

配列からタンパク質の活性を推定する二つの計算手法が示す予測値の差異の検証

満富 健太¹, 木賀 大介²,

^{1,2}早稲田大学大学院先進理工学研究科電気情報生命専攻

Japanese abstract

活性のある祖先型タンパク質を生成するために、二つの計算機ツールを使い、実在した可能性が高く活性が高いと予想される祖先配列を探索する研究に取り組んだ。一つ目は、座位独立にアミノ酸を推定する古典的な祖先配列推定であり、二つ目は、残基間の相互作用を計算する機械学習モデルである。この二つの計算機ツールの癖を知ることが重要であると考え、二つのツールから出力される祖先候補配列群の活性に相関するとされる指標を算出した。一つ目は、各祖先候補配列の存在確率を示す確率尤度であり、二つ目は、機械学習モデルから出力されるELBOである。これら二つの相関グラフを算出したところ、二つのツールにおける祖先配列群の評価に正の相関がありつつも、尤度は多少低くなるが高い活性を維持する祖先配列を見出した。つまり、両ツールの組み合わせが、膨大な祖先候補配列の中から良い配列を見つけるうえでの二重のフィルタとして有効であるといえる。