NATURAL ARCHITECTURE OF THE TRANSCRIPTION UNITS OF *Escherichia coli K-12 MG 1655* AS BASIS FOR DESIGNING SYNTHETIC BIOLOGY

Cynthia Paola-Rangel Chavez and Agustino-Martínez Antonio; CINVESTAV, Genetic Engineering Departmeent, Irapuato Guanajuato, Mexico 36821, cprh_@hotmail.com

Key words: transcriptional unit, synthetic biology, e. coli

Introduction. The information encoded in a genome drives the necessary tasks to keep working properly to all living systems(1). From this encoded information, genetic engineering has extracted DNA elements (promoter regions, gene sequences, ribosome binding site (RBS) sequences, terminators, among others), to further process biological systems to acquire new properties or functions. Under these principles, in this postgenomic era, new technologies have merged to build synthetic DNA elements based on the encoded genomic information. Nevertheless, when it is necessary to build such synthetic elements or compose standard bioparts many questions arise, such as, the distance between coding sequences, the composition of DNA strands, the promoter regions among others. The aim of this work is to analyze the natural organization of transcriptional units in the genome of *E. coli* and define composition and distribution patterns of genetic elements possibly related to biological function. The ultimate purpose is to obtain a guide for the design and assembly of transcription units.

Methods. To achieve this, the current annotation of *E. coli* K-12 genome was obtained from RegulonDB database v8.0. Subsequently, DNA sequence data analysis was performed to obtain the distribution of each of the DNA sequence elements listed: transcription units: promoters, Shine-Dalgarno (SD), genes and transcription terminators. Finally the expression levels and the codon usage of the genes encoded in the transcription unit will be integrated to make a correlation and clustering analysis.

Results. We determined the natural organization tendency of the elements involved in the transcription units. In the promoter region it was found that the distance between recognition boxes to sigma factor has a frequency of 14 nucleotides, and the distance of the -10 box to the transcription start site is 7 nucleotides. The distance from the transcription start site to the RBS was variable but in most cases was in the range of 11 to 40 nucleotides. For the RBS, a consensus analysis was obtained from this short sequences and it was also found that the distance between this site and the beginning of the first gene in the transcription unit was 6 nucleotides, and this result matches with those currently performed synthetically. Regarding to those genes that are inserted in the transcription unit, it was found that most of them have an overlap of four nucleotides. This means, for one first gene its stop codon overlaps with the start codon of the second gene. Due to this overlap, it was also found that the ribosome binding site of the second gene was inside the reading framework of the first one. For the terminators, the stop codon distance of the translation to the start point of the terminator is in the range of 10 to 30 nucleotides and the size of the terminator between 20-32 nucleotides (Fig. 1).

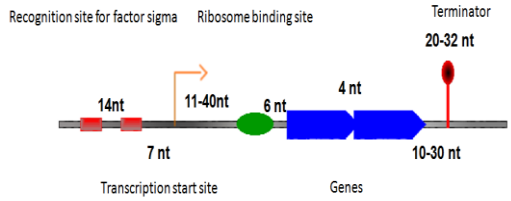

Fig.1 Natural organization tendency of the transcription unit.

Conclusions. The elements that show uniform pattern distances are involved in key steps such as the initiation of transcription and translation. It is pending to compare the average of the elements distribution obtained here with respect to what is done in practice and also synthetically generated. It is also pending to propose a model for one standard transcriptional unit to be validated experimentally.

Acknowledgements.
The authors thanks to CINVESTAV and CONACYT (262969) for the financial support.

References.

1. Byung-Kwan,Yu Qiu, Young, L Barrett, Gao & Bernhard Palsson(2009) Nature biotechnology 27:1043-1052.