Stochastic Modeling of Gene Regulatory Networks in *Escherichia coli*

Rodrigo Santibáñez[1,2,*], Daniel Garrido[2], Tomás Pérez-Acle[1,3] & Alberto J. M. Martin[1,3,4,*]

*rsantibanez@dlab.cl and ajmm@dlab.cl

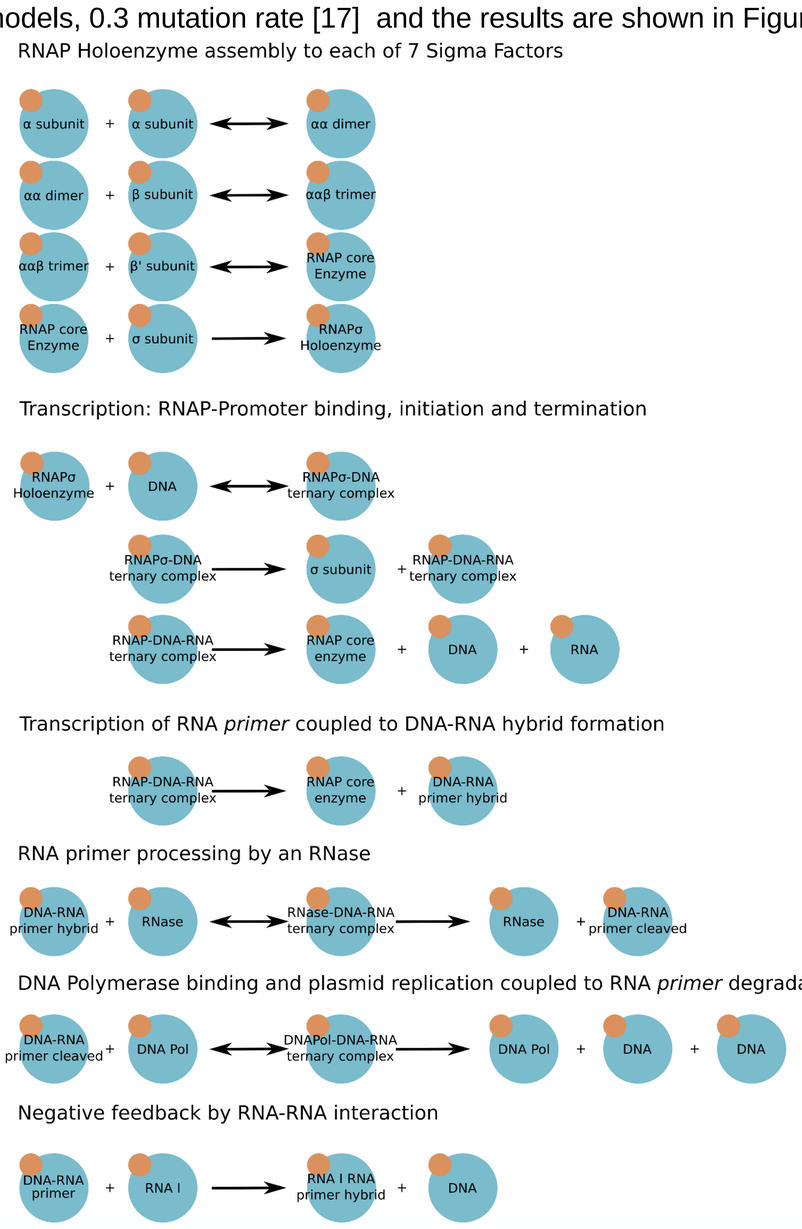
1 Computational Biology Lab, Fundación Ciencia & Vida. Avda. Zañartu 1482, Ñuñoa, Santiago, Chile.
2 Laboratorio de Microbiología de Sistemas, School of Engineering, Pontificia Universidad Católica de Chile. Avda. Vicuña Mackenna 4860, Macul, Santiago, Chile.
3 Centro Interdisciplinario de Neurociencia de Valparaíso, Universidad de Valparaíso. Pasaje Harrington 287, Playa Ancha, Valparaíso, Chile.
4 Centro de Genómica y Bioinformática, Universidad Mayor. Camino La Pirámide 5750, Huechuraba, Santiago, Chile.

Introduction.

Synthetic Biology has the ultimate objective of designing cells, tissues, organisms and communities to produce predictable responses [1]. Interestingly, mathematical and computational modeling has impacted prominently Synthetic Biology, where the manipulation of biological systems is cost-intensive, and computational resources could leverage experimental procedures [2–4]. Traditionally, differential equations (ODEs) have been employed, but their assumptions are not as realistic as those made by other approaches. Particularly, it has been known that biological systems are stochastic, discrete and structurally complex, hampering ODEs to fit these properties [5]. To further resolve a connection between modeling and designing organisms, we present a Rule-Based Model of Gene Regulatory Networks (GRNs) in *Escherichia coli* simulated using the Gillespie's Stochastic Simulation Algorithm [6,7]. Under this approach, rules are macroscopic chemical reactions between entities that recapitulate one or several patterns necessary for a transformation [8].

Methodology. Models were developed from literature GRNs [9,10] within the *Kappa BioBrick Framework* [11], exported to *kappa* with PySB [12] and simulated 1000 Arbitrary Time Units with PISKaS v1.3 [13]. Modeled rules are depicted graphically in following figure [13].

- The **Core GRN Model (Figure 1, left)** recapitulates the interaction of the *E. coli* RNAP with each of the 7 Sigma Factors (σ's) and the control of its transcription [9]. The GRN has 10 nodes and 28 edges. Three settings were simulated: one RNAP and 7 σ's (Figure 2A, blue dots); 7 RNAPs and 7 σ's (Figure 2A, green dots); and 28 RNAPs and 28 σ's (Figure 2A, red dots) distributed as many regulations each σ's has in the GRN. Results are shown in percentage of the total RNAP.
- The **Plasmid Copy Regulation Model (Figure 1, right)** recapitulates the expression and regulatory interaction of RNA I and the RNA *primer* from plasmid ColEI [14], which transcription was assume controlled only by σ70. Two settings were simulated: no interaction between the non-coding RNAs (Figure 2B, blue line) and with interaction between them at an arbitrary rate of 0.5 (Figure 2B, green line).
- A **Genetic Algorithm (GA)** was developed to read a *kappa* model and write new models with modified parameter values in an arbitrary range selected by the user. The *script* calls KaSim [15] or PISKaS [13] to run simulations inside a SLURM *sbatch* task. Finally, all simulations are ranked using a distance-based function and an arbitrary number of models are selected to generate the subsequent model population. The GA was tested using the Core GRN Model and data from [16] with *default* options: 100 individuals, 100 iterations, 10 best models, 0.3 mutation rate [17] and the results are shown in Figure 3.



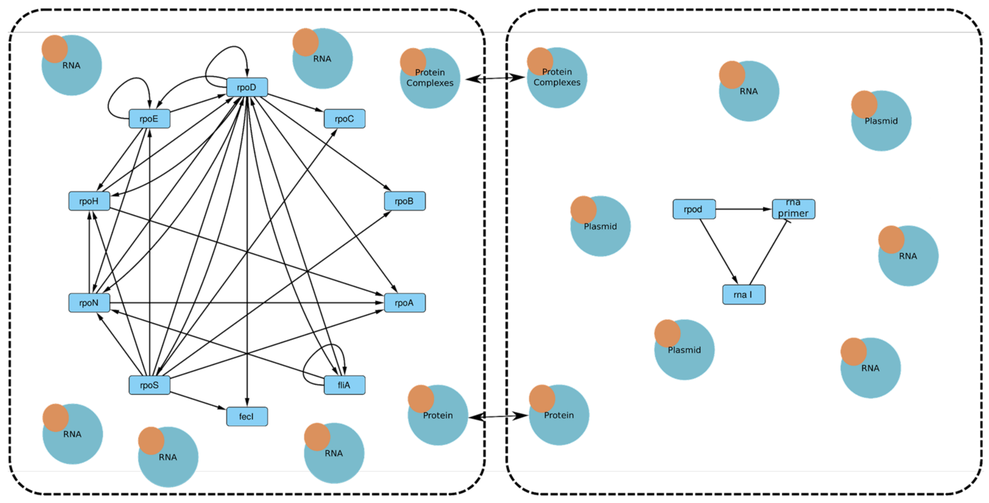


Figure 1. Representation of the Core GRN (Left) and the Plasmid Copy Regulation Model (Right) as a two-compartment model. Only agents that represent free proteins and protein complexes are able to move between compartment as depicted by the solid double arrows.

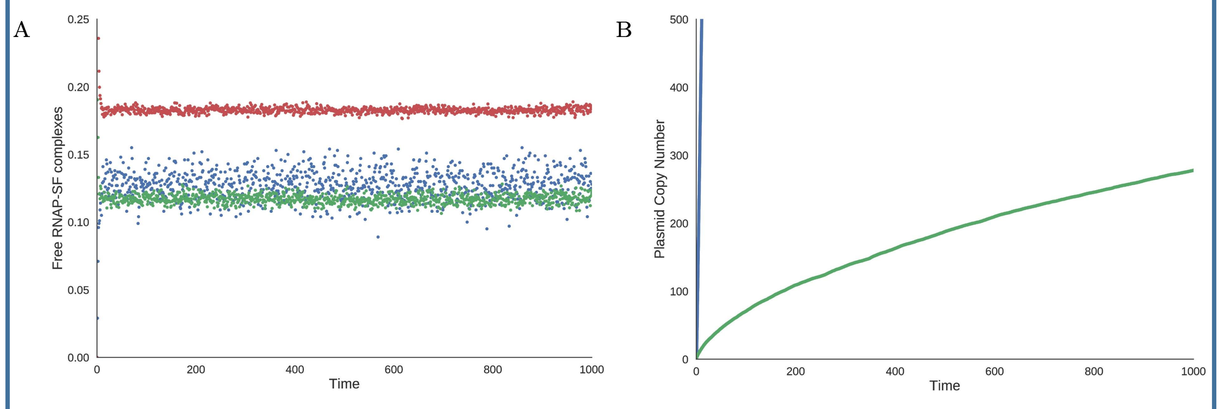


Figure 2. A, The Core GRN Model simulates free RNAP-σ complexes with good agreement with reported values [18] and reduce variability with increasing protein availability (red dots); **B,** the Plasmid Copy Regulation Model predicts a saturation dynamic when there is an explicit interaction between the non-coding RNAs (green line).

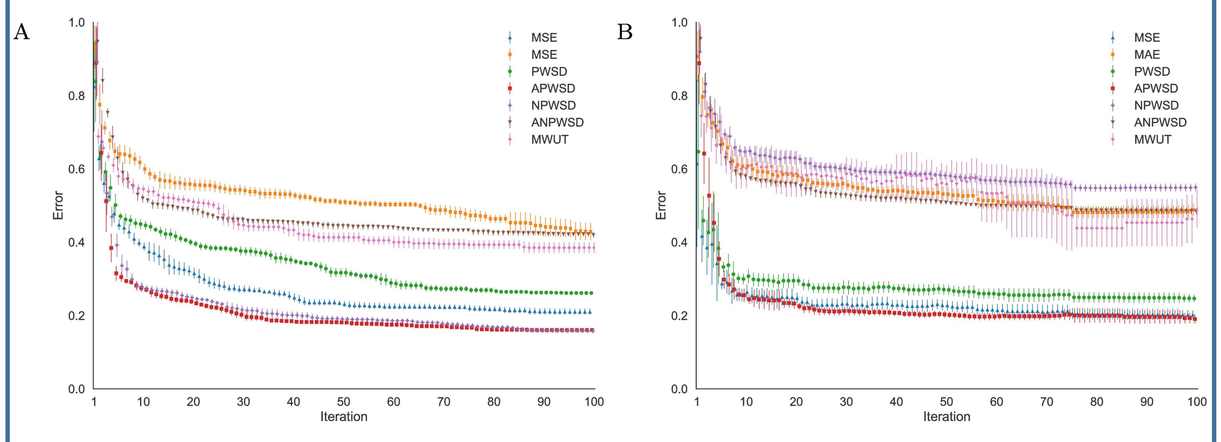
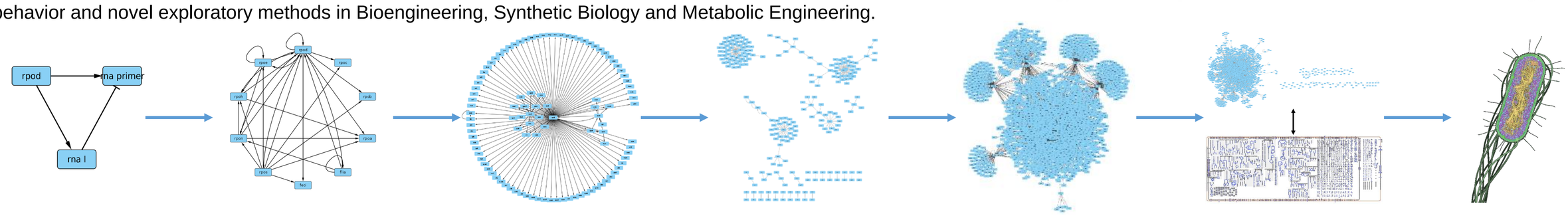


Figure 3. A, Average error for the ten best models per iteration is reduced effectively using common distance-based functions, e.g. Mean Squared Error, or the Mann-Whitney U-test (MWUT); **B,** Average error for the ten best models per iteration using a Multiple Objective Genetic Algorithm combining the U-test and other two common functions [17].

Conclusions and Further Work.

Rule-Based Models are a useful modeling framework to reconstruct known interactions between biological components and to scale-up efficiently the number of agents and processes considered. An ongoing effort to model every known gene, metabolism and processes that encompass the Molecular Biology Central Dogma would lead to a better understanding of cell behavior and novel exploratory methods in Bioengineering, Synthetic Biology and Metabolic Engineering.



Bibliography.
[1] Cardinale & Arkin. Biotechnol. J. 7, 856–66 (2012).
[2] Medema et al. Nat. Rev. Microbiol. 10, 191–202 (2012).
[3] Lim et al. Mol. Cell 49, 202–212 (2013).
[4] McCloskey et al. Mol. Syst. Biol. 9, 661 (2013).
[5] Fisher & Henzinger. Nat. Biotechnol. 25, 1239–49 (2007).
[6] Gillespie. J. Phys. Chem. 81, 2340–2361 (1977).
[7] Hlavacek et al. Phys. Biol. 12, 45007 (2015).
[8] Feret et al. PNAS. 106, 6453–8 (2009).
[9] Cho et al. BMC Biol. 12, 4 (2014).
[10] Keseler et al. Nucleic Acids Res. 41 (2013).
[11] Stewart & Wilson-Kanamori. Electron. Notes Theor. Comput. Sci. 277, 77–87 (2011).
[12] López et al. Mol. Syst. Biol. 9, 646 (2013).
[13] Pérez-Acle et al, manuscript submitted. Available https://github.com/dlab/PISKaS
[14] Merlin & Polisky. J. Mol. Biol. 248, 211–9 (1995).
[15] Available https://github.com/Kappa-Dev/KaSim
[16] Jozefczuk et al. Mol. Syst. Biol. 6, 364 (2010).
[17] Santibáñez et al, manuscript submitted. Available https://github.com/glucksfall/pleiades
[18] Patrick et al. Biochimie 119, 80–91 (2015).

Funding. CONICYT-PCHA/Doctorado Nacional/2014-21140377. CONICYT-FONDECYT de Iniciación Proyecto [11140342]. CONICYT-FONDECYT Regular Proyecto [11130518]. CONICYT-PIA Proyecto Basal Fundación Ciencia & Vida [PFB16]. Iniciativa Científica Milenio/Ministerio de Economía, Fomento y Turismo Proyecto Centro Interdisciplinario de Neurociencia de Valparaíso [ICM-Economía P09-022-F]. US Air Force Office of Scientific Research AFOSR Project FA9550-16-1-0111.