Archaeal RNA polymerase pausing modeling and its gene expression control impacts

Almeida-e-Silva DC[1], Tie Koide[2],Vencio RZN[1]

[1]*Department of Computing and Mathematics FFCLRP-USP, University of Sao Paulo, Ribeirao Preto, Brazil*
[2]*Department of Biochemistry and Immunology FMRP-USP, University of Sao Paulo, Ribeirao Preto, Brazil*

Since the recognition of Archaea as a separate domain of life, there is growing interest in this evolutionary lineage. *Halobacterium salinarum* is an organism used as a model for halophilic archaea study. The NRC1 strain genome was published in 2000, and in 2008 the R1 strain sequencing, comparative analysis and proteomic data were published. These data were essential to the current understanding of haloarchaea as well as on archaeas as a whole.

Recently the massively parallel sequencing technology has enabled the identification of a new class of non-coding RNA, called the "Transcription Start Site-associated RNAs" (TSSaRNAs). The TSSaRNAs are small RNAs generated from the flanking regions of the Transcription Start Sites (TSS). TSSaRNAs have been identified in various eukaryotes and bacteria; and recent analyzes of *H. salinarum* data provide evidence of the presence of TSSaRNAs in Archaea. This suggests that TSSaRNAs are an evolutionarily widespread phenomenon in all three domains of life, and this phenomenon was possibly present in the last universal common ancestor. The conservation and expression regulation of non-coding RNA over all life domains suggests conserved biological functions associated with TSSaRNAs.

Currently there are not many confirmed information about the operation of TSSaRNAs. It is speculated that TSSaRNAs are related to the RNA polymerase pause sites; to elongation factors; to gene expression levels; or the backtracking process. In the present study we propose an in silico analysis of halophilic archaea data in search of answers to some of the unresolved issues that permeate the TSSaRNAs. With this we hope to advance the knowledge of this newly discovered class of small RNAs, in addition to expanding the understanding of archaea biology.

This work main goal is to advance on TSSaRNAs knowledge in archaea through in silico methods. For this, we intend to investigate sequence signatures that may be associated with RNA polymerase pause phenomenon; understand the role of the sequence signatures in archaea transcription; mathematically model the RNA polymerase pause phenomenon in archaea; establish relationships with observed phenomena and perform simulations seeking the relationship between RNA polymerase pause phenomena and gene expression control; investigate the influence of TSSaRNAs on transcription machinery fidelity and its possible consequences on the organism physiology.

In order to develop an automatic TSSaRNAs detection approach for *H. salinarum* and other organisms we are currently performing signal processing of RNA-seq, GC content, sequence conservation and minimal folding energy data.

Funding: Almeida-e-Silva is supported by a CAPES fellowship.