CURVATURE PATTERN ANALYSIS OF *ESCHERICHIA COLI* PROMOTER SEQUENCES REGULATED BY ALTERNATIVE SIGMA FACTORS

Pedro Lenz Casa (BIC-UCS), Scheila de Avila e Silva (Orientador(a))

Promoters constitute DNA sequences upstream of the transcription start site of genes, and are therefore considered key elements in the regulation of gene expression. Sigma factors, particular to prokaryotes, are a group of proteins that can bind into the RNA polymerase and guide it towards specific sites, known as consensual motifs, in promoter sequences. Moreover, they are responsible for melting nucleotide bonds and establishing the transcription open complex. Besides nucleotide composition, the DNA molecule exhibits intrinsic structural properties, which tend to vary between promoters regulated by different sigma factors. The measure of curvature, for instance, determines values of twist, roll and tilt in B-DNA. Therefore, this work aims to analyse the curvature of promoter sequences regulated by alternative sigma factors in *Escherichia coli*, assessing implementation of this characteristic in prediction algorithms. For that, promoter sequences, obtained from the online database RegulonDB, were converted to macroscopic curvature values at each nucleotide. These were then submitted to a clustering algorithm known as K-means in order to attain grouping by similarity. Additionally, the average curvature profile for each sigma factor group was plotted in graphs and compared with a control group of false sequences generated with 25% of each nitrogenous base. The clustered groups displayed a low average purity regarding the number of promoter sequences recognized by the predominant sigma factor in each cluster. Considering that different sigma factor regulated sequences show divergent nucleotide profiles, the low purity in the curvature profiles suggest a similar overall promoter curvature signature. Moreover, it was possible to observe very similar curvature profiles for different sigma groups through the plotted graphs, which further implies that although they have different nucleotide compositions, they originate one curvature signature. In comparison to the false sequences curvature, the real promoters exhibit a higher overall curvature and more peaks and drops in the value along the sequence. Besides indicating that DNA curvature might play an important role in starting the transcription process, the results also reveal that DNA curvature has a high potential for distinguishing promoter sequences in prediction tools.