GEM FLAGSHIP PROJECT

LABS: LIRMM, L2C, IMAG

LEADERS: E. RIVALS (LIRMM), J. PALMERI (L2C)

STARTED JANUARY 2017

PARTNERS: UMR INSTITUT DE GÉNÉTIQUE HUMAIN, UMR INSTITUT DE GÉNÉTIQUE MOLÉCULAIRE DE MONTPELLIER, COMPUTATIONAL BIOLOGY INSTITUTE

TRAINING INTERACTIONS: MASTER OF PHYSIC "PHYSIQUE INGÉNIERIE DU VIVANT", MASTER OF INFORMATIC "BIOINFORMATIQUE, CONNASSANCES, DONNÉES"

BUDGET: 250K€ FUNDED BY NUMEV

LEVERAGE EFFECTS: RECRUITMENT OF A CR CNRS (L2C, BIOPHYSICS), POSTDOC POSITION FUNDED BY INSTITUTE OF COMPUTATIONAL BIOLOGY, PHD FELLOWSHIP FROM DOCTORAL SCHOOL OF BIOLOGY (CBS2)

INTERNATIONAL COOPERATIONS: PRINCETON UNIVERSITY, LMU MUNICH

The way a cell controls how genes are activated to produce RNA and proteins relies on a complex process called gene expression. It involves several levels of control, the two major ones being the transcription from DNA into RNA and the translation from RNA into proteins. Gene expression is a key to understanding cell differentiation and plays an important role in diseases like cancer. **GEM** aims to (1) propose innovative synergistic methods combining bioinformatics, biophysics and biology, and (2) structure the community around these research axes, which are widely studied locally by life sciences research teams.
The GEM project started in January 2017 and was not based on an initially existing tri-disciplinary interaction. NUMEV acted as a catalyst to start this original project at the national level.

The dynamics of ribosome behavior during translation is a key element in gene expression, but it is a complex issue. Three research topics are under study:
1. The difference in ribosome flux between the initiation phase, which results in a relative «slowing down» of the ribosomes, and the elongation phase of of translation (we speak of processivity rates along the RNA)
2. The influence of the nucleotide and di-nucleotide composition on the jump rates of ribosomes.
3. Finally, the influence of the cellular reservoir limits in the tRNA populations that are necessary for decoding mRNA codons by tRNA.
For these three questions about translation, we looked into the Totally Asymmetric Simple Exclusion Process (TASEP) models, the availability of data that allows us to use these models, and the possibilities for controlling results.
For questions 1 and 3, we identified, retrieved and performed the bioinformatics analysis of published data for yeast (a unicellular eukaryote).
In addition, we continued studies on the relationships between genomic rearrangements and chromatin contact points that influence transcription and thus the level of gene activation. This work led to the development of distance calculation algorithms that take into account chromatin contact cards that are experimentally established[2,3].
Another research step was to use bioinformatics tools to explore interactive Ribo-Seq data. Ribo-seq is the type of high-throughput sequencing that can be used to study translation. All fragments covered by ribosomes on the RNAs being translated are isolated, captured, and sequenced. Thus, short sequences of ribosome-covered mRNA portions are obtained irrespective of mRNA in the entire cell sample subjected to sequencing. Bioinformatic analysis allow us to identify the position of ribosomes on mRNA sequences or genes. We

can then explore the variation of the ribosome profile along the mRNAs across different conditions or experiments. However, interactive tools to verify, explore, and interpret these data are missing. For this reason, we develop a bioinformatics tool facilitating the biologist/biophysicist's exploration of these data through an interactive visualization interface; it should interest a growing community and find its usefulness in both research and teaching.

Concerning biophysics models, we proposed a modeling of translation based on the TASEP model. After a thorough study of the literature and extensive discussions between partners, we focused on homogeneous and inhomogeneous TASEP models. The inhomogeneous TASEP model can be used to account for the inhomogeneity of codon-dependent mRNA ribosome elongation rates related to the abundance of the corresponding tRNAs.

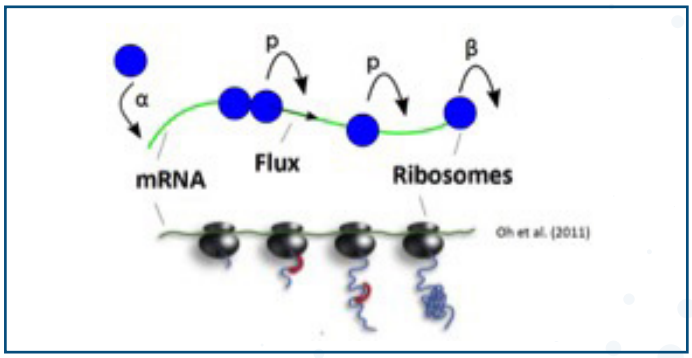

Fig. 1: TASEP model for ribosome elongation on mRNA.

Thanks to the Ribo-seq technique, the study of the codon dependent ribosome speed has become possible. However, there is a major obstacle in the way of successful modeling: Ribo-seq data give access to the local density of ribosomes along the mRNA, but no theoretical method has yet made it possible to reliably obtain ribosome speeds as a function of the density. We propose to resolve this issue.

1. We first used mean-field methods developed by (Shaw et al., 2004) to obtain ribosome speeds. Before applying these methods to biological data, we did a test on data from numerical simulations (Monte Carlo) for which densities and speeds are known. The results are unsatisfactory in some cases, inciting us to develop a method to account for correlation effects between neighboring ribosome sites, which are neglected in the mean field approach.

2. We have worked with bioinformaticians to identify a set of Ribo-seq experimental results that could serve as a «test case» and the analysis of these data is underway.

To address such complex issues, the collaboration between the five involved academic partner laboratories (LIRMM, L2C, IGMM, IGF, IMAG) will strongly structure the local community. The dynamic GEM community that is being established is projected to be the core of a major effort in integrative genomics centered in Montpellier. A better understanding of gene expression lets us foresee a wide spectrum of applications in medicine and biotechnology, including its role in certain diseases, especially cancer. The results of this project should therefore interest other groups (for instance Heidelberg, Munich and especially Princeton University, following the visit of Prof. N. Wingreen in April 2017) and foster new collaborations, including opportunities for technology transfer to industry and hospitals.

Example of Publications related to this project:

[1] *C. Bessière, M. Taha, F. Petitprez, J. Vandel, J.-M. Marin, L. Bréhélin, S. Lèbre, C.-H. Lecellier. Probing instructions for expression regulation in gene nucleotide compositions. PLoS Comput Biol. 2018 Jan2 ;14(1):e1005921. 2018*

[2] *S. Pulicani, P. Simonaitis, E. Rivals, K. M. Swenson. Rearrangement Scenarios Guided by Chromatin Structure. RECOMB-CG LNCS vol 10562 p. 141-155, 2017.*

[3] *P. Simonaitis, K. M. Swenson. Finding Local Genome Rearrangements. Proc. WABI, LIPIcs Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, vol. 88, 24:1-24:13, 2017.*

[4] *J.-C. Walter et al., Looping and clustering model for the organization of protein-DNA complexes on the bacterial genome, New J. Phys. 20,035002 (2018).*