END-TO-END MACHINE LEARNING-DRIVEN DESIGN OF PROTEINS

ABSTRACT

Described herein are techniques for designing proteins for binding to a target. In some embodiments, the techniques include: obtaining an amino acid sequence for a candidate protein that binds to the target with a candidate binding affinity; determining, for proteins in a set of proteins, probabilities that binding affinities between the proteins and the target are greater than the candidate binding affinity, and identifying a subset of the set of proteins based on the determined probabilities. Determining a first probability that a first binding affinity between a first protein and the target is greater than the candidate binding affinity may include: processing a first amino acid sequence of the first protein using a trained machine learning model to obtain a first output indicative of the first binding affinity; and determining the first probability using the first output indicative of the first binding affinity between the first protein and the target.

(FIG. 1A)