

# **TÍTULO: UMA ANÁLISE IN SILICO SOBRE A ESTABILIDADE DE SEQUÊNCIAS PROMOTORAS DE GENES RELACIONADOS A RESPOSTA AO CHOQUE TÉRMICO EM *Escherichia coli***

**Autores:** Gabriel Dall'Alba<sup>1</sup>; Scheila de Avila e Silva<sup>2,3</sup>

Instituição: <sup>1</sup>Universidade de Caxias do Sul- Centro de Ciências Biológicas e da Saúde.

<sup>2</sup>Universidade de Caxias do Sul- Instituto de Biotecnologia.

<sup>3</sup>Universidade de Caxias do Sul- Campus Universitário de Vacaria.

Email: [gdalba@ucs.br](mailto:gdalba@ucs.br); [sasilva6@ucs.br](mailto:sasilva6@ucs.br)

## **Resumo:**

O estresse por choque térmico em uma bactéria provoca mudanças em alguns processos biológicos. A resposta a este estímulo é regulada pela ligação dos fatores sigma ( $\sigma$ ) 24 e 32 na enzima RNA polimerase (RNAP) e o posterior reconhecimento da região promotora. A identificação dos nucleotídeos constituintes do promotor auxiliam o estudo da regulação gênica, ampliando a compreensão dos mecanismos de sobrevivência utilizados. Deste modo, o objetivo deste trabalho foi analisar o conteúdo de estabilidade da dupla fita de DNA dos nucleotídeos dos promotores reconhecidos pelos fatores  $\sigma_{24}$  e  $\sigma_{32}$  de *Escherichia coli*. Foram coletados do banco de dados RegulonDB dois conjuntos: (i) 521 sequências reconhecidas pelo  $\sigma_{24}$ ; (ii) 324 sequências reconhecidas pelo  $\sigma_{32}$ . Os promotores foram codificados em seus valores de estabilidade, conforme metodologia descrita na literatura. Os agrupamentos ocorreram com o uso da técnica computacional de clusterização e o algoritmo K-Means, o qual requer um valor numérico K (atribuído pelo usuário) equivalente ao número de agrupamentos desejados. Como a metodologia para a obtenção do número ótimo de k é empírica, foram realizadas simulações com diferentes valores: 12, 13 e 20. Verificou-se que a simulação com k=12 apresentou maior número de agrupamentos com pureza média superior a 85%. Após esta análise, a composição consensual dos *clusteres* com maior pureza mostrou que a região canônica -10 apresentou conteúdo similar ao consenso biológico estabelecido. Já para a região -35, não foi possível estabelecer relação com o consenso. Estas observações são explicadas pela presença da região -10 estendida, ou seja, a informação da região -35 realocada nesta região, a fim de proporcionar o funcionamento da RNAP. Outra característica foi a prevalência dos nucleotídeos A e G na região +1 (início da transcrição). Conforme a literatura, esta prevalência é responsável pela pausa na transcrição, a qual é um importante mecanismo para reparo de erros de mutação da região codificante. Deste modo, a compreensão dos mecanismos de regulação gênica relacionado ao choque térmico é importante já que este fenômeno altera a forma e composição do genoma. Percebe-se que um certo grau de degeneração das regiões consenso é tolerado no processo de transcrição gênica, porém evidenciam o desafio computacional. Assim, estabelecer os distintos perfis encontrados auxilia na redução do número de falsos positivos em ferramentas *in silico* relacionados à predição de promotores.

**Palavras-chave:** clusterização, promotores, fator sigma, *Escherichia coli*

**Agência de Fomento:** UCS