopI)
               + k_TopI_bySup * s+(x_Gyr, t_Gyr_forSup_forTopI)
* s+(u_ATP, t_ATP_forSup) * s-(x_TopI,t_TopI_forSup_forTopI)
- g_TopI * x_TopI

  threshold-inequalities:
  t_TopI_forSup_forGyr < t_TopI_forSup_forTopI;
  nullcline-inequalities:
      k_TopI_byFis > g_TopI * t_TopI_forSup_forGyr;
      k_TopI_byFis < g_TopI * t_TopI_forSup_forTopI;
      k_TopI_bySup > g_TopI * t_TopI_forSup_forGyr;
      k_TopI_bySup < g_TopI * t_TopI_forSup_forTopI;
      k_TopI_byFis + k_TopI_bySup > g_TopI * t_TopI_forSup_forTopI;
      k_TopI_byFis + k_TopI_bySup < g_TopI * max_TopI;


state-variable: x_Gyr
  zero-parameter: z_Gyr
  box-parameter: max_Gyr
  production-parameters: k_Gyr
  degradation-parameters: g_Gyr
  threshold-parameters: t_Gyr_forSup_forGyr, t_Gyr_forSup_forTopI
  state-equation:
d/dt x_Gyr = k_Gyr * (1 - s+(x_Gyr, t_Gyr_forSup_forGyr) *
     s+(u_ATP, t_ATP_forSup) * s-(x_TopI, t_TopI_forSup_forGyr)) *
     s-(u_Fis, t_Fis_forGyr) - g_Gyr * x_Gyr

  threshold-inequalities:
```

```
   t_Gyr_forSup_forTopI < t_Gyr_forSup_forGyr;
  nullcline-inequalities:
k_Gyr > g_Gyr * t_Gyr_forSup_forGyr;
k_Gyr < g_Gyr * max_Gyr;
```

7.2 GNA model with ATP/ADP ratio as the only input variable

7.2.1 Model

```
input-variable: u_ATP
  zero-parameter: z_ATP
  box-parameter: max_ATP
  threshold-parameters: t_ATP_forSup

state-variable: x_Fis
  zero-parameter: z_Fis
  box-parameter: max_Fis
  production-parameters: k_Fis
  degradation-parameters: g_Fis
  threshold-parameters: t_Fis_forTopI,  t_Fis_forGyr, t_Fis_forFis
  state-equation:
   d/dt x_Fis =  k_Fis * s+(x_Gyr, t_Gyr_forSup_forFis) * s+(u_ATP,t_ATP_forSup)
* s-(x_TopI, t_TopI_forSup_forFis)* s-(x_Fis, t_Fis_forFis)
- g_Fis * x_Fis

  threshold-inequalities:
t_Fis_forTopI < t_Fis_forGyr;
t_Fis_forGyr < t_Fis_forFis;

  nullcline-inequalities:
       k_Fis >g_Fis * t_Fis_forFis;
k_Fis < g_Fis * max_Fis;

state-variable: x_TopI
  zero-parameter: z_TopI
  box-parameter: max_TopI
  production-parameters: k_TopI_byFis, k_TopI_bySup
  degradation-parameters: g_TopI
  threshold-parameters: t_TopI_forSup_forGyr, t_TopI_forSup_forTopI, t_TopI_forSup_forFis
  state-equation:
d/dt x_TopI = k_TopI_byFis * s+(x_Fis, t_Fis_forTopI)
              + k_TopI_bySup * s+(x_Gyr, t_Gyr_forSup_forTopI)
* s+(u_ATP, t_ATP_forSup) * s-(x_TopI,t_TopI_forSup_forTopI)
- g_TopI * x_TopI

  threshold-inequalities:
  t_TopI_forSup_forGyr < t_TopI_forSup_forTopI;
  t_TopI_forSup_forTopI < t_TopI_forSup_forFis;

  nullcline-inequalities:
      k_TopI_byFis > g_TopI * t_TopI_forSup_forGyr;
      k_TopI_byFis < g_TopI * t_TopI_forSup_forTopI;
      k_TopI_bySup > g_TopI * t_TopI_forSup_forGyr;
      k_TopI_bySup < g_TopI * t_TopI_forSup_forTopI;
      k_TopI_byFis + k_TopI_bySup > g_TopI * t_TopI_forSup_forFis;
      k_TopI_byFis + k_TopI_bySup < g_TopI * max_TopI;


state-variable: x_Gyr
  zero-parameter: z_Gyr
  box-parameter: max_Gyr
  production-parameters: k_Gyr
  degradation-parameters: g_Gyr
  threshold-parameters: t_Gyr_forSup_forGyr, t_Gyr_forSup_forTopI, t_Gyr_forSup_forFis
```

```
  state-equation:
d/dt x_Gyr = k_Gyr * (1 - s+(x_Gyr, t_Gyr_forSup_forGyr)
* s+(u_ATP, t_ATP_forSup) * s-(x_TopI, t_TopI_forSup_forGyr))
* s-(x_Fis, t_Fis_forGyr)
- g_Gyr * x_Gyr

  threshold-inequalities:
   t_Gyr_forSup_forFis < t_Gyr_forSup_forTopI;
   t_Gyr_forSup_forTopI < t_Gyr_forSup_forGyr;

  nullcline-inequalities:
k_Gyr > g_Gyr * t_Gyr_forSup_forGyr;
k_Gyr < g_Gyr * max_Gyr;
```

7.2.2 Initial Conditions

```
initial-conditions:
         x_TopI >= z_TopI;
 x_TopI < t_TopI_forSup_forGyr;


 x_Gyr > t_Gyr_forSup_forGyr;
 x_Gyr <= max_Gyr;

 u_ATP <= max_ATP;
 u_ATP > t_ATP_forSup;

         x_Fis >= z_Fis;
 x_Fis < t_Fis_forTopI;
```

References

[1] Ali, Jackson, Howells, and Maxwell. The 43-kilodalton N-terminal fragment of the DNA gyrase B protein hydrolyzes ATP and binds coumarin drugs. *Biochemistry*, 32(10):2717–24, 1993.

[2] Ball, Osuna, Ferguson, and Johnson. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J. Bacteriol.*, 174(24):8043–56, 1992 ⋈.

[3] de Jong, Geiselmann, Hernandez, and Page. Modelling and simulation : Initiation of sporulation in *B. subtilis*. *Bioinformatics*, 19(3):336–344, 2003.

[4] Drlica. Control of bacterial DNA supercoiling - MicroReview. *Molecular Biology*, 6(4):425–433, 1992.

[5] Drlica and Snyder. Superhelical *Escherichia coli* DNA: Relaxation by Coumermycin. *J. Mol. Biol.*, 120:145–154, 1978 ⋈.

[6] Duquenne and Bessis. Fis-Crp Report. *Bimalerie Daily*, 2003.

[7] Gellert et al. *Proc Natl Acad Sci*, 73:4474–4478, 1976.

[8] Jensen, van der Weijden, Jensen, Westerhoff, and Snoep. Extensive regulation compromises the extent to which DNA gyrase controls DNA supercoiling and growth rate of *Escherichia coli*. *Eur. J. of Biochem.*, 266(3):865, 1999.

[9] Kurt W. Kohn. Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. *Molecular Biology of the Cell*, 10:2703–2734, 1999.

[10] Laget and Morin. rpoS Report. *Bimalerie Daily*, 2003.

[11] Massé and Drolet. *Escherichia coli* DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J Biol. Chem.*, 274(23):16659–64, 1999.

[12] Massé and Drolet. Relaxation of transcription-induced negative supercoiling is an essential function of *Escherichia coli* DNA Topoisomerase I. *Journal of Biol. Chem.*, 274(23):16654–16658, 1999.

[13] Menzel and Gellert. Regulation of the Genes for *E. coli* DNA Gyrase: Homeostatic Control of DNA Supercoiling. *Cell*, 34:105–115, 1983.

[14] Ninnemann, Koch, and Kahmann. The *E.coli fis* promoter is subject to stringent control and autoregulation. *The EMBO Journal*, 11(3):1075–1083, 1992.

[15] Pruss, Manes, and Drlica. *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell*, 31(1):35–42, 1982.

[16] Reece and Maxwell. DNA gyrase: structure and function. *Crit. Rev. Biochem. Mol. Biol.*, 26(3-4):335–75, 1991 ⋈.

[17] Reece and Maxwell. The C-terminal domain of the *Escherichia coli* DNA gyrase A subunit is a DNA-binding protein. *Nucleic Acids Res.*, 19(7):1399–405, 1991 ⋈.

[18] Schneider, Travers, Kutateladze, and Muskhelishvili. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.*, 34(5):953–964, 1999.

[19] Schneider, Travers, and Muskhelishvili. FIS modulates growth phase-dependent topological transitions of DNA in *Escherichia coli*. *Mol. Microbiol.*, 26(3):519–530, 1997.

[20] Schneider, Travers, and Muskhelishvili. The expression of the *Escherichia coli fis* gene is strongly dependent on the superhelical density of DNA. *Mol. Microbiol.*, 38(1):167–175, 2000.

[21] Snoep, van der Weijden, Andersen, Westerhoff, and Jensen. DNA supercoiling in *Escherichia coli* is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. *Eur. J. Biochem*, 269:1662–1669, 2002.

[22] Sugino and Cozzarelli. The intrinsic ATPase of DNA gyrase. *J. Biol. Chem.*, 255(13):6299–306, 1980 ⋈.

[23] Travers and Muskhelishvili. DNA microloops and microdomains: A general mechanism for transcription activation by torsional transmission. *Journ. of Mol. Biol.*, 279(5):1027–1043, 1998.

[24] Tse-Dinh. Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Res.*, 13(13):4751–63, 1985.

[25] van Workum, van Dooren, Oldenburg, Molenaar, Jensen, Snoep, and Westerhoff. DNA supercoiling depends on the phosphorylation potential in *Escherichia coli*. *Mol. Mocrobiol.*, 20(2):351–360, 1996.

[26] Weinstein-Fischer, Elgrablyweiss, and Altuvia. *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, Topoisomerase I and Fis. *Mol. Microbiol.*, 35(6):1413, 2000.

[27] Westerhoff, O'Dea, Maxwell, and Gellert. DNA supercoiling by DNA gyrase. A static head analysis. *Cell Biophys.*, 12:157–181, 1988 ⋈.