

# Projeto BacPP

## Predição de elementos regulatórios da expressão de genes de bactérias gram-negativas: aprimoramento da ferramenta computacional BacPP.

Atividade voluntária  
em pesquisa

Jean Lucas Benvenuti, Scheila de Avila e Silva e Daniel Luis Notari.

### Introdução

A expressão gênica de procariotos é regulada de diversas formas. Uma delas é a regulação pré-transcricional, na qual regiões promotoras encontradas “upstream” do sítio de início de transcrição dos genes interagem com a enzima RNA Polimerase (RNAP). A RNAP é formada por cinco subunidades e mais um fator sigma, o qual confere especificidade na expressão gênica de resposta às mudanças ambientais. Os promotores possuem características próprias como a composição de nucleotídeos e características estruturais. Dentre elas, a curvatura do DNA é uma estrutura que atua em importantes processos biológicos, pois ela é formada por ligações proteicas que facilitam a maquinaria de transcrição localizar determinado promotor. Um meio utilizado para a identificação dessas regiões é a simulação de Redes Neurais Artificiais (RN), que são sistemas computacionais constituídos por unidades chamadas neurônios. Esses neurônios possuem a capacidade de se interligam entre si, trabalhando em conjunto para desempenhar determinada tarefa através do aprendizado exercido pela repetição de ações executadas por eles mesmos. Por isso, a bioinformática é uma área que se destaca nas pesquisas da expressão gênica.

### Objetivo

O objetivo deste trabalho é realizar simulações de Redes Neurais Artificiais na identificação de promotores reconhecidos pelo fator sigma 70, que age na maioria dos genes de atuação geral, da bactéria Gram-negativa *Escherichia coli*.

### Materiais e métodos

Para a obtenção das sequências promotoras foi usado o banco de dados RegulonDB. As sequências foram transformadas em valores numéricos de curvatura segundo o modelo AA-Wedge, um dos seis existentes. Além da curvatura, valores do ângulo de curvatura e da maleabilidade foram obtidos. Em seguida, os dados foram suavizados com um filtro passa-baixa, que reduz a amplitude das maiores frequências. Para maior precisão, o processo de suavização foi realizado com 3, 5, 8 e 12 graus. Nas simulações de Redes Neurais, o modelo usado foi a *Multilayer perceptron*, que é formada por três camadas, sendo uma camada de entrada, uma camada oculta e outra camada de saída. Os testes foram realizados com o algoritmo *back-propagation* no ambiente R e para validá-los, a metodologia de validação cruzada foi utilizada. O desempenho de RN foi analisado em modelos matemáticos pelos valores de exatidão (A), de especificidade (S) e de sensibilidade (SN), pelas seguintes fórmulas matemáticas:

$$A = \frac{VP+VN}{VN+VP+FN+FP} \quad S = \frac{VN}{VN+FP} \quad SN = \frac{VP}{VP+FN}$$

onde VP representa verdadeiros positivos, VN verdadeiros negativos, FP falsos positivos e FN falsos negativos.

### Resultados

Como resultado obtido, tem-se a aprendizagem da RN na identificação de promotores, sendo que os parâmetros: ângulo de curvatura e ângulo de maleabilidade não mostraram-se eficientes no processo de simulações de RN. Por outro lado, os valores de curvatura discriminaram os promotores, sendo que a arquitetura ótima de RN apresentou:

Neurônios	Quantidade de neurônios
Camada de entrada	80
Camada oculta	4
Camada de saída	1

Os valores obtidos para as medidas de performance foram:

Performance	Valores
Exatidão	70,59%
Especificidade	74,01%
Sensibilidade	73,16%

### Considerações finais

A próxima etapa será a extração de regras da camada oculta das RN a fim de verificar o aprendizado por ela obtido e assim, promover a ampliação de inferências biológicas relacionadas à estrutura e funcionalidade de promotores. Adicionalmente, pretende-se obter um classificador de promotores que poderá ser usado no melhoramento de ferramentas de predição de promotores bacterianos (BacPP).

### Apoio



### Referências

- KANHERE, Aditi; BANSAL, Manju. Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. **Nucleic Acids Research**, v. 33, n. 10, p.3165-3175, 2 jun. 2005.
- SILVA, Scheila de Avila et al. DNA duplex stability as discriminative characteristic for *Escherichia coli* sigma 54 and sigma 28 dependent promoter sequences. **Biologicals**, v. 42, p.22-28, 2014.