Supercomputing Facility for Bioinformatics and Computational Biology
Indian Institute of Technology Delhi, New Delhi, India

PHYSICOCHEMICAL CHARACTERISATION OF PROMOTERS USING MOLECULAR DYNAMICS SIMULATION-BASED PARAMETERS

Dinesh Sharma[†,§], Aditya Mittal[†,§] and B. Jayaram[†,#,§]

[†]Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Delhi, New Delhi, India

[#]Department of Chemistry, Indian Institute of Technology Delhi, New Delhi, India

[§]Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, New Delhi, India

ABSTRACT

Recruitment of RNA polymerase (RNAP) to specific transcriptional start sites (TSSs) has remained a mystery because prokaryotes and eukaryotes have nearly no consensus promoter sequence. Understanding the fundamental mechanism is essential to comprehend the idea of gene regulation. Working with the hypothesis that genomic DNA sequences must convey their functional roles through their Physico-chemical properties, we have characterised, in this study, the promoter regions (of prokaryotes and eukaryotes) using molecular dynamics simulation derived structural and energy parameters over all the possible unique trinucleotides and tetranucleotides. These Physico-chemical properties (28 properties) show distinct signatures at the TSS.

INTRODUCTION

Elucidation of promoters within the genome of any organism is of prime importance. These elements start the transcription process by joining with the RNAP [1]. Promoter regions aid in correctly identifying and confirming gene boundaries. Having stated the limited role of these entities, it is crucial to annotate these regions within the DNA. The classical molecular biology approach tends to provide us with reliable annotation. However, these are capital-intensive and time-consuming [2]. The need of the hour is to develop an efficient and universal promoter annotation tool/algorithm. The traditional machine learning-based promoter prediction models are accurate but requires training over large sequence data and thus are organism specific [3, 4]. In the present research work, we have investigated the biophysical profiles of promoters from organisms belonging to both the Prokaryotic and Eukaryotic domains of life [2, 5]. 25 Structural and 3 energy parameters obtained through Molecular Dynamics (MD) simulation have been mapped over all the possible tri and tetra nucleotides to give us the initial characterisation. The results presented here strengthen the notion that these Physico-chemical profiles are ubiquitous and can be used for designing a Universal promoter prediction tool in future that will aid in their efficient annotation [2, 5].

METHODOLOGY

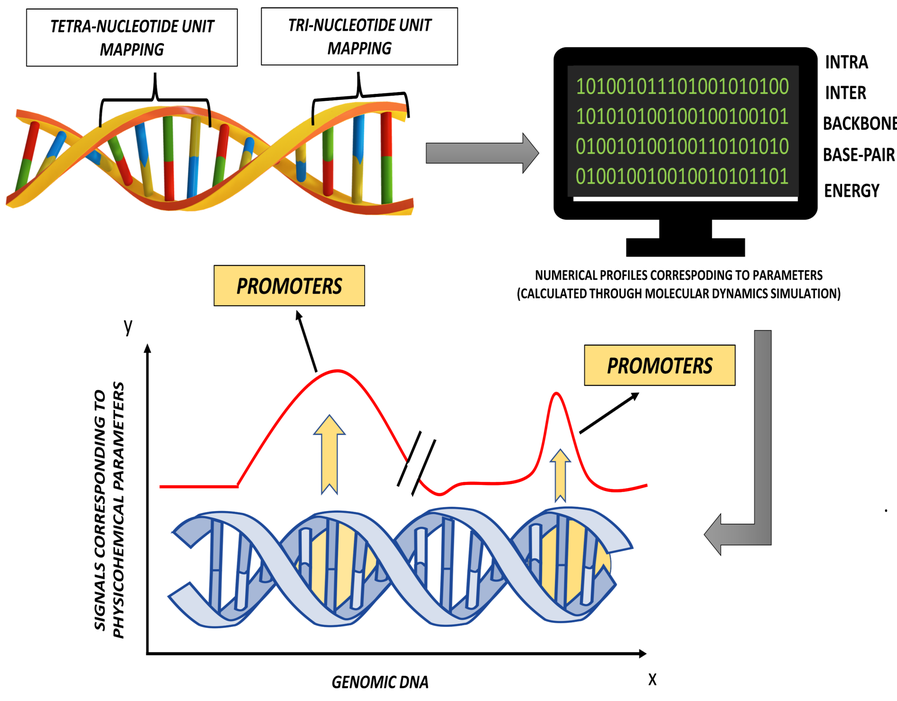

Figure 1: The workflow for characterising the TSS.

- Values of 25 Structures and 3 Energy parameters were obtained for all possible tri-nucleotides and tetra-nucleotides using µ-second MD simulations.
- 16519 promoter sequences were obtained from 12 different prokaryotes.
- 197356 promoter sequences were obtained from 8 different Yeast (Eukaryotes) species.
- Structure and energy profiles of promoter sequences were compared to the Coding sequences (CDS) profiles.
- Physico-chemical Profiles were compared to consensus sequence profiles.
- Individual sequences were analysed using threshold analysis.

RESULTS AND CONCLUSIONS

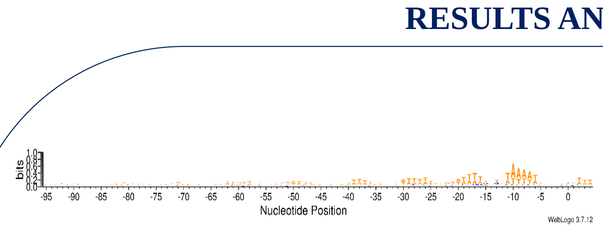

Figure 2: Consensus sequence at TSS for H. pylori

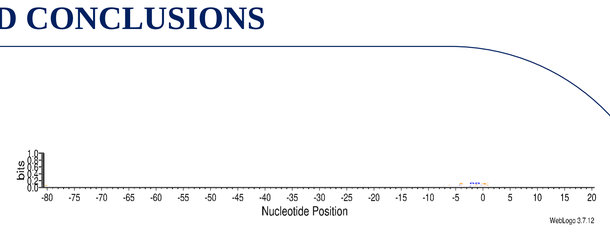

Figure 3: Consensus sequence at TSS for K. lactis

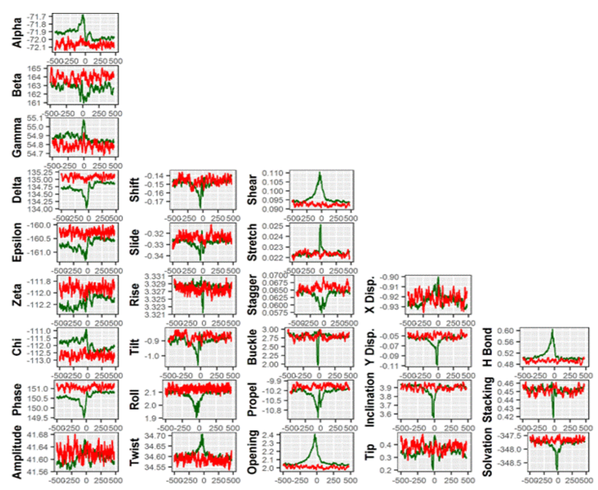

Figure 4: Physicochemical profiles of TSS and CDS for H. pylori

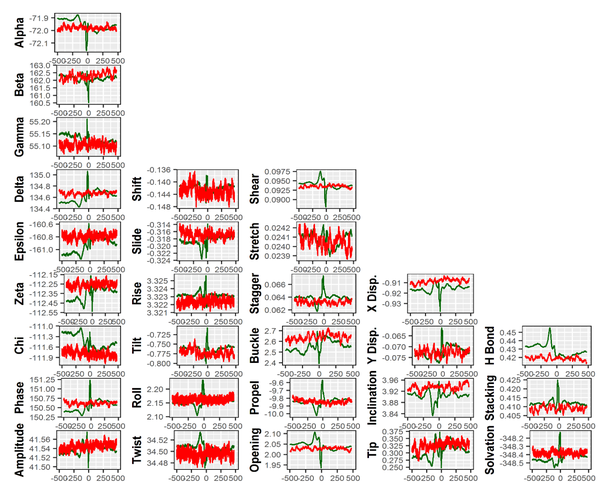

Figure 5: Physicochemical profiles of TSS and CDS for K. lactis

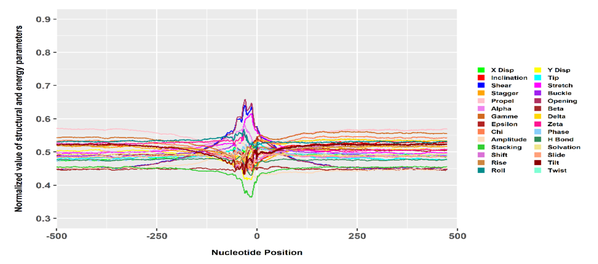

Figure 6: Profiles of the Combined (16519) prokaryotic sequences

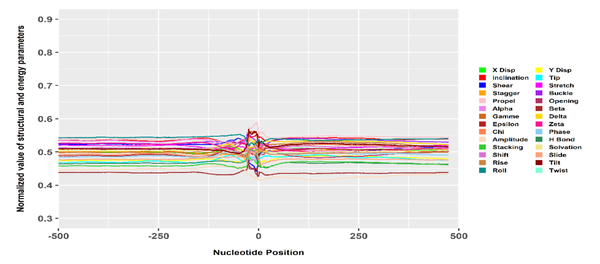

Figure 7: Profiles of the Combined (197356) eukaryotic sequences

- No consensus is evident around TSS in both prokaryotes and eukaryotes.
- Distinct profiles at TSS are obtained for each parameter.
- These Physico-chemical profiles are universal and are not organism-specific.
- Threshold analysis shows the signal corresponding to each parameter in > 98% of sequences.
- Analysis reveals that while sequences may show variations at various sites within the DNA, the structure and energy profiles imparted by these sequences are always conserved.
- These intrinsic signals can assist in efficiently identifying and confirming the desired sites.

ACKNOWLEDGEMENT

We gratefully acknowledge funding to SCFBIO, IIT Delhi, from the Department of Biotechnology, Govt. of India. Dinesh Sharma is a recipient of SRF from the CSIR, Govt. of India.

REFERENCES

1) Watson J, Baker T, Bell S, Gann A, Levine M, Losick R. Molecular Biology of the Gene. 7th ed. Pearson; 2013.
2) Mishra A, Dhanda S, Siwach P, Aggarwal S, Jayaram B. A novel method SEProm for prokaryotic promoter prediction based on DNA structure and energetics. Bioinformatics. 2020;36:2375–84.
3) Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. BMC genomics. 2012;13:1–10.
4) Salamov AA, Solovyev VV. Ab initio gene finding in Drosophila genomic DNA. Genome research. 2000;10:516–22.
5) Mishra A, Siwach P, Misra P, Jayaram B, Bansal M, Olson WK. Toward a universal structural and energetic model for prokaryotic promoters. Biophysical journal. 2018;115:1180–9.