CLAIMS

What is claimed is:

1. A method for designing antibodies for binding to a target, the method comprising:

obtaining an amino acid sequence of a candidate antibody wherein the candidate antibody binds to the target with a candidate binding affinity;

determining, for antibodies in a set of antibodies, probabilities that binding affinities between the antibodies and the target are greater than the candidate binding affinity, the antibodies in the set of antibodies having different amino acid sequences, the antibodies including a first antibody having a first amino acid sequence, and the probabilities including a first probability that a first binding affinity between the first antibody and the target is greater than the candidate binding affinity, wherein determining the first probability comprises:

processing the first amino acid sequence of the first antibody using a trained machine learning model to obtain a first output indicative of the first binding affinity between the first antibody and the target; and

determining the first probability using the first output indicative of the first binding affinity between the first antibody and the target; and

identifying a subset of the set of antibodies based on the determined probabilities that the binding affinities are greater than the candidate binding affinity.

2. The method of claim 1, further comprising:

producing at least one antibody in the identified subset of the set of antibodies.

3. The method of claim 1,

wherein determining the probabilities that the binding affinities between the antibodies and the target are greater than the candidate binding affinity further comprises:

determining a second amino acid sequence of a second antibody in the set of antibodies based on (i) the first probability that the first binding affinity is greater than the candidate binding affinity and (ii) the first amino acid sequence of the first antibody.

4. The method of claim 3, further comprising:

after determining the second amino acid sequence, determining a second probability that

a second binding affinity between the second antibody and the target is greater than the candidate binding affinity.

5. The method of claim 3, wherein determining the second amino acid sequence of the second antibody comprises performing at least a portion of a sampling algorithm to determine the second amino acid sequence.

6. The method of claim 5, wherein the sampling algorithm is a hill climb algorithm, a genetic algorithm, or a Gibbs sampling algorithm.

7. The method of claim 1, further comprising:
identifying the first antibody from among a training set of antibodies having known binding affinities.

8. The method of claim 1, wherein identifying the subset of the set of antibodies comprises:
ranking the antibodies in the set of antibodies by the probabilities determined for the antibodies; and
identifying the subset of the set of antibodies based on the ranking.

9. The method of claim 8,
wherein ranking the antibodies in set of antibodies by the probabilities determined for the antibodies comprises ranking the antibodies from a highest probability of the determined probabilities to a lowest probability of the determined probabilities.

10. The method of claim 8,
wherein identifying the subset of the set of antibodies based on the ranking comprises identifying a predetermined number of antibodies associated with highest probabilities of the determined probabilities.

11. The method of claim 1, further comprising:
identifying, from among the identified subset of the set of antibodies, one or more antibodies having at least one pre-determined property; and

producing the identified one or more antibodies having the at least one pre-determined property.

12. The method of claim 1, wherein the first output indicative of the first binding affinity between the first antibody and the target comprises a mean of the first binding affinity and a standard deviation of the first binding affinity.

13. The method of claim 12, wherein determining the first probability using the first output indicative of the first binding affinity between the first antibody and the target comprises determining the first probability using the mean of the first binding affinity and the standard deviation of the first binding affinity.

14. The method of claim 1, wherein the trained machine learning model comprises at least one regression model.

15. The method of claim 14, wherein the at least one regression model is trained to predict, for an amino acid sequence of an antibody, a binding affinity between the antibody and the target.

16. The method of claim 14,

wherein the at least one regression model comprises multiple regression models, wherein each of the multiple regression models is trained to predict, for an amino acid sequence of an antibody, a binding affinity between the antibody and the target, and

wherein an output of the at least one regression model comprises a mean of the binding affinities predicted by the multiple regression models and a standard deviation of the binding affinities predicted by the multiple regression models.

17. The method of claim 1, wherein the trained machine learning model comprises a Gaussian Process model.

18. The method of claim 17,

wherein the Gaussian Process model is trained to predict, for an amino acid sequence of

an antibody, a binding affinity between the antibody and the target, and

wherein the probabilistic model is trained to output a mean of the predicted binding affinity and standard deviation of the predicted binding affinity.

19. The method of claim 1, wherein the trained machine learning model comprises at least one language model trained to encode amino acid sequences.

20. The method of claim 19, wherein the at least one language model is trained to predict masked amino acids in at least one amino acid sequence.

21. The method of claim 19, wherein the trained machine learning model further comprises:

at least one regression model fine-tuned from the at least one language model, or a probabilistic model fine-tuned from the at least one language model.

22. The method of claim 21, wherein processing the first amino acid sequence of the first antibody using the trained machine learning model comprises:

processing the first amino acid sequence using the at least one language model to obtain an encoded amino acid sequence, and

processing the encoded amino acid sequence using the at least one regression model or the at least one probabilistic model to obtain the first output indicative of the first binding affinity between the first antibody and the target.

23. The method of claim 21, wherein the at least one language model is a bidirectional encoder representations from transformers (BERT) model.

24. The method of claim 1, wherein the antibodies in the set of antibodies include single-chain variable fragments (scFvs).

.

25. A system, comprising:

at least one computer hardware processor; and

at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one computer hardware processor,

cause the at least one computer hardware processor to perform a method for designing antibodies for binding to a target, the method comprising:

obtaining an amino acid sequence of a candidate antibody wherein the candidate antibody binds to the target with a candidate binding affinity;

determining, for antibodies in a set of antibodies, probabilities that binding affinities between the antibodies and the target are greater than the candidate binding affinity, the antibodies in the set of antibodies having different amino acid sequences, the antibodies including a first antibody having a first amino acid sequence, and the probabilities including a first probability that a first binding affinity between the first antibody and the target is greater than the candidate binding affinity, wherein determining the first probability comprises:

processing the first amino acid sequence of the first antibody using a trained machine learning model to obtain a first output indicative of the first binding affinity between the first antibody and the target; and

determining the first probability using the first output indicative of the first binding affinity between the first antibody and the target; and

identifying a subset of the set of antibodies based on the determined probabilities that the binding affinities are greater than the candidate binding affinity.

26. At least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by at least one computer hardware processor, cause the at least one computer hardware processor to perform a method for designing antibodies for binding to a target, the method comprising:

obtaining an amino acid sequence of a candidate antibody, wherein the candidate antibody binds to the target with a candidate binding affinity;

determining, for antibodies in a set of antibodies, probabilities that binding affinities between the antibodies and the target are greater than the candidate binding affinity, the antibodies in the set of antibodies having different amino acid sequences, the antibodies including a first antibody having a first amino acid sequence, and the probabilities including a first probability that a first binding affinity between the first antibody and the target is greater than the candidate binding affinity, wherein determining the first probability comprises:

processing the first amino acid sequence of the first antibody using a trained

machine learning model to obtain a first output indicative of the first binding affinity between the first antibody and the target; and

determining the first probability using the first output indicative of the first binding affinity between the first antibody and the target; and

identifying a subset of the set of antibodies based on the determined probabilities that the binding affinities are greater than the candidate binding affinity.