

Producing a Set of Models for the Iron Homeostasis Network

Nicolas Mobilia

UJF-Grenoble 1 / CNRS
TIMC-IMAG UMR 5525,
Grenoble, F-38041, France
nicolas.mobilia@imag.fr

Alexandre Donzé

EECS Department
University of California Berkeley
Berkeley, CA 94720 USA
donze@eecs.berkeley.edu

Jean Marc Moulis

Univ. Grenoble Alpes, Laboratory of Fundamental and Applied Bioenergetics (LBFA),
and Environmental and Systems Biology (BEeSy), Grenoble, France
Inserm, U1055, Grenoble, France
CEA-iRTSV, Grenoble, France
jean-marc.moulis@cea.fr

Éric Fanchon

UJF-Grenoble 1 / CNRS
TIMC-IMAG UMR 5525,
Grenoble, F-38041, France
eric.fanchon@imag.fr

1 Introduction

The continuous parts of complex biological systems are often modeled by use of Ordinary Differential Equations (ODE). When experimental data are available, they usually have large variability, for example due to variability in cell cultures. Since data on a given system are scarce, one generally uses data from different cell types, different organisms, or different conditions. All this translates into large parameter uncertainty. To cope with this situation and try to integrate all available data in a consistent model, we represent such data by intervals rather than single numerical values. The ranges of the intervals vary depending on the type of experiment and the nature of the experimental system. Some parameters or concentrations are not known at all and are initially defined to belong to the physiological domain. This set of intervals define the search space. Other experimental data are expressed in terms of inequalities involving derived quantities. Our goal is to build a set of models that satisfy all the constraints deduced from experiments, and to analyze the salient features of the dynamics of this set of models.

We present a method for modeling biological systems which combines formal techniques on intervals, numerical simulations and formal verification of STL (Signal Temporal Logic) formula. This allows us to consider intervals for each parameter and to describe the expected behavior of the model. We apply this method to the modeling of the cellular iron homeostasis network in precursors of erythroid cells. A core model [5] has been presented previously. Herein, we describe a more evolved model in which the regulation mechanism acting at the translational level is explicitly considered. This leads to a larger model with more parameters and the integration of newly obtained experimental data. This new model provides a more detailed description of the regulatory mechanism, including quantitative considerations pertaining to the involved species, and it should allow us to more precisely address pending biological questions. The higher level of complexity of this model, compared to the core model, required the development of a method to characterize efficiently steady states.

In Section 2, we describe the iron homeostasis network, and, in Section 3, the corresponding model. Then, in the Section 4, we describe the method used. We finally explain the work that remains to be done and conclude.

2 Biological system

Iron is an essential element for mammalian cells (eg. hemoglobin contains iron), but if present in too high quantity iron has a deleterious effect. The level of available iron is thus finely tuned in mammalian cells. Our goal is to describe and understand this regulatory mechanism. The regulatory network, described in Figure 1, is composed of fifteen species. The species *Fe* (pool of available iron), *IRP* (Iron Regulatory Protein), *Ft* (ferritin), *FPN1a* (ferroportin) and *TfR1* (transferrin receptor) were present in the previous model [5]. The other ones are the mRNA of these proteins either in a free form or in a form complexed with an IRP. The central regulatory mechanism, based on the IRP is described in Mobilia & al [5].

In a nutshell, the available data belong to several categories. A qualitative description of the dynamics, obtained from a large body of biological experiments is the following: if the amount of iron is sufficient, the cell is in a steady state. From this state, if an iron input cut-off occurs, the amount of iron in cells decreases and the IRP are activated, leading to increased binding of IRP to IRE-containing mRNAs. New kinetic parameters have been measured, as well as absolute mRNA concentrations in the iron-replete regime. Our aim is to build models which simulate the behavior of the biological system. From the iron-replete steady state, the evolution of the concentrations of the different species must be qualitatively reproduced upon cutting off the iron supply.

3 Model

We model this system with fifteen differential equations. These equations contain 28 parameters. In the following sub-section, we exhibit the equations related to the ferritin, the transferrin receptor and the IRP. The conventions for the parameter names are the following: a parameter named p_X represents the transcription speed of the mRNA of X (units: mol/(L·s)); t_X represents the reaction rate constant for X mRNA translation (units: s^{-1}); dp_X represents the degradation rate of the protein X (units: s^{-1}); dr_X represents the degradation rate of the mRNA of X (units: s^{-1}); drs_X represents the degradation rate of the mRNA of X , when this mRNA contains an IRE in the 3'-UTR region and is bound to an IRP (units: s^{-1}). The species ending with the subscript $_p$ represent proteins, while the ones ending with the subscript $_f$ (resp. $_b$) represent free (resp. bound) mRNA concentration (units: mol/L).

3.1 Ferritin equations

The ferritin mRNAs contain an IRE in the 5'-UTR region, so the translation speed is proportional to the free mRNA concentration. This is described by the first term in Equation (1). Moreover, as a ferritin is constituted by 24 sub-units, a factor 1/24 appears in this term. The second term represents the spontaneous degradation of the ferritin.

$$\frac{dFt_p}{dt} = (t_{Ft}/24) \cdot Ft_f - dp_{Ft} \cdot Ft_p \quad (1)$$

The free ferritin mRNA concentration, described in Equation (2), depends on four terms. The first one is the transcription speed p_{Ft} . The second one represents the complexation of this mRNA with an IRP (the parameter ka is the complexation second-order reaction rate parameter), while the third one represents the decomplexation of mRNA:IRP complex. The parameter K_d is equal to the ratio kd/ka , where kd is the decomplexation first-order reaction rate parameter. The last term represents the spontaneous

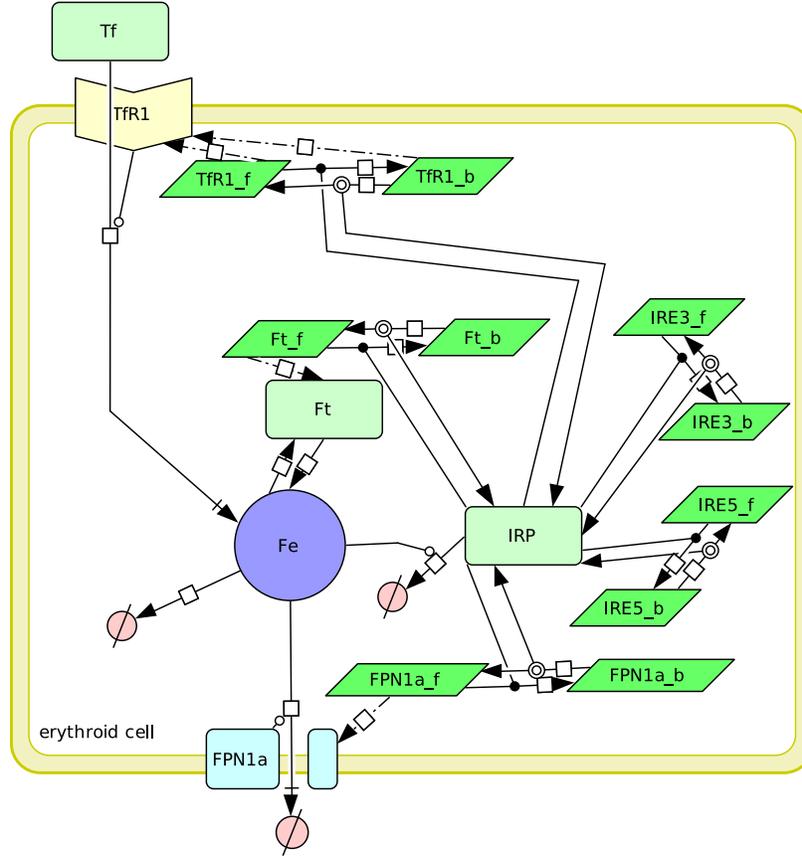


Figure 1: Schematic representation of the main biological processes involved in the cellular control of iron concentration. This diagram was drawn with the software CellDesigner [2]. The dashed arrows represent translation of mRNA into proteins. The lines ending with a combined perpendicular stroke and arrow represent iron transport through membranes. The regular arrows leading to an empty set symbol which indicate either degradation (for *IRP*) or internal consumption (for iron). Moreover, the multi-arrows containing a black dot represent complexation while the ones with two empty circles mean decomplexation. Finally, the two regular arrows represent the loading/unloading of iron into/from the ferritins. The rounded rectangles represent proteins, the parallelograms represent mRNA, and the circle labeled *Fe* represents the pool of available iron. The concave hexagon represents the transferrin receptor. The species *IRE5_{-f}* (resp. *IRE3_{-f}*) and *IRE5_{-b}* (resp. *IRE3_{-b}*) represent all the mRNA having an IRP binding site (called IRE) in the 5'-UTR (resp. 3'-UTR) excepted the ones explicitly drawn.

degradation of the mRNA.

$$\frac{dFt_{-f}}{dt} = p_{Ft} - ka \cdot Ft_{-f} \cdot IRP + ka \cdot K_d \cdot Ft_b - dr_{Ft} \cdot Ft_{-f} \quad (2)$$

Finally, the bound ferritin mRNA concentration is described in Equation (3). This equation is composed of three terms. The first two, describing the complexation and the decomplexation, have the same meaning as in the equation of the free ferritin mRNA. The last term describes the spontaneous degrada-

tion of the mRNA.

$$\frac{dFt_{-b}}{dt} = ka \cdot Ft_{-f} \cdot IRP - ka \cdot K_d \cdot Ft_{-b} - dr_{Ft} \cdot Ft_{-b} \quad (3)$$

3.2 Transferrin receptor equation

The transferrin receptor mRNA contains five IREs in its 3'-UTR region. Here, we make a simplification and consider that TfR1 mRNA contains only one IRE. Equation (4) describes the transferrin receptor concentration. It contains a translation term and a spontaneous degradation term. As the IRE is located in the 3'-UTR region, both free and bound mRNA are translated into proteins. Moreover, because the receptor is a dimer, a factor 1/2 appears in the first term.

$$\frac{dTfR1_{-p}}{dt} = (t_{TfR1}/2) \cdot (TfR1_{-f} + TfR1_{-b}) - dp_{TfR1} \cdot TfR1_{-p} \quad (4)$$

The equation of the free transferrin receptor mRNA is very similar to the free ferritin mRNA. This equation is shown in Equation (5) and is composed of a translation term, a complexation term, a decomplexation term and a spontaneous degradation term.

$$\frac{dTfR1_{-f}}{dt} = p_{TfR1} - ka \cdot TfR1_{-f} \cdot IRP + ka \cdot K_d \cdot TfR1_{-b} - dr_{TfR1} \cdot TfR1_{-f} \quad (5)$$

Equation (6) describes the concentration of the transferrin receptor mRNA complexed with an IRP. This equation is similar to the bound ferritin mRNA, except that the binding of an IRP on this mRNA leads to the mRNA stabilization. To model this mechanism, a specific degradation rate parameter (drs_{TfR1}) is considered. This parameter is lower than the free mRNA degradation rate parameter (dr_{TfR1}).

$$\frac{dTfR1_{-b}}{dt} = ka \cdot TfR1_{-f} \cdot IRP - ka \cdot K_d \cdot TfR1_{-b} - drs_{TfR1} \cdot TfR1_{-b} \quad (6)$$

3.3 IRP equation

The IRP equation is described in Equation (7).

$$\begin{aligned} \frac{dIRP}{dt} = & - (Ft_{-f} + FPN1a_{-f} + TfR1_{-f} + IRE3_{-f} + IRE5_{-f}) \cdot ka \cdot IRP \\ & + (Ft_{-b} + FPN1a_{-b} + TfR1_{-b} + IRE3_{-b} + IRE5_{-b}) \cdot ka \cdot K_d \\ & - k_{Fe \rightarrow IRP} \cdot sig^+(Fe, \theta_{Fe \rightarrow IRP}) \cdot IRP - dp_{IRP} \cdot IRP \end{aligned} \quad (7)$$

The first line of this equation describes the complexation of free mRNAs and IRP for all mRNAs, while the second line describes the decomplexation of the bound mRNAs. The last line describes IRP inactivation. This inactivation is described by a constant basal term for degradation ($dp_{IRP} \cdot IRP$) and the iron-triggered regulation ($k_{Fe \rightarrow IRP} \cdot sig^+(Fe, \theta_{Fe \rightarrow IRP}) \cdot IRP$). Then, if the iron level is significantly below the threshold $\theta_{Fe \rightarrow IRP}$, the degradation rate is $dp_{IRP} \cdot IRP$. Otherwise, if the iron concentration is significantly above this threshold, the degradation rate is $(k_{Fe \rightarrow IRP} + dp_{IRP}) \cdot IRP$, where $k_{Fe \rightarrow IRP}$ is the parameter describing the inhibition of IRP by iron.

3.4 Other equations

The iron equation is the same than in the previous model [5]. The eight remaining equations are very similar to those shown above. The equations for ferroportin and *IRE5* are similar to that of ferritin, and the equations for the *IRE3* species are similar to that of the transferrin receptor.

3.5 Data

For each parameter, we consider an interval deduced from biological data or incorporating meaningful values. For example, in K562 cells, the ferritin half-life is 11 hours [4]. We deduce that the parameter dp_{-Ft} is included in the interval $[3.8e-6, 3.8e-5] s^{-1}$.

Moreover, some data are expressed as relations between parameters. To give an example, our recent data indicate the ferritin mRNA concentration largely exceed that of the other IRP targets in proliferating cells. The total ferritin (resp. IRE5) mRNA concentration at steady state being equal to p_{-Ft}/dr_{-Ft} (resp. p_{-IRE5}/dr_{-IRE5}), it follows the relation described in Equation (8).

$$\frac{p_{-Ft}}{dr_{-Ft}} > \frac{p_{-IRE5}}{dr_{-IRE5}} \quad (8)$$

We can also note that the lower degradation rate due to the binding of IRP on IRE in 3'-UTR mRNA region translates into Equation (9) and Equation (10).

$$drs_{-TfR1} < dr_{-TfR1} \quad (9)$$

$$drs_{-IRE3} < dr_{-IRE3} \quad (10)$$

In addition, some relations describe data related to the stationary state. For example, the degradation rate of total TfR1 mRNA belong to the interval $[7.0 \times 10^{-6}, 7.0 \times 10^{-5}] s^{-1}$ [7][6]. This describe the sum of the degradation of both free and complexed mRNA and translates into the equation (11). The superscript eq indicates that we consider the steady state concentration.

$$\frac{dr_{-TfR1} \cdot TfR1_{-f}^{eq} + drs_{-TfR1} \cdot TfR1_{-b}^{eq}}{TfR1_{-b}^{eq} + TfR1_{-f}^{eq}} \in [7.0 \times 10^{-6}, 7.0 \times 10^{-5}] s^{-1} \quad (11)$$

The last kind of data is related to the dynamic of the system when an iron cut-off happens. The modeling of these data using STL formula is described in Mobilia & al [5].

4 Method

The set of intervals and constraints can be divided in two: those pertaining to the iron-replete steady state, and those pertaining to the cell response to iron shortage. In our previous work [5], we first reduced the search space by using the interval solver Realpaver [3]. Then, we represented formally the whole set of constraints as an STL formula and devised a search algorithm to satisfy it, based on the tool Breach [1]. Basically, a point is randomly drawn in the search space, a simulation is performed and the STL formula is evaluated. In the present more complete model, no iron-replete steady state was initially found following the same procedure. In addition, this failure did not instruct us on the origin of the problem.

To cope with this limitation, we improved the method in two ways. (i) In the first step, the interval solver Realpaver allowed to reduce the intervals by propagating the constraints. Instead of being solely a hyper-rectangle as previously, the search space was allowed to be a union of hyper-rectangles thus reducing it more efficiently. In case the interval solver found an inconsistency, we improved the method by looking for the smallest sets of constraints that have to be lifted to release the inconsistency. This information gives insight when one wants to revisit the model and the data used. (ii) We decomposed the

search algorithm in two parts. We had developed an algorithm to generate efficiently a large number of explicit solutions (steady states concentrations and model parameters) satisfying the constraints of a stable steady state (the set of constraint contains algebraic equations and inequalities involving polynomial expressions). These explicit solutions, that were prerequisite to perform simulation of the dynamics of the system, were then fed into our Breach-based procedure in order to search models satisfying the STL formula specifying the cell response to iron deprivation.

The unknowns of the problem are the model parameters and the concentrations in the iron-replete steady state. The methodology proceeds basically as follows:

1. perform interval reduction with Realpaver;
2. select a subset of unknowns to be sampled (we start with unknowns within a narrow interval, then other criteria are used to decouple the equations and to optimize the following step);
3. for each sample of this subset of unknowns:
4. replace the instantiated unknowns in the algebraic equations and perform deductions (new unknowns can get instantiated);
5. check domain of newly instantiated unknowns;
6. check the validity of inequality constraints as soon as possible;
7. for each sub-problem (i.e.: set of decoupled constraints), apply this algorithm;
8. loop to step 3. until all samples are tried;
9. loop to step 2.

The basic principles underlying this search are to decouple the constraints in order to solve subproblems, and to identify the hardest sub-problems (most constrained) and try to solve them in priority. The aim is of course to trim the branches of the search tree as soon as possible.

The interval solver Realpaver, used during the first step, allows to reduce the intervals by propagating the constraints. The result is an hyper-rectangle (or an union of hyper-rectangles) containing the solutions, if they exist. The existence of solutions is not guaranteed, but it is certain that there are no solutions outside of the volume given by Realpaver. Consequently, this step is important to reduce the search space. Nevertheless, the remaining space may be very large with regard to the solution space. A simple example to illustrate this aspect is the following: consider two unknowns x_1 and x_2 , within the $[0, 1]$ interval, and the constraint $abs(x_1 - x_2) < eps$, with eps small compared to x_1 and x_2 . If the solution space is defined by one box, Realpaver cannot reduce the search space. Considering a union of boxes allows a reduction of the search space. As it is hard to find explicit solutions, we say informally that the constraint is hard to satisfy (the smaller eps , the harder it is).

The constraint system is constituted of algebraic equations and inequalities. For the majority of them, the algebraic equations are used to deduce the values of unknowns, and are thus automatically satisfied. Inequalities are checked *a posteriori*. To be efficient, it is important to check an inequality as soon as all the unknowns in it have been instantiated. For efficiency reasons, redundant constraints are also added. The aim is to add constraints simpler than the initial ones, which can be checked early in the search process. Typically, from the constraints described by Equation (11), we straightforwardly deduce that $t_{TfR1} \cdot TfR1_{-f}^{eq} < 5.5 \times 10^{-13}$ and that $t_{TfR1} \cdot TfR1_{-b}^{eq} < 5.5 \times 10^{-13}$. Even if $TfR1_{-f}$ or $TfR1_{-b}$ is not instantiated, one of these two constraints can be checked and may invalidate this partial instantiation.

When applying this algorithm, we store the result of each verification of domain and check of constraints (whether the constraint or domain is verified or not). When no solution is found, this may be due

to different constraints. This information thus provide the level of difficulty to satisfy each constraint. This may help to manually found inconsistencies between constraints that were not automatically found by Realpaver.

This methodology either provides us valid sets of values, or indicates the hardest subset of constraints to satisfy. We applied it on the iron homeostasis network. No solutions could be found, which is not a proof of nonexistence, but the subset of hard constraints identified allowed us to prove that there was indeed a contradiction. After revision of this part of the model (namely: removing one non-reliable constraint and extending some intervals), the procedure then generated thousands of valid steady states in a short execution time.

5 Conclusion

This evolved model describes in a more realistic way the action of IRP and takes into account the fact that their effect depends on the location of the binding site on mRNA. Moreover, it easily incorporates new information obtained on the system.

Nevertheless, some work still remains to be done in order to completely automate this search, and to interface the steady state search with the part dealing with dynamical behavior (specified with an STL formula).

Here we described our approach to nicely integrate different kinds of biological data, combining an interval solver, simulations, and temporal STL formula verification. We are applying it on a new model of iron homeostasis in mammalian cells.

Acknowledgements

This work was supported by Microsoft Research through its PhD Scholarship Programme and by the following grants: Région Rhône-Alpes Cible 2010, IXXI-Spring 2012, and Univ. J. Fourier-Agir 2013.

References

- [1] A. Donzé (2010): *Breach, A Toolbox for Verification and Parameter Synthesis of Hybrid Systems*. In: *CAV*, pp. 167–170, doi:10.1007/978-3-642-14295-6_17.
- [2] A. Funahashi, M. Morohashi, H. Kitano & N. Tanimura (2003): *CellDesigner: a process diagram editor for gene-regulatory and biochemical networks*. *BIOSILICO* 1(5), pp. 159–162, doi:10.1016/S1478-5382(03)02370-9.
- [3] L. Granvilliers & F. Benhamou (2006): *Algorithm 852: RealPaver: an interval solver using constraint satisfaction techniques*. *ACM Transactions on Mathematical Software* 32(1), pp. 138–156, doi:10.1145/1132973.1132980.
- [4] T. Z. Kidane, E. Sauble & M. C. Linder (2006): *Release of iron from ferritin requires lysosomal activity*. *Am. J. Physiol., Cell Physiol.* 291(3), pp. C445–455, doi:10.1152/ajpcell.00505.2005.
- [5] N. Mobilia, A. Donzé, J. M. Moulis & É. Fanchon (2012): *A Model of the Cellular Iron Homeostasis Network Using Semi-Formal Methods for Parameter Space Exploration*. *EPTCS* 92, pp. 42–57, doi:10.4204/EPTCS.92.4.
- [6] C. Seiser, M. Posch, N. Thompson & L. C. Kuhn (1995): *Effect of transcription inhibitors on the iron-dependent degradation of transferrin receptor mRNA*. *J. Biol. Chem.* 270(49), pp. 29400–29406, doi:10.1074/jbc.270.49.29400.
- [7] L. V. Sharova, A. A. Sharov, T. Nedorezov, Y. Piao, N. Shaik & M. S. Ko (2009): *Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells*. *DNA Res.* 16(1), pp. 45–58, doi:10.1093/dnares/dsn030.