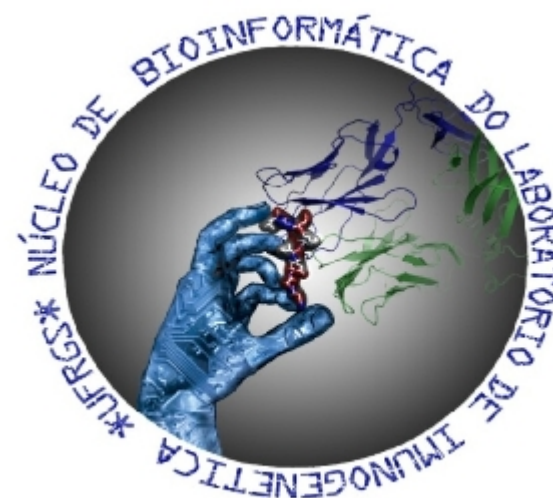


Informação estrutural das proteínas para a predição de eventos proteolíticos



Matheus de Bastos Balbé e Gutierrez, José Artur Bogo Chies
 Universidade Federal do Rio Grande do Sul
 Departamento de Genética
 Núcleo de Bioinformática do Laboratório de Imunogenética



paz no plural

Introdução

O sistema ubiquitina proteossomo compreende a principal via de degradação de proteínas em eucariotos, abrangendo funções que incluem tanto a remoção de proteínas regulatórias ou danificadas quanto a geração de peptídeos para a via de apresentação de antígenos durante a resposta imunológica celular. Neste contexto, células específicas do sistema imune adaptaram-se a reconhecer subprodutos de degradação de proteínas intracelulares na superfície de células apresentadoras. A maior parte destes peptídeos é proveniente da atividade catalítica do proteossomo, um sofisticado complexo proteico formado por uma partícula central com atividade proteolítica, associada a dois componentes regulatórios que reconhecem proteínas poliubiquitinadas, desdobram sua estrutura tridimensional, e as direcionam para degradação no interior. É sabido que o substrato é linearizado para entrar na partícula central do proteossomo, todavia as múltiplas câmaras internas são capazes de armazenar mais de uma molécula de substrato, e, além disso, muitos inibidores de proteossomo mimetizam substratos que formam pequenas estruturas semelhantes a folhas beta. Estas evidências demonstram uma possível relação de clivagem com a estrutura secundária do substrato no interior da câmara catalítica.

Métodos

Um conjunto não redundante de estruturas cristalográficas de proteínas humanas e virais foi obtida a partir do *Protein Data Bank*. Utilizando o programa Stride, anotamos a estrutura secundária para cada resíduo dessas sequências e por meio dos programas NetChop 3.1 e ProteaSMM anotamos a probabilidade de ocorrência de eventos de clivagem nestes respectivos aminoácidos. Para verificar se os eventos proteolíticos estavam associados a alguma estrutura secundária ou ocorriam ao acaso, um teste de qui-quadrado e um posterior cálculo de proporções (O/E) foi aplicado. Desse modo, podemos perceber se há uma sobre-representatividade de alguma estrutura secundária em locais onde a probabilidade de clivagem é alta. A mesma observação é possível em eventos com baixa probabilidade de corte. Em um segundo momento, para validar se a estrutura secundária é uma propriedade independente na discriminação de eventos proteolíticos, realizamos o treinamento de redes neurais utilizando o método *backpropagation*, somente informando as probabilidades de formação de hélices, fitas beta e alças (sem informar a sequência de aminoácidos). Esse conjunto de treinamento foi obtido a partir de dados da literatura de sequências de proteínas digeridas por proteossomo *in vitro*. Para validação do desempenho dessas redes, utilizamos curvas ROC, que é uma representação gráfica de avaliação de um classificador, testando-o sobre um conjunto de peptídeos teste ausente no treinamento.

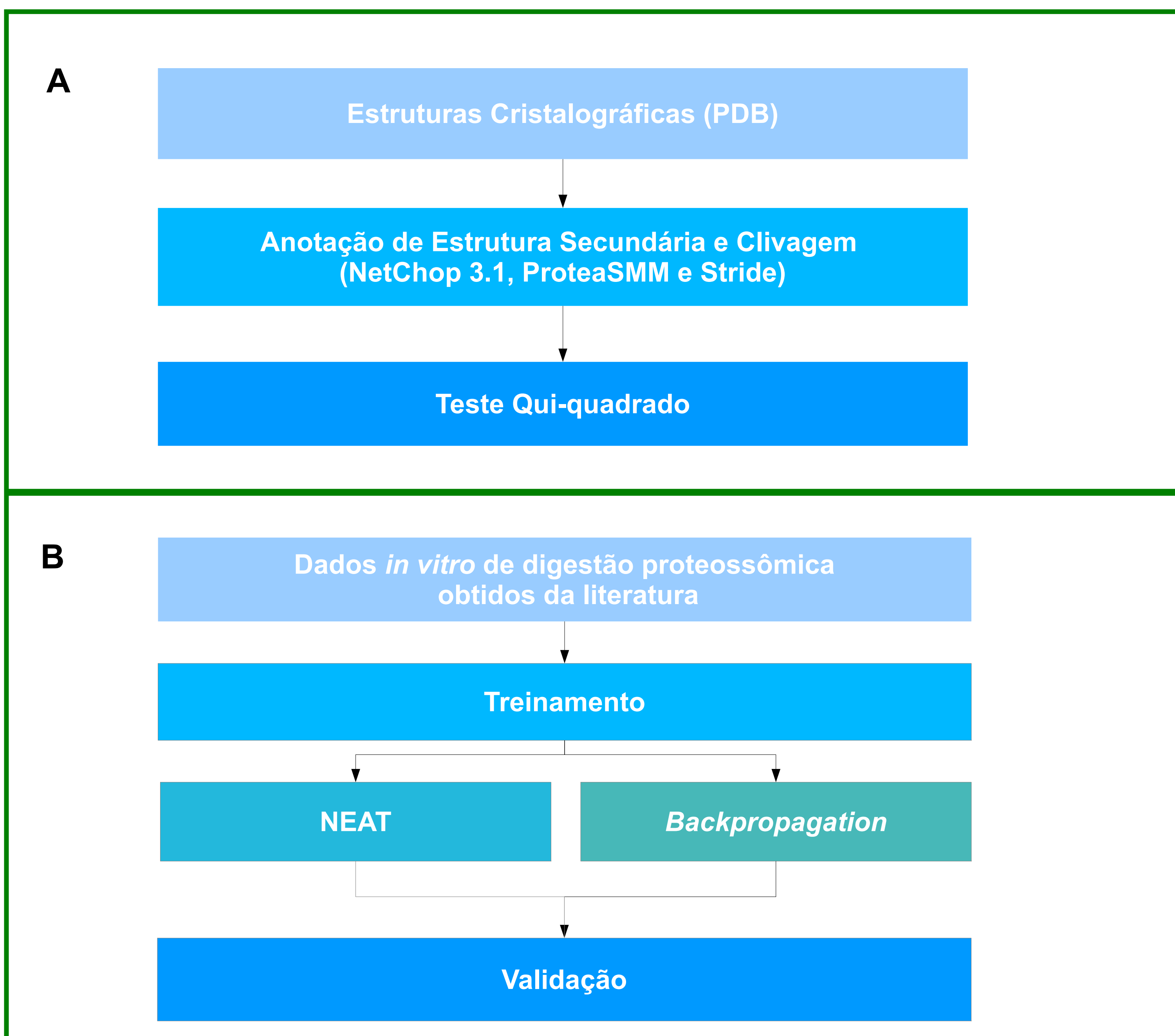


Figura 1. Fluxograma geral do trabalho. **A.** Metodologia proposta para verificar correlação entre estrutura secundária e eventos proteolíticos. **B.** Metodologia empregada para construção de redes neurais para avaliar a estrutura secundária como um elemento preditor independente.

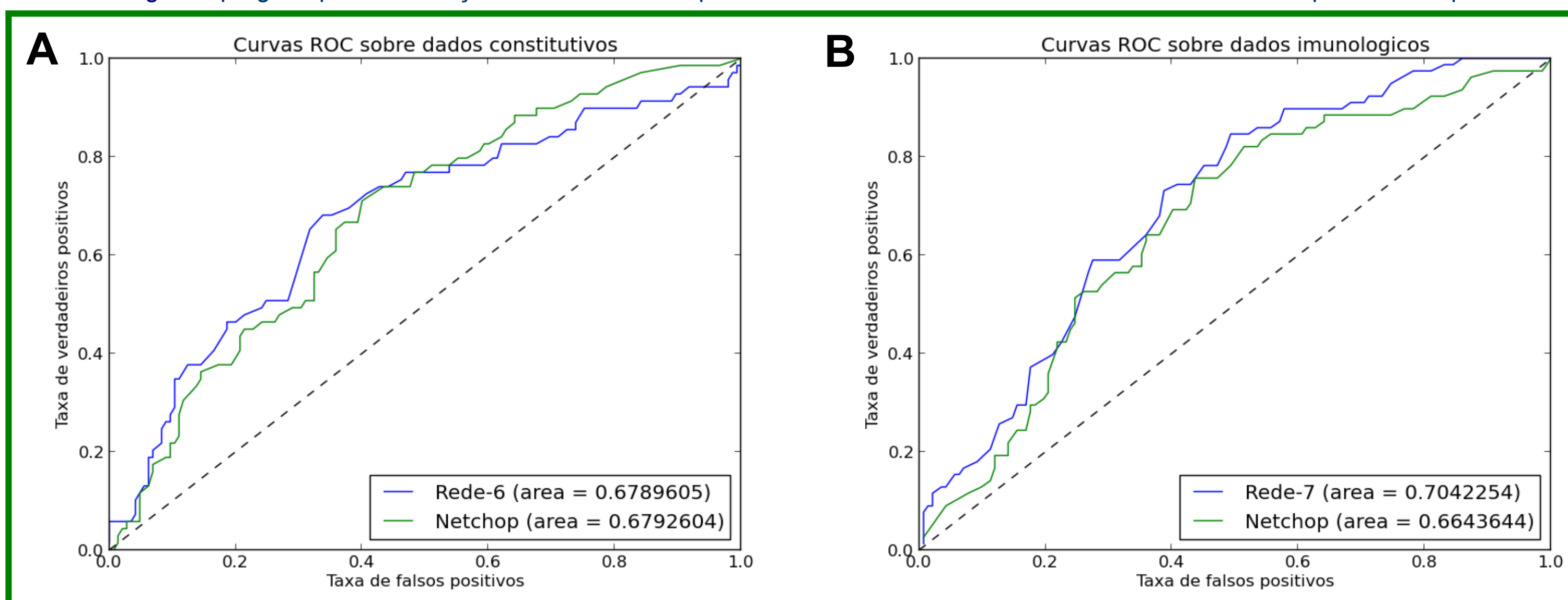


Figura 3. Avaliação das redes por meio de curvas ROC. Redes foram treinadas com sequências peptídicas variando entre 5 e 13 aminoácidos de tamanho. **A.** Teste sobre peptídeos clivados pelo proteossomo constitutivo. Rede apresentando melhor desempenho foi treinada sobre janelas de 6 aminoácidos. **B.** Teste sobre peptídeos clivados pelo proteossomo imunológico. Rede apresentando melhor desempenho foi treinada sobre janelas de 7 aminoácidos.

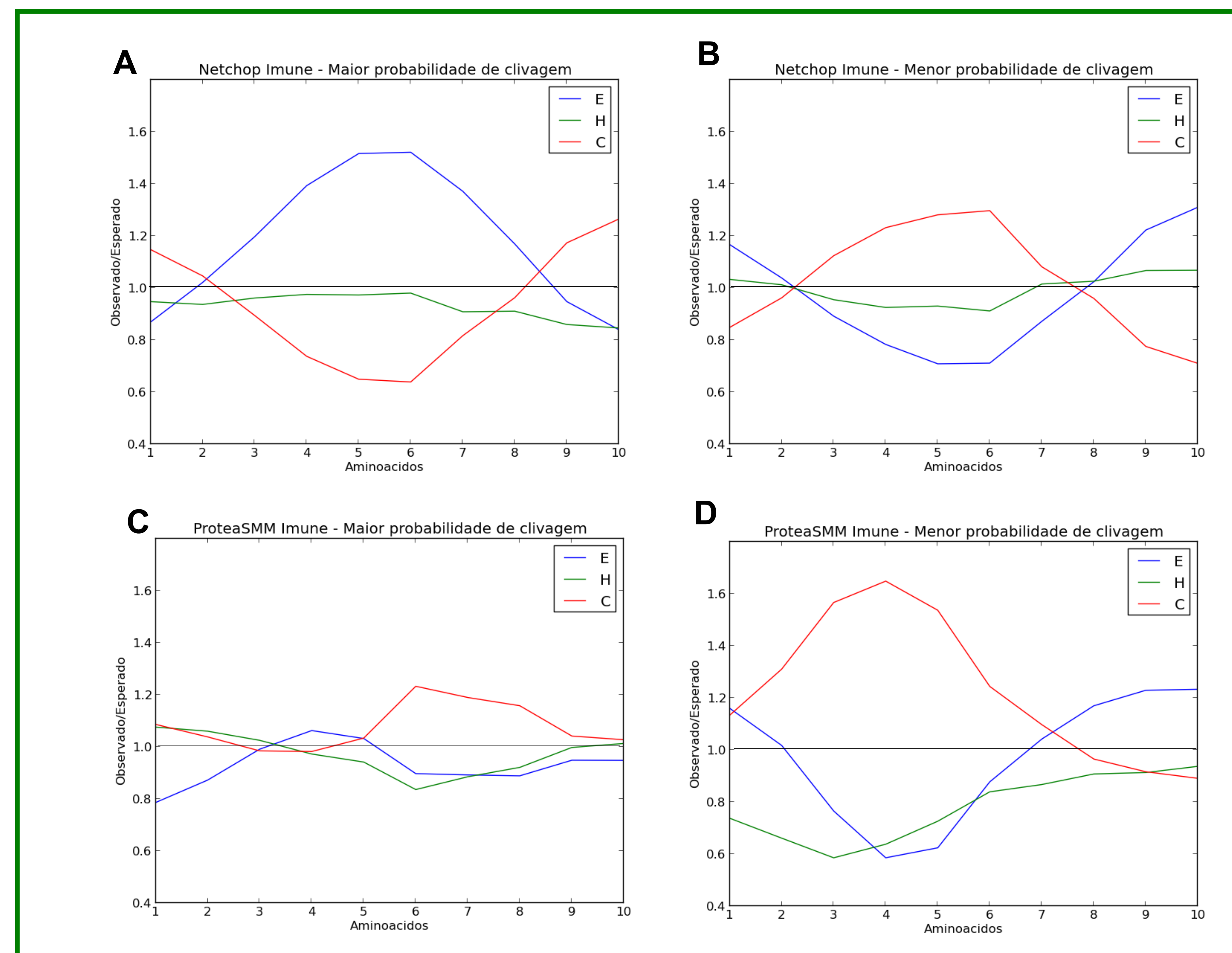


Figura 2. Razão entre a estrutura secundária observada em eventos proteolíticos e o esperado teórico da proporção das estruturas secundárias em todo o conjunto de dados, para as posições que antecedem e sucedem o local da clivagem (entre posições 6 e 7). **A e B.** Razão observada utilizando o preditor Netchop para anotar as probabilidades de clivagem (10% maiores e 10% menores, respectivamente). **C e D.** Razão observada utilizando o preditor ProteaSMM para anotar as probabilidades de clivagem (10% maiores e 10% menores, respectivamente). E, H e C representam as estruturas secundárias de fitas beta, hélice e voltas, respectivamente.

Resultados

As primeiras análises revelaram uma sobre-representação de aminoácidos apresentando estrutura secundária de fitas beta em regiões com alta probabilidade de clivagem, tanto na região do corte como em aminoácidos adjacentes. Por outro lado, há uma sobre-representação de resíduos em regiões de alça quando a probabilidade de corte é muito baixa (figura 2). Estes achados corroboram com a hipótese da estrutura secundária do substrato influenciar eventos proteolíticos. O desempenho equivalente de nossa rede ao de um clássico preditor de clivagem proteossômica confirma a importância da estrutura secundária como um fator independente na ocorrência de eventos proteolíticos.

Perspectivas

- Término da implementação de NEAT
- Comparações com ProteaSMM

